Asymptotics and effectiveness of conditional tests with applications to randomization tests.

Arnold Janssen, Markus Pauly[1]

Department of Mathematics, Heinrich-Heine Universität Düsseldorf,

Universitätsstrasse 1, 40225 Düsseldorf, Germany

Abstract

Conditional tests like two-sample permutation tests establish valid finite sample tests for large classes of nonparametric composite null hypotheses. In particular, we discuss the quality of various conditional distribution free tests in comparison to a (parametric) benchmark test. It is shown that they are often asymptotically effective. This means that conditional tests are asymptotically equivalent to benchmark tests and do not loose asymptotic power under contiguous alternatives. Moreover they turn out to be asymptotically efficient for various models. The results apply to permutation tests, symmetry tests and tests based on U-statistics or maximum likelihood estimators. A counterexample for the Bootstrap shows that the principle is not always applicable for arbitrary resampling tests.

Key words: nonparametric tests, permutation tests, two-sample tests, conditional tests, sufficient statistics, U-statistics, median tests, bootstrap tests, randomization tests

MSC: 62G09, 62G10, 62G20, 62B05

1 Motivation and introduction

Conditional tests (tests with Neyman structure) are powerful classical tools of statistics, see Lehmann and Romano (2005) for further references. By construction they are valid tests with exact nominal level α. These tests have the advantage that they are able to control the error probability of first kind at each sample size for a wide class of nonparametric composite null hypotheses. Since conditional tests are typically distribution free, they are very attractive for practical purposes. For applications we refer to the recent monographs of

[1]This article has been devoloped while the second author was a fellow of the 'Gründerstiftung for promotion of research of young academics' at the Heinrich-Heine Universität Düsseldorf.

Edgington and Onghena (2007) and Good (2005).
Conditional tests are often part of more general two step testing procedures like weighted bootstrap tests (including permutation tests), see Mason and Newton (1992) and Janssen and Pauls (2003). Typically they are based upon an external conditional resampling step which can nowadays be done by high speed computer machines.
In this paper we study the quality of conditional tests in comparison with ideal benchmark tests. As motivation we will discuss the following well known example which illustrates the methodology of our approach.

Example 1.1 (Nonparametric two sample test for the mean)
Consider independent real valued r.v. $X_1, \dots, X_{n_1}, X_{n_1+1}, \dots, X_n$ *with common distribution*

$$\mathcal{L}(X_1, \dots, X_{n_1}, X_{n_1+1}, \dots, X_n) = P^{n_1} \otimes Q^{n_2}, \tag{1.1}$$

where P *and* Q *are distributions on the real line and* $n_2 := n - n_1$ *holds. Suppose that the means* $\mu_1 = E(X_1)$ *and* $\mu_2 = E(X_n)$ *exist. The testing problem is given by*

$$H_0 : \mu_1 \leq \mu_2 \text{ versus } H_1 : \mu_1 > \mu_2.$$

Classical unconditional tests reject the null-hypothesis H_0 *whenever*

$$T_{n,lin} := \sqrt{\frac{n_1 n_2}{n}} \Big(\frac{1}{n_1} \sum_{i=1}^{n_1} X_i - \frac{1}{n_2} \sum_{i=n_1+1}^{n} X_i \Big) > c_n \tag{1.2}$$

holds for some critical value c_n*. It is well known that a proper choice of the critical value* c_n *depends on the model. This model is given by further information about the underlying distributions* P *and* Q*. We will now discuss some typically situations.*

1. *Suppose first that* $P = N(\mu_1, \sigma^2)$ *and* $Q = N(\mu_2, \sigma^2)$ *are normal with the same unknown variance* $\sigma^2 > 0$*. In this case the two-sample t-test with critical value*

$$c_{n,stud} = V_n^{1/2} t_{n-2,1-\alpha} \tag{1.3}$$

is most powerful and of exact level α*.*
Here $V_n = \frac{1}{n-2}(\sum_{i=1}^{n_1}(X_i - \frac{1}{n_1}\sum_{j=1}^{n_1} X_j)^2 + \sum_{i=n_1+1}^{n}(X_i - \frac{1}{n_2}\sum_{j=n_1+1}^{n} X_j)^2$ *and* $t_{n-2,1-\alpha}$ *denote the two-sample variance estimator and the* $(1-\alpha)$*-quantile of the* t_{n-2}*-distribution, respectively.*

2. *If the normality assumption is not realistic we will turn to the following nonparametric extension. We do not make any assumptions about the shape of the distribution of P and the normality is substituted by the shift family*

$$Q = P * \varepsilon_{\mu_2-\mu_1}. \tag{1.4}$$

Here '$*$' denotes the convolution with the Dirac measure with mass at $\mu_2 - \mu_1$. If P posses a finite second moment with non trivial variance the two-sample t-type-test given by (1.2) and (1.3)

$$\varphi_{n,stud} = \mathbf{1}_{(V_n^{1/2} t_{n-2,1-\alpha},\infty)}(T_{n,lin}) \tag{1.5}$$

(or any test with other consistent estimation $c_{n,stud}$ of the critical values) will be asymptotically exact, i.e. $E_{H_0}(\varphi_{n,stud}) \to \alpha$ holds as $n \to \infty$.

3. *However, the test (1.5) may be liberal and not of exact level α at finite sample size. In this case it is well known that the two-sample problem can be made distribution free very easily. Under the null-hypothesis of equal distributions $\mathcal{L}(X_1) = \mathcal{L}(X_2) = \ldots, \mathcal{L}(X_n)$ the order statistics $X_{1:n} \le X_{2:n} \le \cdots \le X_{n:n}$ of $X_1, \ldots, X_n$ are sufficient (and also boundedly complete whenever the model is rich enough). Conditioned under the order statistics the test is now carried out as Pitman's permutation test*

$$\varphi^*_{n,stud} = \mathbf{1}_{(c^*_{n,stud},\infty)}(T_{n,lin}) + \gamma^*_n \mathbf{1}_{\{c^*_{n,stud}\}}(T_{n,lin}). \tag{1.6}$$

Here $c^*_{n,stud}$ is the data dependent $(1-\alpha)$-quantile of the permutation distribution of $T_{n,lin}$, see Example 3.1 below. The construction is given by

$$E_{H_0}(\varphi^*_{n,stud} | X_{1:n}, \ldots, X_{n:n}) = \alpha. \tag{1.7}$$

Obviously, $\varphi^*_{n,stud}$ is of exact level $\alpha = E_{H_0}(\varphi^*_{n,stud})$ at any sample size.

It is not in advance clear that the conditional permutation test has good (asymptotic) power. Our question of interest is now,

- whether there is any price in terms of lower power that we have to pay for making a traditional test (like student's two-sample t-test $\varphi_{n,stud}$) distribution free?

To judge this question in general we will make a power comparison of traditional tests φ_n and corresponding conditional tests φ_n^*.
In the above Example 1.1 the answer is well known. Of course under normality assumptions (part 1 in the above example) Pitman's permutation test $\varphi_{n,stud}^*$ has not the power of the optimal t-test (1.5) at finite sample size. Nevertheless, as it is pointed out in Janssen (1997), Pitmann's permutation test is asymptotically as powerful as the two-sample t-type-test $\varphi_{n,stud}$. Only finite second moments are required. The reason for the asymptotic equivalence of φ_n and Pitmann's permutation test φ_n^* can be explained very easily. In Janssen (1997) and (2005) it is proved that their critical values $c_{n,stud}$ and $c_{n,stud}^*$ are asymptotically equivalent, i.e.

$$|c_{n,stud} - c_{n,stud}^*| \to 0 \text{ in probability} \tag{1.8}$$

holds under H_0 as $n \to \infty$. The same holds true for a large class of tests based on linear statistics.
The main aim of this paper is to extend these results to non-linear statistics T_n instead of $T_{n,stud}$. Important examples are test statistics like two-sample maximum likelihood estimators and certain U-statistics which are not linear but admit an expansion with leading linear term. Notice that a detailed study of conditional distributions is required in order to study the random conditional quantiles $c_{n,stud}^*$. The derivation of the asymptotics in (1.8) is by no means simple although the non-conditional asymptotic of $T_{n,lin}$ is well understood. In general, the problem has to be dealt with care, see the negative result for the Bootstrap (Example 2.3) and Example 3.7 below.

Motivated by the above example we can now describe our general setting. Typically a well motivated sequence of test statistics T_n and upper T_n-tests

$$\mathbb{1}_{(c_n,\infty)}(T_n) \le \varphi_n \le \mathbb{1}_{[c_n,\infty)}(T_n) \tag{1.9}$$

with critical values $c_n \in \mathbb{R}$ are given a priori for a certain composite null hypothesis $\mathcal{P}_{0,n}$. The critical values c_n may be of asymptotic nature such that $E_{P_n}(\varphi_n) \to \alpha$ holds under the null hypothesis (for instance $c_n = V_n^{1/2} q_{1-\alpha}$ given by a consistent variance extimation V_n and some quantile $q_{1-\alpha}$). The test φ_n will be called *benchmark test.* For instance, in the above example our benchmark test would be student's two-sample t-test. Throughout, let S_n be a sufficient statistic for the underlying null hypothesis $\mathcal{P}_{0,n}$. Then a conditional T_n-test (like Pitman's permutation test (1.6))

$$\mathbb{1}_{(c_n^*,\infty)}(T_n) \leq \varphi_n^* \leq \mathbb{1}_{[c_n^*,\infty)}(T_n) \tag{1.10}$$

of nominal level α can be established by the choice of conditional critical values $c_n^* = c_n^*(S_n)$ given by

$$E(\varphi_n^*|S_n) = \alpha \text{ under } \mathcal{P}_{0,n}. \tag{1.11}$$

The general question of interest is now whether we have to pay a price - say in loss of power - for making the benchmark test φ_n distribution free by turning to φ_n^*. Of course, at least along parametric submodels, the conditional test cannot reach the power of the benchmark at finite sample size n. However, we will see that a huge class of conditional tests reach the optimum power (given by the asymptotic power of the benchmark under contiguous alternatives) asymptotically. In this context the crucial property of conditional tests is their effectiveness. We call the sequence φ_n^* *asymptotically effective (with respect to the benchmark test* φ_n*)* if

$$E_{P_n}(|\varphi_n - \varphi_n^*|) \to 0 \tag{1.12}$$

holds for each sequence of members P_n of the null hypothesis $\mathcal{P}_{0,n}$ as $n \to \infty$. In section 2 certain conditions for asymptotic effectiveness of φ_n^* are established. Since permutation tests are the standard example for conditional tests, we will deal with them in its general form in the following example.

Example 1.2 (Permutation tests)
Let $\mathcal{P}$ *be a composite family of distributions on a measurable space* $(\Omega, \mathcal{A})$ *and consider the null-hypothesis* $\mathcal{P}_{0,n} := \{P^n : P \in \mathcal{P}\}$ *given by product measures. Let*

$$T_n : \Omega^n \to \mathbb{R},\ T_n = T_n(X_1, \ldots, X_n)$$

be a general test statistic which may depend on random variables X_i*, which are i.i.d. under the null with common distribution* $P \in \mathcal{P}$*. The upper* T_n*-test can be carried out as a permutation test* φ_n^**, given by (1.10), which may be chosen finite sample distribution free with* $E_{P^n}(\varphi_n^*) = \alpha$ *for all* $P \in \mathcal{P}$*. The critical values* c_n^* *are the data dependent* $(1-\alpha)$*-quantiles of the permutation distribution*

$$\sigma \mapsto T_n = T_n(X_{\sigma(1)}, \ldots, X_{\sigma(n)}). \tag{1.13}$$

Here $\sigma = (\sigma(1), \ldots, \sigma(n))$ is a random permutation of $1, \ldots, n$ which is independent of the data, see Example 3.1 (a) for the general form. For real valued r.v. $X_1, \ldots, X_n$ the construction of the critical values is identically to (1.7).

The conditional tests φ_n^* (like permutation tests) have the advantage that they are finite sample distribution free and of exact nominal level α. Throughout, we will study for general conditional tests

- the quality of the conditional test in the sense whether φ_n^* is asymptotically effective with respect to φ_n, i.e. (1.12) holds, or more generally we may ask
- which power of φ_n^* can be expected at all under (contiguous) alternatives.

The following chart summarizes and exemplifies the reasons for our approach. It opposes the characteristics of parametric benchmark tests φ_n and the associated conditional tests φ_n^*.

null hypothesis $\mathcal{P}_{0,n}$	parametric (e.g. normal)	nonparametric
φ_n parametrical test (e.g. two-sample t-test)	most powerful test or locally best test	may fail: $E_{P_n}(\varphi_n) > \alpha$ for some $P_n \in \mathcal{P}_{0,n}$
φ_n^* conditional test (e.g. Pitman's permutation test)	loss of power ?	distribution free $E_{P_n}(\varphi_n^*) = \alpha$ for all $P_n \in \mathcal{P}_{0,n}$

The finite sample optimality of conditional tests was earlier discussed by Lehmann and Stein (1949), Gebhard and Schmitz (1998) and Janssen and Völker (2007) for instance. The large sample properties of resampling statistics and tests (mainly permutation tests) with non-studentized test-statistics were studied by Romano (1989) and (1990), Praestgaard (1995), Strasser and Weber (1999) as well as by Janssen (1997) and (2005) for studentized versions. Earlier conditional central limit theorems (CLT) were established by Hoeffding (1952). The asymptotics in the general weighted bootstrap case (including permutation procedures) have amongst others been explored by Mason and Newton (1992), Praestgaard and Wellner (1993), Janssen and Pauls (2003) and Janssen (2005). Prepivoted bootstrap tests were discussed by Beran (1988).

The paper is organized as follows. In section 2 we study the asymptotic effectiveness of conditional tests and will give necessary conditions for (1.12) to hold, see our Lemmas 2.4 and 2.5. In this context it will be shown in Lemma 2.2 that a perturbation of the test statistic preserves the convergence (1.12). However, a counterexample for bootstrap tests shows that the principle is not adaptive for all kind of resampling tests. Applications are given in section 3. There we will see that the results hold for a wide class of conditional tests. For example, we will proof the effectiveness of permutation tests based on U-statistics or maximum likelihood estimators. Another application is given for distribution free symmetry tests. The paper closes with a final proof section 4.

2 The asymptotics of conditional tests

Permutation tests are special cases of more general conditional tests which can be described as follows. Below let $(\Omega_n, \mathcal{A}_n, \mathcal{P}_n)$ be a statistical experiment, where $\mathcal{P}_n$ is a set of probability measures on $(\Omega_n, \mathcal{A}_n)$. Let $\mathcal{P}_{0,n} \subset \mathcal{P}_n$ be a composite null-hypothesis of interest. Our testing problem is given by

$$\mathcal{P}_{0,n} \text{ versus } \mathcal{P}_n \backslash \mathcal{P}_{0,n}. \tag{2.1}$$

Throughout we will assume that

$$S_n : (\Omega_n, \mathcal{A}_n) \to (\Omega_{S_n}, \mathcal{A}_{S_n})$$

is a $\mathcal{P}_{0,n}$-sufficient statistic that takes its values in some measurable space $(\Omega_{S_n}, \mathcal{A}_{S_n})$. A test $\psi : \Omega \to [0,1]$ for (2.1) is called a conditional test (or a test with Neyman structure) iff its conditional expectation $E_{\cdot}(\psi|S_n)$ is constant under $\mathcal{P}_{0,n}$, see Lehmann and Romano (2005) for further informations. Obviously, conditional tests given by sufficient σ-fields $\mathcal{F}_n \subset \mathcal{A}_n$ fit into the present approach. In this case we may choose S_n to be the identity $S_n = id : (\Omega_n, \mathcal{A}_n) \to (\Omega_n, \mathcal{F}_n)$. More generally as in section 1 let

$$T_n : (\Omega_n, \mathcal{A}_n) \to \mathbb{R}$$

be a well motivated test statistic. The upper-T_n-test can be carried out as (S_n)-conditional test φ_n^* as follows. There exists a regular conditional distribution of T_n given $S_n = s$ that is independent from $P_n \in \mathcal{P}_{0,n}$, i.e.

$$P_{\cdot}^{T_n|S_n=s}(\cdot) = P_n^{T_n|S_n=s}(\cdot) \text{ for all } P_n \in \mathcal{P}_{0,n}, \tag{2.2}$$

see Dudley (2002, Theorem 10.2.2.). For each $\alpha \in (0,1)$ the conditional test

$$\varphi^*_{n,\alpha} = \mathbb{1}_{(c^*_n(S_n),\infty)}(T_n) + \gamma_n(S_n)\mathbb{1}_{\{c^*_n(S_n)\}}(T_n) \tag{2.3}$$

is based on the $\mathcal{A}_{S_n}$ measurable quantile function c^*_n of (2.2) at level $(1-\alpha)$ and $0 \le \gamma(S_n) \le 1$ which is a solution of $E.(\varphi^*_{n,\alpha}) = \alpha$ given by

$$\int \varphi^*_{n,\alpha} dP_.^{T_n|S_n=s} = \alpha \text{ for } s \in \Omega_{S_n}, \tag{2.4}$$

see Janssen und Völker (2007). Permutation tests (see Example 1.2) like Pitman's permutation test are a special case. There $\mathcal{A}_n = \mathcal{A}^n$ is the product σ-field on Ω^n and $\mathcal{F}_n$ is defined to be the σ-field of coordinate-wise permutation invariant sets of $\mathcal{A}_n$.

First fix a sequence of distributions $P_n \in \mathcal{P}_{0,n}$. We will make one assumption that follows Janssen and Pauls (2003, sect. 2.)

(A) Suppose that under the sequence $P_n \in \mathcal{P}_{0,n}$ we have convergence

$$P_n(T_n \le x) \longrightarrow F(x) \text{ for all } x \in \mathbb{R} \text{ as } n \to \infty, \tag{2.5}$$

where F is a continuous distribution function that is strictly increasing on its support.

For each level $\alpha \in (0,1)$ let $\varphi_{n,\alpha}$ denote as in (1.9) a sequence of unconditional benchmark tests (like the two-sample t-test for instance). We suppose that they are of asymptotic nominal level α, i.e. $E_{P_n}(\varphi_{n,\alpha}) \to \alpha$ holds as $n \to \infty$. The next Lemma is crucial for the comparison and treatment of the conditional counterpart $\varphi^*_{n,\alpha}$ of $\varphi_{n,\alpha}$ under general nonparametric null-hypotheses, for a proof see Janssen and Pauls (2003), Lemma 1.

Let $x \mapsto P_n(T_n \le x|S_n)$ denote the conditional distribution function of T_n given S_n.

Lemma 2.1

If assumption (A) holds, the following statements are equivalent

$$\sup_{x\in\mathbb{R}} \left|P_n(T_n \le x|S_n) - F(x)\right| \xrightarrow{P_n} 0, \tag{2.6}$$

$$E_{P_n}(\left|\varphi_{n,\alpha} - \varphi^*_{n,\alpha}\right|) \to 0 \textit{ for all } \alpha \in (0,1) \tag{2.7}$$

as $n \to \infty$. *Here* $\xrightarrow{P_n}$ *stands for convergence in* P_n*-probability as* $n \to \infty$.

The condition (2.6) has been verified for various linear statistics T_n and their studentized versions, see Janssen and Pauls (2003) or Janssen (2005). For instance, it works for linear two-sample statistics in case of permutation tests. Statistical consequences are sketched in section 3 below.
Various complicated test statistics $\tilde{T}_n$ can be expanded in a treatable (for instance linear) part T_n and an asymptotically negligible part R_n, i.e. $\tilde{T}_n = T_n + R_n$. Assume that $R_n : (\Omega_n, \mathcal{A}_n) \to (\mathbb{R}, \mathcal{B})$ is a statistic which converges in probability to 0, viz. $R_n \xrightarrow{P_n} 0$ for some sequence $P_n \in \mathcal{P}_{0,n}$. Now it is often the case that we know the conditional and unconditional asymptotic behavior of T_n, i.e. (2.5) and (2.6) hold. It is now the question, whether the unconditional convergence in distribution of the more complicated test statistic $\tilde{T}_n$ can be carried over to the conditional case. This is in general not the case - see Examples 2.3 and 3.7 below - but holds under sufficiency of S_n.

Lemma 2.2
Suppose that assumption (A) holds. Then condition (2.6) is equivalent to

$$\sup_{x\in\mathbb{R}} \left|P_n(T_n + R_n \le x|S_n) - F(x)\right| \xrightarrow{P_n} 0. \tag{2.8}$$

The next example shows that Lemma 2.2 does in general not hold for the bootstrap. In contrast to the permutation distribution the remainder R_n has non-negligible influence on the bootstrap.

Example 2.3
Let $X_1, \ldots, X_n$ be i.i.d. real-valued random variables with continuous distribution. Consider the associated Efron's boostrap variables $X_1^*, \ldots, X_n^*$, i.e. they are i.i.d. following the empirical measure F_n of $(X_i)_{i\le n}$. Let $(a_n)_n$ be a sequence with $a_n \to +\infty$ for $n \to \infty$. Define the set

$$A_n := \{(x_1, \ldots, x_n) \in \mathbb{R}^n : \text{ there exist } i \neq j \text{ with } x_i = x_j\}.$$

Then the statistic $R_n = R_n(X_1, \ldots, X_n) := a_n \cdot \mathbf{1}_{A_n}(X_1, \ldots, X_n)$ is P-a.e. equal to 0. However, given the data, the bootstrap version $R_n^* = R_n(X_1^*, \ldots, X_n^*)$ converges in probability to infinity:

$$\begin{aligned} P(R_n^* \ge \kappa|X_1, \ldots, X_n) &= P(\exists i \neq j : X_i^* = X_j^*|X_1, \ldots, X_n) \\ &= 1 - 1\big(1 - \frac{1}{n}\big)\big(1 - \frac{2}{n}\big) \cdots \frac{1}{n} \to 1. \end{aligned}$$

One may have the impression that conditional tests are often asymptotically effective. This is not the case as we will see in Example 3.7 below. The next two Lemmas deal with necessary conditions for (2.6) and give insight in the connection between conditional and unconditional convergence of moments.

Lemma 2.4
Suppose that assumption (A) and (2.6) hold, where T denotes a random variable with distribution function F and expectation $E(T)$. Furthermore assume that the first absolute moments exist and converge, i.e. $E_{P_n}(|T_n|) \to E(|T|) < \infty$ for $n \to \infty$. In this case we have convergence in probability of the conditional expectations to the mean of T

$$E(T_n|S_n) \xrightarrow{P_n} E(T). \tag{2.9}$$

Lemma 2.5
Let $E(T_n) = E(T) = 0$ hold and postulate that the variances exist and converge, i.e. $Var(T_n) \to Var(T) < \infty$. Then the following assumption is necessary for (2.6):

$$Var(T_n - E(T_n|S_n)) \to Var(T), \tag{2.10}$$

viz. $\int E(T_n|S_n)^2 dP_n \to 0$ holds for $n \to \infty$.

3 Applications: Effectiveness of conditional tests

In this section we show that different conditional tests φ_n^* are asymptotically effective in the sense of Definition 1.12 for various nonparametric classes of hypotheses $\mathcal{P}_{0,n}$. In particular we show the effectiveness of different permutation tests, which are a special case of more general randomization tests.

Example 3.1
Let $\mathcal{G}_n = \{g_n : (\Omega_n, \mathcal{A}_n) \to (\Omega_n, \mathcal{A}_n)\}$ be a finite group of transformations and consider a null-hypothesis

$$\mathcal{P}_{0,n} \text{ of } \mathcal{G}_n\text{-invariant distributions,}$$

i.e. $P_n^{g_n} = P_n$ for all $P_n \in \mathcal{P}_{0,n}$ and $g_n \in \mathcal{G}_n$. Let $\mathcal{F}_n := \{A \in \mathcal{A}_n : g_n(A) = A \text{ for all } g_n \in \mathcal{G}_n\}$ be the σ-field of $\mathcal{G}_n$-invariant sets and $S_n = id : (\Omega_n, \mathcal{A}_n) \to$

$(\Omega_n, \mathcal{F}_n)$ *be the identity which is sufficient for* $\mathcal{P}_{0,n}$*. The conditional (randomization) distribution* $P_{\cdot}^{T_n|S_n=\omega} = P_n^{T_n|S_n=\omega}$ *is given by the distribution of*

$$g_n \mapsto T_n(g_n(\omega)), \quad g_n \in \mathcal{G}_n, \tag{3.1}$$

on the orbits with respect to the uniform distribution on $\mathcal{G}_n$*. The conditional tests* $\varphi_{n,\alpha}^*$*, see (2.3), are then called randomization tests. Special cases are*

(a) ***Permutation tests.*** *Here* $\mathcal{P}_{0,n}$ *are (more general as in Ex. 1.2) exchangeable distributions on product spaces* $(\Omega^n, \mathcal{A}_n)$ *and* $\mathcal{G}_n$ *is the group of all permutations of the coordinates of* Ω^n*.*

(b) ***Symmetry tests.*** *In this case we contemplate* $(\Omega_n, \mathcal{A}_n) = (\mathbb{R}^n, \mathcal{B}^n)$ *and the group* $\mathcal{G}_n$ *of reflections* $(x_1, \ldots, x_n) \mapsto (\epsilon_1 x_1, \ldots, \epsilon_n x_n)$ *given by all vectors* $(\epsilon_1, \ldots, \epsilon_n) \in \{-1, 1\}^n$*. A special choice of* $\mathcal{G}_n$*-invariant distributions is then the null-hypothesis of* 0*-symmetric product measures*

$$\mathcal{P}_{0,n} = \{P^n : P \text{ is } 0\text{-symmetric on } (\mathbb{R}, \mathcal{B})\}. \tag{3.2}$$

Nonparametric testing problems are often specified in terms of statistical functionals. We therefore consider tests for a functional

$$\kappa_n : \mathcal{P}_n \to \mathbb{R}. \tag{3.3}$$

Suppose that the null-hypothesis

$$\mathcal{P}_{0,n} \subset \{P_n \in \mathcal{P}_n : \kappa_n(P_n) = 0\} \tag{3.4}$$

is tested versus $\{P_n \in \mathcal{P}_n : \kappa_n(P_n) > 0\}$ (or versus two-sided alternatives). As in Example 3.1 let $\mathcal{P}_{0,n}$ be $\mathcal{G}_n$-invariant and we may condition under the invariant σ-field $\mathcal{F}_n$ which is given by the statistic S_n, see (2.1)ff.. Suppose that the distributions of the model $\mathcal{P}_n$ are given by the vector $(X_1, \ldots, X_n)$ of random variables. Our test statistic is an estimator

$$\hat{\theta}_n := \hat{\theta}_n(X_1, \ldots, X_n) \text{ of our parameter } \kappa_n \tag{3.5}$$

and the unconditional test is of the form

$$\varphi_n = \mathbb{1}_{(c_n, \infty)}(\hat{\theta}_n). \tag{3.6}$$

Consider its conditional counterpart φ_n^*, see (1.10). Let $(P_n)_{n\in\mathbb{N}}, P_n \in \mathcal{P}_{0,n}$, denote a fixed sequence (with $\kappa_n(P_n) = 0$ by definition) such that

$$\sqrt{n}\hat{\theta}_n(X_1, \ldots, X_n) = T_n + o_{P_n}(1) \tag{3.7}$$

holds, where $o_{P_n}(1)$ is a random variable that converges in P_n-probability to 0 as $n \to \infty$. In this case it is enough to know the limit behavior of T_n. This is stated in the following corollary.

Corollary 3.2

Suppose that the above conditions (3.4)- (3.7) hold and that T_n is asymptotically normal, i.e. $P_n(T_n \le x) \longrightarrow \Phi(\frac{x}{\sigma})$ holds for all $x \in \mathbb{R}$ and $\sigma^2 > 0$, where Φ denotes the distrbution function of the standard normal distribution. If the convergence

$$\sup_{x\in\mathbb{R}} \left|P_n(T_n \le x|S_n) - \Phi(\frac{x}{\sigma})\right| \xrightarrow{P_n} 0 \textit{ in } P_n - \textit{probability} \tag{3.8}$$

holds, then the same (3.8) holds for $\sqrt{n}\hat{\theta}_n$ instead of T_n. Thus $E_{P_n}(|\varphi_n - \varphi_n^*|) \to 0$ follows and the sequence of tests φ_n^* is asymptotically effective w.r.t. φ_n.

The statistic T_n is typically a linear statistic, see (3.9) below, which is much easier to deal with and was treated in the literature for various cases.

Example 3.3 (Two-sample permutation tests)

Consider first a functional $\kappa_0 : \mathcal{P} \to \mathbb{R}, \mathcal{P} \subset \mathcal{M}_1(\Omega, \mathcal{A})$, that is of interest for a one-sample problem. Suppose further that there exists an asymptotically optimal estimator $\hat{\kappa}_n := \hat{\kappa}_n(X_1, \ldots, X_n)$ given by independent variables $X_1, \ldots, X_n$ with common distribution P for $P \in \mathcal{P}$ such that the following expansion holds, see (3.7),

$$\sqrt{n}\big(\hat{\kappa}_n - \kappa_0(P)\big) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(X_i) + o_{P_n}(1). \tag{3.9}$$

It is assumed that g is a measurable function (influence function or gradient) that satisfies $E(g(X_1)) = 0$ and $Var(g(X_1)) = \sigma^2 \in (0, \infty)$. Such expansions hold e.g. under mild conditions for maximum likelihood estimators in parametric models, see e.g. van der Vaart (1999, Theorem 5.39.). As generalization of (1.1) let $\mathcal{P}_n = \{P^{n_1} \otimes Q^{n_2} : P, Q \in \mathcal{M}_1(\mathbb{R}, \mathcal{B})\}$ be the set of distributions of the two-sample problem, where $n = n_1 + n_2$ holds. We are now interested in the *functional* $\kappa_n(P^{n_1} \otimes Q^{n_2}) := \kappa_0(P) - \kappa_0(Q)$ *which can be estimated by*

$$\hat{\theta}_n := \hat{\kappa}_{n_1}(X_1, \ldots, X_{n_1}) - \hat{\kappa}_{n_2}(X_{n_1+1}, \ldots, X_n), \tag{3.10}$$

where $X_i, i \leq n$, *denotes the canonical projection. Assume that* $min(n_1, n_2) \to \infty$ *holds. In this case we can expand the estimator under* $\mathcal{P}_{0,n} = \{P^n : P \in \mathcal{P}\}$ *with the help of (3.9) in an asymptotically linear statistic as in (3.7)*

$$\Big(\frac{n_1 n_2}{n}\Big)^{1/2} \hat{\theta}_n = \sum_{i=1}^n c_{ni} g(X_i) + o_{P^n}(1) =: T_n + o_{P^n}(1). \tag{3.11}$$

Here c_{ni} *are the regression-coefficients of the two-sample problem, i.e.*

$$c_{n,i} = \Big(\frac{n_1 n_2}{n}\Big)^{1/2} \cdot \begin{cases} \frac{1}{n_1}, & \text{for } 1 \leq i \leq n_1, \\ -\frac{1}{n_2}, & \text{for } n_1 < i \leq n. \end{cases} \tag{3.12}$$

It follows by the classical CLT and Slutzky's Lemma that

$$\sup_{x \in \mathbb{R}} \Big| P^n\Big(\sqrt{\frac{n_1 n_2}{n}}\, \hat{\theta}_n \leq x\Big) - \Phi(\frac{x}{\sigma})\Big| \longrightarrow 0$$

holds. This gives an unconditional test $\varphi_n = \varphi_n(\hat{\theta}_n)$ *of the form (3.6).* *The calculation of finite sample (or asymptotically) valid critical values for* φ_n *may be a serious problem and* φ_n *may be viewed as a hypothetical benchmark test. Consider now new observations* $Y_i := g(X_i), 1 \leq i \leq n$. *Recall that the conditional permutation distribution of a statistic* $\zeta_n((Y_i)_{i \leq n})$, *see (1.7) and (3.1), is given by conditioning under the order statistics* $Y_{1:n} \leq \cdots \leq Y_{n:n}$ *of* $(Y_i)_{i \leq n}$. *The conditional CLT of Janssen (1997, Theorem 3.1.) for linear permutation statistics proves the convergence*

$$\sup_{x \in \mathbb{R}} \Big| P^n\big(T_n \leq x | Y_{1:n}, \ldots, Y_{n:n}\big) - \Phi(\frac{x}{\sigma})\Big| \xrightarrow{P} 0 \text{ in probability.}$$

Thus the conditional permutation distributions $P_{\cdot}^{\hat{\theta}_n | (Y_{1:n}, \ldots, Y_{n:n})}$ *of (3.1) are also asymptotically normal, see Lemma 2.2 and Corollary 3.2. Hence the permutation tests* $\varphi_n^* = \varphi_n^*(\hat{\theta}_n)$ *for testing the functional* κ_n *in (3.4) are asymptotically effective w.r.t.* φ_n. *This justifies the choice of proper permutation critical values in practice.*

Below we prove the asymptotic effectiveness of further conditional tests $\varphi_n^*(\hat{\theta}_n)$ for other functionals, see (3.3) and (3.4), where now other sufficient statistics may be required. The effectiveness of these tests is again a direct consequence of Lemma 2.2.

Example 3.4 (Difference of two U-statistics test)
Let $h : \mathbb{R}^r \to \mathbb{R}, r \leq \min(n_1, n_2)$, be a symmetric kernel of a non-degenerated U-statistic and denote by $X_i : \Omega^n \to \mathbb{R}$, $(x_1, \dots, x_n) \mapsto x_i$ the canonical projections. We now set $h_1(x) := E(h(x, X_2, \dots, X_r))$ and let $\mathcal{P}_n$ be as in Example 3.3. Suppose that we have non-trivial finite second moments of $h(X_1, \dots, X_r)$ and $h_1(X_1)$ with respect to P^r and Q^r and P and Q respectively. We consider now the difference of two U-statistics

$$U_{n_1,n_2} = \frac{1}{\binom{n_1}{r}} \sum_{1 \leq i_1 < \cdots < i_r \leq n_1} h(X_{i_1}, \dots, X_{i_r}) - \frac{1}{\binom{n_2}{r}} \sum_{1 \leq j_1 < \cdots < j_r \leq n_2} h(X_{n_1+j_1}, \dots, X_{n_1+j_r}).$$

Our functional of interest is now

$$\kappa_n(P^{n_1} \otimes Q^{n_2}) = E_{P^r}(h) - E_{Q^r}(h). \tag{3.13}$$

Since the order statistic is sufficient and complete under the null-hypothesis $\{P = Q\}$ the functional κ_n can be estimated by the UMVU estimator U_{n_1,n_2}. In order to test $\mathcal{P}_{0,n} = \{P^{n_1} \otimes Q^{n_2} : P = Q\}$ we choose $\hat{\theta}_n = (\frac{n_1 n_2}{n})^{1/2} U_{n_1,n_2}$. By the projection method Hoeffding has shown that U_{n_1,n_2} has a decomposition as a weighted linear statistic as in (3.11), i.e.

$$\left(\frac{n_1 n_2}{n}\right)^{1/2} U_{n_1,n_2} = r \cdot \sum_{i=1}^{n} c_{n,i} h_1(X_i) + o_{P^n}(1)$$

holds under $\{P = Q\}$ for the regression coefficients (3.12), see e.g. Serfling (1980, p. 188f.). Hence the two-sample U_{n_1,n_2}-test for testing (3.13) can be carried out as asymptotic effective permutation test $\varphi_n^*(\hat{\theta}_n)$.

Example 3.5 (One-sample symmetry tests)
Let $\mathcal{P}_n = \{P^n : P \in \mathcal{M}_1(\mathbb{R}, \mathcal{B})\}$ be the set of product-distributions and suppose that $\mathcal{P}_{0,n} = \{P^n \in \mathcal{P}_n : P \text{ is } 0\text{-symmetric with } \kappa(P^n) = 0\}$ is, as in (3.4), a subset of the 0-symmetric product-distributions (3.2). In this case we have $S_n(x_1, \dots, x_n) = (|x_1|, \dots, |x_n|))$ as a sufficient statistic.

(a) *We now fix a sequence $P^n \in \mathcal{P}_{0,n}$ and assume that $\hat{\theta}_n$ is an asymptotically linear estimator, viz.*

$$\sqrt{n}\hat{\theta}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} h(X_i) + o_{P^n}(1) =: T_n + o_{P^n}(1) \tag{3.14}$$

holds, where h is a skew symmetric function, i.e. $h(-x) = -h(x)$, with $E(h(X_1)) = 0$ and $0 < \sigma^2 := Var(h(X_1)) < \infty$. Observe that $\hat{\theta}_n$ can *be interpreted as an estimator for a skew symmetric parameter (or functional). Under the above assumptions Janssen (1999, Theorem 4.2.) has already shown that the linear part*

$$T_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} sign(h(X_i))|h(X_i)| \stackrel{\mathcal{D}}{=} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} sign(X_i)|h(X_i)|$$

satisfies a conditional CLT

$$\sup_{x \in \mathbb{R}} \left| P^n\left(T_n \leq x \big| |h(X_1)|, \ldots, |h(X_n)|\right) - \Phi\left(\frac{x}{\sigma}\right) \right| \xrightarrow{P} 0.$$

Hence our Corollary 3.2 shows that the same holds true for $\sqrt{n}\hat{\theta}_n$

$$\sup_{x \in \mathbb{R}} \left| P^n\left(\sqrt{n}\hat{\theta}_n \leq x \big| |h(X_1)|, \ldots, |h(X_n)|\right) - \Phi\left(\frac{x}{\sigma}\right) \right| \xrightarrow{P} 0.$$

This reveals that the corresponding conditional symmetry tests $\varphi_n^*(\hat{\theta}_n)$ *with test statistic* $\hat{\theta}_n$ *are asymptotically effective in the sense of (2.7).*

(b) We now consider the median functional $\kappa_0, \kappa_0(F) := F^{-1}(1/2)$*, and the empirical median* $\hat{\kappa}_n$ *as its natural estimator. As above we are going to test a nonparametric subclass of* 0*-symmetric product measures* $\mathcal{P}_{0,n}$*. We want to detect deviation from the null in terms of the median. Suppose therefore that* $X_1, \ldots, X_n$ *are i.i.d. with common* 0*-symmetric distribution function* F*, which is assumed to be continuously differentiable in a neighborhood of* 0 *with* $F'(0) = f(0) > 0$*. In this case it can be shown that*

$$\sqrt{n}\hat{\kappa}_n = \frac{1}{\sqrt{n}} \frac{1}{2f(0)} \sum_{i=1}^{n} \tilde{g} \circ F(X_i) + o_{P^n}(1) \qquad (3.15)$$

holds true with $\tilde{g}(u) = \mathbb{1}_{(1/2,1)}(u) - \mathbb{1}_{(0,1/2)}(u)$*, see Bickel et.al. (1993, p.303) as well as van der Vaart (1998, p.307f.).Thus (3.9) holds at the origin* $\kappa_0(F) = 0$*. As result we obtain that the* S_n*-conditional median symmetry test* $\varphi_n^*(\sqrt{n}\hat{\kappa}_n)$ *is asymptotically effective for testing* $\mathcal{P}_{0,n}$*.*

Remark 3.6 (Two-sample median permutation test)
Let κ_0 *(*$\hat{\kappa}_n$*, respectively) be again as in Example 3.5(b) the median functional (empirical median). In case of a two-sample problem the difference of the empirical median* $\hat{\theta}_n$*, defined by (3.10), serves as an estimator for the difference of the median given by the functional* $(P, Q) \mapsto \kappa_0(P) - \kappa_0(Q)$*. Under the regularity assumptions similar to the symmetry case of Example 3.5(b) the expansion (3.9)*

holds. Thus Example 3.3 implies the asymptotic effectiveness of the two-sample median permutation test $\varphi_n^*(\hat{\theta}_n)$ for the null-hypothesis $\{P = Q\}$.

The next remarkable example shows that conditional tests are not always asymptotically effective.

Example 3.7
(a) Consider the exponential family $\frac{dP_\vartheta}{dP_0} = C(\vartheta)e^{\vartheta T}, \quad \vartheta \in \Theta \subset \mathbb{R}$ with $0 \in \overset{\circ}{\Theta}$, $E_{P_0}(T) = 0$ and $Var_{P_0}(T) = \sigma^2 \in (0, \infty)$. Let $n = n_1 + n_2$ and suppose that $n_1/n \to p \in (0,1)$ holds for $\min(n_1, n_2) \to \infty$. We now contemplate for the path

$$\vartheta \mapsto Q_{n,\vartheta} := P^{n_1}_{\vartheta/\sqrt{n_1}} \otimes P_0^{n_2}$$

the two-sample problem $H = \{\vartheta = 0\}$ versus $K = \{\vartheta > 0\}$. In this case we get the parametric score-test $\varphi_n = \varphi_n(T_n)$ with test statistic

$$T_n := \frac{1}{\sqrt{n_1}} \sum_{i=1}^{n_1} (T(X_i) - E_{P_0^n}(T(X_i)) = \frac{1}{\sqrt{n_1}} \sum_{i=1}^{n_1} T(X_i),$$

where $X_i, 1 \le i \le n$, are the canonical projections. Under the null $\{\theta = 0\}$ the distribution of T_n converges to a normal-$N(0, \sigma^2)$-distribution, whereas

$$Var_{P_0^n}\big(T_n - E_{P_0^n}(T_n | X_{1:n}, \ldots, X_{n:n})\big) = \frac{n_2}{n}\sigma^2 \to (1-p)\sigma^2$$

hold as $n \to \infty$. By Lemma 2.5 the permutation version φ_n^* of φ_n is not asymptotically effective with respect to this score test φ_n.
(b) The reason for the negative result (a) can be explained as follows. Permutation tests are designed for the composite null hypothesis of product measures $\{P^n\}$. In this case the present test statistic T_n is not adequate. A proper choice for the composite null would be the so called effective score function

$$\tilde{T}_n := T_n - E_{P_0^n}(T_n | X_{1:n}, \ldots, X_{n:n}) = \frac{1}{\sqrt{n_1}} \Big(\frac{n_2}{n} \sum_{i=1}^{n_1} T(X_i) - \frac{n_1}{n} \sum_{i=n_1+1}^{n} T(X_i) \Big)$$

showing up in semiparametric statistics, see Bickel et.al. (1993) or Janssen and Völker (2007, Example 5.1.(a)).
To explain this, observe that T_n and $\tilde{T}_n$ lead to the same permutation tests φ_n^* (since $T_n \ge c_n^*$ iff $T_n - E_{P_0^n}(T_n | X_{1:n}, \ldots, X_{n:n}) \ge c_n^* - E_{P_0^n}(T_n | X_{1:n}, \ldots, X_{n:n}) =: \tilde{c}_n^*$ with new permutation critical values $\tilde{c}_n^*$). It is easy to see that φ_n^* is asymptotically effective with respect to $\tilde{\varphi}_n := \mathbf{1}_{(\tilde{c}_n, \infty)}(\tilde{T}_n)$. Observe that $\tilde{\varphi}_n$ is a score

test for testing $\{\theta = 0\}$ within the family

$$\vartheta \mapsto \tilde{Q}_{n,\vartheta} := P^{n_1}_{n_2\vartheta/(n\sqrt{n_1})} \otimes P^{n_2}_{-n_1\vartheta/(n\sqrt{n_1})},$$

which can be viewed as a least favorable parametrization of the family $\vartheta \mapsto Q_{n,\vartheta}$. In the conditional world T_n and $\tilde{T}_n$ lead to the same test φ^*_n. However, the unconditional models and score tests are different where the power of the score test from (a) can never be reached by permutation tests.

4 Proofs

Throughout let $T : (\Omega_0, \mathcal{A}_0, P_0) \to \mathbb{R}$ denote a test statistic on a further probability space $(\Omega_0, \mathcal{A}_0, P_0)$ with distribution function F.
Notice that we can define all random variables via projections on the product-space

$$(\Omega := \prod_{n=0}^{\infty} \Omega_n, \mathcal{A} := \bigotimes_{n=0}^{\infty} \mathcal{A}_n, P := \bigotimes_{n=0}^{\infty} P_n). \tag{4.1}$$

Proof of Lemma 2.2. Concerning Slutzky's Lemma and Polya's Theorem the assumptions imply $\sup_{x\in\mathbb{R}} |P_n(T_n + R_n \le x) - F(x)| \to 0$ so that it is enough to proof that (2.6) implicates (2.8) (otherwise consider $T_n - R_n$).
Suppose now that (2.6) holds. For $\alpha \in (0,1)$ denote by $c(\alpha)$ the $(1-\alpha)$-quantile of μ and define the test $\psi_n := \mathbb{1}_{(c(\alpha),\infty)}(T_n + R_n)$, so that we get $E_P(\psi_n) \to \alpha$. Like in (2.2) - (2.4) we can now conclude that there exist S_n-measurable functions c_n, γ_n such that $\psi^*_n := \mathbb{1}_{(c_n(\alpha),\infty)}(T_n + R_n) + \gamma_n \mathbb{1}_{\{c_n(\alpha)\}}(T_n + R_n)$ is an S_n-conditional test with

$$\int \psi^*_n dP^{T_n+R_n|S_n=s}_{\cdot} = \alpha \text{ for all } s \in \Omega_{S_n}. \tag{4.2}$$

Due to the equivalence of (2.6) and (2.7) and it remains to show that

$$E_P(|\psi_n - \psi^*_n|) \to 0 \text{ holds for all } \alpha \in (0,1).$$

To verify this we need two more test functions. Define again via (2.2) - (2.4) the function $\phi^*_n := \mathbb{1}_{(b_n(\alpha),\infty)}(T_n) + \tau_n \mathbb{1}_{\{b_n(\alpha)\}}(T_n)$ for measurable functions b_n, τ_n such that $\int \phi^*_n dP^{T_n|S_n=s}_{\cdot} = \alpha$ holds for all $s \in \Omega_{S_n}$ and set $\tilde{\phi}^*_n := \mathbb{1}_{(b_n(\alpha),\infty)}(T_n + R_n) + \tau_n \mathbb{1}_{\{b_n(\alpha)\}}(T_n + R_n)$.
Condition (2.6) now implies the convergence $b_n(\alpha) \xrightarrow{P} c(\alpha)$ for all $P \in \mathcal{P}_0$. Since Slutzky's Lemma deduces $E_P(|\phi^*_n - \tilde{\phi}^*_n|) \to 0$ this shows that $E_P(|\mathbb{1}_{(c(\alpha),\infty)}(T_n) -$

$\tilde{\phi}_n^*|) \to 0$ counts by (2.7). In the same manner it can be shown that $E_P(|\mathbb{1}_{(c(\alpha),\infty)}(T_n) - \psi_n|) \to 0$ holds. Combining these we obtain $E_P(|\psi_n - \tilde{\phi}_n^*|) \to 0$. Hence the proof is completed by showing $E_P(|\psi_n^* - \tilde{\phi}_n^*|) \to 0$.
Because of $T_n - |R_n| \le T_n + R_n \le T_n + |R_n|$ we consider foremost the case that $R_n \ge 0$ holds P-a.e.. Therefore we get the inequality $b_n(\alpha) \le c_n(\alpha)$ and so $\tilde{\phi}_n^* \ge \psi_n^*$. Thence we have with the help of equation (4.2)

$$\begin{aligned} E_P(|\psi_n^* - \tilde{\phi}_n^*|) &= E_P(\tilde{\phi}_n^*) - E_P(\psi_n^*) \\ &= E_P(\tilde{\phi}_n^*) - \int E_P(\psi_n^*|S_n)\, dP \\ &= E_P(\tilde{\phi}_n^*) \\ &\quad - \int dP_{\cdot}^{T_n+R_n|S_n=\cdot}((c_n(\alpha),\infty)) + \gamma_n dP_{\cdot}^{T_n+R_n|S_n=\cdot}(\{c_n(\alpha)\})\, dP \\ &= E_P(\tilde{\phi}_n^*) - \int \alpha\, dP \to 0. \end{aligned}$$

The case $R_n \le 0$ can be treated in an analogous manner.
If we now define tests $\tilde{\phi}_n^+$ and $\tilde{\phi}_n^-$ by substituting R_n in $\tilde{\phi}_n^*$ by $|R_n|$ and $-|R_n|$ respectively, and repeat the above steps for these tests we can conclude again that $E_P(|\psi_n - \tilde{\phi}_n^{+(-)}|) \to 0$ holds. Because of $\tilde{\phi}_n^- \le \psi_n^* \le \tilde{\phi}_n^+$ this implies

$$E_P(|\psi_n - \psi_n^*|) \le E_P(|\psi_n - \tilde{\phi}_n^+|) + E_P(|\psi_n - \tilde{\phi}_n^-|) \to 0,$$

which completes the proof. $\square$

Proof of Lemma 2.4. Recall first that there exists a version $E(T_n|S_n)$ of the conditional expectation $E_P(T_n|S_n) = E_{P_n}(T_n|S_n)$ which is independent of $P_n \in \mathcal{P}_{0,n}$, where P is defined on the product space, see 4.1. We may thus substitute P_n by P during the proof.
Let us first assume that $T_n \ge 0$ holds true for all n. Then by Markoff's inequality we have for all $\delta > 0$

$$P(E(T_n|S_n) \ge \delta) \le \delta^{-1} \sup_n E(T_n).$$

This shows that the sequence of probability measures $(\mathcal{L}(E(T_n|S_n)|P))_n$ is tight. Here we have used the notation $\mathcal{L}(T|P) = P^T$. Hence by Prohorov's Theorem there exist a subsequence $(n_k)_k$ and a random variable T_0 so that we have distributional convergence $E(T_{n_k}|S_{n_k}) \xrightarrow{\mathcal{D}} T_0$. If we now show that $T_0 = E(T)$

counts, than, due to the fact that the weak limit does not depend on the subsequence and is constant, equation (2.9) will follow.
In order to get this equality, consider first that assumption (2.6) implies that there exists another subsequence $(n_m)_m \subset (n_k)_k$ such that the following holds P-a.e. for every $j > 0$

$$\begin{aligned} E(T_{n_m}|S_{n_m}) = \int_{[0,\infty)} t \, dP^{T_{n_m}|S_{n_m}=\cdot}(t) \quad &\geq \quad \int_{[0,\infty)} \min(t,j) dP^{T_{n_m}|S_{n_m}=\cdot}(t) =: Y_m^j \\ &\rightarrow \quad \int_{[0,\infty)} \min(t,j) \, dP^T(t) =: C^j \end{aligned}$$

for $m \to \infty$. By the dominated convergence Theorem the right hand side C^j tends to $E(T)$ for $j \to \infty$. Thus we have for every $\delta > 0$ and all $j \geq j_0(\delta)$

$$\begin{aligned} P(E(T_{n_m}|S_{n_m}) \geq E(T) - 2\delta) \quad &\geq \quad P(Y_m^j \geq E(T) - 2\delta) \\ &\geq \quad P(\{C^j \geq E(T) - \delta\} \cap \{\left|Y_m^j - C^j\right| < \delta\}) \\ &\rightarrow \quad 1 \text{ for } m \to \infty \end{aligned}$$

from which we can conclude $P(T_0 \geq E(T) - 2\delta) = 1$. Due to the fact that this holds for all $\delta > 0$ we get

$$P(T_0 \geq E(T)) = 1. \tag{4.3}$$

Fatou's Lemma and Skohorod's Theorem, see Dudley (2002, Theorem 11.7.2.), now imply that T_0 is constant a.e., namely

$$\begin{aligned} E(T) = \lim_{m\to\infty} E(T_{n_m}) \quad &= \quad \lim_{m\to\infty} \int E(T_{n_m}|S_{n_m}) dP \\ &\geq \quad \int \liminf_{m\to\infty} E(T_{n_m}|S_{n_m}) dP = E(T_0). \end{aligned}$$

Together with (4.3) this means $P^{T_0} = \varepsilon_{E(T)}$ and therefore (2.9).
For the general case write $T_n = T_n^+ - T_n^-$ for $T_n^+ := \max(T_n, 0)$ and $T_n^- := \max(-T_n, 0)$ respectively. By the continuous mapping theorem we have again

$$\sup_{x\in\mathbb{R}} \left|P(T^{+(-)} \leq x) - P(T_n^{+(-)} \leq x|S_n))\right| \xrightarrow{P} 0.$$

Since

$$E(|T_n|) = E(T_n^+) + E(T_n^-) \to E(T^+) + E(T^-) = E(|T|)$$

and $\liminf_{n\to\infty} E(T_n^{+(-)}) \geq E(T^{+(-)})$ hold, the second assumption $E(T_n^{+(-)}) \to E(T^{+(-)})$ holds too. Thus we can apply the first case for showing

$$E(T_n^{+(-)}|S_n) \xrightarrow{P} E(T^{+(-)}),$$

which completes the proof. $\square$

Proof of Lemma 2.5. Denote by $\|\cdot\|_2$ the $L_2(P)$ norm. Because of

$$\|T_n\|_2^2 = \|T_n - E(T_n|S_n)\|_2^2 + \|E(T_n|S_n)\|_2^2$$

we have to establish that $\|E(T_n|S_n)\|_2^2 \to 0$ holds. Recall that the conditional convergence (2.6) implies the unconditional convergence $T_n \xrightarrow{\mathcal{D}} T$. Hence by Skohorod's Theorem there exist random variables $\hat{T}_n$ on another probability space $(\hat{\Omega}, \hat{\mathcal{A}}, \hat{P})$ with $\mathcal{L}(T_n) = \mathcal{L}(\hat{T}_n)$ and $\mathcal{L}(T) = \mathcal{L}(\hat{T}_0)$, such that $\hat{T}_n$ converges $\hat{P}$-a.e. to $\hat{T}_0$. Because of

$$\int \hat{T}_n^2 d\hat{P} = \int T_n^2 dP \to \int T^2 dP = \int \hat{T}^2 d\hat{P} < \infty \tag{4.4}$$

and $0 \leq |\hat{T}_n| \leq 1 + \hat{T}_n^2$ we get by Pratt's Lemma that

$$\int |T_n| dP = \int |\hat{T}_n| d\hat{P} \to \int |\hat{T}| d\hat{P} = \int |T| dP$$

holds. Thus Lemma 2.4 deduces the convergence of the conditional expectations

$$E(T_n|S_n) \xrightarrow{P} E(T) = 0.$$

Hence it remains to show that the sequence $(E(T_n|S_n)^2)_n$ is uniformly integrable. For this purpose we show foremost that $(T_n^2)_n$ is also uniformly integrable.
By (4.4) we get that $(\hat{T}_n^2)_n$ converges in $L_1(\hat{P})$ to $\hat{T}^2$. Thus $\hat{T}_n^2$ is uniformly integrable. As this property only depends on the law of $\hat{T}_n^2$, we can implicate that $(T_n^2)_n$ is also uniformly integrable. Furthermore we have by Markoff's inequality that

$$\mathbb{1}_{\{E(T_n^2|S_n) \geq a_n\}} \xrightarrow{P} 0$$

holds for every sequence a_n that tends to ∞ for $n \to \infty$. Thus we conclude by Jensen's inequality

$$\begin{aligned} \int \mathbb{1}_{\{E(T_n|S_n)^2 \geq a_n\}} E(T_n|S_n)^2 dP &\leq \int \mathbb{1}_{\{E(T_n^2|S_n) \geq a_n\}} E(T_n^2|S_n) dP \\ &= \int \mathbb{1}_{\{E(T_n^2|S_n) \geq a_n\}} T_n^2 dP \to 0 \end{aligned}$$

which completes the proof. $\square$

References

[1] Beran, R. (1988). *Prepivoting test statistics: A bootstrap view of asymptotic refinements.* J. Am. Stat. Assoc. 83, 687-697.

[2] Bickel, P. J.; Klaassen, C. A.J.; Ritov, Y. and Wellner, J. A. (1993). *Efficient and adaptive estimation for semiparametric models.* Johns Hopkins Series in the Mathematical Sciences, Johns Hopkins University Press, Baltimore.

[3] Dudley, R.M. (2002). *Real Analysis and Probability.* Cambridge Studies in Advanced Mathematics, Cambridge University Press, Cambridge.

[4] Edgington, E.S. and Onghena, P. (2007). *Randomization tests.* Statistics: Textbooks and Monographs. Boca Raton, FL: Chapman & Hall.

[5] Gebhard, J. and Schmitz, N. (1998). *Permutation tests - a revival?! I and II: Optimum properties and an efficient algorithm for computing the critical region.* Stat. Pap. 39, 75-96.

[6] Good P. (2005). *Permutation, parametric and bootstrap tests of hypotheses. 3rd ed.* Springer Series in Statistics, New York.

[7] Hoeffding, W. (1952). *The large-sample power of tests based on permutations of observations.* Ann. Math. Stat. 23. 169-192 (1952).

[8] Janssen, A. (1997). *Studentized permutation tests for non-i.i.d. hypotheses and the generalized Behrens-Fisher problem.* Stat. Probab. Lett. 36, 9-21.

[9] Janssen, A. (1999). *Nonparametric symmetry tests for statistical functionals.* Math. Methods Stat. 8, 320-343.

[10] Janssen, A. and Pauls, T. (2003). *How do Bootstrap and permutation tests work?* Ann. Stat.31, 768-806.

[11] Janssen, A. (2005). *Resampling Student's t-Type Statistics.* Ann. Inst. Statist. Math., 507-529.

[12] Janssen, A. and Völker, D. (2007). *Most powerful conditional tests.* Stat.&Dec. 25, 41-62.

[13] Lehmann, E.L. and Stein, C (1949). *On the theory of some non-parametric hypotheses.* Ann. Math. Stat. 20, 20-45.

[14] Lehmann, E.L. and Romano, J.P. (2005). *Testing Statistical Hypotheses.* Springer, New York.

[15] Mason, D. M. and Newton, M. A. (1992). *A Rank Statistics Approach to the Consistency of a General Bootstrap.* Ann. Stat. 20, 1611-1624.

[16] Praestgaard, J.T. (1995). *Permutation and bootstrap Kolmogorov-Smirnov tests for the equality of two distributions.* Scand. J. Stat. 22, 305-322 (1995).

[17] Praestgaard, J. and Wellner, J.A. (1993). *Exchangeably weighted Bootstraps of the general empirical process.* Ann. Probab. 21, 2053-2086.

[18] Romano, J.P. (1989). *Bootstrap and randomization tests of some nonparametric hypotheses.* Ann. Stat. 17, 141-159.

[19] Romano, J.P. (1990). *On the Behavior of Randomization Tests Without a Group Invariance Assumption.* J. Am. Stat. Assoc. 85, 686-692.

[20] Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics.* John Wiley, New York.

[21] Strasser, H. and Weber, C. (1999). *The asymptotic theory of permutation statistics.* Math. Methods Stat. 8, 220-250.

[22] van der Vaart, A.W. (1998). *Asymptotic Statistics.* Cambridge University Press, Cambridge.