Helmut Strasser and Christian Weber

On the Asymptotic Theory of Permutation Statistics

Working Paper

Adaptive Information Systems and Modelling in Economics and Management Science

Report Series

On the Asymptotic Theory of Permutation Statistics

Helmut Strasser, Christian Weber

Report No. 27
January 1999

Report Series

January 1999

SFB
'Adaptive Information Systems and Modelling in Economics and Management Science'

Vienna University of Economics
and Business Administration
Augasse 2–6, 1090 Wien, Austria

in cooperation with
University of Vienna
Vienna University of Technology

http://www.wu-wien.ac.at/am

Papers published in this report series
are preliminary versions of journal articles.

This paper was accepted for publication in:
Mathematical Methods of Statistics, **8**, 220–250

This piece of research was supported by the Austrian Science Foundation (FWF) under grant SFB#010 ('Adaptive Information Systems and Modelling in Economics and Management Science').

On the asymptotic theory of permutation statistics

Helmut Strasser and Christian Weber

June 1999

Abstract

In this paper limit theorems for the conditional distributions of linear test statistics are proved. The assertions are conditioned on the σ-field of permutation symmetric sets. Limit theorems are proved both for the conditional distributions under the hypothesis of randomness and under general contiguous alternatives with independent but not identically distributed observations. The proofs are based on results on limit theorems for exchangeable random variables by Strasser and Weber, [20]. The limit theorems under contiguous alternatives are consequences of a LAN-result for likelihood ratios of symmetrized product measures. The results of the paper have implications for statistical applications. By example it is shown that minimum variance partitions which are defined by observed data (e.g. by LVQ) lead to asymptotically optimal adaptive tests for the k-sample problem. As another application it is shown that conditional k-sample tests which are based on data-driven partitions lead to simple confidence sets which can be used for the simultaneous analysis of linear contrasts.

1 Introduction

Let $(\Omega, \mathcal{A})$ be a measurable space and let $\mathbb{P}$ and $\mathbb{Q}$ be probability measures on $\mathcal{A}$. Let further $\underline{X}_n = (X_{n1}, X_{n2}, \ldots, X_{nn})\,, n \in \mathbb{N}$, be a triangular array of random variables with values in a sample space $(E, \mathcal{B})$. Assume that under $\mathbb{P}$ the random variables of each row of the triangular array are independent und identically distributed. The distribution $\mathbb{P} * X_{ni}$ of any random variable is denoted by P. Assume that under $\mathbb{Q}$ the random variables of each row are independent but not necessarily identically distributed. Let us denote these distributions by $Q_{n,i} := \mathbb{Q} * X_{ni}$.

If a simple hypothesis $\mathbb{P}$ is tested against any alternative $\mathbb{Q}$, the linear statistics

$$T_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} w_{ni} f_n(X_{ni}) \tag{1}$$

with influence functions f_n and weights w_{ni} are typical test statistics. Limit theorems for test statistics of such kind under $\mathbb{P}$ and $\mathbb{Q}$ are a well-established part of the asymptotic statistical theory. This theory can be treated in the framework of tangent spaces and tangent vectors, which has been developed by Pfanzagl and Wefelmeyer, [13]. Originally, these concepts have been established for independent and identically distributed observations, but they have been extended by Strasser, [15], to the case of not necessarily identically distributed observations.

If the null-hypothesis is composite, i.e. if it consists of a set H_0 of probability measures, then for the construction of tests we have to consider how the distributions of the test statistics depend on $\mathbb{P} \in H_0$. On one hand the unknown probability measure $\mathbb{P} \in H_0$ is a nuisance parameter for the construction of critical values and must be estimated. On the other hand, the tangent vectors describing those alternatives where the test should have large power, may depend on $\mathbb{P} \in H_0$, too. As a consequence, the shape of the

[1] *AMS 1991 subject classifications.* Primary: 62F05, 62G10, 62G20, secondary: 65H15, 60F17, 60G09.

[2] *Key words and phrases:* permutation test, limit distribution under hypothesis, limit distribution under alternative, conditional limit theorem, exchangeable random variables, conditional local asymptotic normality, multivariate.

influence functions f_n and sometimes even the weights w_{ni} of linear test statistics (1) have to be chosen in a data-dependent way. These problems are related to the so–called adaptivity problem of test statistics.

Now, let us consider the null-hypothesis H_0 of all probability measures such that the rows of the triangular array are independent and identically distributed. This set H_0 is the so-called hypothesis of randomness. In general, the distributions of the test statistics depend on $\mathbb{P} \in H_0$. For this null-hypothesis, however, we can dispose of this dependence on $\mathbb{P} \in H_0$ by conditioning. This leads to a family of tests which are known as permutation tests.

In the following the symmetric σ-field which is defined by $\underline{X}_n$ plays a central role and is thus isolated by a definition. Further information on this σ-field is given in section 6.

(1.1) DEFINITION *Let $\mathcal{S}_n$ be the σ-field of all permutation symmetric sets in $\mathcal{B}^n$. Then $\mathcal{S}(\underline{X}_n) = \underline{X}_n^{-1}(\mathcal{S}_n)$ is called the symmetric $\sigma-$field defined by $\underline{X}_n$.*

Permutation tests rely on the fact that under $\mathbb{P} \in H_0$ the conditional distribution of $\underline{X}_n$ under the symmetric σ-field $\mathcal{S}(\underline{X}_n)$ is the uniform distribution on the symmetric group Π_n, i.e.

$$\mathbb{P}(\underline{X}_n \in B|\mathcal{S}(\underline{X}_n)) = \frac{1}{n!} \sum_{\pi \in \Pi_n} 1_B((X_{n\pi(1)}, X_{n\pi(2)}, \ldots X_{n\pi(n)}), \quad B \in \mathcal{B}^n.$$

Let $T_n(\underline{X}_n)$ by any statistic and let $\alpha \in [0,1]$. Then for $\mathbb{P} \in H_0$ we may compute critical values such that the test attains a fixed significance level α exactly, i.e.

$$\mathbb{P}(T_n(\underline{X}_n) > c_\alpha(\underline{X}_n)|\mathcal{S}(\underline{X}_n)) + p_\alpha(\underline{X}_n)\,\mathbb{P}(T_n(\underline{X}_n) = c_\alpha(\underline{X}_n)|\mathcal{S}(\underline{X}_n)) \equiv \alpha,$$

where $c_\alpha(\underline{X}_n)$ and $0 \le p_\alpha(\underline{X}_n) < 1$ depend on $\underline{X}_n$ in a permutation symmetric way. Such a test is called a permutation test. The family of all permutation tests has several attractive properties (cf. Lehmann, [6], Theorem 5.6, and Lehmann and Stein, [7]).

Since H_0 is a composite hypothesis, for reasons explained previously it is natural to extend the family of linear test statistics (1) by admitting data-dependent influence functions f_n and weights w_{ni}. Thus, we generalize the notion of linear test statistics (1) considering statistics of the form

$$T_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} w_{ni}(\underline{X}_n) f_n(X_{ni}, \underline{X}_n), \tag{2}$$

where we suppose that $w_{ni}(\underline{X}_n)$ and $f_n(X_{ni}, \underline{X}_n)$ depend on $\underline{X}_n$ in a permutation symmetric way. Permutation tests which are based on test statistics of the form (2) cover interesting special cases:

1. Linear rank tests are of this kind since for the ranks R_{ni} we have $R_{ni} = nF_{n,\underline{X}_n}(X_{ni})$. Here $F_{n,\underline{X}_n}$ denotes the empirical distribution function of X_{n1}, X_{n2},..., X_{nn}.
2. Dealing with explorative multivariate data analysis and aiming at a complexity reduction of the data we often apply a clustering algorithm which is based on the observed data themselves. If after such a procedure test statistics are constructed by analogy to classical procedures, then the shapes of the arising influence functions are data-dependent. Typically, the sequential order of the data is not used by the clustering algorithm. In such cases the test statistics are of the kind given in (2).

Already in his early textbook on statistical methods, [11], Pfanzagl noted the importance of permutation tests for statistical applications. However, until recently permutation tests have played only a minor role in applications (with the important exception of rank tests). Since the critical values $c_\alpha(\underline{X}_n)$ of permutations tests depend on the observed data, it is (apart from particularly simple cases like the exact test by Fisher, the test by McNemar or rank tests) impossible to compute tables for critical values which can be applied in general situations. On the other hand, the time required for the actual computation of such critical values increases very fast with the sample size n. Hence, in the past the exact compuation of critical values for permutation tests was practically not feasible. Recently, however, the performance of modern computers leads to an increasing interest in permutation tests for practical applications.

The present paper is devoted to a theoretical and asymptotic analysis of permutation tests of a certain kind. Such an analysis has to be based on limit theorems for the distributions of the test statistics under the null-hypothesis and under alternatives. For the special case of linear rank tests this is the subject of the theory by Hájek and Sidak, [3]. More general results for larger classes of permutation tests are due to Neuhaus, [10] and Janssen and Neuhaus, [5]. These results are devoted to one-dimensional data and the asymptotic distributions under the null-hypothesis. However, the results by Neuhaus, [10], and Janssen and Neuhaus, [5], are stronger than those given by the theory by Hájek und Sidak, [3], as these authors prove limit theorems for the conditional distributions under $\mathcal{S}(\underline{X}_n)$.

Limit theorems for conditional distributions contain more detailed information than limit theorems for unconditional distributions. If we consider the conditional distributions under the null-hypothesis then we may apply such limit theorems for the approximation of data-dependent critical values. Even more interesting are the conditional distributions under alternatives. If these conditional distributions are known or at least approximately known, then we may study the power of permutation tests, conditioned on the data observed. Such a conditional power analysis is far better suited to the experimental situation than a classical unconditional analysis. For a further discussion of conditional versus unconditional statistical analysis see Lehmann, [6], Chapter 10 and pages 150ff.

In the present paper we prove limit theorems for the conditional distributions of test statistics of the form (2). We state our results for multivariate test statistics with influence functions f_n valued in $\mathbb{R}^d$ and matrix valued weights w_n. Our results cover both the conditional distributions under the null-hypothesis H_0 and under contiguous alternatives.

The organization of our proofs is different from what has been done previously. The techniques applied by Neuhaus, [10] and Janssen and Neuhaus, [5], are based on a martingale argument in the course of the proof of the statistical limit theorem. Our way of proving conditional limit theorems consists in isolating the martingale part in a probabilistic limit theorem which is of interest of its own. This limit theorem is an extension of the invariance principle by Billingsley for exchangeable random variables, [1], to multivariate and not necessarily standardized partial sum processes. There is a proof of this invariance principle by Hoffmann, [4], which applies a martingale argument resambling those methods applied by Neuhaus, [10], and by Janssen and Neuhaus, [5]. The extension of Billingsley's theorem in Strasser and Weber, [20], is also proved by martingale methods.

Our main result on conditional distributions under contiguous alternatives is based on a LAN-type assertion for likelihood ratios restricted to the symmetric σ-fields $\mathcal{S}(\underline{X}_n)$. This LAN-result shows that it is possible to develop a theory of the asymptotic efficiency of tests, where conditional distributions replace the unconditional distributions of the classical theory. The present paper is a first step into that direction.

The paper is organized as follows. In section 2 we present our main results with detailed comments but without proofs. Section 3 contains statistical applications. In section 4 we start with the technical part preparing some notions from probability theory. The proofs of our main results are presented in section 5, and in section 6 we collect auxiliary lemmas.

2 Main Results

This section contains the statement of the main results. Proofs are collected in section 5.

From the mathematical point of view our results are statistical limit theorems for conditional distributions under symmetric $\sigma-$fields $\mathcal{S}(\underline{X}_n)$. For the precise statement we require a concept of convergence for conditional distributions. A general concept of this kind is considered in Definition (4.5). At this point we only consider conditional asymptotic normality.

Let $(Y_n)_{n\in\mathbb{N}}$ be a sequence of random variables with values in a Euclidean space and let $(P_n)_{n\in\mathbb{N}}$ be a sequence of probability measures on that Euclidean space. Let further $(\mathcal{F}_n)_{n\in\mathbb{N}}$, $\mathcal{F}_n \subseteq \mathcal{A}$, be a sequence of sub-$\sigma$-fields, and consider sequences of $\mathcal{F}_n$-measurable random elements $(\mu_n)_{n\in\mathbb{N}}$ and random operators

$(\sigma_n^2)_{n\in\mathbb{N}}$, $\sigma_n^2 \geq 0$. Denote by Φ the set of all bounded, real-valued and Lipschitz-continuous functions ϕ on the Euclidean space.

(2.1) DEFINITION *The sequence of random variables (Y_n) is asymptotically normal under (P_n) and conditioned on $(\mathcal{F}_n)$, denoted by $P_n \star Y_n \sim N(\mu_n, \sigma_n^2)$ $(\mathcal{F}_n)$, if*

$$\mathbb{E}_{P_n}(\phi(Y_n)|\mathcal{F}_n) - \int \phi(x)\, dN(\mu_n, \sigma_n^2) \xrightarrow{P_n} 0, \quad \phi \in \Phi. \tag{3}$$

Let us give some brief remarks concerning our notation. By $L^p(\Omega, \mathcal{A}, \mu; E)$ we denote as usual the space of p-integrable functions on $(\Omega, \mathcal{A}, \mu)$ with values in E. If E is omitted then it is to be understood as $E = \mathbb{R}$. Similarly, if Ω is a Borel set of $\mathbb{R}^d$ then $\mathcal{A}$ is by default to be understood as the Borel σ-field and μ as the Lebesgue measure.

Now, let us return to the situation considered in (1).

(2.2) ASSUMPTIONS

1. *Under $\mathbb{P}$ the random variables $X_{n1}, X_{n2}, \ldots, X_{nn}$ are independent and identically distributed, i.e. $\mathbb{P} \in H_0$. We denote $\mathbb{P} \star \underline{X}_n =: P^n$.*
2. *Under $\mathbb{Q}$ the random variables $X_{n1}, X_{n2}, \ldots, X_{nn}$ are independent, but not necessarily identically distributed. The distributions are denoted by $Q_{n,i} := \mathbb{Q} * X_{ni}$.*
3. *The functions $f_n : (E, E^n) \to \mathbb{R}^m$, $(x, \underline{y}) \mapsto f_n(x, \underline{y})$, $n \in \mathbb{N}$, are measurable and such that they depend on $\underline{y} \in E^n$ in a permutation symmetric way.*

These assumptions define the situation for which the main results of this paper are proved. They are not repeated when the results are stated explicitely. If nothing else is mentioned then we denote by $\mathbb{E}$, $\mathbb{V}$ and $\mathbb{CV}$ the expectation, the variance and the covariance of random variables under some probability measure $\mathbb{P} \in H_0$.

Given the assumptions (2.2) the conditional expectations $\mathbb{E}(f_n\,(X_{ni}, \underline{X}_n)\,|\mathcal{S}(\underline{X}_n))$ and the conditional variances $\mathbb{V}(f_n\,(X_{ni}, \underline{X}_n)\,|\mathcal{S}(\underline{X}_n))$ are independent of i (see Remarks (6.4)). For notational convenience let us denote them by $\mathbb{E}(f_n|\mathcal{S}(\underline{X}_n))$ and $\mathbb{V}(f_n|\mathcal{S}(\underline{X}_n))$, i.e.

$$\mathbb{E}(f_n|\mathcal{S}(\underline{X}_n)) := \frac{1}{n}\sum_{i=1}^{n} f_n(X_{ni}, \underline{X}_n),$$

$$\mathbb{V}(f_n|\mathcal{S}(\underline{X}_n)) := \frac{1}{n}\sum_{i=1}^{n}(f_n(X_{ni}, \underline{X}_n) - \mathbb{E}(f_n|\mathcal{S}(\underline{X}_n)))(f_n(X_{ni}, \underline{X}_n) - \mathbb{E}(f_n|\mathcal{S}(\underline{X}_n)))^t$$

Let (T_n) be a sequence of linear statistics given by

$$T_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} w_{ni} f_n(X_{ni}, \underline{X}_n), \tag{4}$$

where (w_{ni}) is a triangular array of nonrandom weights in $\mathbb{R}^{m\times d}$. (The extension of our results to random weights is discussed in Remark (2.11).) Every triangular array of such weights can be generated by a sequence of functions $w_n \in L^2([0,1]; \mathbb{R}^{m\times d})$ defining

$$w_{ni} := n\int_{\frac{i-1}{n}}^{\frac{i}{n}} w_n(t)dt, \quad i = 1, 2, \ldots, n.$$

Throughout the following we suppose that the weights are generated by a sequence of weight functions (w_n).

The following theorem is our first main result. It is concerned with the convergence of the conditional distributions of (T_n) under H_0.

(2.3) THEOREM *(Asymptotic normality under the null hypothesis.) Suppose that the triangular array*

$$\frac{1}{\sqrt{n}}\left(f_n(X_{ni},\underline{X}_n)-\mathbb{E}(f_n|\mathcal{S}(\underline{X}_n))\right),\quad 1\le i\le n,\, n\in\mathbb{N},$$

is infinitesimal and that $\mathbb{V}(f_n|\mathcal{S}(\underline{X}_n))$ *is stochastically bounded. Assume further that* $(w_n)\subseteq L^2([0,1];\mathbb{R}^{m\times d})$ *is a relatively compact sequence of weight functions. Then* $(T_n-\mathbb{E}(T_n|\mathcal{S}))$ *is asymptotically normal under* $\mathbb{P}\in H_0$ *and conditioned on the symmetric* σ*-fields* $\mathcal{S}(\underline{X}_n)$*, i.e.*

$$\mathbb{P}\star(T_n-\mathbb{E}(T_n|\mathcal{S}(\underline{X}_n)))\sim\mathcal{N}(0,\mathbb{V}(T_n|\mathcal{S}(\underline{X}_n)))\ (\mathcal{S}(\underline{X}_n)) \tag{5}$$

Theorem (2.3) contains an assertion on multivariate statistics. Such a multivariate version is essential for several reasons. First, the proof of Theorem (2.9) which deals with the asymptotics of conditional distributions under contiguous alternatives is based on a multivariate version of Theorem (2.3). Another reason is, that many statistical applications, e.g. those of section 3, require multivariate versions of Theorem (2.3).

(2.4) REMARKS 1. The only property of the distributions of $X_{n1},X_{n2},\ldots,X_{nn}$ which is needed in the proof of Theorem (2.3) is the fact that the conditional distributions are uniform on the set of permuted realisations. For this property we do not need that the random variables $X_{n1},X_{n2},\ldots,X_{nn}$ are independent and identically distributed. In fact, the conditional distribution is uniform iff the random variables $X_{n1},X_{n2},\ldots,X_{nn}$ are exchangeable. (See Definition (4.1).)

2. The assertion of Theorem (2.3) does not imply that

$$\mathbb{P}\star T_n\sim\mathcal{N}(\mathbb{E}(T_n|\mathcal{S}(\underline{X}_n)),\mathbb{V}(T_n|\mathcal{S}(\underline{X}_n)))\ (\mathcal{S}(\underline{X}_n)) \tag{6}$$

However, if the conditional expectations $\mathbb{E}(T_n|\mathcal{S}(\underline{X}_n))$ are stochastically bounded, then assertions (5) and (6) are equivalent.

3. For reasons of completeness let us give explicit expressions for the conditional expectations and variances of T_n under $\mathcal{S}(\underline{X}_n)$:

$$\begin{aligned}\mathbb{E}(T_n|\mathcal{S}(\underline{X}_n))&=\frac{1}{\sqrt{n}}\sum_i w_{ni}\mathbb{E}(f_n|\mathcal{S}(\underline{X}_n))\\ \mathbb{V}(T_n|\mathcal{S}(\underline{X}_n))&=\frac{1}{n-1}\sum_i w_{ni}\mathbb{V}(f_n|\mathcal{S}(\underline{X}_n))w_{ni}^t\\ &\quad-\frac{1}{n\,(n-1)}\sum_{i,j}w_{ni}\mathbb{V}(f_n|\mathcal{S}(\underline{X}_n))w_{nj}^t\end{aligned} \tag{7}$$

The validity of the first formula is obvious. The proof of the second formula is given in Remarks (6.4), paragraph 2.

Our further results are concerned with contiguous alternatives. We apply the framework of Strasser, [15], for the definition of tangent vectors for alternatives with independent but not identically distributed observations.

(2.5) DISCUSSION Let

$$L^*(P):=\left\{h\in L^2\left([0,1]\times E,\lambda\otimes P\right):\int h(t,.)\,dP=0\quad\text{whenever } t\in[0,1]\right\},$$

and

$$Q_{n,t} := \sum_{i=1}^{n} Q_{n,i} 1_{\left(\frac{i-1}{n}, \frac{i}{n}\right]}(t), \quad t \in [0,1],\, n \in \mathbb{N}.$$

We say, the families $(Q_{n,t})$ have a sequence $(h_n) \subseteq L^*(P)$ as tangent vectors if

$$\sqrt{\frac{dQ_{n,t}}{dP}(x)} = 1 + \frac{1}{2\sqrt{n}} h_n(t,x) + \frac{1}{\sqrt{n}} r_n(t,x),$$

where the residuals (r_n) converge to zero in L^2, i.e.

$$\lim_{n\to\infty} \int \int r_n(t,.)^2 \, dP \, dt = 0$$

The tangent space $T_P\,(H_0)$ consists of all $h \in L^*(P)$, which do not depend on t. Thus, the orthogonal space $T_P\,(H_0)^{\perp}$ consists of all $h \in L^*(P)$, such that $\int_0^1 h(t,x)\,dt = 0 \quad P$-a.e. Now, let (h_{ni}) be that triangular array of functions in $L^2(P)$, which is defined by

$$h_{ni}(x) := n \int_{\frac{i-1}{n}}^{\frac{i}{n}} h_n(t,x)\,dt.$$

If the sequence (h_n) of tangent vectors is relatively compact in $L^*(P)$, then the likelihood ratios satisfy the LAN-condition, i.e.

$$\prod_{i=1}^{n} \frac{dQ_{n,i}}{dP}(\underline{x}) = \exp\Big(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} h_{ni}(x_i) - \frac{1}{2}\|h_n\|^2 + r_n\,(\underline{x})\Big), \tag{8}$$

where the residuals (r_n) converge to zero in P^n-probability (cf. Strasser, [15], Theorem (2.2)).

(2.6) EXAMPLE By example let us consider alternatives of the k-sample problem. A general k-sample alternative is defined in the following way: If $R_{n,j}$, $1 \le j \le k$, $n \in \mathbb{N}$, are probability measures on $\mathcal{A}$, and if $0 = t_0 < t_1 < \cdots < t_k = 1$, then we define $Q_{n,i} = R_{n,j}$ whenever $nt_{j-1} < i \le nt_j$, $1 \le j \le k$. The tangent vectors for such alternatives are of the form

$$h(t,x) = \sum_{j=1}^{k} g_j(x) 1_{(t_{j-1},t_j]}(t)$$

where $g_j \in L^2\,(E,P)$, $\int g_j dP = 0$, and $\|h\|^2 = \sum_{j=1}^{k}(t_j - t_{j-1}) \int g_j^2 \, dP$.

Our second main result gives a stochastic expansion of the likelihood ratios of $\mathbb{P}$ and $\mathbb{Q}$ if they are restricted to $\mathcal{S}(\underline{X}_n)$. On the symmetric σ-fields $\mathcal{S}(\underline{X}_n)$ these measures are not product measures and thus a direct application of known LAN-assertions is not possible.

Let us prepare the result by some explanations. First, we note that

$$\frac{d\mathbb{Q}|\mathcal{S}(\underline{X}_n)}{d\mathbb{P}|\mathcal{S}(\underline{X}_n)} = \mathbb{E}\left(\frac{d\mathbb{Q}}{d\mathbb{P}}\middle|\, \mathcal{S}(\underline{X}_n)\right).$$

For the likelihood ratio $\dfrac{d\mathbb{Q}}{d\mathbb{P}}$ the expansion (8) is valid. Next we decompose the exponent of this expansion into two parts

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} h_{ni}(X_{ni}) - \frac{1}{2} \int_0^1 \int h_n(t,x)^2 \, dP \, dt =: \alpha_n + \beta_n, \tag{9}$$

where the first part

$$\alpha_n := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \bar{h}_n(X_{ni}) - \frac{1}{2} \int \bar{h}_n(x)^2 \, dP \tag{10}$$

is $\mathcal{S}(\underline{X}_n)$-measurable. This part coincides with the loglikelihood of the orthogonal projection

$$\bar{h}_n(x) := \int_0^1 h_n(t,x)\, dt = \frac{1}{n} \sum_{i=1}^{n} h_{ni}(x)$$

of the tangent vector h_n onto the tangent space $T_P\,(H_0)$ of the hypothesis of randomness. The second part is

$$\beta_n := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (h_{ni} - \bar{h}_n)(X_{ni}) - \frac{1}{2} \int_0^1 \int (h_n(t,.) - \bar{h}_n(.))^2 \, dP \, dt, \tag{11}$$

and this coincides with the loglikelihood of the orthogonal projection of the tangent vector h_n onto the orthogonal complement $T_P\,(H_0)^\perp$ of the hypothesis of randomness. Putting terms together we obtain

$$\frac{d\mathbb{Q}|\mathcal{S}(\underline{X}_n)}{d\mathbb{P}|\mathcal{S}(\underline{X}_n)} = \mathbb{E}\left(\frac{d\mathbb{Q}}{d\mathbb{P}} \middle| \mathcal{S}(\underline{X}_n) \right) = \exp(\alpha_n)\mathbb{E}(\exp(\beta_n + r_n)|\mathcal{S}(\underline{X}_n)). \tag{12}$$

Our second main result states that in this expression the second factor can be disposed of.

(2.7) THEOREM *(Local asymptotic normality of symmetrized products.) If the sequence of tangent vectors* (h_n) *is relatively compact in* $L^*(P)$, *then*

$$\frac{d\mathbb{Q}|\mathcal{S}(\underline{X}_n)}{d\mathbb{P}|\mathcal{S}(\underline{X}_n)} = \exp\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \bar{h}_n(X_{ni}) - \frac{1}{2} \int \bar{h}_n(x)^2 \, dP + r_n(\underline{X}_n) \right),$$

where the residuals (r_n) *converge to zero in* P^n*-probability.*

The proof of Theorem (2.7) is based on Theorem (2.3). It is an essential point for this proof that under $\mathbb{P}$ the random variables are identically distributed.

(2.8) REMARK If the random variables are not identically distributed under $\mathbb{P}$ then the expansion of Theorem (2.7) is still valid under some additional conditions. E.g., this is an easy consequence of Theorem (2.7) if $\mathbb{P}$ is contiguous to product measures with identical factors. However, if this is not the case then any extensions of Theorem (2.7) require more sophisticated techniques. LAN-expansions of that kind are related to investigations by Pfanzagl, [12], concerning the asymptotic efficiency of permutation symmetric estimator sequences in the presence of nuisance parameters. A problem posed by Pfanzagl has been partially solved by Strasser, [18] and [19]. The general answer is equivalent to a LAN-assertion for likelihood ratios of symmetrized product measures which contains Theorem (2.7) as a special case but needs a completely different proof. For details see Strasser, [17].

Our last main result is Theorem (2.9) which states that under contiguous alternatives $\mathbb{Q}$ with any tangent vector $h \in L^*(P)$ the sequence of statistics T_n is asymptotically normal, conditioned on $\mathcal{S}(\underline{X}_n)$, i.e.

$$\mathbb{Q} \star (T_n - \mathbb{E}(T_n|\mathcal{S}(\underline{X}_n))) \sim \mathcal{N}(\mathbb{CV}(T_n, \beta_n|\mathcal{S}(\underline{X}_n)), \mathbb{V}(T_n|\mathcal{S}(\underline{X}_n)))\ (\mathcal{S}(\underline{X}_n)) \tag{13}$$

This limit theorem is similar to the corresponding classical assertion for unconditional distributions. The essential point is that the limit distribution under the alternative differs from the limit distribution under the null hypothesis only by a sequence of translations. Similarly to the unconditional case the amount of the translations is given by the covariances $\mathbb{CV}(T_n, \beta_n|\mathcal{S}(\underline{X}_n))$ between the statistics T_n and the loglikelihoods β_n of a sequence of tangent vectors. These are not the tangent vectors h_n of the alternatives themselves but as a result of conditioning by $\mathcal{S}(\underline{X}_n)$ those are replaced by their projections $h_n(t,.) - \bar{h}_n$ to the orthogonal complement $T_P\,(H_0)^\perp$ of the null-hypothesis.

(2.9) THEOREM *(Asymptotic normality under contiguous alternatives.) Suppose that the triangular array*

$$\frac{1}{\sqrt{n}}\left(f_n(X_{ni},\underline{X}_n) - \mathbb{E}(f_n|\mathcal{S}(\underline{X}_n))\right), \quad 1 \le i \le n,\, n \in \mathbb{N},$$

is infinitesimal and that the variances $\mathbb{V}(f_n|\mathcal{S}(\underline{X}_n))$ are stochastically bounded. Assume further that $(w_n) \subseteq L^2([0,1];\mathbb{R}^{m\times d})$ is a relatively compact sequence of weight functions. If the sequence of tangent vectors (h_n) is relatively compact in $L^*(P)$, then (T_n) is asymptotically normal under $\mathbb{Q}$, conditioned on $\mathcal{S}(\underline{X}_n)$, i.e.

$$\mathbb{Q} \star (T_n - \mathbb{E}(T_n|\mathcal{S}(\underline{X}_n))) \sim N(\mathbb{CV}(T_n,\beta_n|\mathcal{S}(\underline{X}_n)), \mathbb{V}(T_n|\mathcal{S}(\underline{X}_n)))\ (\mathcal{S}(\underline{X}_n)),$$

where the sequence (β_n) is defined by (11).

(2.10) REMARK An explicit expression for the covariances is

$$\mathbb{CV}(T_n,\beta_n|\mathcal{S}(\underline{X}_n)) = \frac{1}{n-1}\sum_{i=1}^{n} w_{ni}\mathbb{CV}(f_n, h_{ni} - \bar{h}|\mathcal{S}(\underline{X}_n)) \tag{14}$$

which is proved in Remarks (6.4), paragraph 3.

For some applications we need versions of Theorems (2.3) and (2.9) where the weight functions may depend both on $t \in [0,1]$ and the observed data. This is the concern of the following remark.

(2.11) REMARK Let $w_n : [0,1] \times E^n \to \mathbb{R}^{m\times d}$ be weight functions. We are going to discuss the question whether the assertions of Theorems (2.3) and (2.9) are also valid for linear statistics of the form

$$T_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} w_{ni}(\underline{X}_n) f_n(X_{ni},\underline{X}_n),$$

where

$$w_{ni}(x) = n\int_{\frac{i-1}{n}}^{\frac{i}{n}} w_n(t,x)\,dt.$$

It is clear that a minimal requirement is that the functions $x \mapsto w_n(t,x)$ are permutation symmetric in $x \in E^n$. Moreover, we need a condition which corresponds to the compactness condition for the sequence (w_n) in Theorems (2.3) and (2.9). A preliminary idea comes from the fact that the desired extension is immediate for weight functions of the type

$$w_n(t,x) := \sum_{j=1}^{M} \alpha_{nj}(t) v_{nj}(x), \tag{15}$$

if the functions $v_{nj} : E^n \to \mathbb{R}^{k\times d}$ are permutation symmetric and stochastically bounded, and if the sequences of functions $\alpha_{nj} : [0,1] \to \mathbb{R}^{m\times d}$ are relatively compact in $L^2([0,1],\lambda)$. In this case the functions f_n can be replaced by $(v_{n1}f_n, v_{n2}f_n, \ldots, v_{nM}f_n)$ and we may apply Theorems (2.3) and (2.9) as they are. Moreover, it follows that the extension to more general weight functions (w_n) can be performed, if the weight functions can be approximated by linear combinations of the form (15) in a suitable way. To be precise, such an approximation should be possible with respect to the conditional variances $\mathbb{V}(\,.\,|\mathcal{S}(\underline{X}_n))$. The details are left to the reader.

3 Applications

By way of example let us consider tests for the k-sample problem in situations where the test statistics are based on data-driven partitions. In such cases the indicators of the cells of the partition are random variables which depend on the whole data set.

In the field of goodness-of-fit tests there has been interest in tests based on random partitions for a long time, e.g. see Witting, [21], Chibisov, [2], Moore and Spruill, [9], Pollard, [14] und Li and Doss, [8]. However, the k-sample problem is basically different from a goodness of fit situation. For the k-sample problem, conditioning under $\mathcal{S}(\underline{X}_n)$ leads to asymptotically optimal tests, but for goodness-of-fit problems, conditioning under $\mathcal{S}(\underline{X}_n)$ makes no sense at all, since permutation tests have no power for goodness of fit problems. As a consequence papers on goodness-of-fit with random partitions deal with approximations of the unconditional distributions, whereas in our case we are interested in the conditional distributions under $\mathcal{S}(\underline{X}_n)$.

There is a plenty of partitioning algorithms used by explorative data analysis, e.g. CART, MARS, Principal Points. Related to artificial neural networks Learning Vector Quantization (LVQ) and Self Organizing Feature Maps (SOFM) are well-known partitioning algorithms. Each of this methods does not regard the sequential order of the data. Hence, the partitions obtained depend on the data in a permutation symmetric way and we may apply the theoretical results of the present paper.

Let $B_n(\underline{X}_n) = (B_{n,1}(\underline{X}_n), B_{n,2}(\underline{X}_n), \dots, B_{n,m}(\underline{X}_n))$ be a partition of the sample space, where $B_n(\underline{x})$ depends on $\underline{x} \in E^n$ in a permutation symmetric way. Let further $G = (G_1, G_2, \dots, G_k)$ be a partition of the unit interval $[0, 1]$ into adjacent subintervals $G_j = (t_{j-1}, t_j]$, such that $1_{G_j}(\frac{i}{n})$ indicates that the i-th observation X_{ni} is in G_j. By

$$p_n(B_{n,s} \times G_j) := \frac{1}{n} \sum_{i=1}^{n} 1_{G_j} \left(\frac{i}{n} \right) 1_{B_{n,s}(\underline{X}_n)}(X_{ni})$$

we denote the relative frequency of cell $B_{n,s}$ in group G_j, similarly by

$$p_n(B_{n,s}) := \mathbb{E}(1_{B_{n,s}} | \mathcal{S}(\underline{X}_n)) = \frac{1}{n} \sum_{i=1}^{n} 1_{B_{n,s}(\underline{X}_n)}(X_{ni})$$

the relative frequency of cell $B_{n,s}$ and by $p_n(G_j) := \frac{1}{n} \sum_{i=1}^{n} 1_{G_j}(\frac{i}{n})$ the percentage of observations in group G_j. Moreover, we denote by

$$Z_n := \left(\sqrt{n} \, \frac{p_n(B_{n,s} \times G_j) - p_n(B_{n,s}) p_n(G_j)}{\sqrt{p_n(B_{n,s}) p_n(G_j)}} \right)_{\substack{s = 1, 2, \dots, m, \\ j = 1, 2, \dots, k,}}$$

the matrix of centered and normed cell frequencies (the so-called standardized contingency table). It follows from our main results (2.3) and (2.9), that for the conditional distributions of the sequence (Z_n) the same assertions are true which are well-known for the unconditional distributions of (Z_n) in the case of deterministic partitions.

(3.1) THEOREM *Suppose that the partitions* $B_n(\underline{X}_n) = (B_{n,1}(\underline{X}_n), B_{n,2}(\underline{X}_n), \dots, B_{n,m}(\underline{X}_n))$ *are such that the cell frequencies* $p_n(B_n)$ *are stochastically bounded away from zero. Assume further that for* $G = (G_1, G_2, \dots, G_k)$ *and for the alternatives the conditions of Example (2.6) are satisfied. Then*

$$\mathbb{P} \star Z_n \sim N(0, C_{p_n}) \quad (\mathcal{S}(\underline{X}_n))$$
$$\mathbb{Q} \star Z_n \sim N(\mu_n, C_{p_n}) \quad (\mathcal{S}(\underline{X}_n))$$

where the variance-covariance matrix is given by the Kronecker product

$$C_{p_n} = (I_k - \sqrt{p_n(G)} \sqrt{p_n(G)}^t) \otimes (I_m - \sqrt{p_n(B_n)} \sqrt{p_n(B_n)}^t)$$

The means μ_n of the conditional distributions under contiguous $k-$sample alternatives are given by

$$\mu_n = \left(\sqrt{p_n(G_j)p_n(B_{n,s})}\mathbb{E}_{\underline{X}_n}(g_j - \bar{h}|B_{n,s})\right)_{\substack{s=1,2,\ldots,m,\\ j=1,2,\ldots,k,}}$$

where

$$\mathbb{E}_{\underline{X}_n}(g_j - \bar{h}|B_{n,s}) = \frac{1}{p_n(B_{n,s})}\frac{1}{n}\sum_{i=1}^{n} 1_{B_{n,s}(\underline{X}_n)}(X_{ni})\cdot(g_j - \bar{h})(X_{ni}).$$

The proof of Theorem (3.1) is given in section 5.

In the remaining part of this section we will show at hand of two examples how the theoretical assertion of Theorem (3.1) can be used to obtain information which is relevant of the application of statistical methods. As a first example we consider the question how to construct partitions in order to maximize the power of the χ^2-test.

From Theorem (3.1) we obtain easily the conditional limit distribution of the χ^2-test statistic. The conditional limit distributions of (Z_n) are $\mathcal{S}(\underline{X}_n)$-measurable functions of the random partitions, but the conditional limit distribution under the null-hypothesis H_0 of the statistic $\chi^2 := \sum_{s,j} Z^2_{n,sj}$ is independent of $\underline{X}_n$. Similarly to the unconditional case it is a $\chi^2_{(k-1)(m-1)}$-distribution. The number of degrees of freedom follows from the rank of the matrix C_{p_n}. Under contiguous k-sample alternatives the distribution of χ^2 conditioned on $\mathcal{S}(\underline{X}_n)$ can be approximated by a $\chi^2_{(k-1)(m-1)}(\delta^2)$-distribution, where the noncentrality parameter is given by $\delta^2 = \sum_{s,j}\mu^2_{n,sj}$. The noncentrality parameter depends on the random partitions $B_n(\underline{X}_n)$. Therefore the power of the χ^2-test depends on the method of partitioning. We are now going to analyze the influence of the partitioning method on the power of the test.

The test statistic of a χ^2-test is based on a finite-dimensional subspace of $T(H_0)^{\perp}$. Let us assume that the random variables X_{ni} take values in $E = \mathbb{R}^d$ and are centered under $\mathbb{P}$. As a finite-dimensional subspace of tangent vectors $h(t,x) = \sum_{j=1}^{k} g_j(x)1_{G_j}(t)$ for k-sample alternatives let us consider the set of those vectors, which for each sample are linear functions of the data, i.e. $g_j(x) = a_j^t x$ for some $a_j \in \mathbb{R}^d$. Then the orthogonal projection of h onto $T(H_0)^{\perp}$ is given by

$$h(t,x) - \overline{h}(x) = \sum_{j=1}^{k}(a_j - \overline{a})^t\, x \cdot 1_{G_j}(t) \quad \text{where } \overline{a} = \sum_{l=1}^{k} a_l\lambda(G_l)\,.$$

For such a tangent vector the power of the χ^2-test at any significance level is an isotonic function of the noncentrality parameter

$$\delta^2 = \sum_{s,j}\mu^2_{n,sj} = \sum_{s,j} p_n(G_j)p_n(B_{n,s})((a_j - \overline{a})^t\,\mathbb{E}_{\underline{X}_n}(x|B_{n,s}))^2. \tag{16}$$

Thus, this expression can be viewed as a simple indicator of the power of test for a particular alternative. Since the tangent vector of the actual alternative is not known it makes sense that the average of those expressions taken over all directions in the tangent space should be large. Accepting this as a reasonable goal we compute the average of the noncentrality parameters over all choices of vectors $a_j \in \mathbb{R}^d$ with $\|a_j\| = 1$ and obtain by routine calculations

$$\int\cdots\int_{a_j\in\mathbb{R}^d,\|a_j\|=1}\Big(\sum_{s,j}\mu^2_{n,sj}\Big)da_1\ldots da_k$$
$$= \frac{1}{2}\Big(1 - \sum_{l=1}^{k}\lambda(G_l)^2\Big)\Big(\sum_{s} p_n(B_{n,s})\left\|\mathbb{E}_{\underline{X}_n}(x|B_{n,s})\right\|^2\Big) + O\big(\frac{1}{n}\big). \tag{17}$$

The interpretation of this formula gives us hints for the practical application of the χ^2-test. The first factor illustrates the fact that equal sample sizes maximize the average noncentrality parameter since the sum of squares $\sum_{l=1}^{k} \lambda\,(G_j)^2$ is a measure of concentration which is minimal for $\lambda\,(G1) = \lambda\,(G_2) = \ldots = \lambda\,(G_k)$. Even more interesting is the role of the second factor

$$\sum_s p_n(B_{n,s}) \left\|\mathbb{E}_{\underline{X}_n}(x|B_{n,s})\right\|^2 . \tag{18}$$

This second factor shows how the partition $B_n(\underline{X}_n)$ governs the power of the χ^2-test. Maximizing this factor is a familiar optimization problem of statistical cluster analysis, in fact it is equivalent to constructing a minimal variance partition. This is a combinatorial optimization problem for which several computer intensive algorithms are available. From (17) it follows that constructing a minimal variance partition maximizes the average noncentrality parameter for a plausible subspace of contiguous alternatives. The interesting feature of these results is the fact that we are dealing with distributions conditioned on $\mathcal{S}(\underline{X}_n)$. Thus, the partitions may be constructed without applying any statistical model, only using the information contained in the observed data. This is an adaptive approach which does not affect the power of the test, as long as the data-driven partitions depend on the data in a permutation symmetric way.

The second example is concerned with the question where the deviations between the k samples are located in case the null-hypothesis is rejected. We will give an answer in the form of confidence intervals which we obtain from our result on the conditional distributions under contiguous alternatives. For this kind of answer it is essential that the conditional distributions under the null-hypothesis and under contiguous alternatives differ only by a translation.

From the limit theorems for Z_n we can obtain critical values and even the power function of tests based on the statistic $\max_{s,j} |Z_{n,sj}|$. If $c = c(\underline{X}_n)$ is an approximate critical value for level α, i.e.

$$\mathbb{P}(\max_{s,j} |Z_{n,sj}| \le c(\underline{X}_n)|\mathcal{S}(\underline{X}_n)) \ge 1 - \alpha + o(1),$$

then it follows from Theorem (2.9) that

$$\mathbb{Q}(\max_{s,j} |Z_{n,sj} - \mu_{nsj}| \le c(\underline{X}_n)|\mathcal{S}(\underline{X}_n)) \ge 1 - \alpha + o(1),$$

or in other words, we obtain the confidence set

$$\mathbb{Q}\left(\bigcap_{s,j}\left\{Z_{n,sj} - c(\underline{X}_n) \le \mu_{nsj} \le Z_{n,sj} + c(\underline{X}_n)\right\}\middle|\, \mathcal{S}(\underline{X}_n)\right) \ge 1 - \alpha + o(1). \tag{19}$$

This confidence set makes it possible to identify some of the k samples whose tangent vector components differ from the average, i.e. where $\mu_{nj} \neq 0$, by testing the set of hypotheses $\mu_{nj} = 0$, $j = 1, 2, \ldots, k$, simultaneously. In fact, let

$$J_0 := \{j : \exists\, s \,\text{with}\, |\mu_{nsj}| > 0\}$$

be the set of samples where the hypothesis $\mu_{nj} = 0$ is false and let

$$J(\underline{X}_n) := \{j : \exists\, s \,\text{with}\, |\hat{Z}_{n,sj}| > c(\underline{X}_n)\}$$

be the random set of samples where the data indicate a deviation. Then we have

$$\bigcap_{s,j}\{Z_{n,sj} - c(\underline{X}_n) \le \mu_{nsj} \le Z_{n,sj} + c(\underline{X}_n)\} \subseteq \{J(\underline{X}_n) \subseteq J_0\}$$

and it follows from (19) that

$$\mathbb{Q}(J(\underline{X}_n) \subseteq J_0|\mathcal{S}(\underline{X}_n)) \ge 1 - \alpha + o(1).$$

Hence, the probability of no false rejection is at least $1 - \alpha$.

4 Probabilistic Prerequisites

In this section we collect some probabilistic facts which are needed for the proofs of our main results.

Let $(\Omega, \mathcal{A}, P)$ be a probability space and let $\xi_1, \xi_2, \ldots, \xi_k$ be random variables from $(\Omega, \mathcal{A})$ to $(\mathbb{R}^d, \mathcal{B}^d)$. Let us denote the vector of random variables by $\underline{\xi}_k = (\xi_1, \xi_2, \ldots, \xi_k)^t$ and by $\mathcal{F}_k = \underline{\xi}_k^{-1}(\mathcal{B}^d)$ the generated σ-field. A permutation of $1, 2, \ldots, k$ is denoted by π. If $\underline{x}_k = (x_1, x_2, \ldots, x_k)$, then $\pi(\underline{x}_k) = (x_{\pi(1)}, x_{\pi(2)}, \ldots, x_{\pi(k)})$ is the vector permuted by π. Let $\pi(A)$ be the set of all vectors in $A \subseteq \mathbb{R}^{k \times d}$ permuted by π. The group of all permutations of order k will be denoted by Π_k. A measurable function f is said to be permutation symmetric (with respect to $\underline{\xi}_k$) if

$$f(\underline{\xi}_k) = f(\pi(\underline{\xi}_k)) \quad \text{for all } \pi \in \Pi_k.$$

It is easy to see that f is permutation symmetric (with respect to $\underline{\xi}_k$) iff $f(\underline{\xi}_k)$ is $\mathcal{S}(\underline{\xi}_k)$-measurable.

(4.1) DEFINITION *The random variables $\xi_1, \xi_2, \ldots \xi_k$ are exchangeable under P if the random vectors $\pi(\underline{\xi}_k)$ have the same distribution under P for all $\pi \in \Pi_k$.*

If $f : (\mathbb{R}^d, \mathcal{B}^d)^k \to (\mathbb{R}, \mathcal{B})$ is a measurable function and if the random variables $\xi_1, \xi_2, \ldots, \xi_k$ are exchangeable under P, then we have by Lemma (6.2)

$$\mathbb{E}(f(\underline{\xi}_k)|\mathcal{S}(\underline{\xi}_k)) = \frac{1}{k!} \sum_{\pi} f(\pi(\underline{\xi}_k)). \tag{20}$$

Similar combinatorial formulas are valid for conditional variances and covariances (see Remarks (6.4)).

Let $\underline{\xi}_n = (\xi_{n1}, \xi_{n2}, \ldots, \xi_{nn})^t$ be a triangular array of random variables with values in $\mathbb{R}^d$. The triangular array is called infinitesimal if

$$\max_{1 \le i \le n} \|\xi_{ni}\| \xrightarrow{P} 0 \quad \text{as } n \to \infty$$

Let further $\mathcal{S}(\underline{\xi}_n)$ be the symmetric σ-field generated by $\underline{\xi}_n$ and let $S_n(t) = \sum_{i=1}^{[nt]} \xi_{ni}$ be the partial sum process. The partial sum process is called standardized if $\sum_{i=1}^{n} \xi_{ni} = 0$ and $\sum_{i=1}^{n} \xi_{ni}\xi_{ni}^t = I_d$. Throughout the following $(B_t^0)_{0 \le t \le 1}$ is a $d-$dimensional Brownian bridge on $(\Omega, \mathcal{A}, P)$ which is independent of $(\underline{\xi}_n)_{n \in \mathbb{N}}$.

(4.2) THEOREM *Suppose that the triangular array is infinitesimal and standardized. If the triangular array is exchangeable under P, then we have*

$$(S_n(t))_{0 \le t \le 1} \overset{w}{\sim} (B_t^0)_{0 \le t \le 1}.$$

The assertion of Theorem (4.2) is a multivariate version of the classical invariance principle by Billingsley, [1], for exchangeable random variables. This multivariate version is a nontrivial extension of the one-dimensional case. It seems not to be an immediate consequence of the one-dimensional result by Billingsley, [1]. For details see Strasser and Weber, [20].

Theorem (4.2) can easily be formulated as an invariance principle for conditional distributions under $\mathcal{S}(\underline{\xi}_n)$. We are going a step further, stating an invariance principle for distributions which are conditioned even on super-σ-fields $\mathcal{C}_n \supseteq \mathcal{S}(\underline{\xi}_n)$.

(4.3) REMARK At this point we have to explain why the $\sigma-$fields $\mathcal{S}(\underline{\xi}_n)$ have to be replaced by super-σ-fields $\mathcal{C}_n \supseteq \mathcal{S}(\underline{\xi}_n)$. The reason is that for the proof of our main results we will apply Theorem (4.2) to composed functions of the form $\xi_i = f(X_i, \underline{X})$, $i = 1, 2, \ldots, n$. In section 6, Lemma (6.5), ff., we consider such situations a little more detailed. In particular, we show that in such cases one has $\mathcal{S}(\underline{\xi}_n) \subseteq \mathcal{S}(\underline{X})$.

Since we are interested in limit theorems conditioned on $\mathcal{S}(\underline{X})$ we require a version of Theorem (4.2), which is conditioned on a sequence of super-σ-fields $\mathcal{C}_n \supseteq \mathcal{S}(\underline{X}) \supseteq \mathcal{S}(\underline{\xi}_n)$.

For this let us extend the notion of exchangeability.

(4.4) DEFINITION *The random variables $\xi_{n1}, \xi_{n2}, \ldots, \xi_{nk}$ are exchangeable under $P(\,.\, |\mathcal{C}_n)$, if for all permutations $\pi \in \Pi_k$ the equation*

$$P(\underline{\xi}_n \in B|\mathcal{C}_n) = P(\pi(\underline{\xi}_n) \in B|\mathcal{C}_n) \quad P-\text{a.e.}$$

is valid for all $B \in \mathcal{B}^k$.

Obviously, the random variables are exchangeable under P iff they are exchangeable under $P(\,.\, |\mathcal{S}(\underline{\xi}_k))$ (Corollary (6.3)). Moreover, it is clear that exchangeability remains true if the conditioning σ-field is replaced by a smaller one. Hence, increasing the σ-field gives a stronger form of conditional exchangeability.

(4.5) REMARK Let us introduce a concept of convergence for conditional distributions. Assume that S is a metric space, $\mathcal{S}$ is the Borel σ-field on S, and Φ is the set of all bounded, real-valued and Lipschitz-continuous functions $\phi : S \to \mathbb{R}$. Let further $(Y_n)_{n\in\mathbb{N}}$ and $(Z_n)_{n\in\mathbb{N}}$ be sequences of random variables with values in S, and let $(\mathcal{F}_n)_{n\in\mathbb{N}}$, $\mathcal{F}_n \subseteq \mathcal{A}$, be a sequence of sub-$\sigma$-fields. We say that the sequences (Y_n) and (Z_n) are weakly asymptotically equal, conditioned on $(\mathcal{F}_n)_{n\in\mathbb{N}}$, in symbols $Y_n \overset{w}{\sim} Z_n$ $(\mathcal{F}_n)$,if

$$\mathbb{E}(\phi(Y_n)|\mathcal{F}_n) - E(\phi(Z_n)|\mathcal{F}_n) \xrightarrow{P} 0, \quad \text{for every } \phi \in \Phi.$$

This concept will be applied both for $S = \mathbb{R}^d$ and for $S = C([0,1])$. Weak asymptotic equality contains as special cases the familiar concept of weak convergence and Definition (2.1) of conditional asymptotic normality.

The following invariance principle is the probabilistic basis of the statistical results of the present paper.

(4.6) THEOREM *Suppose that the triangular array is infinitesimal and that the random variables*

$$\textstyle\sum_{i=1}^{n}\xi_{ni} \quad \text{and} \quad s^2(\underline{\xi}_n) = \displaystyle\sum_{i=1}^{n}(\xi_{ni} - \bar{\xi}_n)(\xi_{ni} - \bar{\xi}_n)^t \tag{21}$$

are stochastically bounded. If $\mathcal{C}_n \supseteq \mathcal{S}(\underline{\xi}_n)$, $n \in \mathbb{N}$, is a sequence of σ-fields and if the rows of the triangular array are exchangeable under $P(\,.\, |\mathcal{C}_n)$, then

$$(S_n(t))_{0\le t\le 1} \overset{w}{\sim} \left(\sqrt{s^2(\underline{\xi}_n)}\, B_t^0 + t\textstyle\sum_{i=1}^{n}\xi_{ni}\right)_{0\le t\le 1} \quad (\mathcal{C}_n)$$

This Theorem (4.6) is the main result of Strasser and Weber, [20].

Next we are going to consider weighted partial sums of the triangular array. Let (w_{ni}) be a triangular array of weights in $\mathbb{R}^{m\times d}$. Each of those arrays can be generated by a sequence of functions $w_n \in L^2([0,1];\mathbb{R}^{m\times d})$ defining

$$w_{ni} := n\int_{\frac{i-1}{n}}^{\frac{i}{n}} w_n(t)dt, \quad i = 1, 2, \ldots, n.$$

With this construction in mind we may conveniently write weighted partial sums as path integrals

$$\sum_{i=1}^{n} w_{ni}\xi_{ni} = \int_0^1 w_n(t)\, dS_n'(t) \tag{22}$$

of the linearly interpolated partial sum processes defined by

$$S_n'(t) := S_n\left(\frac{i-1}{n}\right) + n\left(t - \frac{i-1}{n}\right)\left(S_n\left(\frac{i}{n}\right) - S_n\left(\frac{i-1}{n}\right)\right),$$

whenever $\frac{i-1}{n} \leq t < \frac{i}{n}$. We will apply this integral representation in the proof of the following theorem.

(4.7) THEOREM *Suppose that the triangular array is infinitesimal and that the random variables (21) are stochastically bounded. Assume further that* $(w_n) \subseteq L^2([0,1];\mathbb{R}^{m\times d})$ *is a relatively compact sequence of weight functions. If* $\mathcal{C}_n \supseteq \mathcal{S}(\underline{\xi}_n)$, $n \in \mathbb{N}$, *is a sequence of* σ*-fields and if the rows of the triangular array are exchangeable under* $P(\,.\,|\mathcal{C}_n)$, *then*

$$P * \sum_{i=1}^{n} w_{ni}\xi_{ni} \sim \mathcal{N}(\mu_n, \Sigma_n) \quad (\mathcal{C}_n),$$

where

$$\mu_n = \int_0^1 w_n(t)dt \cdot \sum_{i=1}^{n} \xi_{ni},$$

and

$$\Sigma_n = \int_0^1 w_n(t)\cdot s^2(\underline{\xi}_n)\cdot w_n^t(t)\,dt - \int_0^1 w_n(t)\,dt\cdot\, s^2(\underline{\xi}_n)\cdot \int_0^1 w_n^t(t)\,dt. \tag{23}$$

Proof: In the following $||\cdot||$ denotes a matrix norm. It is clear by inifinitesimality that the linearly interpolated processes $S_n'(t)$ are asymptotically indistinguishable from the partial sum processes themselves. The weighted partial sums (22) are linear function of the weights w_n. If the triangular array is exchangeable under $P(\,.\,|\mathcal{C}_n)$, then from Remarks (6.4), (31), it follows that

$$E(\|\int_0^1 w_n(t)\,dS_n'(t)\|^2|\mathcal{S}(\underline{X}_n)) \leq \frac{1}{n-1}\sum_{i=1}^{n}\|w_{ni}\|^2 \leq \frac{n}{n-1}\int_0^1 \|w_n(t)\|^2\,dt,$$

which implies that the weighted partial sums (22) are even uniformly equicontinuous in w_n. A similar assertion is true of the corresponding path integrals of the Brownian bridge (Remark (6.9)). Hence it is sufficient to prove the assertion of the theorem for fixed weight functions w in a set $M \subseteq L^2([0,1];\mathbb{R}^{m\times d})$ such that $\overline{\mathrm{span}}(M) = L^2([0,1];\mathbb{R}^{m\times d})$. Let us take for M the set of all weight functions of the kind $w(s) = x1_{(t_1,t_2]}(s)$ with $x \in \mathbb{R}^{m\times d}$. Thus, suppose that $w(s) = x1_{(t_1,t_2]}(s)$. From Theorem (4.6) it follows that

$$S_n'(t_2) - S_n'(t_1) \overset{w}{\sim} \sqrt{s^2(\underline{\xi}_n)}\,(B_{t_2}^0 - B_{t_1}^0) + (t_2 - t_1)n\bar{\xi}_n \quad (\mathcal{C}_n)$$

and therefore

$$\int_0^1 w(t)\,dS_n'(t) \overset{w}{\sim} x\sqrt{s^2(\underline{\xi}_n)}\,(B_{t_2}^0 - B_{t_1}^0) + x(t_2 - t_1)n\bar{\xi}_n \quad (\mathcal{C}_n)$$

This implies the assertion for $w(s) = x1_{(t_1,t_2]}(s)$. □

(4.8) REMARK The parameters μ_n and Σ_n can also be expressed through the weights w_{ni} themselves. It is clear that

$$\int_0^1 w_n(t)dt = \frac{1}{n}\sum_{i=1}^{n} w_{ni}.$$

Since the sequence $(w_n) \subseteq L^2([0,1];\mathbb{R}^{m\times d})$ is relatively compact, it follows from theorems on the L^2-convergence of martingales (noting that the w_{ni} are conditional means of w_n) that

$$\sum_{i=1}^{n}\int_{\frac{i-1}{n}}^{\frac{i}{n}} \|w_{ni} - w_n(t)\|^2\,dt \to 0,$$

and this implies

$$\Sigma_n = \frac{1}{n}\sum_{i=1}^{n} w_{ni}\cdot s^2(\underline{\xi}_n)\cdot w_{ni}^t - \frac{1}{n}\sum_{i=1}^{n} w_{ni}\cdot\ s^2(\underline{\xi}_n)\cdot\frac{1}{n}\sum_{l=1}^{n} w_{nl} + o(1).$$

The extension of Theorem (4.7) to random weights runs along the lines of Remark (2.11).

5 Proofs

Let us start with the situation of section 2. In particular, we suppose the validity of the assumptions (2.2). For the convenient representation of linear statistics as path integrals we rely on the vector valued process

$$F_n(t) = \frac{1}{\sqrt{n}}\sum_{i=1}^{[nt]}(f_n(X_{ni},\underline{X}_n) - \mathbb{E}(f_n|\mathcal{S}(\underline{X}_n))),\quad 0\le t\le 1, \tag{24}$$

and on the corresponding interpolated process $(F_n'(t))$. By means of these processes we may write the weighted sums

$$T_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} w_{ni} f_n(X_{ni},\underline{X}_n),$$

as path integrals

$$T_n - \mathbb{E}(T_n|\mathcal{S}(\underline{X}_n)) = \int_0^1 w_n(t)\, dF_n'(t) \tag{25}$$

and apply Theorem (4.7).

Let us start with the proof of Theorem (2.3).

Proof: (of Theorem (2.3)) Let

$$\xi_{ni} := \frac{1}{\sqrt{n}}(f_n(X_{ni},\underline{X}_n) - \mathbb{E}(f_n|\mathcal{S}(\underline{X}_n)).$$

By (6.6) the random variables $\xi_{n1},\xi_{n2},\dots,\xi_{nn}$ are exchangeable under $\mathbb{P}(\ .\ |\mathcal{S}(\underline{X}_n))$. Using notations (24) and (25), it follows from Theorem (4.7) with $\mathcal{C}_n = \mathcal{S}(\underline{X}_n)$, that

$$\mathbb{P}\star(T_n - \mathbb{E}(T_n|\mathcal{S}(\underline{X}_n)) \sim \mathcal{N}(0,\Sigma_n)\quad (\mathcal{S}(\underline{X}_n)),$$

where

$$\begin{aligned}\Sigma_n = &\int_0^1 w_n(t)\cdot\mathbb{V}(f_n|\mathcal{S}(\underline{X}_n))\cdot w_n^t(t)\, dt\\ &-\int_0^1 w_n(s)\, ds\cdot\mathbb{V}(f_n|\mathcal{S}(\underline{X}_n))\cdot\int_0^1 w_n^t(t)\, dt.\end{aligned}$$

By (7) we have

$$\begin{aligned}\mathbb{V}(T_n|\mathcal{S}(\underline{X}_n)) = &\frac{1}{n-1}\sum_{i=1}^{n} w_{ni}\cdot\mathbb{V}(f_n|\mathcal{S}(\underline{X}_n))\cdot {w_{ni}}^t\\ &-\frac{1}{n-1}\sum_{i=1}^{n} w_{ni}\cdot\mathbb{V}(f_n|\mathcal{S}(\underline{X}_n))\cdot\frac{1}{n}\sum_{l=1}^{n} {w_{nl}}^t.\end{aligned}$$

Similarly to Remark (4.8) it follows that $\mathbb{V}(T_n|\mathcal{S}(\underline{X}_n)) - \Sigma_n \xrightarrow{\mathbb{P}} 0$. □

As a next step we apply Theorem (2.3) in order to study the joint distribution of the statistics T_n and another linear statistic U_n. The following lemma will lead us to a generalization of the so-called 3rd Lemma by LeCam, if we replace the statistics U_n by the loglikelihoods of contiguous alternatives.

(5.1) LEMMA *Suppose that the triangular array*

$$\frac{1}{\sqrt{n}}\left(f_n(X_{ni},\underline{X}_n) - \mathbb{E}(f_n|\mathcal{S}(\underline{X}_n))\right), \quad 1 \le i \le n,\, n \in \mathbb{N},$$

is infinitesimal and that the conditional variances $\mathbb{V}(f_n|\mathcal{S}(\underline{X}_n))$ are stochastically bounded. Assume further that the sequence of weight functions $(w_n) \subseteq L^2([0,1];\mathbb{R}^d)$ is relatively compact and let (h_n) be a relatively compact sequence of functions in $L^2([0,1]\times E, \lambda\times P)$. Then the joint distributions of

$$U_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(h_{ni}(X_{ni}) - \overline{h}_n(X_{ni})),$$

and T_n, conditioned on $\mathcal{S}(\underline{X}_n)$, are asymptotically normal, i.e.

$$\mathbb{P}\star\begin{pmatrix} T_n - \mathbb{E}(T_n|\mathcal{S}(\underline{X}_n)) \\ U_n \end{pmatrix} \sim \mathcal{N}\left(0, \begin{pmatrix} \mathbb{V}(T_n|\mathcal{S}(\underline{X}_n)) & \mathbb{C}\mathbb{V}(T_n,U_n|\mathcal{S}(\underline{X}_n)) \\ \mathbb{C}\mathbb{V}(T_n,U_n|\mathcal{S}(\underline{X}_n))^t & \mathbb{V}(U_n|\mathcal{S}(\underline{X}_n)) \end{pmatrix}\right) \quad (\mathcal{S}(\underline{X}_n))\,.$$

Proof: It is sufficient to prove the assertion for a subset of functions h which is dense in $L^2([0,1]\times E, \lambda\otimes P)$. The extension to relatively compact sequences $(h_n) \subseteq L^2([0,1]\times E, \lambda\otimes P)$ is then a matter of routine. As a dense subset let us take the set of all bounded step functions on interval partitions. Thus, let

$$h(t,.) = \sum_{j=1}^{k} g_j(.)1_{(t_{j-1},t_j]}(t),$$

where g_j are bounded functions in $L^2(E,P)$. Let us denote $v_j(t) := 1_{(t_{j-1},t_j]}(t)$ and

$$v_{j,ni} := n\int_{\frac{i-1}{n}}^{\frac{i}{n}} 1_{(t_{j-1},t_j]}(t)\,dt \quad \text{and} \quad \overline{v}_j := \frac{1}{n}\sum_{i=1}^{n} v_{j,ni} = t_j - t_{j-1}.$$

Next, we are going to simplify the random variables

$$U_n - \mathbb{E}(U_n|\mathcal{S}(\underline{X}_n)) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(h_{ni} - \overline{h})(X_{ni})$$

Easy calculations show us that

$$h_{ni}(X_{ni}) = \sum_{j=1}^{k} v_{j,ni}g_j(X_{ni}) \quad \text{and} \quad \overline{h}(X_{ni}) = \sum_{j=1}^{k}\overline{v}_j g_j(X_{ni}),$$

whence we obtain

$$\sum_{i=1}^{n}\overline{h}(X_{ni}) = n\sum_{j=1}^{k}\overline{v}_j\mathbb{E}(g_j|\mathcal{S}(\underline{X}_n)).$$

This gives

$$U_n - \mathbb{E}(U_n|\mathcal{S}(\underline{X}_n)) = \sum_{i=1}^{n}\sum_{j=1}^{k} v_{j,ni}\frac{1}{\sqrt{n}}\left(g_j(X_{ni}) - \mathbb{E}(g_j|\mathcal{S}(\underline{X}_n))\right)$$
$$= \int_0^1 v^t(t)\, dG'_n(t),$$

where $v = (v_1, v_2, \dots v_k)^t$, $g = (g_1, g_2, \dots g_k)^t$ and $G'_n(t)$ denotes the interpolated version of

$$G_n(t) := \frac{1}{\sqrt{n}}\sum_{i=1}^{[nt]}\left(g_j(X_{ni}) - \mathbb{E}(g_j|\mathcal{S}(\underline{X}_n))\right).$$

Thus, we arrive at

$$\begin{pmatrix} T_n - \mathbb{E}(T_n|\mathcal{S}(\underline{X}_n)) \\ U_n - \mathbb{E}(U_n|\mathcal{S}(\underline{X}_n)) \end{pmatrix} = \int_0^1 \begin{pmatrix} w_n(t) & 0 \\ 0 & v^t(t) \end{pmatrix} d\begin{pmatrix} F'_n(t) \\ G'_n(t) \end{pmatrix}$$

Defining

$$\xi_{ni} := \frac{1}{\sqrt{n}}\begin{pmatrix} f_n(X_{ni}, \underline{X}_n) - \mathbb{E}(f_n|\mathcal{S}(\underline{X}_n)) \\ g(X_{ni}) - \mathbb{E}(g|\mathcal{S}(\underline{X}_n)) \end{pmatrix}$$

and $\mathcal{C}_n := \mathcal{S}(\underline{X}_n)$ the assumptions of Theorem (4.7) are satisfied (cf. Lemma (6.6) and Lemma (6.5). This gives

$$\mathbb{P}\star\begin{pmatrix} T_n - \mathbb{E}(T_n|\mathcal{S}(\underline{X}_n)) \\ U_n - \mathbb{E}(U_n|\mathcal{S}(\underline{X}_n)) \end{pmatrix} \sim \mathcal{N}\left(0, \begin{pmatrix} \Sigma_{n(11)} & \Sigma_{n(12)} \\ \Sigma^t_{n(12)} & \Sigma_{n(22)} \end{pmatrix}\right) \quad (\mathcal{S}(\underline{X}_n)),$$

where

$$\Sigma_{n(11)} = \int_0^1 w_n(t)\,\mathbb{V}(f_n|\mathcal{S}(\underline{X}_n))w_n^t(t)\,dt$$
$$- \int_0^1 w_n(t)\,dt\,\mathbb{V}(f_n|\mathcal{S}(\underline{X}_n))\int_0^1 w_n^t(s)\,ds,$$

$$\Sigma_{n(12)} = \int_0^1 w_n(t)\,\mathbb{CV}(f_n, g|\mathcal{S}(\underline{X}_n))v(t)\,dt$$
$$- \int_0^1 w_n(t)\,dt\,\mathbb{CV}(f_n, g|\mathcal{S}(\underline{X}_n))\int_0^1 v(s)\,ds,$$

$$\Sigma_{n(22)} = \int_0^1 v^t(t)\,\mathbb{V}(g|\mathcal{S}(\underline{X}_n))v(t)\,dt$$
$$- \int_0^1 v^t(t)\,dt\,\mathbb{V}(g|\mathcal{S}(\underline{X}_n))\int_0^1 v(s)\,ds.$$

Since $\mathbb{E}(U_n|\mathcal{S}(\underline{X}_n)) = 0$, it remains to be shown that

$$\mathbb{V}(T_n|\mathcal{S}(\underline{X}_n)) - \Sigma_{n,11} \xrightarrow{\mathbb{P}} 0, \tag{26}$$

$$\mathbb{CV}(T_n, U_n|\mathcal{S}(\underline{X}_n)) - \Sigma_{n,12} \xrightarrow{\mathbb{P}} 0, \tag{27}$$

$$\mathbb{V}(U_n|\mathcal{S}(\underline{X}_n)) - \Sigma_{n,22} \xrightarrow{\mathbb{P}} 0. \tag{28}$$

The assertions (26) and (28) are a consequence of Remark (4.8) and are proved in a similar way as in the proof of Theorem (2.3). The assertion (27) requires some additional arguments.

To begin with, we simplify $\mathbb{CV}(T_n, U_n|\mathcal{S}(\underline{X}_n))$ replacing in (14) means by integrals. This gives

$$\mathbb{CV}(T_n, U_n|\mathcal{S}(\underline{X}_n)) = \frac{1}{n-1}\sum_{i=1}^{n} w_{ni}\mathbb{CV}(f_n, h_{ni} - \bar{h}|\mathcal{S}(\underline{X}_n))$$
$$= \int_0^1 w_n(t)\mathbb{CV}(f_n, h(t,.) - \bar{h}|\mathcal{S}(\underline{X}_n))\,dt + o_{\mathbb{P}}(1).$$

Next we simplify $\Sigma_{n(12)}$. For the first term of the formula for $\Sigma_{n(12)}$ we have

$$\int_0^1 w_n(t)\,\mathbb{CV}(f_n, g|\mathcal{S}(\underline{X}_n))v(t)\,dt$$
$$= \int_0^1 w_n(t)\,\mathbb{E}(f_n(g - \mathbb{E}(g|\mathcal{S}(\underline{X}_n)))^t v(t)|\mathcal{S}(\underline{X}_n))\,dt$$
$$= \int_0^1 w_n(t)\,\mathbb{E}(f_n(h(t,.) - \mathbb{E}(h(t,.)|\mathcal{S}(\underline{X}_n)))^t|\mathcal{S}(\underline{X}_n))\,dt$$
$$= \int_0^1 w_n(t)\,\mathbb{CV}(f_n, h(t,.)|\mathcal{S}(\underline{X}_n))\,dt.$$

Similarly, for the second term of the formula of $\Sigma_{n(12)}$ we have

$$\int_0^1 w_n(t)\,dt\ \ \mathbb{CV}(f_n, g|\mathcal{S}(\underline{X}_n))\int_0^1 v(t)\,dt$$
$$= \int_0^1 w_n(t)\,dt\ \ \mathbb{E}(f_n(g - \mathbb{E}(g|\mathcal{S}(\underline{X}_n)))^t\int_0^1 v(t)\,dt|\mathcal{S}(\underline{X}_n))$$
$$= \int_0^1 w_n(t)\,dt\ \ \mathbb{E}(f_n(\overline{h} - \mathbb{E}(\overline{h}|\mathcal{S}(\underline{X}_n)))^t|\mathcal{S}(\underline{X}_n))\,dt$$
$$= \int_0^1 w_n(t)\,dt\ \ \mathbb{CV}(f_n, \overline{h}|\mathcal{S}(\underline{X}_n))\,dt.$$

□

Now, we are in the position to prove Theorem (2.7). We will use the notation of (2.5).

Proof: (of Theorem (2.7).)Recall that by (12) we have to prove that

$$\mathbb{E}(\exp(\beta_n + r_n(\underline{X}_n)|\mathcal{S}(\underline{X}_n)) \xrightarrow{\mathbb{P}} 1.$$

It is sufficient to prove this for those pairs $(\mathbb{P}, \mathbb{Q})$ for which $Q_{ni} = (1 + \frac{1}{\sqrt{n}}h_{ni})P$ where $h \in L^*(P)$ is bounded (cf. Strasser, [16]). Under these assumptions the remainders $r_n(\underline{X}_n)$ are even uniformly bounded (cf. Strasser, [16]) and $\exp(\beta_n)$ is uniformly $\mathbb{P}-$integrable (Theorem (8.6) in Strasser, [16]). Thus, it remains to show that

$$\mathbb{E}(\exp(\beta_n)|\mathcal{S}(\underline{X}_n)) \xrightarrow{\mathbb{P}} 1. \tag{29}$$

For the proof we apply Lemma (5.1). Observing that $\mathbb{E}(\beta_n|\mathcal{S}(\underline{X}_n)) - \frac{1}{2}\mathbb{V}(\beta_n|\mathcal{S}(\underline{X}_n)) \xrightarrow{\mathbb{P}} 0$ it follows from Lemma (5.1) that

$$\mathbb{P} \star \beta_n \sim \mathcal{N}(-\frac{1}{2}\mathbb{V}(\beta_n|\mathcal{S}(\underline{X}_n), \mathbb{V}(\beta_n|\mathcal{S}(\underline{X}_n)) \quad (\mathcal{S}(\underline{X}_n))\,.$$

Then the assertion (29) follows from

$$\int \exp(x)\,d\mathcal{N}(-\frac{1}{2}a^2, a^2)(dx) = 1$$

by straightforward arguments. □

Our next proof is concerned with Theorem (2.9).

Proof: (of Theorem (2.9).) Again it is sufficient to prove the theorem for such pairs $(\mathbb{P},\mathbb{Q})$ where $Q_{ni} = (1+\frac{1}{\sqrt{n}}h_{ni})P$ with bounded $h \in L^*(P)$ (cf. Strasser, [16]). Let ϕ be a continuous and bounded function on $\mathbb{R}^d$. Then we have

$$\mathbb{E}_{\mathbb{Q}}(\phi(T_n - \mathbb{E}(T_n|\mathcal{S}(\underline{X}_n))|\mathcal{S}(\underline{X}_n)) = \frac{\mathbb{E}\left(\phi(T_n - \mathbb{E}(T_n|\mathcal{S}(\underline{X}_n))\frac{d\mathbb{Q}}{d\mathbb{P}}\middle|\,\mathcal{S}(\underline{X}_n)\right)}{\mathbb{E}\left(\frac{d\mathbb{Q}}{d\mathbb{P}}\middle|\,\mathcal{S}(\underline{X}_n)\right)}$$

and it follows in a similar way as in the proof of Theorem (2.7), that

$$\mathbb{E}_{\mathbb{Q}}(\phi(T_n - \mathbb{E}(T_n|\mathcal{S}(\underline{X}_n))|\mathcal{S}(\underline{X}_n)) - \mathbb{E}((\phi(T_n - \mathbb{E}(T_n|\mathcal{S}(\underline{X}_n)))\exp(\beta_n)|\mathcal{S}(\underline{X}_n)) \xrightarrow{\mathbb{P}} 0.$$

From Lemma (5.1) and from the fact that $\exp(\beta_n)$ is uniformly $\mathbb{P}-$integrable we obtain by calculations which are familiar from the 3rd Lemma by LeCam, that

$$\begin{aligned}
&\mathbb{E}(\phi(T_n - \mathbb{E}(T_n|\mathcal{S}(\underline{X}_n)))\exp(\beta_n)|\mathcal{S}(\underline{X}_n)) \\
&\qquad - \exp(\mathbb{E}(\beta_n|\mathcal{S}(\underline{X}_n)) - \frac{1}{2}\mathbb{V}(\beta_n|\mathcal{S}(\underline{X}_n))) \\
&\qquad\qquad \cdot \int \phi\, d\mathcal{N}\left(\mathbb{CV}(T_n,\beta_n|\mathcal{S}(\underline{X}_n)), \mathbb{V}(T_n|\mathcal{S}(\underline{X}_n))\right) \xrightarrow{\mathbb{P}} 0.
\end{aligned}$$

Observing that $\mathbb{E}(\beta_n|\mathcal{S}(\underline{X}_n)) - \frac{1}{2}\mathbb{V}(\beta_n|\mathcal{S}(\underline{X}_n)) \xrightarrow{\mathbb{P}} 0$ the assertion follows. □

Proof: (of Theorem 3.1.) In order to embed the assertion into the notation of Theorems (2.3) and (2.9), we define

$$f_{n,sj}(X_{ni},\underline{X}_n) := \frac{1_{B_{ns}(\underline{X}_n)}(X_{ni})}{\sqrt{p_n(B_{ns}(\underline{X}_n))}}, \quad 1 \le s \le m, 1 \le j \le k,$$

and

$$w_{ni,sj,tl} := \frac{1_{G_j(\frac{i}{n})}}{\sqrt{p_n(G_j)}}\delta_{st}\delta_{jl}, \quad 1 \le s,t \le m, 1 \le j,l \le k.$$

The functions f_n have values in $\mathbb{R}^{mk}$ and the weights w_{ni} are diagonal matrices with mk rows and columns. Let

$$T_n := \frac{1}{\sqrt{n}}\sum_{i=1}^n w_{ni} f_n(X_{ni},\underline{X}_n).$$

Then we have $Z_n = T_n - \mathbb{E}(T_n|\mathcal{S}(\underline{X}_n))$. The assumptions of Theorems (2.3) and (2.9) are satisfied. It remains to be shown that the expressions for $\mathbb{V}(T_n|\mathcal{S}(\underline{X}_n))$ and $\mathbb{CV}(T_n,\beta_n|\mathcal{S}(\underline{X}_n))$ are of the asserted form.

Let us start with the computation of (see (7))

$$\begin{aligned}
\mathbb{V}(T_n|\mathcal{S}(\underline{X}_n)) &= \frac{1}{n-1}\sum_i w_{ni}\mathbb{V}(f_n|\mathcal{S}(\underline{X}_n))w_{ni}^t \\
&\qquad - \frac{1}{n\,(n-1)}\sum_{i,j} w_{ni}\mathbb{V}(f_n|\mathcal{S}(\underline{X}_n))w_{nj}^t.
\end{aligned}$$

First, we obtain by easy computations that

$$\begin{aligned}
&\mathbb{CV}\left(\frac{1_{B_{ns}(\underline{X}_n)}(X_{ni})}{\sqrt{p_n(B_{ns}(\underline{X}_n))}} \frac{1_{B_{nt}(\underline{X}_n)}(X_{ni})}{\sqrt{p_n(B_{nt}(\underline{X}_n))}} \middle| \mathcal{S}(\underline{X}_n) \right)_{st} \\
&\qquad = \frac{1}{n}\sum_{i=1}^{n} \frac{1_{B_{ns}(\underline{X}_n)}(X_{ni}) - p_n(B_{ns}(\underline{X}_n))}{\sqrt{p_n(B_{ns}(\underline{X}_n))}} \frac{1_{B_{nt}(\underline{X}_n)}(X_{ni}) - p_n(B_{nt}(\underline{X}_n))}{\sqrt{p_n(B_{nt}(\underline{X}_n))}} \\
&\qquad = \frac{1}{n}\sum_{i=1}^{n} \frac{1_{B_{ns}(\underline{X}_n)\cap B_{nt}(\underline{X}_n)}(X_{ni})}{\sqrt{p_n(B_{ns}(\underline{X}_n))p_n(B_{nt}(\underline{X}_n))}} - \sqrt{p_n(B_{ns}(\underline{X}_n))p_n(B_{nt}(\underline{X}_n))},
\end{aligned}$$

whence (with $e_k = (1,1,\ldots,1)'$)

$$\mathbb{V}(f_n|\mathcal{S}(\underline{X}_n)) = e_k e_k' \otimes (I_m - \sqrt{p_n(B_n)}\sqrt{p_n(B_n)}'.$$

Morever, denoting

$$v_{ni} := \begin{pmatrix} 1_{G_1}(\frac{i}{n})/\sqrt{p_n(G_1)} \\ \vdots \\ 1_{G_k}(\frac{i}{n})/\sqrt{p_n(G_k)} \end{pmatrix},$$

we obtain that

$$w_{ni}\mathbb{V}(f_n|\mathcal{S}(\underline{X}_n))w_{nj}^t = v_{ni}v_{nj}' \otimes (I_m - \sqrt{p_n(B_n)}\sqrt{p_n(B_n)}').$$

This implies

$$\begin{aligned}
&\mathbb{V}(T_n|\mathcal{S}(\underline{X}_n)) \\
&\qquad = \left(\frac{1}{n-1}\sum_i v_{ni}v_{ni}^t - \frac{1}{n\,(n-1)}\sum_{i,j} v_{ni}v_{nj}^t \right) \otimes (I_m - \sqrt{p_n(B_n)}\sqrt{p_n(B_n)}') \\
&\qquad = \frac{n}{n-1}(I_k - \sqrt{p_n(G)}\sqrt{p_n(G)}') \otimes (I_m - \sqrt{p_n(B_n)}\sqrt{p_n(B_n)}').
\end{aligned}$$

This is the expression for $\mathbb{V}(T_n|\mathcal{S}(\underline{X}_n))$ up to a null sequence.

Now, let us turn to the computation of $\mathbb{CV}(T_n, \beta_n|\mathcal{S}(\underline{X}_n))$ The starting point is equation (14), saying that

$$\begin{aligned}
\mathbb{CV}(T_n, \beta_n|\mathcal{S}(\underline{X}_n)) &= \frac{1}{n-1}\sum_{i=1}^{n} w_{ni}\mathbb{CV}(f_n, h_{ni} - \bar{h}|\mathcal{S}(\underline{X}_n)) \\
&= \frac{1}{n-1}\sum_{i=1}^{n} w_{ni}\frac{1}{n}\sum_{l=1}^{n}(f_n(X_{nl}) - \mathbb{E}(f_n|\mathcal{S}(\underline{X}_n))(h_{ni} - \overline{h})(X_{nl}).
\end{aligned}$$

This gives us for the component with index (js)

$$\begin{aligned}
&\mathbb{CV}(T_n, \beta_n|\mathcal{S}(\underline{X}_n))_{js} \\
&\qquad = \frac{1}{n-1}\sum_{i=1}^{n} \frac{1_{G_j(\frac{i}{n})}}{\sqrt{p_n(G_j)}} \frac{1}{n}\sum_{l=1}^{n} \frac{1_{B_{ns}(\underline{X}_n)}(X_{nl}) - p_n(B_{ns}(\underline{X}_n))}{\sqrt{p_n(B_{ns}(\underline{X}_n))}}(h_{ni} - \overline{h})(X_{nl}) \\
&\qquad = \frac{n}{n-1}\sqrt{p_n(G_j)}\frac{1}{n}\sum_{l=1}^{n} \frac{1_{B_{ns}(\underline{X}_n)}(X_{nl}) - p_n(B_{ns}(\underline{X}_n))}{\sqrt{p_n(B_{ns}(\underline{X}_n))}}(g_j - \overline{h})(X_{nl}) \\
&\qquad = \frac{n}{n-1}\sqrt{p_n(G_j)p_n(B_{ns}(\underline{X}_n))}\frac{1}{p_n(B_{ns}(\underline{X}_n))} \\
&\qquad\quad \cdot \frac{1}{n}\sum_{l=1}^{n}(1_{B_{ns}(\underline{X}_n)}(X_{nl}) - p_n(B_{ns}(\underline{X}_n)))(g_j - \overline{h})(X_{nl}).
\end{aligned}$$

Since

$$\frac{1}{n}\sum_{l=1}^{n} g_j(X_{nl}) \to 0 \quad (\mathbb{P}),$$

it follows

$$\mathbb{CV}(T_n,\beta_n|\mathcal{S}(\underline{X}_n))_{js} = \sqrt{p_n(G_j)p_n(B_{ns}(\underline{X}_n))}\frac{1}{p_n(B_{ns}(\underline{X}_n))} \cdot \frac{1}{n}\sum_{l=1}^{n} 1_{B_{ns}(\underline{X}_n)}(X_{nl})(g_j-\overline{h})(X_{nl}) + o(1).$$

□

6 Auxiliary Lemmas

Let $(\Omega,\mathcal{A},P)$ be a probability space and let $\xi_1,\xi_2,\dots,\xi_k$ be random variables from $(\Omega,\mathcal{A})$ to $(\mathbb{R}^d,\mathcal{B}^d)$. The vector of random variables is denoted by $\underline{\xi}_k = (\xi_1,\xi_2,\dots\xi_k)^t$ and $\mathcal{F}_k$=$\underline{\xi}_k^{-1}(\mathcal{B}^d)$ denotes the generated σ-field.

We continue the considerations at the beginning of section 4.

Assume that $\mathcal{C}$ is a sub-σ-field of $\mathcal{A}$. Let $\mathcal{S}(\underline{\xi}_k,\mathcal{C}) := \sigma(\mathcal{S}(\underline{\xi}_k),\mathcal{C})$ be the σ-field which is generated by $\mathcal{S}(\underline{\xi}_k)$ and $\mathcal{C}$. Dealing with assertions on distributions conditioned on $\mathcal{S}(\underline{\xi}_k,\mathcal{C})$, we cover the special cases $\mathcal{C}=\{\Omega,\emptyset\}$, i.e. $\mathcal{S}(\underline{\xi}_k,\mathcal{C})=\mathcal{S}(\underline{\xi}_k)$, and $\mathcal{C}\supseteq\mathcal{S}(\underline{\xi}_k)$, i.e. $\mathcal{S}(\underline{\xi}_k,\mathcal{C})=\mathcal{C}$. The first case is a familiar one. The second case is that we are interested in (cf. Remark (4.3)).

(6.1) LEMMA *Let $f:(\mathbb{R}^d,\mathcal{B}^d)^k\to(\mathbb{R},\mathcal{B})$ be a measurable function such that $\mathbb{E}(f(\underline{\xi}_k))$ is well-defined. If $\xi_1,\xi_2,\dots,\xi_k$ are exchangeable under $P(\,.\,|\mathcal{C})$, then*

$$\mathbb{E}(1_{A\cap C}f(\pi(\underline{\xi}_k))) = \mathbb{E}(1_{A\cap C}f(\underline{\xi}_k)) \quad \textit{for all } A\in\mathcal{S}(\underline{\xi}_k) \textit{ and } C\in\mathcal{C}.$$

Proof: Let $A\in\mathcal{S}(\underline{\xi}_k)$ and $C\in\mathcal{C}$. Then there is $\tilde{A}\in B^{dk}$ such that $A=\{\underline{\xi}_k\in\tilde{A}\}$ and $\pi(\tilde{A})=\tilde{A}$, hence $1_{\tilde{A}}(\pi(\underline{\xi}_k))=1_{\tilde{A}}(\underline{\xi}_k)$. For $f=1_{\tilde{B}}$ we have

$$\begin{aligned}\mathbb{E}(1_{A\cap C}1_{\tilde{B}}(\pi(\underline{\xi}_k))) &= \mathbb{E}(1_{\tilde{A}}(\underline{\xi}_k)1_{\tilde{B}}(\pi(\underline{\xi}_k))1_C) = \mathbb{E}(1_{\tilde{A}}(\pi(\underline{\xi}_k))1_{\tilde{B}}(\pi(\underline{\xi}_k))1_C)\\ &= \mathbb{E}(1_{\tilde{A}\cap\tilde{B}}(\pi(\underline{\xi}_k))1_C) = \mathbb{E}(\mathbb{E}(1_{\tilde{A}\cap\tilde{B}}(\pi(\underline{\xi}_k))|\mathcal{C})1_C)\\ &= \mathbb{E}(\mathbb{E}(1_{\tilde{A}\cap\tilde{B}}(\underline{\xi}_k)|\mathcal{C})1_C) = \mathbb{E}(1_{\tilde{A}\cap\tilde{B}}(\underline{\xi}_k)1_C) = \mathbb{E}(1_{A\cap C}1_{\tilde{B}}(\underline{\xi}_k))\end{aligned}$$

This implies the assertion for indicator functions. The extension to measurable functions is obvious. □

In the following lemma we use the notion of the conditional expectation in the extended sense: The conditional expectation is well-defined for a non-integrable function if the conditional expectation of the positive or negative part is finite.

(6.2) LEMMA *Let $f:(\mathbb{R}^d,\mathcal{B}^d)^k\to(\mathbb{R},\mathcal{B})$ be a measurable function. If $\xi_1,\xi_2,\dots,\xi_k$ are exchangable under $P(\,.\,|\mathcal{C})$, then*

$$\mathbb{E}(f(\underline{\xi}_k)|\mathcal{S}(\underline{\xi}_k,\mathcal{C})) = \frac{1}{k!}\sum_{\pi} f(\pi(\underline{\xi}_k)).$$

Proof: To begin with, let f be bounded. Then $g(\underline{\xi}_k) = \frac{1}{k!}\sum_\pi f(\pi(\underline{\xi}_k))$ is a permutation symmetric function of $\underline{\xi}_k$. Hence it is $\mathcal{S}(\underline{\xi}_k)$-measurable and also $\mathcal{S}(\underline{\xi}_k, \mathcal{C})$-measurable. Let $A \in \mathcal{S}(\underline{\xi}_k)$ and $C \in \mathcal{C}$. By Lemma (6.1) it follows that

$$\mathbb{E}(1_{A\cap C}\frac{1}{k!}\sum_\pi f(\pi(\underline{\xi}_k))) = \frac{1}{k!}\sum_\pi \mathbb{E}(1_{A\cap C} f(\pi(\underline{\xi}_k)))$$
$$= \frac{1}{k!}\sum_\pi \mathbb{E}(1_{A\cap C} f(\underline{\xi}_k)) = \mathbb{E}(1_{A\cap C} f(\underline{\xi}_k))$$

Since $\mathcal{S}(\underline{\xi}_k, \mathcal{C})$ is generated by $\{A\cap C : A \in \mathcal{S}(\underline{\xi}_k), C \in \mathcal{C}\}$, we have

$$\mathbb{E}(1_D\frac{1}{k!}\sum_\pi f(\pi(\underline{\xi}_k))) = \mathbb{E}(1_D f(\underline{\xi}_k)) \quad \text{whenever } D \in \mathcal{S}(\underline{\xi}_k, \mathcal{C})$$

which implies the assertion for f. The extension to arbitrary measurable functions is obvious. □

If $\xi_1, \xi_2, \ldots, \xi_k$ are exchangeable under $P(\,.\,|\mathcal{C})$, then the conditional probabilities of $A \in \mathcal{F}_k$ under $\mathcal{S}(\underline{\xi}_k, \mathcal{C})$ can be given by explicit expressions: There is some $\tilde{A}$ such that $A = \underline{\xi}_k^{-1}(\tilde{A})$ and we have

$$P(A|\mathcal{S}(\underline{\xi}_k, \mathcal{C})) = \mathbb{E}(1_A|\mathcal{S}(\underline{\xi}_k, \mathcal{C})) = \mathbb{E}(1_{\tilde{A}}(\underline{\xi}_k)|\mathcal{S}(\underline{\xi}_k, \mathcal{C})) = \frac{1}{k!}\sum_\pi 1_{\tilde{A}}(\pi(\underline{\xi}_k)). \tag{30}$$

This is a regular conditional probability on $\mathcal{F}_k$ which will serve as the canonical version of $P(\,.\,|\mathcal{S}(\underline{\xi}_k, \mathcal{C}))$.

(6.3) COROLLARY *The random variables* $\xi_1, \xi_2, \ldots, \xi_k$ *are exchangeable under* $P(\,.\,|\mathcal{S}(\underline{\xi}_k, \mathcal{C}))$ *iff they are exchangeable under* $P(\,.\,|\mathcal{C})$.

Proof: From Lemma (6.2) it follows that

$$\mathbb{E}(f(\underline{\xi}_k)|\mathcal{S}(\underline{\xi}_k, \mathcal{C})) = \frac{1}{k!}\sum_\pi \mathbb{E}(f(\pi(\underline{\xi}_k))) = \mathbb{E}(f(\pi(\underline{\xi}_k))|\mathcal{S}(\underline{\xi}_k, \mathcal{C}))$$

Thus, exchangeability under $P(\,.\,|\mathcal{C})$ implies exchangeability under $P(\,.\,|\mathcal{S}(\underline{\xi}_k, \mathcal{C}))$. □

(6.4) REMARKS

1. Let $f, g : \mathbb{R}^d \to \mathbb{R}^m$ be measurable functions. If the random variables $\xi_1, \xi_2, \ldots, \xi_k$ are exchangeable under $P(\,.\,|\mathcal{C})$, then we obtain easily that

$$\mathbb{E}(f(\xi_i)|\mathcal{S}(\underline{\xi}_k, \mathcal{C})) = \frac{1}{k}\sum_{l=1}^k f(\xi_l)$$

and

$$\mathbb{E}(f(\xi_i)g(\xi_j)^t|\mathcal{S}(\underline{\xi}_k, \mathcal{C})) = \begin{cases} \frac{1}{k}\sum_{l=1}^k f(\xi_l)g(\xi_l)^t, & i = j, \\ \frac{1}{k(k-1)}\sum_{1\le l_1,l_2\le k, l_1\ne l_2} f(\xi_{l_1})g(\xi_{l_2})^t, & i \ne j. \end{cases}$$

2. Let us prove formula (7). First we obtain by easy calculations

$$\begin{aligned}
\mathbb{V}(T_n|\mathcal{S}(\underline{X}_n)) &= \mathbb{V}\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} w_{ni} f_n(X_{ni},\underline{X}_n)\Big|\mathcal{S}(\underline{X}_n)\Big) \\
&= \frac{1}{n}\sum_{i=1}^{n} w_{ni}\mathbb{V}\left(f_n|\mathcal{S}(\underline{X}_n)\right) w_{ni}^t \\
&\quad + \frac{1}{n}\sum_{i\neq j} w_{ni}\mathbb{E}\Big(\big(f_n(X_{ni},\underline{X}_n) - \mathbb{E}\left(f_n|\mathcal{S}(\underline{X}_n)\right)\big) \\
&\qquad \cdot \big(f_n(X_{nj},\underline{X}_n) - \mathbb{E}\left(f_n|\mathcal{S}(\underline{X}_n)\right)\big)^t\Big|\mathcal{S}(\underline{X}_n)\Big) w_{nj}^t
\end{aligned}$$

Moreover, by paragraph 1 we have for $i \neq j$

$$\begin{aligned}
&\mathbb{E}\Big(\big(f_n(X_{ni},\underline{X}_n) - \mathbb{E}(f_n|\mathcal{S}(\underline{X}_n))\big)\big(f_n(X_{nj},\underline{X}_n) - \mathbb{E}(f_n|\mathcal{S}(\underline{X}_n))\big)^t|\mathcal{S}(\underline{X}_n)\Big) \\
&\quad = \frac{1}{n(n-1)}\sum_{r\neq s}\big(f_n(X_{nr},\underline{X}_n) - \mathbb{E}(f_n|\mathcal{S}(\underline{X}_n))\big)\big(f_n(X_{ns},\underline{X}_n) - \mathbb{E}\left(f_n|\mathcal{S}(\underline{X}_n)\right)\big)^t \\
&\quad = \frac{1}{n(n-1)}\sum_{r,s}\big(f_n(X_{nr},\underline{X}_n) - \mathbb{E}\left(f_n|\mathcal{S}(\underline{X}_n)\right)\big)\big(f_n(X_{ns},\underline{X}_n) - \mathbb{E}\left(f_n|\mathcal{S}(\underline{X}_n)\right)\big)^t \\
&\qquad - \frac{1}{n(n-1)}\sum_{r}\big(f_n(X_{nr},\underline{X}_n) - \mathbb{E}\left(f_n|\mathcal{S}(\underline{X}_n)\right)\big)\big(f_n(X_{nr},\underline{X}_n) - \mathbb{E}\left(f_n|\mathcal{S}(\underline{X}_n)\right)\big)^t \\
&\quad = 0 - \frac{1}{n-1}\mathbb{V}\left(f_n|\mathcal{S}(\underline{X}_n)\right).
\end{aligned}$$

3. Next, let us prove formula (14). First we observe that

$$\begin{aligned}
&\mathbb{CV}(T_n,\beta_n|\mathcal{S}(\underline{X}_n)) \\
&\quad = \mathbb{E}\Big(\frac{1}{n}\sum_{i=1}^{n} w_{ni}\big(f_n(X_{ni},\underline{X}_n) - \mathbb{E}(f_n|\mathcal{S}(\underline{X}_n))\big)\sum_{j=1}^{n}(h_{nj}-\overline{h})(X_{nj})\Big|\mathcal{S}(\underline{X}_n)\Big) \\
&\quad = \frac{1}{n}\sum_{i=1}^{n} w_{ni}\mathbb{E}\big((f_n(X_{ni},\underline{X}_n) - \mathbb{E}(f_n|\mathcal{S}(\underline{X}_n)))(h_{ni}-\overline{h})(X_{ni})|\mathcal{S}(\underline{X}_n)\big) \\
&\qquad + \frac{1}{n}\sum_{i\neq j} w_{ni}\mathbb{E}\left(\left(f_n(X_{ni},\underline{X}_n) - \mathbb{E}\left(f_n|\mathcal{S}(\underline{X}_n)\right)\right)(h_{nj}-\overline{h})(X_{nj})\big|\,\mathcal{S}(\underline{X}_n)\right).
\end{aligned}$$

The first term of the sum is

$$\frac{1}{n}\sum_{i=1}^{n} w_{ni}\mathbb{CV}(f_n, h_{ni}-\overline{h}|\mathcal{S}(\underline{X}_n)).$$

The second term can be simplified to

$$\begin{aligned}
&\frac{1}{n}\sum_{i\neq j} w_{ni}\frac{1}{n(n-1)}\sum_{r\neq s}\left(f_n(X_{nr},\underline{X}_n) - \mathbb{E}\left(f_n|\mathcal{S}(\underline{X}_n)\right)\right)(h_{nj}-\overline{h})(X_{ns}) \\
&\quad = -\frac{1}{n}\sum_{i} w_{ni}\frac{1}{n(n-1)}\sum_{r\neq s}\left(f_n(X_{nr},\underline{X}_n) - \mathbb{E}\left(f_n|\mathcal{S}(\underline{X}_n)\right)\right)(h_{ni}-\overline{h})(X_{ns}) \\
&\quad = \frac{1}{n}\sum_{i} w_{ni}\frac{1}{n(n-1)}\sum_{r}\left(f_n(X_{nr},\underline{X}_n) - \mathbb{E}\left(f_n|\mathcal{S}(\underline{X}_n)\right)\right)(h_{ni}-\overline{h})(X_{nr}) \\
&\quad = \frac{1}{n(n-1)}\sum_{i=1}^{n} w_{ni}\mathbb{CV}(f_n, h_{ni}-\overline{h}|\mathcal{S}(\underline{X}_n)).
\end{aligned}$$

4. Suppose that the random variables $\xi_1, \xi_2, \dots, \xi_k$ are exchangable under $P(\,.\,|\mathcal{C})$ and let $\sum_{i=1}^k \xi_i = 0, \quad \sum_{i=1}^k \xi_i \xi_i^t = I_d$. Then we have

$$\mathbb{E}(\xi_i \xi_j^t | \mathcal{S}(\underline{\xi}_k, \mathcal{C})) = \begin{cases} \frac{1}{k} I_d, & i = j, \\ -\frac{1}{k(k-1)} I_d, & i \neq j, \end{cases}$$

and if $w \in L^2([0,1])$ then it follows that for every matrix norm $|| \cdot ||$

$$\mathbb{E}(\| \int_0^1 w \, dS'_k \|^2 | \mathcal{S}(\underline{\xi}_k, \mathcal{C})) = \frac{1}{k-1} \sum_{i=1}^k \|w_{ni}\|^2 - \frac{1}{k(k-1)} \| \sum_{i=1}^k w_{ni} \|^2. \tag{31}$$

Now, let us turn to the situation considered in Remark (4.3).

(6.5) LEMMA *Let $f : E \times E^k \to \mathbb{R}^m$ be a measurable function, which is permutation symmetric in its second variable, and let $\xi_i = f(X_i, \underline{X}_k)$, $i = 1, 2 \dots, k$. Then $\mathcal{S}(\underline{\xi}_k) \subseteq \mathcal{S}(\underline{X}_k)$.*

Proof: Suppose that $h : (\mathbb{R}^m)^k \to \mathbb{R}$ is given by $h(x) = (f(x_1, x), \dots, f(x_k, x))^t$. Then we have $\pi(h(x)) = h(\pi(x))$. Let us prove that for any measurable function $g : (\mathbb{R}^m)^k \to \mathbb{R}$ the implication

$$g(\underline{\xi}_k) \quad \mathcal{S}(\underline{\xi}_k)\text{-measurable} \Longrightarrow g(\underline{\xi}_k) \quad \mathcal{S}(\underline{X}_k)\text{-measurable}$$

is valid. Let $g(\underline{\xi}_k)$ be $\mathcal{S}(\underline{\xi}_k)$-measurable. Then $g(\underline{\xi}_k) = g(\pi(\underline{\xi}_k))$ for any permutation π. We want to show that $g(\underline{\xi}_k)$ is $\mathcal{S}(\underline{X}_k)$-measurable. For that, define $\tilde{g}(\underline{X}_k) := g(h(\underline{X}_k)) = g(\underline{\xi}_k)$. It follows that

$$\tilde{g}(\pi(\underline{X}_k)) = g(h(\pi(\underline{X}_k))) = g(\pi(h(\underline{X}_k))) = g(\pi(\underline{\xi}_k)) = g(\underline{\xi}_k) = \tilde{g}(\underline{X}_k)$$

Thus, $\tilde{g}$ is permutation symmetric and $\mathcal{S}(\underline{X}_k)$-measurable. □

(6.6) LEMMA *Let $f : E \times E^k \to \mathbb{R}^m$ be a measurable function, which is permutation symmetric in its second variable. If $X_1, X_2, \dots, X_k$ are exchangeable under $P(\,.\,|\mathcal{C})$, then $f(X_1, \underline{X}_k), f(X_2, \underline{X}_k), \dots, f(X_k, \underline{X}_k)$ are exchangeable under $P(\,.\,|\mathcal{C})$, too.*

Proof: Given any measurable function $g : (\mathbb{R}^m, \mathcal{B}^m)^k \to (\mathbb{R}, \mathcal{B})$, we show that

$$\mathbb{E}(g(\pi(f(X_1, \underline{X}_k), \dots, f(X_k, \underline{X}_k)))|\mathcal{C}) = \mathbb{E}(g(f(X_1, \underline{X}_k), \dots, f(X_k, \underline{X}_k))|\mathcal{C}) \quad P-\text{a.e.}$$

Let $h : (\mathbb{R}^m)^k \to \mathbb{R}$ be defined by $h(x) = (f(x_1, x), \dots f(x_k, x))^t$. By assumption we have

$$\mathbb{E}(g(h(\pi(\underline{X}_k)))|\mathcal{C}) = \mathbb{E}(g(h(\underline{X}_k))|\mathcal{C}).$$

From $\pi(h(x)) = h(\pi(x))$ it follows that

$$\mathbb{E}(g(\pi(h(\underline{X}_k)))|\mathcal{C}) = \mathbb{E}(g(h(\pi(\underline{X}_k)))|\mathcal{C}) = \mathbb{E}(g(h(\underline{X}_k))|\mathcal{C}).$$

□

(6.7) REMARK Suppose that $f : E \times E^k \to \mathbb{R}^m$ and $g : \mathbb{R}^m \times (\mathbb{R}^m)^k \to \mathbb{R}^r$ are permutation symmetric with respect to the second variable. Then the function $h : E \times E^k \to \mathbb{R}^r$, which is defined by $h(x, y) = g(f(x, y), f(y_1, y), f(y_2, y), \dots f(y_k, y))$, is also permutation symmetric with respect to the second variable.

(6.8) LEMMA *Let $\bar{x} := \frac{1}{k} \sum_{i=1}^k x_i$, $x = (x_1, x_2, \dots, x_k) \in (\mathbb{R}^m)^k$, and assume that $V(x) := \sum_{i=1}^k (x_i - \bar{x})(x_i - \bar{x})^t$ is positive definite. Let $g : \mathbb{R}^m \times (\mathbb{R}^m)^k \to \mathbb{R}^m$ be defined by $(y, x) \mapsto V(x)^{-\frac{1}{2}}(y - \bar{x})$ and* *let* $\eta_i = g(\xi_i, \underline{\xi}_k)$, $i = 1, 2, \ldots k$. *If* $\xi_1, \xi_2, \ldots, \xi_k$ *are exchangeable under* $P(. \,|\mathcal{C})$ *and if* $\bar{\xi}$ *and* $V(\underline{\xi}_k)$ *are* $\mathcal{C}$*-measurable, then* $\mathcal{S}(\underline{\xi}_k, \mathcal{C}) = \mathcal{S}(\underline{\eta}_k, \mathcal{C})$.

Proof: Note that $\bar{x}$ und $V(x)$ are permutation symmetric functions of x. Thus, g is permutation symmetric with respect to the second variable and by Lemma (6.5) we have $\mathcal{S}(\underline{\eta}_k) \subseteq \mathcal{S}(\underline{\xi}_k)$, which implies $\mathcal{S}(\underline{\eta}_k, \mathcal{C}) \subseteq \mathcal{S}(\underline{\xi}_k, \mathcal{C})$.
Let us show that $\mathcal{S}(\underline{\xi}_k) \subseteq \mathcal{S}(\underline{\eta}_k, \mathcal{C})$. For this, let f be any measurable function such that $f(\underline{\xi}_k)$ is $\mathcal{S}(\underline{\xi}_k)$-measurable. We have to show that in this case $f(\underline{\xi}_k)$ has to be $\mathcal{S}(\underline{\eta}_k, \mathcal{C})-$measurable.
If $z \in (\mathbb{R}^m)^k$, $a \in \mathbb{R}^m$, $B \in \mathbb{R}^{m\times m}$, and if B is positive semidefinite and symmetric, then define

$$\tilde{h}(z, a, B) := (B^{\frac{1}{2}} z_1 + a, B^{\frac{1}{2}} z_2 + a, \ldots B^{\frac{1}{2}} z_k + a)^t.$$

It is clear that $\tilde{h}(\pi(z), a, B) = \pi(\tilde{h}(z, a, B))$ and $\underline{\xi}_k = \tilde{h}(\underline{\eta}_k, \bar{\xi}, V(\underline{\xi}_k))$. It follows that $f(\underline{\xi}_k) = f(\tilde{h}(\underline{\eta}_k, \bar{\xi}, V(\underline{\xi}_k))$. Moreover, we have

$$\begin{aligned} f(\tilde{h}(\pi(\underline{\eta}_k), \bar{\underline{\xi}}_k, V(\underline{\xi}_k)) &= f(\pi(\tilde{h}(\underline{\eta}_k, \bar{\underline{\xi}}_k, V(\underline{\xi}_k)))) \\ &= f(\pi(\underline{\xi}_k)) = f(\underline{\xi}_k) = f(\tilde{h}(\underline{\eta}_k, \bar{\underline{\xi}}_k, V(\underline{\xi}_k)). \end{aligned}$$

Since $\bar{\xi}$ und $V(\underline{\xi}_k)$ are $\mathcal{C}$-measurable, it can easily be seen that $f(\tilde{h}(\underline{\eta}_k, \bar{\xi}, V(\underline{\xi}_k))$ is $\mathcal{S}(\underline{\eta}_k, \mathcal{C})$-measurable. It follows that $\mathcal{S}(\underline{\xi}_k) \subseteq \mathcal{S}(\underline{\eta}_k, \mathcal{C})$ and $\mathcal{S}(\underline{\xi}_k, \mathcal{C}) \subseteq \mathcal{S}(\underline{\eta}_k, \mathcal{C})$. □

Our last remark is concerned with a Brownian bridge.

(6.9) REMARK Let B_t be a d-dimensional Brownian motion and let $B^0_t = B_t - tB_1$ be the corresponding Brownian bridge. If $w : [0, 1] \to \mathbb{R}^{m\times d}$ is square integrable, i.e. $\mathbb{E}(\int_0^1 \|w(t)\|^2\, dt) < \infty$, then the stochastic integral $\int_0^1 w(t)\, dB^0_t$ is well-defined and we have

$$\begin{aligned} \mathbb{E}(\| \int_0^1 w(t)\, dB^0_t \|^2) &= \mathbb{E}(\| \int_0^1 (w(t) - \overline{w})\, dB_t \|^2) \\ &= \int_0^1 \|w(t) - \overline{w}\|^2\, dt \le \int_0^1 \|w(t)\|^2\, dt. \end{aligned}$$

References

[1] P. Billingsley. *Convergence of Probability Measures*. John Wiley and Sons, New York, 1968.

[2] D. M. Chibisov. Certain chi-square type tests for continuous distributions. *Theor. Probability Appl.*, 16:1–21, 1971.

[3] J. Hájek and Z. Sidák. *Theory of rank tests*. Academic Press, 1967.

[4] K. Hoffmann. *Approximation von stochastischen Integralgleichungen durch Martingaldifferenzschemata*. PhD thesis, University of Bayreuth, 1987.

[5] A. Janssen and G. Neuhaus. Two sample rank tests for censored data with non-predictable weights. *To appear in J. Statist. Planning and Inference*.

[6] E.L. Lehmann. *Testing Statistical Hypotheses*. John Wiley and Sons, New York, 1986.

[7] E.L. Lehmann and C. Stein. On the theory of some non-parametric hypotheses. *Annals of Mathematical Statistics*, 20:28–45, 1949.

[8] G. Li and H. Doss. Generalized Pearson-Fisher chi-square goodness-of-fit tests, with applications to models with life history data. *Annals of Statistics*, 21:772–797, 1993.

[9] D. S. Moore and M. C. Spruill. Unified large-sample theory of general chi-squared statistics for tests of fit. *Annals of Statistics*, 3:599–616, 1975.

[10] G. Neuhaus. Conditional and unconditional two-sample rank tests for truncated data. *Statistics and Decisions*, 1998.

[11] J. Pfanzagl. *Allgemeine Methodenlehre der Statistik*. Walter de Gruyter, Berlin, 1974.

[12] J. Pfanzagl. Incidental versus random nuisance parameters. *Annals of Statistics*, 21:1663–1691, 1993.

[13] J. Pfanzagl and W. Wefelmeyer. *Contributions to a general asymptotic statistical theory*, volume 13 of *Lecture Notes in Statistics*. Springer, 1982.

[14] D. Pollard. General chi-square goodness-of-fit tests with data-dependent cells. *Z. Wahrsch. Verw. Gebiete 50*, 50:317–331, 1979.

[15] H. Strasser. Tangent vectors for models with independent but non-identically distributed observations. *Statistics and decisions*, 7:127–152, 1989.

[16] H. Strasser. Asymptotic normality toolbox. Technical report, Department of Statistics, Vienna University of Economics and Business Administration, 1994.

[17] H. Strasser. Local asymptotic normality of symmetrized product experiments. Technical report, Department of Statistics, Vienna University of Economics and Business Administration, 1994.

[18] H. Strasser. Asymptotic efficiency of estimates for models with incidental nuisance parameters. *Annals of Statistics*, 24, 1996.

[19] H. Strasser. Perturbation invariant estimates and incidental nuisance parameters. *Mathematical Methods of Statistics*, 7:1–26, 1998.

[20] H. Strasser and C. Weber. Invariance principles for exchangeable random vectors. Technical report, Department of Statistics, Vienna University of Economics and Business Administration, 1999.

[21] H. Witting. Über einen χ^2-Test, dessen Klassen durch geordnete Stichprobenfunktionen festgelegt werden. *Arch. Math.*, 10:468–479, 1959.