

HOW DO BOOTSTRAP AND PERMUTATION TESTS WORK?

BY ARNOLD JANSSEN AND THORSTEN PAULS¹

Universität Düsseldorf

Resampling methods are frequently used in practice to adjust critical values of nonparametric tests. In the present paper a comprehensive and unified approach for the conditional and unconditional analysis of linear resampling statistics is presented. Under fairly mild assumptions we prove tightness and an asymptotic series representation for their weak accumulation points. From this series it becomes clear which part of the resampling statistic is responsible for asymptotic normality. The results lead to a discussion of the asymptotic correctness of resampling methods as well as their applications in testing hypotheses. They are conditionally correct iff a central limit theorem holds for the original test statistic. We prove unconditional correctness iff the central limit theorem holds or when symmetric random variables are resampled by a scheme of asymptotically random signs. Special cases are the $m(n)$ out of $k(n)$ bootstrap, the weighted bootstrap, the wild bootstrap and all kinds of permutation statistics. The program is carried out for convergent partial sums of rowwise independent infinitesimal triangular arrays in detail. These results are used to compare power functions of conditional resampling tests and their unconditional counterparts. The proof uses the method of random scores for permutation type statistics.

1. Introduction. Over the last 20 years nonparametric resampling procedures have become a powerful tool for setting confidence intervals and critical values of tests for composite hypotheses. In practice nonparametric two-step testing procedures benefit from the strong computational efforts of the new computer generation. Special resampling methods are Efron's bootstrap or Fisher's permutation tests. A justification is largely given by asymptotic considerations in order to compare the quality of resampling procedures with other tests. Throughout, we offer a unified asymptotic approach for the treatment of conditional resampling tests mainly given by partial sums of arbitrary arrays. It is based on a new setup for linear resampling statistics which is of independent interest. We will have a careful look at all kinds of permutation statistics, the $m(n)$ out of $k(n)$ -bootstrap and the weighted bootstrap (including the wild bootstrap); see Section 5.

Conditional limit theorems can be motivated by the following nonparametric testing problem. Let H_0 denote a composite null hypothesis of distributions (or the

Received March 2001; revised January 2002.

¹Supported by a DFG grant.

AMS 2000 subject classifications. 62G09, 62G10.

Key words and phrases. Sample mean, bootstrap, weighted bootstrap, wild bootstrap, permutation statistics, resampling tests, conditional tests, random scores, rank statistics, infinitely divisible laws, two-sample test, exchangeable variables.

boundary of an even larger null hypothesis). As special example consider as null hypothesis H_0 a subset of the set of product measures

$$(1.1) \quad H_0 \subset H_{\text{prod}} = \{P^n : P \in \mathcal{M}_1(\Omega, \mathcal{A})\}$$

where $\mathcal{M}_1(\Omega, \mathcal{A})$ denotes the set of probability measures on some measurable space.

Suppose that $T_n(X_1, \dots, X_n)$ is a given test statistic $T_n : \Omega^n \rightarrow \mathbb{R}$ based on Ω -valued random variables. In practice often a central limit theorem holds for T_n under H_0 and a consistent variance estimator V_n for its unknown asymptotic variance is available. Then the one-sided upper T_n -test φ_n for H_0 can be carried out as asymptotic level α test

$$(1.2) \quad \varphi_n = \mathbf{1}_{(c_n, \infty)}(T_n)$$

with asymptotically correct critical values $c_n = u_{1-\alpha} V_n^{1/2}$, $\Phi(u_{1-\alpha}) = 1 - \alpha$. Here Φ stands for the standard normal distribution function. Similarly, quadratic forms of asymptotically multivariate normal distributed vectors can be treated. This approach has the disadvantage that the statistician has in general no control over the real nominal level of φ_n on H_0 for fixed sample size n . Also available rates of convergence may depend on unknown parameters.

At this stage typically resampling methods are applied in order to establish more accurate data dependent critical values $c_n^* = c_n^*(x_1, \dots, x_n)$ of (1.2) where x_i denotes the outcomes of X_i . The most popular resampling methods are permutation and bootstrap procedures where the critical values $c_n^*(x_1, \dots, x_n)$ are determined by the quantiles of the permutation or bootstrap distribution of T_n ; see Beran (1988) for more refined bootstrap tests. To fix ideas draw a resampling sample X_1^*, \dots, X_n^* given X_1, \dots, X_n and consider a resampling variable

$$(1.3) \quad T_n^* = T_n(X_1^*, \dots, X_n^*).$$

Here X_1^*, \dots, X_n^* may be the bootstrap sample of X_1, \dots, X_n or in the case of permutation tests they are defined by random permutations $X_{\sigma(1)}, \dots, X_{\sigma(n)}$. The (conditional) distribution functions of the resampling variable T_n^* and their $(1 - \alpha)$ -quantiles $c_n^*(X_1, \dots, X_n)$ are (in principle) known or easy to simulate. This procedure defines a resampling test version φ_n^* of $\varphi_n = \mathbf{1}_{(c_n, \infty)}(T_n)$ by

$$(1.4) \quad \varphi_n^* = \begin{cases} 1, & > \\ \gamma, & T_n = c_n^*(X_1, \dots, X_n), \\ 0, & < \end{cases}$$

where $\gamma = \gamma(X_1, \dots, X_n)$ may be a suitable random variable.

In the case of H_{prod} the permutation test is of exact level α for each sample size n . This idea is rather old and goes back to Fisher. Our main question is now:

- *How does resampling work in testing hypotheses?*

The answer relies heavily on the asymptotic behaviour of the conditional distributions of T_n which is studied in detail in the next sections. It is closely related to the work of Mammen (1992a) who showed that the bootstrap works iff the central limit theorem holds for T_n . However, our question is a little bit different since permutation tests always work on H_{prod} . Relevant is now how the resampling tests work in comparison to traditional tests φ_n (1.2) under H_0 and under local alternatives. It is shown that under various conditions, when the central limit theorem holds for T_n the permutation and bootstrap tests share the optimality conditions with the asymptotic tests φ_n and their resampling version can be recommended.

The structure of our paper is as follows. Lemma 1 discusses the asymptotic equivalence of conditional and unconditional tests in general. Section 3 introduces tightness and convergence results for conditional (and unconditional) resampling statistics under fairly general assumptions. Their weak accumulation points can be represented by the sum of an infinite series of extremes and an additional infinitely divisible random variable. Special attention is devoted to the so-called L_2 -convergent resampling schemes where now the infinitely divisible part is normal; see Theorem 3. The extremes explain which part of the resampling statistics is responsible for the normal part. This reveals the well-known fact that the bootstrap does not work for extremes. A special case is the classification of all limit laws of two-sample partial sums of exchangeable random variables. In Section 4 partial sums of infinitesimal triangular arrays of independent random variables are resampled. The scheme is not restricted to rowwise i.i.d. random variables since we like to apply the results to two-sample alternatives; see Section 6. Here the question is answered when resampling conditionally works, that is, when the asymptotic conditional resampling distribution and the present limit variable ξ coincide. Resampling conditionally works iff the limit variable is normal. The same question is treated for unconditional limit variables. Then resampling works iff either ξ is normal or ξ is symmetric with (asymptotically) random signs as the resampling scheme. This result has a very natural statistical explanation; see Example 4.

In Section 5 it is shown that the bootstrap resampling scheme is L_2 -convergent if $\lim_{n \rightarrow \infty} \frac{m(n)}{k(n)} > 0$ holds. Thus all results mentioned above hold for the bootstrap, too. This leads to a discussion about the asymptotic correctness of the bootstrap for partial sums; see Remark 5. Section 6 introduces power functions of resampling tests when outcomes under alternatives are resampled. Again the procedures work well in the asymptotically normal case. Otherwise different things may happen. Example 5 explains what conditional tests are really doing in case of two-sample tests. Here the two-sample permutation tests are typically superior. The results of Section 7 about permutation statistics are of independent interest. It turns out that most of our proofs can be reduced to this important case. All other proofs are presented in Section 8.

There exists a huge amount of literature about the bootstrap of the sample mean which is cited briefly. Various kinds of bootstrap central limit theorems (in the almost sure or convergence in probability setup) were obtained by Giné and Zinn (1989), Csörgő and Mason (1989), Hall (1990), Arcones and Giné (1989, 1991), Mason and Newton (1992) and Mammen (1992a, b), among other authors.

A comprehensive treatment of the bootstrap including Edgeworth expansion and a review of related references can be found in the monograph of Hall (1992). A central limit theorem and Edgeworth expansions concerning the weighted bootstrap of differentiable functionals is due to Barbe and Bertail (1995). van der Vaart and Wellner (1996) applied advanced empirical process techniques to the bootstrap. Converse bootstrap central limit theorems are discussed by Wellner (2001). The performance of different kinds of bootstrap procedures is reviewed by Bickel, Götze and van Zwet (1997) in terms of expansions. Roughly speaking, it is known that the bootstrap works in the i.i.d. case iff the central limit theorem holds for the random variable under consideration.

Beyond these regular cases conditional limit theorems exist for the bootstrapped partial sums for random variables in the domain of attraction of stable laws; see Athreya (1987), Knight (1989), Kinatered (1992) and del Barrio and Matrán (2000). The approach of Knight and Kinatered uses the series representation of LePage, Woodroffe and Zinn (1981) for stable laws which is similar to our series construction below. The recent paper of Cuesta-Albertos and Matrán (1998) about the mean of infinitesimal rowwise i.i.d. arrays of random variables leads to conditional infinitely divisible limit laws for the bootstrap. Their results are related to the applications of our resampling results given in Section 5 where general non-i.i.d. triangular arrays are bootstrapped. Necessary conditions for the unconditional convergence of the bootstrap for triangular arrays can be found in del Barrio, Matrán and Cuesta-Albertos (1999).

Our method of proof treats resampling statistics as rank statistics with random scores. The present method extends the L_2 -convergence analysis of Hájek for score functions. In connection with permutation statistics this idea is rather old. Chernoff and Teicher (1958) and Witting and Nölle [(1970), Section 4.1] derived central limit theorems. It also served as a technical tool in Janssen and Mason (1990). Conditional central limit theorems for (permutation) survival statistics based on censored data were established by Neuhaus (1988, 1993) and Janssen (1989, 1991) along these lines. The present approach is closely related to Mason and Newton (1992) who used this principle for the weighted bootstrap. Among other authors Shorack (1996, 2000) studied the central limit theorem for permutation and bootstrap statistics. Strasser and Weber (1999) discussed multivariate central limit theorems for permutation statistics. Romano (1989) applied central limit theorems to randomization tests. Recent results about Studentized permutation statistics can be found in Janssen (1997) and in Mason and Shao (2001) for Studentized bootstrap statistics. The present technique is closely related to the treatment of

exchangeable random variables; see Billingsley (1971) and Einmahl and Mason (1992).

In the following let $X_{1:k(n)} \leq X_{2:k(n)} \leq \dots \leq X_{k(n):k(n)}$ denote the order statistics of real random variables $X_{n,1}, \dots, X_{n,k(n)}$ where $X_{i:k(n)} := 0$ for $i \notin \{1, \dots, k(n)\}$. Basic work about ranks and antiranks can be found in Hájek, Šidák and Sen (1999). In connection with infinitely divisible laws (including Lévy measures) we refer to Araujo and Giné (1980) and Petrov (1995). We use the notion of tightness as defined in Billingsley (1971).

2. Asymptotic equivalence of conditional and unconditional tests. To motivate the analysis let us first consider the following situation. Let $X_{n,i} : (\Omega, \mathcal{A}, P) \rightarrow \mathbb{R}$ be an arbitrary triangular array of random variables for $1 \leq i \leq k(n)$ (not necessarily independent) with $\bar{X}_n := \frac{1}{k(n)} \sum_{i=1}^{k(n)} X_{n,i}$. Let $\{k(n)\}_{n \in \mathbb{N}}$ be a sequence of integers, $k(n) \geq 2$, with $k(n) \rightarrow \infty$ as $n \rightarrow \infty$. Additional assumptions are discussed later.

EXAMPLE 1. Linear test statistics

$$(2.1) \quad T_n = k(n)^{1/2} \sum_{i=1}^{k(n)} c_{ni} X_{n,i}$$

given by a scheme of real regression coefficients $(c_{ni})_{i \leq k(n)}$ are of significant importance. Notice that if all $X_{n,i}$ are different (only assumed for a moment) T_n can be expressed via the antiranks $(D_{ni})_{i \leq k(n)}$ of the $X_{n,i}$ (given by $X_{n,D_{ni}} = X_{i:k(n)}$) via

$$(2.2) \quad T_n = k(n)^{1/2} \sum_{i=1}^{k(n)} c_{nD_{ni}} X_{i:k(n)}.$$

This form is closely related to resampling statistics. Two cases are of special interest.

(a) *Two-sample regression coefficients.* Consider sample sizes n_1, n_2 of two groups with $n_1 + n_2 = k(n)$. Define

$$(2.3) \quad c_{ni} = \left(\frac{n_1 n_2}{k(n)}\right)^{1/2} \begin{cases} -\frac{1}{n_1}, & i \leq n_1 & \text{(Sample 1),} \\ \frac{1}{n_2}, & n_1 < i \leq k(n) & \text{(Sample 2).} \end{cases}$$

A typical null hypothesis is given by

$$(2.4) \quad H_0 : X_{n,1}, \dots, X_{n,k(n)} \text{ are exchangeable.}$$

(b) *One-sample case.* The choice $c_{ni} = k(n)^{-1/2}$ defines linear statistics

$$(2.5) \quad T_n = \sum_{i=1}^{k(n)} X_{n,i}.$$

Typically, we will assume that (2.1) converges in distribution. The additional factor $k(n)^{1/2}$ is used to compensate the coefficients (2.3). The choice of the test statistics depends of course on the underlying alternatives. Statistics given by (2.1) are for instance appropriate for two-sample alternatives; see Hájek, Šidák and Sen (1999). However, modified resampling methods (permutation tests) sometimes also work when H_0 is larger than (2.4); see Janssen (1997) for treatment of Behrens–Fisher type two-sample testing problems.

The comparison of unconditional tests of type (1.2) and resampling tests (1.4) is based on the following lemma. It applies under fairly general circumstances and it is not only restricted to resampling tests. Let d denote any metric on the set of probability measures $\mathcal{M}_1(\mathbb{R})$ on \mathbb{R} such that convergence in $(\mathcal{M}_1(\mathbb{R}), d)$ is equivalent to weak convergence. We will consider the following setup.

- (I) (*Unconditional convergence*). Let P stand for a member of the null hypothesis (or its boundary). The real test statistics T_n are convergent in distribution $\mathcal{L}(T_n | P) \rightarrow \mathcal{L}(T | P)$ where T has a continuous distribution function F_T which is strictly increasing on its support.
- (II) (*Asymptotically unconditional upper T_n level α tests*). Let $\varphi_{n,\alpha}$ be a test given by unconditional critical values $c_n(\alpha)$ with

$$(2.6) \quad \varphi_{n,\alpha} = 1 \quad \text{if } T_n > c_n(\alpha), \quad \varphi_{n,\alpha} = 0 \quad \text{if } T_n < c_n(\alpha)$$

such that $E_P(\varphi_{n,\alpha}) \rightarrow \alpha$ holds as $n \rightarrow \infty$.

- (III) (*Upper T_n conditional tests*). Let T_n^* by any statistic (in our case a resampling statistic) with conditional distribution function F_n^* of $\mathcal{L}(T_n^* | X_{n,1}, \dots, X_{n,k(n)})$ given the data. Let $F_n^{*-1}(1 - \alpha) = c_n^*(\alpha) = c_n^*(\alpha, X_{n,1}, \dots, X_{n,k(n)})$ denote the conditional $(1 - \alpha)$ -quantile of that distribution given $X_{n,1}, \dots, X_{n,k(n)}$. A conditional (resampling) upper T_n -test is then any test with

$$(2.7) \quad \varphi_{n,\alpha}^* = 1 \quad \text{if } T_n > c_n^*(\alpha), \quad \varphi_{n,\alpha}^* = 0 \quad \text{if } T_n < c_n^*(\alpha).$$

LEMMA 1. *Under conditions (I)–(III) the following statements (a) and (b) are equivalent:*

- (a) $E_P(|\varphi_{n,\alpha} - \varphi_{n,\alpha}^*|) \rightarrow 0$ for all $\alpha \in (0, 1)$ as $n \rightarrow \infty$,
- (b) $d(\mathcal{L}(T_n), \mathcal{L}(T_n^* | X_{n,1}, \dots, X_{n,k(n)})) \rightarrow 0$ in P -probability as $n \rightarrow \infty$.

For the proof see Section 8.

The equivalence of unconditional and conditional test sequences holds iff the resampling asymptotically reproduces (in the conditional sense) the limit law. Applications of this fact are discussed in Section 6. This observation motivates the treatment of conditional distributions given in the next sections.

REMARK 1. The limit distribution of (2.1) is typically infinitely divisible; see Sections 4 and 6. Tucker (1975) showed that the supports of infinitely divisible laws are unbounded intervals if they are not shifts of compound Poisson distributions. In this case condition (I) holds since the distribution function of T is then continuous; see Hartman and Wintner (1942) or Zolotarev and Kruglov (1975).

3. Resampling linear statistics. Consider again as above an arbitrary triangular array of real random variables. Specify the set

$$(3.1) \quad H_n := \left\{ \sum_{i=1}^{k(n)} (X_{n,i} - \bar{X}_n)^2 = 0 \right\}$$

which is typically a set of small probability. Throughout, let us consider statistics

$$(3.2) \quad T_n^* = k(n)^{1/2} \sum_{i=1}^{k(n)} W_{n,i} (X_{n,i} - \bar{X}_n),$$

denoted as weighted resampling statistics, where $W_{n,i} : (\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{P}) \rightarrow \mathbb{R}$ is a triangular array of weight functions for $1 \leq i \leq k(n)$ on another probability space. With respect to $P \otimes \tilde{P}$ the X 's and the W 's become independent on the joint space $\Omega \times \tilde{\Omega}$. Consider the following conditions for the weights:

$$(3.3) \quad (W_{n,1}, \dots, W_{n,k(n)}) \quad \text{is exchangeable,}$$

$$(3.4) \quad \max_{1 \leq i \leq k(n)} |W_{n,i} - \bar{W}_n| \rightarrow 0 \quad \text{in } \tilde{P}\text{-probability,}$$

$$(3.5) \quad \sum_{i=1}^{k(n)} (W_{n,i} - \bar{W}_n)^2 \rightarrow 1 \quad \text{in } \tilde{P}\text{-probability.}$$

Notice that if in addition the condition $\bar{W}_n = \frac{1}{k(n)} \sum_{i=1}^{k(n)} W_{n,i} = 0$ holds \tilde{P} a.e. then $T_n^* = k(n)^{1/2} \sum_{i=1}^{k(n)} W_{n,i} X_{n,i}$ follows a.e.

The random weight functions $W_{n,i}$ stand for different resampling procedures. Various kinds of permutation and bootstrap statistics are discussed below. The central limit theorem concerning T_n^* was for instance discussed—among other authors—by Mason and Newton (1992), Barbe and Bertail (1995) and van der Vaart and Wellner (1996). Regularity conditions which ensure (3.3)–(3.5) are discussed in the next sections; see also Lemma 4.

EXAMPLES 2. (a) [*Permutation statistics, resampling T_n given in (2.1)*]. In this case the choice

$$(3.6) \quad W_{n,i} = c_{n\sigma(i)}$$

based on uniformly distributed permutations $\sigma(\cdot)$ of the indices $1, \dots, k(n)$ is appropriate. The regularity conditions (3.4), (3.5) are then equivalent to the same conditions for the c_{ni} instead of $W_{n,i}$. The two-sample coefficients (2.3) fulfill (3.4) and (3.5) iff $\min(n_1, n_2) \rightarrow \infty$. For centered c_{ni} the statistics (3.2) and (2.2) coincide in distribution whenever the antiranks are uniformly distributed and independent of the order statistics, for instance for H_{prod} (1.1).

(b) [*m(n)-bootstrap of partial sums*]. Let $(M_{n,1}, \dots, M_{n,k(n)})$ be a multinomial distributed random variable with sample size $m(n) = \sum_{i=1}^{k(n)} M_{n,i}$ and equal success probability $1/k(n)$ for each cell. The $m(n)$ -bootstrap weights are given by

$$(3.7) \quad W_{n,i} = m(n)^{1/2} \left(\frac{1}{m(n)} M_{n,i} - \frac{1}{k(n)} \right).$$

In case $m(n) = k(n)$ we arrive at Efron’s ordinary bootstrap statistic.

(c) (*Wild bootstrap*). Consider a triangular array of rowwise i.i.d. normalized random variables $Z_{n,1}, \dots, Z_{n,k(n)}$ with $E(Z_{n,1}) = 0$ and $\text{Var}(Z_{n,1}) = 1$. Let

$$(3.8) \quad W_{n,i} = k(n)^{-1/2} Z_{n,i}.$$

A special case is given by the bootstrap with Poisson sample size; see Mammen [(1992a), page 15] for details.

In the next step the conditional limit distributions of T_n^* given the data $X_{n,1}, \dots, X_{n,k(n)}$ will be studied in detail. We will see that the sequence of functions from Ω in $(\mathcal{M}_1(\mathbb{R}), d)$

$$(3.9) \quad \mathcal{L} \left(\frac{T_n^*}{(\sum_{i=1}^{k(n)} (X_{n,i} - \bar{X}_n)^2)^{1/2}} \mid X_{n,1}, \dots, X_{n,k(n)} \right)$$

is tight under (3.3) and (3.5) (with the convention $0/0 := 0$ on H_n). To prove this check that (7.8) below can be used to control its conditional variance. The class of accumulation points of (3.9) will be studied via limit theorems for rank statistics with random scores; see Section 7. For convenience define normalized random variables

$$(3.10) \quad Y_{n,i} := \frac{X_{n,i} - \bar{X}_n}{(\sum_{i=1}^{k(n)} (X_{n,i} - \bar{X}_n)^2)^{1/2}}.$$

Below let $(\Omega', \mathcal{A}', P')$ be a further probability space and let $Z_i, \tilde{Z}_j, Z^{(\omega)} : \Omega' \rightarrow \mathbb{R}$ be random variables for fixed $\omega \in \Omega$ whereas $\zeta_i, \tilde{\zeta}_j, Y_{n,i}, \Pi$ are random variables on Ω with $\Pi(\omega) := \omega$. All random variables can be defined in an obvious manner via projections on the joint probability space $(\Omega \times \Omega', \mathcal{A} \otimes \mathcal{A}', P \otimes P')$ where now random variables based on different projections only become independent. As in Lemma 1 above let the metric d describe weak convergence.

THEOREM 1. *For each subsequence there exists a further subsequence $\{m\} \subset \mathbb{N}$ such that the following distributions and conditional distributions are convergent.*

(a) *The sequences of order statistics*

$$(3.11) \quad (Y_{i:k(m)})_{i \in \mathbb{N}} \rightarrow (\zeta_i)_{i \in \mathbb{N}}, \quad (Y_{k(m)+1-j:k(m)})_{j \in \mathbb{N}} \rightarrow (\tilde{\zeta}_j)_{j \in \mathbb{N}}$$

are convergent in distribution on $[-1, 1]^{\mathbb{N}}$, where $(\zeta_i)_{i \in \mathbb{N}}$ and $(\tilde{\zeta}_j)_{j \in \mathbb{N}}$ denote certain limit variables. Moreover

$$(3.12) \quad \sum_{i=1}^{\infty} \zeta_i^2 + \sum_{j=1}^{\infty} \tilde{\zeta}_j^2 \leq 1$$

holds a.e. and $\lim_{m \rightarrow \infty} P(H_m)$ exists.

(b) *There exists another subsequence $m \rightarrow \infty$ of that given in (a) such that we have convergence of the conditional distributions*

$$(3.13) \quad d\left(\mathcal{L}\left(\frac{T_m^*}{(\sum_{i=1}^{k(m)} (X_{m,i} - \bar{X}_m)^2)^{1/2}} \mid X_{m,1}, \dots, X_{m,k(m)}\right), \mathcal{L}(T_0 \mid \Pi)\right) \rightarrow 0$$

with respect to the distance d in P -probability. The appertaining random variable $T_0: \Omega \times \Omega' \rightarrow \mathbb{R}$ is given by the series

$$(3.14) \quad T_0 = \sum_{i=1}^{\infty} Z_i \zeta_i + \sum_{j=1}^{\infty} \tilde{Z}_j \tilde{\zeta}_j + Z^{(\Pi)}$$

and certain i.i.d. random variables $Z_i, \tilde{Z}_j, i, j \in \mathbb{N}$, with $E(Z_1) = 0, \text{Var}(Z_1) \leq 1$. Moreover $Z^{(\Pi)}$ is given by a family of infinitely divisible random variables $Z^{(\omega)}$ with $E((Z^{(\omega)})^2) \leq 1$, for each $\omega \in \Omega$.

For the proof see Section 8.

REMARK 2. (a) According to (3.12) the series (3.14) is P' a.e. convergent for fixed ω [via $L_2(P')$ -convergence and the three series theorem].

(b) Under extra assumptions discussed below the subsequences $\{m\}$ of part (a) and part (b) coincide.

Before applications are discussed the construction principle of the Z_i, \tilde{Z}_j will be established. For these reasons consider the new subspace

$$(3.15) \quad S := \{\varphi \in L_2(0, 1) : \varphi \text{ nondecreasing, } \|\varphi\|_2 \leq 1\}$$

of square integrable functions $L_2(0, 1)$ of the uniform distribution. For $p = 1, 2$ the set S becomes by restriction a metric space as a subset of the normed spaces $(L_p(0, 1), \|\cdot\|_p)$. For convenience let $(S, \|\cdot\|_1)$ and $(S, \|\cdot\|_2)$ denote the different

metric spaces. According to Lemma 6 of Section 8 below $(S, \|\cdot\|_1)$ is a compact metric space whereas $(S, \|\cdot\|_2)$ is not compact.

The weights define a random variable $\tilde{\omega} \mapsto \varphi_n = \varphi_n(\tilde{\omega}, \cdot)$ from $\tilde{\Omega}$ into S by the step functions

$$(3.16) \quad \varphi_n(\cdot, u) := \frac{(k(n) - 1)^{1/2}}{(\sum_{i=1}^{k(n)} (W_{n,i} - \bar{W}_n)^2)^{1/2}} (W_{1+[k(n)u]:k(n)} - \bar{W}_n)$$

for $0 < u < 1$ where $[\cdot]$ denotes the entire function. Thus Lemma 6 implies:

LEMMA 2. *For each subsequence there exist a further subsequence $\{m\} \subset \mathbb{N}$ and a random variable $\varphi : \tilde{\Omega} \rightarrow S$ with $\varphi_m \rightarrow \varphi$ in distribution on $(S, \|\cdot\|_1)$ as $m \rightarrow \infty$.*

The limit variable of (3.16) now defines the distribution of Z_1 as follows. Let $\varphi : (\Omega_1, \mathcal{A}_1, P_1) \rightarrow S$ be an arbitrary random variable and let $U : (\Omega_2, \mathcal{A}_2, P_2) \rightarrow (0, 1)$ be a uniformly distributed random variable. Obviously then

$$(3.17) \quad (\omega_1, \omega_2) \mapsto \varphi(\omega_1, U(\omega_2))$$

is a real random variable w.r.t. the completion of $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2, P_1 \otimes P_2)$, briefly $\varphi(U)$, with conditional distribution $\mathcal{L}(\varphi(\omega_1, U) \mid \omega_1)$ given ω_1 and second moment

$$(3.18) \quad \int \varphi(\omega_1, U(\omega_2))^2 dP_1 \otimes P_2(\omega_1, \omega_2) \leq 1.$$

This simple observation is used to establish new random variables Z_i, \tilde{Z}_j which are part of T_0 . For this purpose let $(V_i)_{i \in \mathbb{N}}, (\tilde{V}_j)_{j \in \mathbb{N}}$ be two independent sequences of jointly uniformly distributed random variables with values in $(0, 1)$ jointly independent of all other variables. They may be defined on $(\Omega_2, \mathcal{A}_2, P_2)$.

REMARK 3. The variables Z_i, \tilde{Z}_j are given by

$$(3.19) \quad Z_i = \varphi(V_i), \quad \tilde{Z}_j = \varphi(\tilde{V}_j), \quad i, j \in \mathbb{N}$$

where $\varphi : \tilde{\Omega} \rightarrow S$ is a limit variable of (3.16) for a suitable subsequence.

In general we only have $(S, \|\cdot\|_1)$ cluster points of (3.16). The important stronger $(S, \|\cdot\|_2)$ convergence can be characterized as follows.

LEMMA 3. *Suppose that a subsequence $(\varphi_m)_m$ given by (3.16) converges in distribution on $(S, \|\cdot\|_1)$ to a random variable $\varphi : \tilde{\Omega} \rightarrow S$. Let $\varphi(U)$ denote the associated random variable. Then $E(\varphi(U)) = 0$ holds and*

$$(3.20) \quad \varphi_m \rightarrow \varphi \text{ in distribution on } (S, \|\cdot\|_2) \quad \text{iff} \quad \text{Var}(\varphi(U)) = 1.$$

For the proof see Section 8.

The structure of the infinitely divisible random variable $Z^{(\omega)}$ heavily relies on the form of the W 's. Fortunately, they are normally distributed for most of all practical situations. Recall that each infinitely divisible random variable X can uniquely be decomposed into the sum of two independent infinitely divisible random variables X_1 and X_2 , $X \stackrel{D}{=} X_1 + X_2$, such that X_2 is a centered normal random variable (or zero) and X_1 is the Poisson part of X without normal factors of its distribution.

THEOREM 2. (a) For each ω let $\bar{\sigma}^2(\omega)$ denote the variance of the normal part of $Z^{(\omega)}$ given by (3.14). For each $\omega \in \Omega$ we have

$$(3.21) \quad \bar{\sigma}^2(\omega) \geq \text{Var}(Z_1) \left(1 - \sum_{i=1}^{\infty} \zeta_i^2(\omega) - \sum_{j=1}^{\infty} \tilde{\zeta}_j^2(\omega) \right).$$

(b) Suppose that φ_m is convergent in $(S, \|\cdot\|_2)$. Then $\text{Var}(Z_1) = 1$ holds and $Z^{(\omega)}$ is a centered normal random variable for each ω with variance

$$(3.22) \quad \bar{\sigma}^2(\omega) = \text{Var}(Z^{(\omega)}) = 1 - \sum_{i=1}^{\infty} \zeta_i^2(\omega) - \sum_{j=1}^{\infty} \tilde{\zeta}_j^2(\omega).$$

In this case the distributional convergence of the order statistics (3.11) and of $P(H_m)$ along $\{m\}$ implies the conditional convergence (3.13) along the same sequence.

Under various conditions the results also hold for T_n^* and not only for Studentized versions (3.9).

THEOREM 3. Let $\varphi_n \rightarrow \varphi$ be convergent in $(S, \|\cdot\|_2)$ and let $\bar{X}_n \rightarrow 0$ hold in probability. Suppose that we have distributional convergence

$$(3.23) \quad \left((X_{i:k(n)})_{i \in \mathbb{N}}, (X_{k(n)+1-j:k(n)})_{j \in \mathbb{N}}, \left(\sum_{i=1}^{k(n)} (X_{n,i} - \bar{X}_n)^2 \right)^{1/2} \right) \\ \xrightarrow{D} ((\xi_i)_{i \in \mathbb{N}}, (\tilde{\xi}_j)_{j \in \mathbb{N}}, \xi_0)$$

of the joint distributions on $\mathbb{R}^{\mathbb{N}} \times \mathbb{R}^{\mathbb{N}} \times \mathbb{R}$. Then

$$(3.24) \quad d(\mathcal{L}(T_n^* | X_{n,1}, \dots, X_{n,k(n)}), \mathcal{L}(\tilde{T}_0 | \Pi)) \rightarrow 0$$

holds in P -probability, where \tilde{T}_0 is given similar to (3.14) by

$$\tilde{T}_0 = \sum_{i=1}^{\infty} Z_i \xi_i + \sum_{j=1}^{\infty} \tilde{Z}_j \tilde{\xi}_j + \tilde{Z}^{(\Pi)}.$$

The infinitely divisible part $\tilde{Z}^{(\omega)}$ of \tilde{T}_0 is a centered normal distribution with variance

$$\tilde{\sigma}^2(\omega) = \xi_0^2(\omega) - \sum_{i=1}^{\infty} \xi_i^2(\omega) - \sum_{j=1}^{\infty} \tilde{\xi}_j^2(\omega).$$

For the proof see Section 8.

Next let us state that conditional convergence in the sense of (3.13) or (3.24) implies distributional convergence of the (Studentized) resampling statistics T_n^* in an unconditional sense.

COROLLARY 1. *Suppose that under the conditions of Theorem 1 or 3 we have conditional convergence (3.13) or (3.24), respectively. For the same (sub)sequence we have unconditional convergence*

$$(3.25) \quad \frac{T_m^*}{(\sum_{i=1}^{k(m)} (X_{m,i} - \bar{X}_m)^2)^{1/2}} \rightarrow T_0$$

for (3.13) and

$$(3.26) \quad T_m^* \rightarrow \tilde{T}_0$$

for (3.24), respectively, in distribution as $m \rightarrow \infty$.

For the proof apply Lemma 8.

The results can now be used to classify all cluster points of Studentized (nontrivial) partial sums (2.1) of two-sample type for rowwise exchangeable random variables.

COROLLARY 2. *Consider T_n (2.1) given by rowwise exchangeable random variables $X_{n,1}, \dots, X_{n,k(n)}$. Let T_n^* be the permutation statistic induced by the weights $W_{n,i} = c_{n\sigma(i)}$ (3.6). Then $T_n \stackrel{\mathcal{D}}{=} T_n^*$ holds and (3.25) and (3.26) hold with T_m^* replaced by T_m .*

The unconditional convergence case is related to recent results of Mason and Shao (2001) for Studentized statistics when the denominator is also subject to the bootstrap. They showed that the unconditional bootstrap central limit theorem is valid iff the same holds for the original Studentized statistic. Their paper is based on a discussion similar to (3.22) for the variance given by the denominator.

EXAMPLE 3 (Two-sample permutation statistics; see also Example 6). Consider the permutation resampling scheme (3.6) given by the two-sample coefficients (2.3) with $n_1/k(n) \rightarrow \kappa \in (0, 1)$. Then the sequence $(\varphi_n)_n$ (3.16) converges in $(S, \|\cdot\|_2)$ and

$$(3.27) \quad \mathcal{L}(Z_1) = \kappa \varepsilon_{-((1-\kappa)/\kappa)^{1/2}} + (1 - \kappa) \varepsilon_{(\kappa/(1-\kappa))^{1/2}}$$

is a two-point distribution and the results above apply as follows. Similar results remain true for non-Studentized versions.

(a) For arbitrary triangular arrays $(X_{n,i})_i$ our Theorems 1 and 3 classify all (conditional or unconditional) cluster points of the Studentized permutation statistics

$$(3.28) \quad \frac{k(n)^{1/2} \sum_{i=1}^{k(n)} c_{n\sigma(i)} X_{n,i}}{(\sum_{i=1}^{k(n)} (X_{n,i} - \bar{X}_n)^2)^{1/2}}.$$

(b) If we start with a rowwise exchangeable array $(X_{n,i})_{1 \leq i \leq k(n)}$ there is no difference between (3.28) and the Studentized version of T_n (2.1) and we end up with all cluster points of Studentized two-sample statistics.

4. Resampling sums of independent variables. For the preceding results independence of the scheme is not needed. In this section let now $(X_{n,i})_{1 \leq i \leq k(n)}$ denote a triangular array of rowwise independent real random variables with distributional convergent partial sums

$$(4.1) \quad T_n = \sum_{i=1}^{k(n)} X_{n,i} \rightarrow \xi.$$

The array is assumed to be infinitesimal, that is

$$(4.2) \quad \max_{1 \leq i \leq k(n)} P(|X_{n,i}| > \varepsilon) \rightarrow 0 \quad \text{for each } \varepsilon > 0.$$

Then the law of ξ is infinitely divisible and ξ can be represented by mutually independent ‘‘Poisson series’’ Δ^+ , Δ^- , a centered normal random variable N with $\text{Var}(N) =: \sigma^2$, and a constant K , namely

$$(4.3) \quad \xi \stackrel{\mathcal{D}}{=} \Delta^- + N + \Delta^+ + K;$$

see Csörgő, Häusler and Mason (1988) and Janssen (1994) for details. The series Δ^+ , Δ^- are defined via partial sums $S_n = \sum_{i=1}^n Y_i$, $\tilde{S}_n = \sum_{j=1}^n \tilde{Y}_j$ of mutually independent standard exponential random variables $(Y_i)_i$, $(\tilde{Y}_j)_j$ with mean one. For some $\tau > 0$ they are given by

$$(4.4) \quad \Delta^- := \sum_{i=1}^{\infty} (\psi_1(S_i) - E(\psi_1(S_i) \mathbf{1}_{(-\tau, 0]}(\psi_1(S_i))))),$$

$$(4.5) \quad \Delta^+ := \sum_{i=1}^{\infty} (\psi_2(\tilde{S}_i) - E(\psi_2(\tilde{S}_i) \mathbf{1}_{[0, \tau)}(\psi_2(\tilde{S}_i))))).$$

The representation (4.3) is called the quantile representation since the functions $\psi_1, \psi_2 : (0, \infty) \rightarrow \mathbb{R}$ are just the quantile functions of the Lévy measure η of ξ . They are defined by

$$(4.6) \quad \psi_1(y) := \inf\{t : \eta(-\infty, t] \geq y\} \wedge 0,$$

$$(4.7) \quad \psi_2(y) := \sup\{t : \eta[t, \infty) \geq y\} \vee 0;$$

for details compare with Janssen [(1994), Sections 2 and 4 and Remark 8 below]. Notice first that $\bar{X}_n \rightarrow 0$ holds. According to Lemma 7 we need the following centering condition for $a_{n,i} := E(X_{n,i} \mathbf{1}_{(-\tau, \tau)}(X_{n,i}))$:

$$(4.8) \quad \sum_{i=1}^{k(n)} a_{n,i}^2 \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

We will assume (4.8) throughout. Notice that a change of centering in (4.1) is irrelevant for the underlying tests. On the other hand recall from Gnedenko and Kolmogorov [(1954), page 118] that (4.1) can always be centered by $X'_{n,i} := X_{n,i} - a_{n,i}$. The corresponding truncated means $a'_{n,i}$ of $X'_{n,i}$ then converge $\sum_{i=1}^{k(n)} a'_{n,i} \rightarrow 0$ and (4.8) holds for $X'_{n,i}$.

Now we are well prepared to apply Theorem 3 to the original statistic T_n^* (3.2). As in Theorem 1 the S_i, \tilde{S}_j and Π variables are defined on Ω whereas the Z 's are defined on Ω' .

THEOREM 4. *Consider convergent partial sums (4.1)–(4.8) and $(S, \|\cdot\|_2)$ convergent step functions $(\varphi_n)_n$, (3.16). Then we have*

$$(4.9) \quad d(\mathcal{L}(T_n^* | X_{n,1}, \dots, X_{n,k(n)}), \mathcal{L}(X_0 | \Pi)) \rightarrow 0$$

in P -probability. The random variable X_0 is defined via the random variables Z_i, \tilde{Z}_j , given in Remark 3, a standard normal random variable Z , jointly independent, and (4.6), (4.7) by

$$(4.10) \quad X_0 = \sum_{i=1}^{\infty} \psi_1(S_i)Z_i + \sum_{j=1}^{\infty} \psi_2(\tilde{S}_j)\tilde{Z}_j + \sigma Z.$$

The proof follows from Theorem 3 and Lemma 7.

REMARK 4. (a) For constant ξ we have $X_0 = 0$.

(b) If ξ is not constant we obtain three equivalent statements (4.11)–(4.13) concerning the conditional limit distribution $\mathcal{L}(X_0 | \Pi)$ in comparison with (4.1).

$$(4.11) \quad \mathcal{L}(X_0 | \Pi = \omega) \text{ is independent of } \omega \quad P \text{ a.e.}$$

$$(4.12) \quad E(\xi) \text{ exists and } \mathcal{L}(X_0 | \Pi = \omega) \stackrel{\mathcal{D}}{=} \xi - E(\xi) \quad P \text{ a.e.}$$

$$(4.13) \quad \xi \text{ is a normal random variable.}$$

Next we will compare the unconditional distributions of ξ and X_0 , see (4.1) and (4.10). The characteristic function of ξ is given by the Lévy–Hinčîn formula

$$(4.14) \quad E(\exp(it\xi)) = \exp\left(-\sigma^2 t^2/2 + iat + \int \rho(t, x) d\eta(x)\right)$$

with some $a \in \mathbb{R}, \sigma^2 \geq 0$ and Lévy measure η . Let $\rho(t, x) := \exp(itx) - 1 - \frac{itx}{1+x^2}$.

THEOREM 5. Consider X_0 , (4.10), with the Z_i 's having common distribution ν . Suppose that $\int x \, d\nu(x) = 0$ and $\int x^2 \, d\nu(x) \leq 1$ holds which may be more general as in Theorem 4. Then:

(a) The law of X_0 is infinitely divisible with Lévy measure

$$(4.15) \quad \eta_0(A) = \int_{\mathbb{R} \setminus \{0\}} \nu(s^{-1}A) \, d\eta(s), \quad A \in \mathcal{B}(\mathbb{R} \setminus \{0\}).$$

Its characteristic function is given by

$$(4.16) \quad E(\exp(itX_0)) = \exp\left(-\sigma^2 t^2/2 + itb + \int \rho(t, x) \, d\eta_0(x)\right)$$

with

$$b = \int \int \left(\frac{xz}{1+(xz)^2} - \frac{xz}{1+x^2}\right) \, d\nu(z) \, d\eta(x).$$

(b) The random variables $\xi \stackrel{\mathcal{D}}{=} X_0$ coincide in law iff (i) or (ii) holds.

- (i) ξ is a centered normal random variable or $\xi = 0$.
- (ii) ξ is a symmetric random variable and $\nu = \frac{1}{2}(\varepsilon_{-1} + \varepsilon_1)$.

For the proof see Section 8.

Compared with the conditional characterization (4.11)–(4.13) the additional case (ii), occurs which has the following natural interpretation in terms of symmetric random variables.

EXAMPLE 4. Consider symmetric random variables $X_{n,i}$ with convergent partial sums (4.1). Let $W_{n,i} = k(n)^{-1/2}Z_{n,i}$ be as in (3.8) a resampling scheme given by independent uniformly distributed signs $Z_{n,i} \in \{+1, -1\}$ (Rademacher random variables); see Theorem 5(b)(ii). Then we have equality in distribution of

$$(4.17) \quad \sum_{i=1}^{k(n)} X_{n,i} \stackrel{\mathcal{D}}{=} \sum_{i=1}^{k(n)} Z_{n,i} X_{n,i} = T_n^*$$

and $\xi \stackrel{\mathcal{D}}{=} X_0$ trivially holds. The Rademacher resampling scheme is of practical importance for the following cases:

- (a) When symmetric random variables are conditioned under their absolute values and conditional tests are considered for the null hypothesis of symmetry; see Janssen (1999).
- (b) For two-sample testing problems with balanced sample sizes $n_1/n_2 \rightarrow 1$, see Examples 1 and 6.
- (c) For the wild bootstrap (3.8) with distribution ν , see Theorem 5(b)(ii).

5. Bootstrapping the mean. This section establishes conditional limit theorems for bootstrap statistics (see Example 1) as application of Sections 3 and 4.

We again refer to related work cited in Section 1 which is mostly concerned with special limit laws.

LEMMA 4. *Let $k(n) \rightarrow \infty$. Then the conditions (3.3)–(3.5) hold for the bootstrap weights $W_{n,i}$ under the following circumstances:*

(a) *For the $m(n)$ -bootstrap whenever $m(n) \rightarrow \infty$ holds.*

(b) *For the wild bootstrap, see Example 2(c), if $k(n)^{-1/2} \sum_{i=1}^{k(n)} Z_{n,i}$ is distributional convergent to a standard normal random variable. This condition holds when*

$$(5.1) \quad Z_{n,1} \rightarrow Z_1 \quad \text{as } n \rightarrow \infty$$

in distribution where Z_1 is a random variable with $E(Z_1) = 0$ and $\text{Var}(Z_1) = 1$.

For the proof see Section 8.

The key is again the consideration of the bootstrap step functions.

THEOREM 6. *Let $m(n) \rightarrow \infty$. For the following cases the step functions $\varphi_n : \Omega \rightarrow S$, given by (3.16) and their bootstrap weights, are convergent, $\varphi_n \rightarrow \varphi$. The corresponding random variables $Z_1 \stackrel{\mathcal{D}}{=} \varphi(U)$ can be specified as follows.*

(I) *We have $(S, \|\cdot\|_2)$ -convergence for the following examples:*

(a) *For the $m(n)$ -bootstrap, whenever*

$$(5.2) \quad \frac{m(n)}{k(n)} \rightarrow c, \quad 0 < c \leq \infty$$

holds. Then the crucial random variables (3.19) are given for $c < \infty$ by

$$(5.3) \quad Z_1 \stackrel{\mathcal{D}}{=} c^{-1/2}(X - c),$$

where X denotes a Poisson random variable with mean 1, and

$$(5.4) \quad Z_1 \text{ is standard normal for } c = \infty.$$

(b) *For the wild bootstrap, Example 2(c), when condition (5.1) holds.*

(II) *In the case of the $m(n)$ -bootstrap with limit $c = 0$ of (5.2) we have $\varphi_n \rightarrow 0$ in $(S, \|\cdot\|_1)$ and $(\varphi_n)_n$ is not $\|\cdot\|_2$ -convergent. Then $Z_1 = 0$ holds.*

For the proof see Section 8.

REMARK 5. To explain how the bootstrap works let us again consider convergent partial sums T_n of independent random variables (4.1) and (4.8). Let $T_n^* = k(n)^{1/2} \sum_{i=1}^{k(n)} W_{n,i}(X_{n,i} - \bar{X}_n)$ denote the bootstrap statistic.

(I) In comparison with the unconditional series (4.3) of ξ and the series representation of the limit variable X_0 of T_n^* given by (4.10) we find extra multiplier factors Z_i, \tilde{Z}_j . According to Theorem 6 the Z_i 's are:

- of Poisson type ($m(n)$ -bootstrap $c > 0$),
- standard normal ($m(n)$ -bootstrap $c = \infty$),
- or given by the wild bootstrap limit variable.

(a) In the sense of Remark 4(b) the bootstrap is then asymptotically conditionally correct iff ξ is a centered normal random variable.

(b) The $m(n)$ -bootstrap with $c = \infty$ always produces centered normal conditional limit distributions $\mathcal{L}(X_0 | \Pi)$ of T_n^* with random conditional variance

$$(5.5) \quad \sigma^2(\omega) = \sum_{i=1}^{\infty} \psi_1(S_i(\omega))^2 + \sum_{j=1}^{\infty} \psi_2(\tilde{S}_j(\omega))^2 + \sigma^2.$$

(II) The results of Section 4 cannot be applied to the $m(n)$ -bootstrap with $c = 0$. What only can be said is that $\tilde{T}_0 = Z^{(\Pi)}$ is conditionally infinitely divisible given Π , see Theorem 3. For further results we refer to del Barrio and Matrán (2000) who proved correctness of the bootstrap with low resampling intensity for i.i.d. schemes attracted to stable laws.

(III) The unconditional correctness of the bootstrap, specified by $X_0 \stackrel{D}{=} \xi$, is now a special case of Theorem 5 of Section 4.

(IV) The present results based on arbitrary triangular arrays generalize earlier work of Cuesta-Albertos and Matrán (1998) for rowwise i.i.d. schemes. In particular, they obtained conditional normal laws for $c = \infty$ and in the case of $0 < c < \infty$ conditional infinitely divisible limit laws in terms of conditional characteristic functions.

6. Resampling tests under alternatives. The present results have a lot of applications for conditional resampling tests. Some of them—mostly to permutation and bootstrap tests—are discussed in this section. Further results will be considered forthcoming. Throughout, we will restrict ourselves to linear statistics T_n (2.1) and resampling statistics T_n^* (3.2) given by $(S, \|\cdot\|_2)$ convergent step functions. We suppose that $\bar{W}_n = 0$ holds in this whole section. The null hypothesis may consist of certain product measures (1.1). Assume for a moment in (6.1) and (6.2) that $T_n \rightarrow Z$ is asymptotically normal under H_0 . Then we have asymptotic normality of $T_n^* \rightarrow Z$, and $\mathcal{L}(T_n^* | X_{n,1}, \dots, X_{n,k(n)})$, respectively, and the equivalence of unconditional and conditional tests $\varphi_n - \varphi_n^* \xrightarrow{H_0} 0$ follows. The (asymptotic) level α of φ_n and φ_n^* is always assumed to be fixed, $0 < \alpha < 1$. Consider now a sequence of relevant contiguous alternatives Q_n

$$(6.1) \quad Q_n = \mathcal{L}(X_{n,1}, \dots, X_{n,k(n)}) \triangleleft P^{k(n)}$$

for some $P^{k(n)} \in H_0$; see Hájek, Šidák and Sen (1999). Then by Lemma 1,

$$(6.2) \quad \lim_{n \rightarrow \infty} (E_{Q_n}(\varphi_n^*) - E_{Q_n}(\varphi_n)) = 0$$

follows and φ_n^* is asymptotically efficient for $\{P^{k(n)}\}$ against $\{Q_n\}$ whenever φ_n is asymptotically efficient. Under mild regularity assumptions the consistency of φ_n^* can be derived under noncontiguous alternatives in the general case.

THEOREM 7. *Consider again $P^{k(n)}$ and Q_n similar to (6.1) where Q_n may be a sequence of noncontiguous alternatives. Let $(\mathcal{L}(T_n | P^{k(n)}))_n$ be tight and assume that $T_n \rightarrow \infty$ holds in Q_n distribution. Suppose also tightness of the distributions given by the random variable from Ω into $(\mathcal{M}_1(\mathbb{R}), d)$*

$$(6.3) \quad \omega \mapsto \mathcal{L}(T_n^* | X_{n,1}(\omega), \dots, X_{n,k(n)}(\omega))$$

as well as under $P^{k(n)}$ and Q_n . Under these conditions we have consistency,

$$(6.4) \quad \lim_{n \rightarrow \infty} E_{Q_n}(\varphi_n) = \lim_{n \rightarrow \infty} E_{Q_n}(\varphi_n^*) = 1.$$

For the proof see Section 8.

The meaning of the results will now be explained for two-sample tests.

EXAMPLE 5 (Example 1 continued).

Consider a two sample test statistic $T_n = k(n)^{1/2} \sum_{i=1}^{k(n)} c_{ni} X_{n,i}$ given by regression coefficients (2.3) with $n_1/k(n) \rightarrow \kappa \in (0, 1)$; see also Examples 1 and 6.

(a) (*Null hypothesis*). Under the null hypothesis let $(X_{n,i})_{i \leq k(n)}$ be rowwise i.i.d. with convergent partial sums $\sum_{i=1}^{k(n)} X_{n,i} \rightarrow \xi$; see (4.1), where ξ is a nontrivial infinitely divisible random variable being not of compound Poisson type; see Remark 1. According to Corollary 2 we have

$$(6.5) \quad T_n \rightarrow X_0 = \sum_{i=1}^{\infty} \psi_1(S_i) Z_i + \sum_{j=1}^{\infty} \psi_2(\tilde{S}_j) \tilde{Z}_j + \sigma Z$$

where Z_i, Z_j are i.i.d. with joint distribution $\mathcal{L}(Z_1)$ given in (3.27).

(b) (*Permutation tests*). For each n the upper T_n -permutation tests φ_n^* (1.4) are exact level α tests and their conditional quantiles $c_n^*(\alpha, X_{n,1}(\omega), \dots, X_{n,k(n)}(\omega))$ converge to the quantiles of $\mathcal{L}(X_0 | \Pi = \omega)$ given ω for all α [except for a countable number of values $\alpha \in (0, 1)$]. The power of φ_n and φ_n^* can be compared under the following local alternatives. Let $X'_{n,i}$ (i.i.d.) stand for the null hypothesis (a) and consider two-sample alternatives

$$(6.6) \quad X_{n,i} := X'_{n,i} + \frac{\theta}{c_{ni}k(n)^{3/2}}, \quad \theta \in \mathbb{R},$$

where now $T_n((X_{n,i})_i) = T_n((X'_{n,i})_i) + \theta$ holds. However, the permutation statistics (3.2) with coefficients (3.6) of T_n are asymptotically the same under the null hypothesis and the alternatives, namely

$$(6.7) \quad T_n^*((X_{n,i})_i) - T_n^*((X'_{n,i})_i) = \sum_{i=1}^{k(n)} c_{ni} \frac{\theta}{k(n)} \rightarrow 0$$

in probability (apply Hájek’s variance formula, Section 7). Thus

$$(6.8) \quad d(\mathcal{L}(T_n^* | X_{n,1}, \dots, X_{n,k(n)}), \mathcal{L}(X_0 | \Pi)) \rightarrow 0$$

also follows under alternatives (6.6). We see that φ_n is asymptotically testing

$$(6.9) \quad \mathcal{L}(X_0) \text{ against } \mathcal{L}(X_0 + \theta)$$

and the permutation tests φ_n^* are asymptotic tests for the conditional distributions

$$(6.10) \quad \mathcal{L}(X_0 | \Pi = \omega) \text{ against } \mathcal{L}(X_0 + \theta | \Pi = \omega).$$

In the normal case $X_0 = \sigma Z$ (Remark 4) the testing problems (6.9) and (6.10) are equivalent and Lemma 1 can be applied to local alternatives. In the nonnormal case φ_n and φ_n^* will not have the same asymptotic power in general. Examples of this type show up for nonstandard rank tests; see Janssen and Mason (1990), which is discussed in Example 6 below.

(c) (*Bootstrap two-sample tests*). From the previous comments it is clear that permutation tests should always be preferred for two-sample problems. From the mathematical point of view we may ask what happens when the critical values c_n^* are evaluated by other resampling statistics T_n^* , for instance by bootstrap procedures. We will indicate the steps which are different. Actually, there are two ways to organize the two-sample bootstrap. First, if $m(n) = k(n)$ holds, one may bootstrap the $X_{n,i}$ themselves and replace them by bootstrap quantities and sum them up in the two-sample manner $k(n) \sum_{i=1}^{k(n)} c_{ni}(W_{n,i} X_{n,i})$. Second, the bootstrap sample can be taken from new variables $X'_{n,i} = c_{ni} X_{n,i}$ which establishes a bootstrap partial sum. In both cases, our bootstrap statistic is $T_n^* = k(n) \sum_{i=1}^{k(n)} W_{n,i}(c_{ni} X_{n,i})$, where now an additional factor $k(n)^{1/2}$ is needed for reasons of normalization. It should be mentioned that the second procedure is not restricted to the ordinary bootstrap.

According to Theorem 4 we have again convergence (6.8) where now $\mathcal{L}(X_0 | \Pi)$ above has to be substituted by some $\mathcal{L}(\tilde{T}_0 | \Pi)$. In the normal case we typically have $X_0 \stackrel{D}{=} \tilde{T}_0$ under our regularity conditions; see Remarks 4 and 5 for the bootstrap. Thus we have again equivalence of the unconditional and conditional tests. However, it is easy to see that $X_0 \neq \tilde{T}_0$ holds in general. In comparison with (6.5), the Z ’s of \tilde{T}_0 may be of Poisson type and also ψ_1, ψ_2 may be different (given by the limit law X_0 of T_n). The bootstrap tests φ_n^* are asymptotically testing

$$(6.11) \quad \mathcal{L}(\tilde{T}_0 | \Pi = \omega) \text{ against } \mathcal{L}(\tilde{T}_0 + \theta | \Pi = \omega)$$

and it is even not clear what the asymptotic level of φ_n^* is! Also the relation of the testing problems (6.9) and (6.11) is unclear. As explained in part (b) above permutation tests are much more pleasant.

7. Simple linear permutation statistics. This section is devoted to rank type statistics which play a central role for the subsequent proofs.

In the first step our main theorem is treated for deterministic X 's and a sort of permutation weights. This essential step is of separate interest since T_n^* is a simple linear rank statistic in special cases. It is also the key to all other proofs. Given $k(n)$ let $(R_1, \dots, R_{k(n)})$ be uniformly distributed ranks, that is, a random variable with uniform distribution on the set of permutations $\mathfrak{S}_{k(n)}$ of $\{1, \dots, k(n)\}$. An additional index n concerning the ranks is suppressed throughout. Consider a sequence of simple linear permutation statistics

$$(7.1) \quad S_n = \sum_{i=1}^{k(n)} (c_{ni} - \bar{c}_n) d_n(R_i)$$

where c_{ni} are regression coefficients, $1 \leq i \leq k(n)$, with

$$(7.2) \quad \sum_{i=1}^{k(n)} (c_{ni} - \bar{c}_n)^2 = 1, \quad \bar{c}_n := \frac{1}{k(n)} \sum_{i=1}^{k(n)} c_{ni}$$

and random scores $d_n(i) : \tilde{\Omega} \rightarrow \mathbb{R}$, $1 \leq i \leq k(n)$ with

$$(7.3) \quad \frac{1}{k(n) - 1} \sum_{i=1}^{k(n)} (d_n(i) - \bar{d}_n)^2 = 1 \quad \text{on } A_n := \left\{ \sum_{i=1}^{k(n)} (d_n(i) - \bar{d}_n)^2 > 0 \right\}.$$

The c 's are here considered to be fixed and the d 's are allowed to be random variables independent of the ranks R_j . Without restrictions we may assume ordered regression coefficients

$$(7.4) \quad c_{n1} \leq c_{n2} \leq \dots \leq c_{nk(n)}.$$

For a moment let $d_{i:k(n)}$ denote the order statistics of (7.3). Then we have equality in distribution for $(d_n(R_i))_{i \leq k(n)} \stackrel{\mathcal{D}}{=} (d_{R_i:k(n)})_{i \leq k(n)}$ which is a consequence of the exchangeability of these variables. In these cases ranks and order statistics are independent. Without restriction we may thus assume that the score functions are also ordered

$$(7.5) \quad d_n(1) \leq d_n(2) \leq \dots \leq d_n(k(n)).$$

In our context the d 's are typically given by

$$(7.6) \quad d_n(i) = (k(n) - 1)^{1/2} \frac{W_{i:k(n)}}{(\sum_{i=1}^{k(n)} (W_{n,i} - \bar{W}_n)^2)^{1/2}}.$$

Conversely, if (7.3) holds then $W_{n,i} := d_n(R_i)/(k(n) - 1)^{1/2}$ defines an exchangeable scheme. Obviously, the conditions (3.4) and (3.5) are equivalent to $\tilde{P}(A_n) \rightarrow 1$ and

$$(7.7) \quad (k(n) - 1)^{-1/2} \max_{1 \leq i \leq k(n)} |d_n(i) - \bar{d}_n| \rightarrow 0 \quad \text{in } \tilde{P}\text{-probability.}$$

In our case with $Y_{n,i} = c_{ni} - \bar{c}_n$ we have equality in distribution of (3.2) and (7.1) via (7.6), namely

$$(7.8) \quad S_n \stackrel{D}{=} \left(\frac{k(n) - 1}{k(n)} \right)^{1/2} \frac{T_n^*}{(\sum_{i=1}^{k(n)} (W_{n,i} - \bar{W}_n)^2)^{1/2} (\sum_{i=1}^{k(n)} (X_{n,i} - \bar{X}_n)^2)^{1/2}}.$$

As an abbreviation let us always denote conditional expectations and conditional variances with respect to $(W_{i:k(n)})_{i \leq k(n)}$ [including $d_n(i)$] by $E(\cdot|W)$ and $\text{Var}(\cdot|W)$.

According to Hájek, Šidák and Sen (1999) we have $\text{Var}(S_n|W) = \mathbf{1}_{A_n}$. Since $E(S_n|W) = 0$ holds always convergent subsequences of S_n exist and under some regularity conditions we can classify all possible limit distributions of subsequences with respect to the distributional convergence. It is well known from Hájek’s work that the corresponding step functions φ_n introduced in (3.16) play a central role; see also Janssen (1997), Lemma 3.4, for a discussion of the central limit theorem for rank statistics. They are given by

$$(7.9) \quad \varphi_n(u) = d_n(1 + [k(n)u]) - \bar{d}_n, \quad u \in (0, 1),$$

which agrees with (3.16) in the case of (7.6).

REMARK 6. In the special case (7.1)–(7.9) the cluster points of S_n given by (7.1) are now

$$(7.10) \quad T_0 = \sum_{i=1}^{\infty} c_i Z_i + \sum_{j=1}^{\infty} \tilde{c}_j \tilde{Z}_j + Z$$

where Z is infinitely divisible and c_i (\tilde{c}_j) are cluster points of $(c_{ni} - \bar{c}_n)_n$, $(c_{n(n+1-j)} - \bar{c}_n)_n$, respectively.

As a technical tool we will frequently use the following well-known lemma of Skorohod; see Dudley [(1989), page 325] for a general version.

LEMMA 5. *Let ξ_n be a distributionally convergent sequence of random variables with values in a complete separable metric space. Then there exist almost surely convergent versions of ξ_n on a suitable probability space.*

PROOF OF THEOREM 1 FOR DETERMINISTIC $X_{n,i}$ WITH $c_{ni} = Y_{n,i}$, $\bar{c}_n = 0$. For each n consider the following triangular array of rowwise i.i.d. uniformly distributed random variables on $(0, 1)$. Let the V ’s be as in Remark 3 and let

$$(7.11) \quad (U_{n,1}, \dots, U_{n,k(n)})$$

be given by $U_{n,i} = V_i$ for $i \leq k(n)/2$ and $U_{n,k(n)} = \tilde{V}_1, \dots, U_{n,k(n)-j} = \tilde{V}_{j+1}$ otherwise. Without restriction we may assume that $(R_1, \dots, R_{k(n)})$ are the ranks of (7.11). At this stage a variance inequality of Hájek [(1961), Lemma 2.1] can be applied to the conditional variance given W . Statement (3.11) of his Theorem 3.1 applied to fixed scores $(W_{i:k(n)}(\tilde{\omega}))_i$ yields

$$(7.12) \quad \text{Var} \left(S_n - \sum_{i=1}^{k(n)} c_{ni} \varphi_n(U_{n,i}) \mid W \right) \leq 2\sqrt{2} \max_{1 \leq i \leq k(n)} \frac{|W_{n,i} - \overline{W}_n|}{(\sum_{j=1}^{k(n)} (W_{n,j} - \overline{W}_n)^2)^{1/2}} \rightarrow 0 \quad \text{in } \tilde{P}\text{-probability.}$$

Notice that condition (3.4) corresponds to Hájek’s assumption (3.3). This statement immediately implies according to Lemma 8 of the Appendix,

$$(7.13) \quad S_n - \sum_{i=1}^{k(n)} c_{ni} \varphi_n(U_{n,i}) \rightarrow 0$$

in probability. Thus the problem is reduced to weighted partial sums of the independent random variables. Consider now the three sequences (φ_n) in $(S, \|\cdot\|_1)$, $(c_{n1}, c_{n2}, \dots, c_{nk(n)}, 0, \dots)$ and $(c_{nk(n)}, c_{nk(n)-1}, \dots, c_{n1}, 0, \dots)$ in $[-1, 1]^{\mathbb{N}}$.

According to Lemma 2 we can choose a subsequence $\{m\} \subset \mathbb{N}$ so that each of them converges along the subsequence in $(S, \|\cdot\|_1)$, $[-1, 1]^{\mathbb{N}}$, respectively. Hence there exist $c_i, \tilde{c}_j \in [-1, 1]$ with

$$(7.14) \quad c_{mi} \rightarrow c_i, \quad c_{mk(m)+1-j} \rightarrow \tilde{c}_j \quad \text{for each } i, j$$

as $m \rightarrow \infty$ and our norming condition (7.2) implies (3.12). There exists a random variable $\varphi: \tilde{\Omega} \rightarrow S$ such that $\varphi_m \rightarrow \varphi$ holds in distribution. By Skorohod’s lemma we can find another probability space such that $\varphi_m \rightarrow \varphi$ converges almost everywhere in S . This fact implies that for these versions

$$\|\varphi_m - \varphi\| := \int \int |\varphi_m(\tilde{\omega}, x) - \varphi(\tilde{\omega}, x)| dx d\tilde{P}(\tilde{\omega}) \rightarrow 0$$

holds as $m \rightarrow \infty$. This norm obviously coincides with the L_1 -norm for the random variables (3.17) denoted by $\varphi(U)$ throughout. By the convergence of the three subsequences above we can find a sequence $r_m \uparrow \infty, r_m < k(m)/2$, with

$$(7.15) \quad a_m := \sum_{i=1}^{r_m} |c_{mi} - c_i| + r_m \|\varphi_m - \varphi\| \rightarrow 0.$$

Since $\|\varphi_m\| \leq 1, \|\varphi\| \leq 1$ hold we have

$$\begin{aligned}
 (7.16) \quad & \left\| \sum_{i=1}^{r_m} c_{mi} \varphi_m(U_{m,i}) - \sum_{i=1}^{r_m} c_i \varphi(V_i) \right\| \\
 & \leq \sum_{i=1}^{r_m} \|\varphi_m(U_{m,i})\| |c_{mi} - c_i| \\
 & \quad + \sum_{i=1}^{r_m} |c_i| \|\varphi_m(U_{m,i}) - \varphi(V_i)\| \leq a_m \rightarrow 0.
 \end{aligned}$$

For these reasons the lower part $\sum_{i=1}^{r_m} c_{mi} \varphi_m(U_{m,i})$ converges to the first series of (3.14) in distribution; see also (7.10). The conditional L_2 -convergence of that series T_0 is obvious; cf. Remark 3. Similarly, we find a sequence $s_m \uparrow \infty, s_m < k(m)/2$ with

$$(7.17) \quad \sum_{i=k(m)-s_m+1}^{k(m)} c_{mi} \varphi_m(U_{m,i}) \rightarrow \sum_{j=1}^{\infty} \tilde{c}_j \varphi(\tilde{V}_j)$$

in distribution. The remaining middle part

$$(7.18) \quad \sum_{i=r_m+1}^{k(m)-s_m} c_{mi} \varphi_m(U_{m,i})$$

is tight since (7.18) is centered and its variance is bounded by one. The corresponding array is infinitesimal since

$$(7.19) \quad \|c_{mi} \varphi_m(U_{m,i})\| \leq \max\{|c_{mi}| : r_m + 1 \leq i \leq k(m) - s_m\} \rightarrow 0.$$

This follows as in the proof of Lemma 6 from condition (7.4). Thus (7.18) has a cluster point Z which is infinitely divisible. Since the second moment of (7.18) is bounded by 1 then Fatou’s lemma implies $E(Z^2) \leq 1$. \square

PROOF OF THEOREM 2 FOR DETERMINISTIC $Y_{n,i} = c_{ni}$. (b) Again the choice of r_m and s_m given in the proof above by (7.15) and (7.17) is appropriate. Notice that these conditions together with

$$\left| \sum_{i=1}^{r_m} (c_{mi}^2 - c_i^2) \right| \leq \sum_{i=1}^{r_m} |c_{mi} - c_i| |c_{mi} + c_i| \leq 2 \sum_{i=1}^{r_m} |c_{mi} - c_i| \rightarrow 0$$

imply

$$(7.20) \quad \sum_{i=r_m+1}^{k(m)-s_m} c_{mi}^2 \rightarrow \sigma^2 := 1 - \sum_{i=1}^{\infty} c_i^2 - \sum_{j=1}^{\infty} \tilde{c}_j^2.$$

Now the central limit theorem for triangular arrays can be applied to the middle part (7.18) since (7.19) together with the L_2 -convergence of the sequence φ_m implies Lindeberg’s condition; see Lemma 3 of Section 3.

(a) Again we may restrict ourselves to the middle part (7.18) which is assumed to be convergent along the subsequence $\{m\}$. Recall from Gnedenko and Kolmogorov [(1954), page 116] that the limit variance of the normal part is given by

$$(7.21) \quad \tilde{\sigma}^2 = \lim_{\varepsilon \rightarrow 0} \limsup_{m \rightarrow \infty} \sum_{i=r_m+1}^{k(m)-s_m} \text{Var}(c_{mi} \varphi_m(U_{m,i}) \mathbf{1}_{(-\varepsilon, \varepsilon)}(c_{mi} \varphi_m(U_{m,i}))).$$

In a first step assume that $\varphi_m(U_{m,i})$ are symmetric random variables. Since without restriction $\tilde{\omega} \mapsto \|\varphi_m(\tilde{\omega}, \cdot) - \varphi(\tilde{\omega}, \cdot)\|_1$ converges in \tilde{P} -probability to 0 on a suitable probability space there exist for each $\delta > 0$ some $K > 0$ and m_0 with

$$(7.22) \quad \begin{aligned} &\text{Var}(\varphi_m(U_{m,1}) \mathbf{1}_{(-K, K)}(\varphi_m(U_{m,1}))) \\ &\geq \int \int \varphi(\tilde{\omega}, x)^2 dx d\tilde{P}(\tilde{\omega}) - \delta =: \beta(\delta) \end{aligned}$$

for all $m \geq m_0$.

It is easy to see that for any random variable X with $E(X) = 0$ we have $\text{Var}(X) \geq \text{Var}(X \mathbf{1}_A(X))$ for all measurable sets $A \subset \mathbb{R}$.

On the other hand for each i , $r(m) + 1 \leq i \leq k(m) - s_m$, finally

$$(7.23) \quad |\varepsilon/c_{mi}| > K$$

holds according to (7.19) whenever $c_{mi} \neq 0$. Thus finally

$$(7.24) \quad \begin{aligned} &\text{Var}(c_{mi} \varphi_m(U_{m,i}) \mathbf{1}_{(-\varepsilon, \varepsilon)}(c_{mi} \varphi_m(U_{m,i}))) \\ &\geq c_{mi}^2 \text{Var}(\varphi_m(U_{m,i}) \mathbf{1}_{(-K, K)}(\varphi_m(U_{m,i}))) \geq c_{mi}^2 \beta(\delta) \end{aligned}$$

holds. The result now follows from (7.20), (7.21) and

$$(7.25) \quad \sum_{i=r_m+1}^{k(m)-s_m} c_{mi}^2 \beta(\delta) \rightarrow \sigma^2 \beta(\delta).$$

The proof of the general case is similar. Choose independent copies $\tilde{\varphi}_m(\tilde{U}_{m,i})$ and consider $\varphi_m(U_{m,i}) - \tilde{\varphi}_m(\tilde{U}_{m,i})$. Here the bound of the normal variance can be established as above. \square

We will now discuss applications for rank statistics. It is sometimes convenient to interchange the role of regression coefficients and scores; see (7.26) below and Janssen (1997). That approach can be applied to the two-sample problem, Example 1(a); see (2.2) for the notation of antiranks D_i .

EXAMPLE 6 (*Two-sample rank statistics*). Consider a two-sample problem of size n_1 and n_2 with $n_1 + n_2 = k(n)$ and deterministic scores; see Example 1. Let

$$(7.26) \quad S_n = \sum_{i=1}^{k(n)} b_{nD_i} a_n(i)$$

be a two-sample statistic with real scores $a_n(i)$, $i \leq k(n)$. Let b_{ni} be the regression coefficients given by (2.3) which are centered and for which (7.2) holds. If we now set

$$(7.27) \quad d_n(i) = (k(n) - 1)^{1/2} b_{ni} \quad \text{and} \quad c_{ni} = (k(n) - 1)^{-1/2} a_n(i),$$

the new scores $d_n(i)$ are of the form (7.3) and (7.5) and the meaning of Theorem 1 can be discussed; see also Remark 6. Suppose that the c 's of (7.27) satisfy the required conditions (7.2) and (7.4) with convergent schemes $c_{ni} - \bar{c}_n \rightarrow c_i$ and $c_{nk(n)+1-j} - \bar{c}_n \rightarrow \tilde{c}_j$ as $n \rightarrow \infty$ for all $i, j \in \mathbb{N}$; see (3.11). Check that the scheme $d_n(D_i)/(k(n) - 1)^{1/2}$ fulfills (3.3)–(3.5) whenever $\min(n_1, n_2) \rightarrow \infty$. Under these assumptions the cluster points of S_n are given by T_0 (7.10).

(a) Suppose that $n_1/k(n) \rightarrow \kappa \in (0, 1)$ holds for $n \rightarrow \infty$. Then $S_n \rightarrow T_0$ converges in distribution where Z is a centered normal random variable with variance σ^2 given by (7.20). In this case obviously φ_n is $L_2(0, 1)$ -convergent with limit

$$(7.28) \quad \varphi(u) = -\left(\frac{1 - \kappa}{\kappa}\right)^{1/2} \mathbf{1}_{[0, \kappa]}(u) + \left(\frac{\kappa}{1 - \kappa}\right)^{1/2} \mathbf{1}_{(\kappa, 1)}(u).$$

Thus we have convergence of S_n .

However, it is known from the nonregular two-sample Weibull location model that $Z = 0$ and $\sigma^2 = 0$ may occur and that T_0 is given by the remaining sum; see Janssen and Mason [(1990), Theorem 5.4, Example 5.7, Corollary 6.5 and Example 6.6].

(b) If $n_1/k(n)$ is bounded away from 0 and 1 for $n \rightarrow \infty$ the condition

$$(7.29) \quad \max_{1 \leq i \leq k(n)} |a_n(i) - \bar{a}_n| / \left(\sum_{i=1}^{k(n)} (a_n(i) - \bar{a}_n)^2 \right)^{1/2} \rightarrow 0$$

implies asymptotic normality $S_n \rightarrow T_0 = Z$. To see this assume first that $n_1/k(n)$ is convergent. Thus S_n has only standard normal cluster points.

(c) In the remaining case $\min(n_1, n_2) \rightarrow \infty$ and $n_1/k(n) \rightarrow 0$ the limit $L_2(0, 1)$ -function vanishes, $\varphi = 0$. Thus all cluster points $T_0 = Z$ are infinitely divisible and examples are known when T_0 is normal or when T_0 has nontrivial Poisson parts. Its behaviour depends on regularity conditions concerning the a 's.

8. Proofs.

PROOF OF LEMMA 1. It is well known that condition (b) implies (a); see Witting and Nölle (1970), page 58.

Suppose now that (a) holds for all $\alpha \in (0, 1)$. Let $c(\alpha) = F_T^{-1}(1 - \alpha)$ be the $(1 - \alpha)$ -quantile of the distribution function of T . Then it is easy to see that

$$(8.1) \quad E(|\varphi_{n, \alpha} - \mathbf{1}_{(c(\alpha), \infty)}(T_n)|) \rightarrow 0$$

holds and we may choose $c_n(\alpha) = c(\alpha)$. Let now F_n denote the distribution function of T_n and let U be a uniformly distributed random variable on $(0, 1)$ which is independent of the data. Thus condition (a) implies

$$\begin{aligned} & E(|F_n(c(U)) - F_n(c_n^*(U))|) \\ & \leq P(\{T_n \in (c(U), c_n^*(U)]\} \cup \{T_n \in (c_n^*(U), c(U)]\}) \\ & \leq \int_0^1 E(|\varphi_{n,\alpha} - \varphi_{n,\alpha}^*|) d\alpha + 2 \sup_{x \in \mathbb{R}} P(\{T_n = x\}) \rightarrow 0. \end{aligned}$$

For each subsequence there exists a further subsequence $\{m\}$ such that

$$(8.2) \quad F_m(c(U)) - F_m(c_m^*(U)) \rightarrow 0 \quad P \text{ a.e.}$$

The present assumptions imply $F_m(x) \rightarrow F_T(x)$ and $F_m^{-1}(u) \rightarrow F_T^{-1}(u)$ for all $x \in \mathbb{R}, u \in (0, 1)$, respectively. Thus

$$(8.3) \quad c_m^*(U) \rightarrow c(U)$$

holds P a.e. Observe now that $c(U) \stackrel{\mathcal{D}}{=} T$ and $c_n^*(U)$ has conditional distribution $\mathcal{L}(T_n^* | X_{n,1}, \dots, X_{n,k(n)})$ given the X 's. Altogether we have convergence

$$(8.4) \quad d(\mathcal{L}(T), \mathcal{L}(T_m^* | X_{m,1}, \dots, X_{m,k(m)})) \rightarrow 0$$

P a.e. along the present subsequence $\{m\}$. This statement implies the convergence of (b) in probability. \square

The key for our investigations is the following compactness lemma.

LEMMA 6. *The space $(S, \|\cdot\|_1)$, see (3.15), is a compact metric space.*

PROOF. Let φ_n be a sequence in S . Similarly to Janssen [(1997), Lemma 3.4] notice that the inequality

$$(8.5) \quad |\varphi_n(u)| \leq \max(u^{-1/2}, (1-u)^{-1/2}) =: \varphi_0(u), \quad 0 < u < 1,$$

holds, where $\varphi_0 \in L_1(0, 1)$ dominates the whole sequence. For fixed u the sequence $(\varphi_n(u))_n$ is bounded. Thus there exists a subsequence $\{m\} \subset \mathbb{N}$ so that $(\varphi_m(u))_m$ is convergent for all $u \in \mathbb{Q} \cap (0, 1)$. Since φ_m is nondecreasing we have

$$(8.6) \quad \varphi_m(u) \rightarrow \varphi(u)$$

for some function $\varphi : (0, 1) \rightarrow \mathbb{R}$ for almost every $u \in (0, 1)$. The dominated convergence theorem now implies $L_1(0, 1)$ -convergence of (8.6). \square

PROOF OF THEOREMS 1 AND 2 IN GENERAL FORM. Part (a) of Theorem 1 is a consequence of common tightness arguments.

(b) This proof extends the proof of Section 7 given for fixed X 's. Consider again the ranks $(R_i)_{i \leq k(n)} : (\Omega', \mathcal{A}', P') \rightarrow \mathcal{S}_{k(n)}$ of (7.11) jointly independent of the X - and W -variables. By (3.3)–(3.5), we have equality in distribution of

$$(8.7) \quad T_n^* \stackrel{\mathcal{D}}{=} k(n)^{1/2} \sum_{i=1}^{k(n)} W_{R_i:k(n)}(X_{n,i} - \bar{X}_n) =: \tilde{T}_n^*$$

on $\Omega \times \tilde{\Omega} \times \Omega'$. Since it is enough to prove distributional convergence of (3.13) under P , by Lemma 5 we may replace the jointly distributionally convergent order statistics (3.11) and indicators $\mathbf{1}_{H_n}$, see (3.1), by P almost surely convergent versions. Here we apply Lemma 5 to the compact space $([-1, 1]^{\mathbb{N}})^2 \times \{0, 1\}$.

On the new probability space we have by the dominated convergence theorem $L_1(P)$ -convergence of the components of (3.11), namely

$$(8.8) \quad \|Y_{i:k(m)} - \zeta_i\|_1 \rightarrow 0, \quad \|Y_{k(m)+1-j:k(m)} - \tilde{\zeta}_j\|_1 \rightarrow 0$$

along the subsequence for all $i, j \in \mathbb{N}$. It is easy to see that there exist sequences $r_m \uparrow \infty, s_m \uparrow \infty$ in \mathbb{N} with

$$(8.9) \quad \sum_{i=1}^{r_m} |Y_{i:k(m)} - \zeta_i| \rightarrow 0 \quad \text{and} \quad \sum_{j=1}^{s_m} |Y_{k(m)+1-j:k(m)} - \tilde{\zeta}_j| \rightarrow 0$$

both in $L_1(P)$. Turning again to subsequences we may assume that in addition to (8.8), the sequences (3.11) and (8.9) are P a.e. convergent, say for $\omega \in M, P(M) = 1$. By our construction we have $\mathbf{1}_{H_m} \rightarrow V$ a.e. where V is a random variable with values 0 or 1.

At this stage we have prepared all ingredients to adapt the proof of Section 7 for deterministic $X_{n,i}$ where now the choice $c_{ni} = Y_{i:k(n)}(\omega), \omega \in M \cap \{V = 0\}$, given by (3.10), is appropriate. If ω is fixed then

$$(8.10) \quad \tilde{\omega} \mapsto \frac{\tilde{T}_m^*}{(\sum_{i=1}^{k(m)} (X_{m,i}(\omega) - \bar{X}_m(\omega))^2)^{1/2}} = k(m)^{1/2} \sum_{i=1}^{k(m)} c_{mi} W_{R_i:k(m)}$$

is a permutation statistic with random scores. For fixed ω the proof of the deterministic case can be repeated, see Section 7. Thus we have convergence of the r_m lower and the s_m upper parts of (8.10) along a subsequence. These subsequences are independent of $\omega \in M \cap \{V = 0\}$. On the other hand, the central part (7.18) defines a tight sequence

$$(8.11) \quad \omega \mapsto \mathcal{L} \left(\sum_{i=r_m+1}^{k(m)-s_m} c_{mi} \varphi_m(U_{m,i}) \mid X_{m,1}, \dots, X_{m,k(m)} \right)$$

on Ω into $(\mathcal{M}_1(\mathbb{R}), d)$. Recall that the set of centered distributions with bound 1 for its second moments is relatively compact in $\mathcal{M}_1(\mathbb{R})$. For these reasons we may assume that (8.11) is P a.e. convergent along a further subsequence (also denoted

by $\{m\}$). For fixed ω the sequence is infinitesimal and the limit variables $Z^{(\omega)}$ are infinitely divisible.

For $\omega \in M \cap \{V = 1\}$ we may define $Z^{(\omega)} = 0$. On the new probability space we have $\mathbf{1}_{H_m}(\omega) = 1$ and $\tilde{T}_m^*(\omega) = 0$ finally. Since also $\zeta_i(\omega) = \tilde{\zeta}_j(\omega) = 0$ holds we have $T_0(\omega) = 0$. The same arguments imply Theorem 2 in its general form. As pointed out above $c_{ni} = Y_{ik(n)}(\omega)$ may again be fixed and the proof of Section 7 works. \square

PROOF OF LEMMA 3. Consider almost everywhere convergent versions of φ_m on $(S, \|\cdot\|_1)$ for a suitable probability space with $\|\varphi_m(\tilde{\omega}, \cdot) - \varphi(\tilde{\omega}, \cdot)\|_1 \rightarrow 0$ \tilde{P} a.e. Clearly we have $E(\varphi(U)) = E(\varphi_m(U)) = 0$ and $\|\varphi_m(\tilde{\omega}, \cdot)\|_2 \rightarrow 1$ in \tilde{P} -probability. If $E(\varphi(U)^2) = 1$ holds then obviously $\|\varphi(\tilde{\omega}, \cdot)\|_2 = 1$ a.e. follows and in conclusion we have $\|\varphi_m(\tilde{\omega}, \cdot)\|_2 \rightarrow \|\varphi(\tilde{\omega}, \cdot)\|_2$ in \tilde{P} -probability. Via a.e. convergent subsequences we may apply Vitali's theorem for fixed $\tilde{\omega}$; see Shorack [(2000), page 55] which yields

$$\|\varphi_m(\tilde{\omega}, \cdot) - \varphi(\tilde{\omega}, \cdot)\|_2 \rightarrow 0$$

in \tilde{P} -probability and distributional convergence in $(S, \|\cdot\|_2)$.

Conversely, we may assume that we have $\varphi_m \rightarrow \varphi$ a.e. in $(S, \|\cdot\|_2)$. Then again Vitali's theorem implies $\|\varphi(\omega, \cdot)\|_2 = 1$ a.e. and $E(\varphi(U)^2) = 1$ holds. \square

PROOF OF THEOREM 3. Suppose first that $(P(H_n))_n$ is convergent. Again we may assume that (3.23) and $\mathbf{1}_{H_n}$ are almost surely convergent on another probability space. Thus we may divide by $(\sum_{i=1}^{k(n)} (X_{n,i} - \bar{X}_n)^2)^{1/2} \mathbf{1}_{H_n^c}$ on the set $\{\xi_0 > 0\}$ and the result follows from Theorem 2(b) on $\{\xi_0 > 0\}$ (where $H_n^c = \Omega \setminus H_n$).

On the complement $\{\xi_0 = 0\}$ the limit variable \tilde{T}_0 vanishes. It is now easy to prove that the conditional distributions (3.24) converge to ε_0 under the restriction $\xi_0 = 0$. For this purpose define similarly to (8.10) regression coefficients $c_{ni} = X_{i:k(n)}(\omega) - \bar{X}_n(\omega)$ for fixed ω . The condition $\xi_0(\omega) = 0$ together with almost sure convergence of (3.23) now implies $\sum_{i=1}^{k(n)} c_{ni}^2 \rightarrow 0$ and

$$\text{Var} \left(k(n)^{1/2} \sum_{i=1}^{k(n)} c_{ni} W_{R_i:k(n)} \right) \rightarrow 0.$$

If $(P(H_n))_n$ is not convergent we may turn to convergent subsequences and proceed as above. However, the limit variable \tilde{T}_0 is the same for all subsequences. This implies the desired result (3.24). \square

LEMMA 7. Let $(X_{n,i})_{i \leq k(n)}$ as in Section 4 be such that the conditions (4.1), (4.2) and (4.8) hold. Taking (4.3)–(4.7) into account we have distributional

convergence on $\mathbb{R}^{\mathbb{N}} \times \mathbb{R}^{\mathbb{N}} \times \mathbb{R}$ of the following random variables:

$$(8.12) \quad \left((X_{i:k(n)})_{i \in \mathbb{N}}, (X_{k(n)+1-j:k(n)})_{j \in \mathbb{N}}, \sum_{k=1}^{k(n)} X_{n,k}^2 \right) \longrightarrow \left((\psi_1(S_i))_{i \in \mathbb{N}}, (\psi_2(\tilde{S}_j))_{j \in \mathbb{N}}, \sum_{i=1}^{\infty} \psi_1(S_i)^2 + \sum_{j=1}^{\infty} \psi_2(\tilde{S}_j)^2 + \sigma^2 \right).$$

REMARK 7. (a) The lemma extends Raikov’s result, see Gnedenko and Kolmogorov (1954) and Raikov (1938), which states $\sum_{k=1}^{k(n)} X_{n,k}^2 \rightarrow \sigma^2$ whenever $\xi = N$ is normal and the $X_{n,k}$ are centered.

(b) The series

$$(8.13) \quad \sum_{i=1}^{\infty} \psi_1(S_i)^2 \quad \text{and} \quad \sum_{j=1}^{\infty} \psi_2(\tilde{S}_j)^2$$

are almost surely convergent infinitely divisible random variables with characteristic functions

$$(8.14) \quad t \mapsto \exp \left(\int_{(0,\infty)} ((\exp(iut) - 1) dv_i(u)) \right), \quad i = 1, 2,$$

given by the following Lévy measures ν_i on $(0, \infty)$. They are defined by

$$(8.15) \quad \begin{aligned} x &\mapsto \nu_1([x, \infty)) := \eta((-\infty, -\sqrt{x}]), \\ x &\mapsto \nu_2([x, \infty)) := \eta([\sqrt{x}, \infty)) \end{aligned}$$

for $x > 0$; see Janssen [(1994), Section 4] for details about convergent series.

PROOF OF LEMMA 7.

Step 1. In a first step the distributional convergence of the squares given by the third component of (8.12) is treated. Suppose first that our $X_{n,i}$ are symmetric. The proof relies on truncation arguments similar to Janssen [(1994), Section 5]. Choose $\delta > 0$ such that $\pm\delta$ are continuity points of the Lévy measure η of ξ . The joint convergence of the order statistics, see Janssen [(1994), (2.13)] implies

$$(8.16) \quad \begin{aligned} &\sum_{i=1}^{k(n)} X_{n,i}^2 \mathbf{1}_{[\delta, \infty)}(|X_{n,i}|) \\ &= \sum_{k=1}^{k(n)} X_{k:k(n)}^2 \mathbf{1}_{[\delta, \infty)}(|X_{k:k(n)}|) \\ &\rightarrow \sum_{i=1}^{\infty} \psi_1(S_i)^2 \mathbf{1}_{(-\infty, -\delta]}(\psi_1(S_i)) + \sum_{j=1}^{\infty} \psi_2(\tilde{S}_j)^2 \mathbf{1}_{[\delta, \infty)}(\psi_2(\tilde{S}_j)). \end{aligned}$$

The asymptotic independence of our upper and lower order statistic parts follows as in Janssen [(1994), Section 5, Lemma 5.3], which is referred to as the

splitting lemma. Similar arguments are well known in the literature; see Araujo and Giné (1980). For $\delta \downarrow 0$ the right-hand side is almost surely convergent, see (8.13). Thus there exists a sequence $\delta_n \downarrow 0$ with

$$(8.17) \quad \sum_{i=1}^{k(n)} X_{n,i}^2 \mathbf{1}_{[\delta_n, \infty)}(|X_{n,i}|) \rightarrow \sum_{i=1}^{\infty} \psi_1(S_i)^2 + \sum_{j=1}^{\infty} \psi_2(\tilde{S}_j)^2$$

in distribution. We may choose $\delta_n \rightarrow 0$ slow enough such that the splitting Lemma 5.3 may be applied and

$$(8.18) \quad \sum_{i=1}^{k(n)} X_{n,i} \mathbf{1}_{(-\delta_n, \delta_n)}(X_{n,i}) \rightarrow N$$

is convergent to the normal part N of ξ . Thus Raikov’s result implies

$$(8.19) \quad \sum_{i=1}^{k(n)} X_{n,i}^2 \mathbf{1}_{(-\delta_n, \delta_n)}(X_{n,i}) \rightarrow \sigma^2$$

and (8.17) combined with (8.19) establishes the results for the symmetric case.

The convergence for arbitrary sums of squares can be established as follows. Choose an i.i.d. sequence $(\varepsilon_i)_{i \in \mathbb{N}}$, $\varepsilon_i \in \{+1, -1\}$, of uniformly distributed signs which is independent of the triangular array. We claim that

$$(8.20) \quad \sum_{i=1}^{k(n)} \varepsilon_i (X_{n,i} - a_{n,i}) \rightarrow T$$

converges in distribution to a symmetric infinitely divisible random variable T with variance σ^2 for its normal part and Lévy measure $\bar{\eta}$,

$$(8.21) \quad \bar{\eta}(A) := \frac{1}{2}(\eta(A) + \eta(-A)), \quad A \subset \mathbb{R} \setminus \{0\}.$$

The proof follows from Petrov [(1995), Theorem 3.3]. Using his notation with $\bar{F}_{n,i}$ for the distribution function of $X_{n,i} - a_{n,i}$ the symmetric random variable $\varepsilon_i(X_{n,i} - a_{n,i})$ has then the distribution function

$$(8.22) \quad x \mapsto \frac{1}{2}(\bar{F}_{n,i}(x) + 1 - \bar{F}_{n,i}((-x)-)).$$

The convergence criterion for triangular arrays now implies the result. Condition (4.8) ensures

$$(8.23) \quad \sum_{i=1}^{k(n)} \varepsilon_i X_{n,i} \rightarrow T$$

and $\sum_{i=1}^{k(n)} X_{n,i}^2 \rightarrow \xi_0$ is convergent in distribution according to our first result. Thus the Lévy measure of ξ_0 is just $x \mapsto 2\bar{\eta}([\sqrt{x}, \infty))$, $x > 0$ on $(0, \infty)$. According to Remark 7 the sum of the series (8.13) has the Lévy measure $\nu_1 + \nu_2$, see (8.15).

This is just the Lévy measure of ξ_0 . These arguments complete the treatment of the third component of (8.12).

Step 2. The joint convergence of the first two components on $\mathbb{R}^{\mathbb{N}} \times \mathbb{R}^{\mathbb{N}}$ follows from Janssen (1994). The joint convergence of the triple (8.12) can be treated as follows. The splitting Lemma 5.3 of Janssen (1994) together with Raikov’s theorem implies

$$(8.24) \quad \sum_{i=1}^{k(n)} (\varepsilon_i X_{n,i})^2 \mathbf{1}_{(-\tau_n, \tau_n)}(\varepsilon_i X_{n,i}) \rightarrow \sigma^2$$

for some sequence $\tau_n \downarrow 0$. Since (8.12) is tight it remains to identify the cluster points. Let

$$(8.25) \quad ((\zeta_i)_{i \in \mathbb{N}}, (\tilde{\zeta}_j)_{j \in \mathbb{N}}, R)$$

be a cluster point of (8.12). By Lemma 5 we may change the probability space such that (8.24) and (8.25) are almost sure limits. For fixed m we have

$$(8.26) \quad \sum_{i=1}^m \zeta_i^2 + \sum_{j=1}^m \tilde{\zeta}_j^2 + \sigma^2 \leq R \quad \text{a.e.}$$

and

$$(8.27) \quad \sum_{i=1}^{\infty} \zeta_i^2 + \sum_{j=1}^{\infty} \tilde{\zeta}_j^2 + \sigma^2 \leq R \quad \text{a.e.}$$

However, we have equality in distribution of the latter random variables (8.27). It is easy to see that this already implies equality a.e. for (8.27). Thus the proof of Lemma 7 is complete. \square

REMARK 8. In case of rowwise i.i.d. random variables the change of the underlying probability space as well as the representation (3.3) have a nice interpretation in terms of order statistics and their well-known quantile representation. Almost sure convergence of the order statistics in (3.23) can be proved by using a common representation of uniform order statistics given by ratios of partial sums of exponential random variables. For details see Csörgő, Csörgő, Horváth and Mason (1986) or Janssen [(2000), (2.7)]. Observe in addition that $F_n^{-1}(u/k(n)) \rightarrow \psi_1(u)$ is convergent where F_n denotes the distribution function of $X_{n,1}$.

PROOF OF THEOREM 5. (a) Obviously, it is enough to determine the characteristic function of the series part of X_0 . For these reasons we may assume that $\sigma^2 = 0$ holds. Since $\mathcal{L}(\xi)$ is infinitely divisible we may choose a rowwise i.i.d. infinitesimal triangular scheme with $X_{n,1} + \dots + X_{n,n} \stackrel{\mathcal{D}}{=} \xi$ for each n . Theorem 1 of Gnedenko and Kolmogorov [(1954), page 116] implies that

$nE(X_{n,i} \mathbf{1}_{(-\tau, \tau)}(X_{n,i}))$ is convergent. Thus (4.8) holds. Hence Theorem 4 applies to the present wild bootstrap scheme (3.8) with $Z_{n,i} = Z_i$ and

$$(8.28) \quad \sum_{i=1}^n Z_i X_{n,i} \rightarrow X_0$$

is convergent in distribution where X_0 is given by the series (4.10). After an obvious renormalization the result also holds when $0 < \text{Var}(Z_1) < 1$ occurs. The limit distribution of (8.28) can now be identified via characteristic functions. Write (4.14) in the form $\exp(\Psi(t))$ where $\sigma = 0$ is kept in mind. Fubini's theorem implies that $\int |\Psi| d\nu$ is finite. The characteristic function of the left-hand side of (8.28) is

$$(8.29) \quad t \mapsto \left(\int \exp(\Psi(tz)/n) d\nu(z) \right)^n.$$

A Taylor expansion yields for fixed t

$$(8.30) \quad \int \exp(\Psi(tz)/n) d\nu(z) = 1 + \int \Psi(tz) d\nu(z)/n + o\left(\frac{1}{n}\right).$$

To prove this let $\Psi = \Psi_1 + i\Psi_2$ be its representation given by real functions Ψ_j with $\Psi_1 \leq 0$. Uniformly in $s > 0$ we have

$$\begin{aligned} \left| \frac{\exp(s\Psi(tz)) - 1}{s} \right| &\leq \left| \frac{\exp(s\Psi_1(tz)) - 1}{s} \right| + \left| \frac{\exp(is\Psi_2(tz)) - 1}{s} \right| \\ &\leq |\Psi_1(tz)| + |\Psi_2(tz)|. \end{aligned}$$

The dominated convergence theorem establishes (8.30) and (8.29) tends to $\exp(\int \Psi(tz) d\nu(z))$. Taking $\int x d\nu(x) = 0$ and Fubini's theorem into account we arrive at (4.16) and the Lévy measure can be identified to be η_0 . The shift b is given by

$$itb = \int \Psi(tz) d\nu(z) - \int \rho(t, x) d\eta_0(x).$$

(b) Suppose that (b)(i) or (b)(ii) is true. It is obvious that $\xi \stackrel{D}{=} X_0$ holds if $\eta = \eta_0$. Let $\bar{\eta}(A) = \eta(-A)$ be the reflection of η . Then $\eta_0 = (\eta + \bar{\eta})/2 = \eta$ easily follows since $\nu = \frac{1}{2}(\varepsilon_1 + \varepsilon_{-1})$.

Conversely, assume now that $\xi \stackrel{D}{=} X_0$ is not a centered normal variable and is not $\xi = 0$. A shifted normal random variable is not of type (4.16). Also if ξ is constant then $X_0 = 0$ follows. Thus the Lévy measure is nontrivial $\eta \neq 0$ and $\eta = \eta_0$ holds. We may assume that $1 \in \text{supp}(\eta)$ belongs to the support of η . Otherwise the whole partial sum (4.1) can be multiplied by a factor and it can be reconsidered. Recall that

$$(8.31) \quad \int x d\nu(x) = 0 \quad \text{and} \quad \int x^2 d\nu(x) \leq 1$$

hold. For convenience set $\nu_s(A) := \nu(s^{-1}A)$ for $s \neq 0$ and $a \wedge b = \min(a, b)$. Thus we have

$$(8.32) \quad \int y^2 \wedge 1 d\eta_0(y) = \int \int x^2 \wedge 1 d\nu_s(x) d\eta(s) \leq \int s^2 \wedge 1 d\eta(s).$$

Notice that $f(s) = s^2 \wedge 1 - \int x^2 \wedge 1 d\nu_s(x) \geq 0$ is a continuous function with $\int f d\eta = 0$ if $\eta = \eta_0$. Since 1 belongs to the support of η we have $f(1) = 0$ and

$$(8.33) \quad \int x^2 \wedge 1 d\nu(x) = 1.$$

Together with (8.31) we see that ν must be supported by $[-1, 1]$ with $\int x^2 \times d\nu(x) = 1$. Hence $x^2 = 1 \nu$ a.e. follows. Since the first moment of ν vanishes it must be symmetric and ν has the desired form. Consequently, $\xi \stackrel{\mathcal{D}}{=} X_0$ is symmetric. \square

PROOF OF LEMMA 4. (a) This result is well known. Check that condition (3.5) holds since

$$(8.34) \quad \mathbb{E} \left(\sum_{i=1}^{k(n)} W_{n,i}^2 \right) = 1 - \frac{1}{k(n)} \rightarrow 1$$

and

$$(8.35) \quad \text{Var} \left(\sum_{i=1}^{k(n)} W_{n,i}^2 \right) = \frac{(m(n) - 1)(k(n) - 1)}{m(n)k(n)^2} \rightarrow 0.$$

Condition (3.4) can be proven as follows. For $\varepsilon > 0$,

$$\begin{aligned} P \left(\max_{1 \leq i \leq k(n)} |W_{n,i}| \geq \varepsilon \right) &\leq k(n) P \left(\left| M_{n,1} - \frac{m(n)}{k(n)} \right| \geq m(n)^{1/2} \varepsilon \right) \\ &\leq k(n) m(n)^{-2} \varepsilon^{-4} \mathbb{E} \left(\left(M_{n,1} - \frac{m(n)}{k(n)} \right)^4 \right) \end{aligned}$$

holds. Elementary computations show that the right-hand side converges to zero.

(b) It is well known that the central limit theorem, applied to the scheme W_{ni} (3.8), implies $\overline{W}_n \rightarrow 0$ and $\max_{1 \leq i \leq k(n)} |W_{ni}| \rightarrow 0$. Moreover, Raikov's theorem proves $\sum_{i=1}^{k(n)} W_{ni}^2 \rightarrow 1$ in probability; see Remark 7. Suppose now that condition (5.1) holds. Consider the quantile representation

$$(8.36) \quad Z_{n,1} = F_n^{-1}(U), \quad Z_1 = F_0^{-1}(U)$$

via inverse distribution functions and uniformly distributed random variables U . Then $F_n^{-1} \rightarrow F_0^{-1}$ holds almost everywhere and $F_n^{-1}(U) - F_0^{-1}(U) \rightarrow 0$ is convergent in $L_2(0, 1)$ by Vitali's theorem. Thus $\text{Var}(F_n^{-1}(U) - F_0^{-1}(U)) \rightarrow 0$

holds which establishes a central limit theorem for the scheme $Z_{n,i}$ given by its quantile representation (8.36). \square

PROOF OF THEOREM 6. (I)(a) The choice of the bootstrap weights (3.7) implies $\bar{W}_n = 0$. The $(S, \|\cdot\|_2)$ convergence of φ_n will be established for a special construction of the multinomial variables. Let U_1, U_2, \dots be i.i.d. uniformly distributed random variables on $(0, 1)$ and let $e_1 = (1, 0, \dots), e_2 = (0, 1, 0, \dots), \dots$ be the unit vectors in $\mathbb{R}^{k(n)}$. Then

$$(8.37) \quad M_{n,i} = \sum_{k=1}^{m(n)} \mathbf{1}_{\{e_i\}}(Z_{n,k}), \quad 1 \leq i \leq k(n),$$

has a multinomial distribution if

$$(8.38) \quad Z_{n,k} = (\mathbf{1}_{((j-1)/k(n), j/k(n)]}(U_k))_{1 \leq j \leq k(n)}.$$

The functions φ_n are much the same as

$$(8.39) \quad \tilde{\varphi}_n(u) := \frac{(k(n) - 1)^{1/2}}{m(n)^{1/2}} \left(M_{1+[k(n)u]:k(n)} - \frac{m(n)}{k(n)} \right)$$

which will turn out to be convergent. First the case $0 < c < \infty$ is treated.

Let F_n denote the distribution functions of the $\mathcal{B}(m(n), \frac{1}{k(n)})$ binomial distribution of $M_{n,1}$. For fixed real x we have $F_n(x) \rightarrow G(x)$ where G is the distribution function of a Poisson random variable with mean c . Consider now the variables

$$(8.40) \quad (U_i)_i \mapsto \mathbf{1}_{(-\infty, x]}(M_{n,1}) \quad \text{and} \quad (U_i)_i \mapsto -\mathbf{1}_{(-\infty, x]}(M_{n,k(n)})$$

which are both nondecreasing in each argument U_i . Thus Hájek's inequality, Hájek [(1968), Lemma 3.1] implies

$$(8.41) \quad \text{Cov}(\mathbf{1}_{(-\infty, x]}(M_{n,1}), -\mathbf{1}_{(-\infty, x]}(M_{n,k(n)})) \geq 0.$$

Since the $M_{n,i}$ are exchangeable we have $\text{Cov}(\mathbf{1}_{(-\infty, x]}(M_{n,i}), \mathbf{1}_{(-\infty, x]}(M_{n,j})) \leq 0$ for each pair $i < j$ and

$$(8.42) \quad \text{Var} \left(\frac{1}{k(n)} \sum_{k=1}^{k(n)} \mathbf{1}_{(-\infty, x]}(M_{n,k}) \right) \leq \frac{1}{k(n)^2} \sum_{k=1}^{k(n)} \text{Var}(\mathbf{1}_{(-\infty, x]}(M_{n,k})) \rightarrow 0.$$

This result implies first for fixed x and then uniformly in x the convergence of

$$(8.43) \quad \sup_{x \in \mathbb{R}} \left| \frac{1}{k(n)} \sum_{k=1}^{k(n)} \mathbf{1}_{(-\infty, x]}(M_{n,k}) - G(x) \right| \rightarrow 0$$

in P -probability.

Let now $u \in (0, 1)$ be fixed. For each integer $r \geq 0$ we have convergence of

$$(8.44) \quad P(M_{1+[k(n)u]:k(n)} \leq r) = P\left(\frac{1}{k(n)} \sum_{k=1}^{k(n)} \mathbf{1}_{(-\infty, r]}(M_{n,k}) \geq \frac{1 + [k(n)u]}{k(n)}\right).$$

The limit is 1 for $u > G(r)$ and 0 for $u < G(r)$. This result implies

$$(8.45) \quad M_{1+[k(n)u]:k(n)} \rightarrow G^{-1}(u)$$

in P -probability for all $u \in S(G^{-1}) = \{v : \text{continuity point of } G^{-1}\}$. Now we can proceed as in the proof of Lemmas 3 and 6 and we have pointwise convergence along suitable subsequences of

$$(8.46) \quad \tilde{\varphi}_m(u) \rightarrow c^{-1/2}(G^{-1}(u) - c) := \varphi(u), \quad u \in S(G^{-1}), P \text{ a.e.}$$

For a further subsequence (3.4) holds P a.e. and (8.46) also follows for φ_m . Since $E(\varphi(U_1)) = 0$ and $\text{Var}(\varphi(U_1)) = 1$ we have $(S, \|\cdot\|_2)$ -convergence of $\varphi_m \rightarrow \varphi$. Thus we have only one cluster point and tightness yields the result in case $0 < c < \infty$.

For $c = \infty$ the present proof can be modified as follows and we end up with the inverse normal distribution function $\varphi = \Phi^{-1}$. Recall first that a sequence ζ_n of $\mathcal{B}(m(n), \frac{1}{k(n)})$ binomial distributed random variables is now normal, that is,

$$(8.47) \quad \left(\frac{m(n)}{k(n)}\right)^{-1/2} \left(\zeta_n - \frac{m(n)}{k(n)}\right) \rightarrow Y$$

in distribution with standard normal Y . Instead of $\mathbf{1}_{(-\infty, x]}(M_{n,k})$ we will now consider

$$(8.48) \quad \mathbf{1}_{(-\infty, x]} \left(\left(\frac{k(n)}{m(n)}\right)^{-1/2} \left(M_{n,k} - \frac{m(n)}{k(n)}\right) \right).$$

Analogous results (8.41) and (8.42) imply

$$(8.49) \quad \sup_{x \in \mathbb{R}} \left| \frac{1}{k(n)} \sum_{k=1}^{k(n)} \mathbf{1}_{(-\infty, x]} \left(\left(\frac{k(n)}{m(n)}\right)^{-1/2} \left(M_{n,k} - \frac{m(n)}{k(n)}\right) \right) - \Phi(x) \right| \rightarrow 0.$$

Similarly as in (8.44), we have

$$P\left(\left(\frac{k(n)}{m(n)}\right)^{-1/2} \left(M_{1+[k(n)u]:k(n)} - \frac{m(n)}{k(n)}\right) \leq r\right) \rightarrow 1$$

if $u > \Phi(r)$ and the limit is zero in case $u < \Phi(r)$. Thus (8.45) can be substituted by

$$(8.50) \quad \left(\frac{k(n)}{m(n)}\right)^{-1/2} \left(M_{1+[k(n)u]:k(n)} - \frac{m(n)}{k(n)}\right) \rightarrow \Phi^{-1}(u)$$

and the proof for $c = \infty$ is complete.

(b) The proof for the wild bootstrap follows the device above. Using the notation of the proof of Lemma 4(b) we again have

$$(8.51) \quad \sup_{x \in \mathbb{R}} \left| \frac{1}{k(n)} \sum_{k=1}^{k(n)} \mathbf{1}_{(-\infty, x]}(Z_{n,k}) - F_0(x) \right| \rightarrow 0$$

which implies $Z_{1+[k(n)u]:k(n)} \rightarrow F_0^{-1}(u)$ except for a countable number of u 's, $u \in (0, 1)$, and $\varphi = F_0^{-1}$ follows since $\bar{Z}_n \rightarrow 0$ holds.

(II) We will show that in case $c = 0$ the limit φ vanishes. Notice first that the limit variable of (8.47) is $Y = 0$. To prove this we may apply Poisson's limit law and we may substitute ζ_n by a sequence of Poisson random variables η_n with mean $E(\eta_n) = m(n)/k(n)$. The characteristic function

$$\rho_n(t) = E\left(\exp\left(it\left(\frac{k(n)}{m(n)}\right)^{1/2}\left(\eta_n - \frac{m(n)}{k(n)}\right)\right)\right)$$

is equal to

$$\rho_n(t) = \exp\left(\frac{m(n)}{k(n)}\left[\exp\left(it\left(\frac{k(n)}{m(n)}\right)^{1/2}\right) - 1\right] - \left(\frac{m(n)}{k(n)}\right)^{1/2} it\right)$$

and $\rho_n(t) \rightarrow 1$ holds for all t . Thus (8.49) holds with Φ substituted by $F_0 = \mathbf{1}_{[0, \infty)}$ and (8.50) converges to the inverse $F_0^{-1} = 0$. Obviously, the sequence $(\varphi_n)_n$ is $(S, \|\cdot\|_1)$ -convergent but not with respect to $\|\cdot\|_2$. Consult Lemma 3. \square

PROOF OF THEOREM 7. Under tightness of $(\mathcal{L}(T_n | P^{k(n)}))_n$ the unconditional critical values c_n of φ_n (1.2) remain bounded above. By the tightness of (6.3) we may turn to subsequences and a.e. convergent versions of $\mathcal{L}(T_n | X_{n,1}, \dots, X_{n,k(n)})$ as well as for $P^{k(n)}$ and Q_n . Along subsequences then

$$(8.52) \quad \left\{ \omega : \limsup_{n \rightarrow \infty} c_n^*(\alpha, X_{n,1}(\omega), \dots, X_{n,k(n)}(\omega)) = \infty \right\}$$

is of probability 0 as well for $P^{k(n)}$ as for Q_n . Since now $T_n \rightarrow \infty$ follows we have $\mathbf{1}_{(c_n^*, \infty)}(T_n) \rightarrow 1$ in Q_n -probability first along subsequences. \square

APPENDIX

The following technical tool is frequently used. The proof is obvious and left to the reader.

LEMMA 8. *Let $R_n(S, T)$ be a sequence of real statistics given by independent random variables T and S . Then:*

(a) *Weak convergence of the conditional distributions*

$$d(\mathcal{L}(R_n(S, T) | T = \cdot), \mathcal{L}(R_0(S, T) | T = \cdot)) \rightarrow 0$$

in $\mathcal{L}(T)$ probability implies unconditional weak convergence $\mathcal{L}(R_n) \rightarrow \mathcal{L}(R_0)$ as $n \rightarrow \infty$.

(b) *Suppose that $E(R_n | T) = 0$ and $\text{Var}(R_n | T) := E(R_n^2 | T) \rightarrow 0$ hold in probability. Then $R_n \rightarrow 0$ follows in probability. If in addition $\text{Var}(R_n | T)$ is uniformly bounded then $\text{Var}(R_n) \rightarrow 0$ holds.*

REFERENCES

- ARAUJO, A. and GINÉ, E. (1980). *The Central Limit Theorem for Real and Banach Valued Random Variables*. Wiley, New York.
- ARCONES, M. and GINÉ, E. (1989). The bootstrap of the mean with arbitrary bootstrap sample size. *Ann. Inst. H. Poincaré Probab. Statist.* **25** 457–481.
- ARCONES, M. and GINÉ, E. (1991). Additions and corrections to “The bootstrap of the mean with arbitrary bootstrap sample size.” *Ann. Inst. H. Poincaré Probab. Statist.* **27** 583–595.
- ATHREYA, K. B. (1987). Bootstrap of the mean in the infinite variance case. *Ann. Statist.* **15** 724–731.
- BARBE, P. and BERTAIL, P. (1995). *The Weighted Bootstrap. Lecture Notes in Statist.* **98**. Springer, New York.
- BERAN, R. (1988). Prepivoting test statistics: A bootstrap view of asymptotic refinements. *J. Amer. Statist. Assoc.* **83** 687–697.
- BICKEL, P. J., GÖTZE, F. and VAN ZWET, W. R. (1997). Resampling fewer than n observations: Gains, losses, and remedies for losses. *Statist. Sinica* **7** 1–31.
- BILLINGSLEY, P. (1971). *Weak Convergence of Measures: Applications in Probability*. SIAM, Philadelphia.
- CHERNOFF, H. and TEICHER, H. (1958). A central limit theorem for sums of interchangeable random variables. *Ann. Math. Statist.* **29** 118–130.
- CSÖRGŐ, M., CSÖRGŐ, S., HORVÁTH, L. and MASON, D. M. (1986). Normal and stable convergence of integral functions of the empirical distribution function. *Ann. Probab.* **14** 86–118.
- CSÖRGŐ, S., HÄUSLER, E. and MASON, D. M. (1988). A probabilistic approach to the asymptotic distribution of sums of independent, identically distributed random variables. *Adv. in Appl. Math.* **9** 259–333.
- CSÖRGŐ, S. and MASON, D. M. (1989). Bootstrapping empirical functions. *Ann. Statist.* **17** 1447–1471.
- CUESTA-ALBERTOS, J. A. and MATRÁN, C. (1998). The asymptotic distribution of the bootstrap sample mean of an infinitesimal array. *Ann. Inst. H. Poincaré Probab. Statist.* **34** 23–48.
- DEHEUELS, P., MASON, D. M. and SHORACK, G. R. (1993). Some results on the influence of extremes on the bootstrap. *Ann. Inst. H. Poincaré Probab. Statist.* **29** 83–103.
- DEL BARRIO, E. and MATRÁN, C. (2000). The weighted bootstrap mean for heavy-tailed distributions. *J. Theoret. Probab.* **13** 547–569.
- DEL BARRIO, E., MATRÁN, C. and CUESTA-ALBERTOS, J. A. (1999). Necessary conditions for the bootstrap of the mean of a triangular array. *Ann. Inst. H. Poincaré Probab. Statist.* **35** 371–386.
- DUDLEY, R. M. (1989). *Real Analysis and Probability*. Wadsworth, Pacific Grove, CA.
- EINMAHL, U. and MASON, D. M. (1992). Approximations to permutation and exchangeable processes. *J. Theoret. Probab.* **5** 101–126.

- GINÉ, E. and ZINN, J. (1989). Necessary conditions for the bootstrap of the mean. *Ann. Statist.* **17** 684–691.
- GNEDENKO, B. V. and KOLMOGOROV, A. N. (1954). *Limit Distributions for Sums of Independent Random Variables*. Addison-Wesley, Cambridge, MA.
- HÁJEK, J. (1961). Some extensions of the Wald–Wolfowitz–Noether theorem. *Ann. Math. Statist.* **32** 506–523.
- HÁJEK, J. (1968). Asymptotic normality of simple linear rank statistics under alternatives. *Ann. Math. Statist.* **39** 325–346.
- HÁJEK, J., ŠIDÁK, Z. and SEN, P. K. (1999). *Theory of Rank Tests*, 2nd ed. Academic Press, Orlando, FL.
- HALL, P. (1990). Asymptotic properties of the bootstrap for heavy-tailed distributions. *Ann. Probab.* **18** 1342–1360.
- HALL, P. (1992). *The Bootstrap and Edgeworth Expansion*. Springer, New York.
- HARTMAN, P. and WINTNER, A. (1942). On the infinitesimal generators of integral convolutions. *Amer. J. Math.* **64** 273–298.
- JANSSEN, A. (1989). Local asymptotic normality for randomly censored models with applications to rank tests. *Statist. Neerlandica* **43** 109–125.
- JANSSEN, A. (1991). Conditional rank tests for randomly censored data. *Ann. Statist.* **19** 1434–1456.
- JANSSEN, A. (1994). Sums of independent triangular arrays and extreme order statistics. *Ann. Probab.* **22** 1766–1793.
- JANSSEN, A. (1997). Studentized permutation tests for non-i.i.d. hypotheses and the generalized Behrens–Fisher problem. *Statist. Probab. Lett.* **36** 9–21.
- JANSSEN, A. (1999). Nonparametric symmetry tests for statistical functionals. *Math. Methods Statist.* **8** 320–343.
- JANSSEN, A. (2000). Invariance principles for sums of extreme sequential order statistics attracted to Lévy processes. *Stochastic Process. Appl.* **85** 255–277.
- JANSSEN, A. and MASON, D. M. (1990). *Non-standard Rank Tests. Lecture Notes in Statist.* **65**. Springer, New York.
- KINATEDER, J. (1992). An invariance principle applicable to the bootstrap. In *Exploring the Limits of Bootstrap* (R. LePage and L. Billard, eds.) 157–181. Wiley, New York.
- KNIGHT, K. (1989). On the bootstrap of the sample mean in the infinite variance case. *Ann. Statist.* **17** 1168–1175.
- LEPAGE, R., WOODROOFE, M. and ZINN, J. (1981). Convergence to a stable distribution via order statistics. *Ann. Probab.* **9** 624–632.
- MAMMEN, E. (1992a). *When Does Bootstrap Work? Asymptotic Results and Simulations. Lecture Notes in Statist.* **77**. Springer, New York.
- MAMMEN, E. (1992b). Bootstrap, wild bootstrap, and asymptotic normality. *Probab. Theory Related Fields* **93** 439–455.
- MASON, D. M. and NEWTON, M. A. (1992). A rank statistics approach to the consistency of a general bootstrap. *Ann. Statist.* **20** 1611–1624.
- MASON, D. M. and SHAO, Q.-M. (2001). Bootstrapping the Student t -statistic. *Ann. Probab.* **29** 1435–1450.
- NEUHAUS, G. (1988). Asymptotically optimal rank tests for the two-sample problem with randomly censored data. *Comm. Statist. Theory Methods* **17** 2037–2058.
- NEUHAUS, G. (1993). Conditional rank tests for the two-sample problem under random censorship. *Ann. Statist.* **21** 1760–1779.
- PETROV, V. V. (1995). *Limit Theorems of Probability Theory*. Clarendon Press, Oxford.
- RAIKOV, D. A. (1938). On a connection between the central limit law of the theory of probability and the law of large numbers. *Bull. Acad. Sci. URSS Ser. Math.* **2** 323–338.

- ROMANO, J. P. (1989). Bootstrap and randomization tests of some nonparametric hypotheses. *Ann. Statist.* **17** 141–159.
- SHORACK, G. R. (1996). Linear rank statistics, finite sampling, permutation tests and winsorizing. *Ann. Statist.* **24** 1371–1385.
- SHORACK, G. R. (2000). *Probability for Statisticians*. Springer, New York.
- STRASSER, H. and WEBER, C. (1999). The asymptotic theory of permutation statistics. *Math. Methods Statist.* **8** 220–250.
- TUCKER, H. G. (1975). The supports of infinitely divisible distribution functions. *Proc. Amer. Math. Soc.* **49** 436–440.
- VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes. With Applications to Statistics*. Springer, New York.
- WELLNER, J. (2001). Some converse limit theorems for exchangeable bootstraps. In *State of the Art in Probability and Statistics. Festschrift for Willem R. van Zwet* (M. de Gunst et al., eds.) 593–606. IMS, Beachwood, OA.
- WITTING, H. and NÖLLE, G. (1970). *Angewandte Mathematische Statistik. Optimale finite und asymptotische Verfahren*. Teubner, Stuttgart.
- ZOLOTAREV, V. M. and KRUGLOV, V. M. (1975). The structure of infinitely divisible distributions on a bicomact Abelian group. *Theory Probab. Appl.* **20** 698–709.

MATHEMATISCHES INSTITUT
UNIVERSITÄT DÜSSELDORF
UNIVERSITÄTSSTR. 1
40225 DÜSSELDORF
GERMANY
E-MAIL: janssena@uni-duesseldorf.de
paulst@uni-duesseldorf.de