

Combinatorial stochastic processes

Lutz Dümbgen

Institut für Angewandte Mathematik, Universität Heidelberg, Im Neuenheimer Feld 294, 69120 Heidelberg, Germany

Received 4 December 1992; revised 27 August 1993

Abstract

Well-known results for sums of independent stochastic processes are extended to processes $S = \sum_{i=1}^n \phi_{i\Pi(i)}$, where $\phi = (\phi_{ij})_{1 \leq i, j \leq n}$ is a collection of independent stochastic processes ϕ_{ij} on some set \mathcal{T} , and Π is a random permutation of $\{1, 2, \dots, n\}$ such that Π, ϕ are independent. The general results, a uniform Law of Large Numbers and a functional Central Limit Theorem, are applied to permutation processes and randomized trials.

Key words: Random permutation; Symmetrization; Permutation process; Randomized trials; Conservative procedures

AMS Subject Classifications: Primary 60F05, 60F15; Secondary 62G10, 62G15.

1. Introduction

The starting point of the present paper are results of Alexander (1987) on stochastic processes $\tilde{S} = (\tilde{S}(t))_{t \in \mathcal{T}}$ of the form

$$\tilde{S} = \sum_{i=1}^n \tilde{\phi}_i$$

with independent processes $\tilde{\phi}_1, \tilde{\phi}_2, \dots, \tilde{\phi}_n$ on a set \mathcal{T} . Special cases are empirical processes and set-indexed partial sum processes. Another useful reference for this topic is Pollard (1990).

There are various statistical applications involving experimental randomization, in which the $\tilde{\phi}_i$ can be viewed as a random subsample of a larger family of processes. Precisely, let $\phi = (\phi_{ij})_{1 \leq i, j \leq n}$ be a collection of independent stochastic processes

* Corresponding author.

¹ Work supported by the Miller Institute, University of California at Berkeley, and Sonderforschungsbereich 123, Universität Heidelberg.

ϕ_{ij} on \mathcal{T} . Let Π be a uniformly distributed random permutation of $\{1, 2, \dots, n\}$, where Π, ϕ are independent. In what follows we investigate the stochastic process

$$S := \sum_{i=1}^n \phi_{i\Pi(i)}.$$

Conditional on Π , this process S has the same structure as \tilde{S} , but we are interested in its unconditional behavior and thus have to account for the dependence of the summands $\phi_{i\Pi(i)}$. At first glance S seems to have a rather special form, but one can represent quite different processes in this way. To give a simple example, let M be a uniformly distributed random subset of $\{1, 2, \dots, n\}$ with fixed cardinality $\#M = m$ such that M and $\tilde{\phi}_1, \tilde{\phi}_2, \dots, \tilde{\phi}_n$ are independent. Then $\sum_{i \in M} \tilde{\phi}_i$ is distributed as $\sum_i \phi_{i\Pi(i)}$, if $\mathcal{L}(\phi_{ij}) = \mathcal{L}(\tilde{\phi}_i)$ for $j \leq m$ and $\phi_{ij} \equiv 0$ for $j > m$.

An early reference in this context is Hoeffding (1951), who considered statistics of the form $\sum_i r_{i\Pi(i)}$ with some fixed $r = (r_{ij})_{1 \leq i, j \leq n}$ in $\mathbb{R}^{n \times n}$. His combinatorial central limit theorem has been extended by numerous authors. In particular, Chen (1978) considers $\sum_i R_{i\Pi(i)}$ with a random $n \times n$ matrix R . Classical applications are rank tests and permutation tests.

Section 2 contains general asymptotic results. That means, n, \mathcal{T} , and ϕ depend on an additional, but hidden index $\nu = 1, 2, \dots$, and all asymptotic statements refer to $\nu \rightarrow \infty$. The results and proofs are similar to parts of Alexander (1987). The main contribution is a new method of symmetrization, which is suitable for the present situation. Throughout this paper it is assumed that $2 \leq n = n_\nu \rightarrow \infty$ and that the mean function $\mu: \mathcal{T} \rightarrow \mathbb{R}^{n \times n}$ of ϕ exists, namely $\mu_{ij}(t) := \mathbb{P}\phi_{ij}(t)$. An elementary calculation shows that

$$\mathbb{P}S(t) = n^{-1} \mu_{++}(t), \tag{1}$$

where $r_{++} := \sum_{ij} r_{ij}$, $r_{i+} := \sum_j r_{ij}$ and $r_{+j} := \sum_i r_{ij}$ for $r \in \mathbb{R}^{n \times n}$. Theorem 1 and Corollary 1 give sufficient conditions for

$$\|S - \mathbb{P}S\| \rightarrow_p 0,$$

where $\|f\| = \|f\|_{\mathcal{T}} := \sup_{t \in \mathcal{T}} |f(t)|$ for $f: \mathcal{T} \rightarrow \mathbb{R}$. In addition assume the existence of the covariance function $\gamma: \mathcal{T} \times \mathcal{T} \rightarrow \mathbb{R}^{n \times n}$ of ϕ , namely $\gamma_{ij}(s, t) := \text{Cov}(\phi_{ij}(s), \phi_{ij}(t))$. Let

$$Gr := (r_{ij} - n^{-1}r_{i+} - n^{-1}r_{+j} + n^{-2}r_{++})_{1 \leq i, j \leq n}$$

and let $\langle r, s \rangle := \sum_{ij} r_{ij}s_{ij}$ be the usual inner product on $\mathbb{R}^{n \times n}$. Then standard calculations such as in Hoeffding (1951) show that

$$\text{Cov}(S(s), S(t)) = n^{-1} \gamma_{++}(s, t) + (n - 1)^{-1} \langle G\mu(s), G\mu(t) \rangle. \tag{2}$$

Theorem 2 gives sufficient conditions, under which the process $S - \mathbb{P}S$ can be approximated by a centered Gaussian process having uniformly continuous paths with respect to the pseudo-distance ρ on \mathcal{T} , which is defined by

$$\rho(s, t)^2 := n^{-1} \sum_{ij} \mathbb{P}(\phi_{ij}(s) - \phi_{ij}(t))^2 \geq (1 - n^{-1}) \text{Var}(S(s) - S(t)). \tag{3}$$

Two different applications are given in Sections 3 and 4. The first one is an analysis of permutation processes, which arise in connection with certain permutation tests. The second one concerns randomized trials, in which subjects with subject- and treatment-dependent distributions of responses are randomly assigned to different treatments.

2. Asymptotic results

In order to avoid measurability problems, we assume that the index sets $\mathcal{T} = \mathcal{T}_v$ are countable; see Pollard (1984) for other conditions that guarantee measurability of various suprema. Alternatively one may assume that the underlying probability space is discrete (as it is in Section 3). The results are defined in terms of the random pseudo-distances d_1, d_2, d_2^* on \mathcal{T} , where

$$d_p(s, t)^p = d_p(s, t | \phi_{i\pi(i)} : 1 \leq i \leq n)^p := \sum_i |\phi_{i\pi(i)}(s) - \phi_{i\pi(i)}(t)|^p, \quad p = 1, 2,$$

$$d_2^*(s, t)^2 = d_2^*(s, t | \phi_{i\pi(i)} : 1 \leq i \leq n)^2 := \sum_i (|\phi_{i\pi(i)}(s)|^{1/2} - |\phi_{i\pi(i)}(t)|^{1/2})^2.$$

Note that $\mathbb{P}d_2^2$ is just ρ^2 as defined in (3). For any pseudo-distance d on \mathcal{T} define the covering numbers

$$N(u, \mathcal{T}, d) := \min \left\{ \# \mathcal{T}_0 : \mathcal{T}_0 \subset \mathcal{T}, \inf_{t_0 \in \mathcal{T}_0} d(t_0, t) \leq u \quad \forall t \in \mathcal{T} \right\}.$$

In the sequel $\{condition\}$ stands for 1 if *condition* holds and 0 otherwise.

Theorem 1. *Suppose that all ϕ_{ij} are nonnegative and the following conditions are satisfied:*

$$\sup_{t \in \mathcal{T}} \mu_{++}(t) = O(n); \tag{4}$$

$$\sum_{ij} \mathbb{P} \|\phi_{ij}\| \{ \|\phi_{ij}\| > u \} = o(n) \quad \forall u > 0; \tag{5}$$

$$N(u, \mathcal{T}, d_2^*) = O_p(1) \quad \forall u > 0. \tag{6}$$

Then $\|S - \mathbb{P}S\|$ tends to zero in probability.

Corollary 1. *Suppose that (4) and (5) hold with $\mathbb{P}|\phi_{ij}|$ in place of μ_{ij} . If $N(u, \mathcal{T}, d_1) = O_p(1)$ for arbitrary $u > 0$, then $\|S - \mathbb{P}S\|$ tends to zero in probability.*

Corollary 1 follows from Theorem 1, applied to the processes $\phi_{ij}^{(1)} := \phi_{ij} \vee 0$ and $\phi_{ij}^{(2)} := (-\phi_{ij}) \vee 0$ separately, because

$$d_2^*(\cdot, \cdot | \phi_{i\pi(i)}^{(l)} : 1 \leq i \leq n)^2 \leq d_1(\cdot, \cdot | \phi_{i\pi(i)}^{(l)} : 1 \leq i \leq n) \leq d_1 \quad \text{for } l = 1, 2.$$

Theorem 2. Suppose that the following conditions are satisfied;

$$\sup_{t \in \mathcal{T}} \sum_{ij} \mathbb{P} \phi_{ij}(t)^2 = O(n); \tag{7}$$

$$\sum_{ij} \mathbb{P} \|\phi_{ij}\|^2 \{ \|\phi_{ij}\|^2 > u \} = o(n) \quad \forall u > 0; \tag{8}$$

$$\int_0^\alpha (\log N(u, \mathcal{T}, d_2))^{1/2} \rightarrow_p 0 \quad (v \rightarrow \infty, \alpha \downarrow 0). \tag{9}$$

Then

$$\|d_2^2 - \rho^2\|_{\mathcal{F} \times \mathcal{F}} \rightarrow_p 0 \quad \text{and} \quad N(u, \mathcal{F}, \rho) = O(1) \quad \forall u > 0, \tag{10}$$

$$\sup_{s, t \in \mathcal{F}: \rho(s, t) \leq \alpha} |(S - \mathbb{P}S)(s) - (S - \mathbb{P}S)(t)| \rightarrow_p 0 \quad (v \rightarrow \infty, \alpha \downarrow 0), \tag{11}$$

$$\|S - \mathbb{P}S\| = O_p(1). \tag{12}$$

Moreover, let $a_v: \mathcal{T} \rightarrow \mathbb{R}$ be arbitrary functions such that $\sum_{t \in \mathcal{T}} |a_v(t)| = O(1)$. Then

$$D\left(\mathcal{L}\left(\sum_{t \in \mathcal{T}} a_v(t)(S - \mathbb{P}S)(t)\right), \mathcal{N}\left(0, \text{Var}\left(\sum_{t \in \mathcal{T}} a_v(t)S(t)\right)\right)\right) \rightarrow 0, \tag{13}$$

where $D(\cdot, \cdot)$ is any metric on the space of Borel-probability measures on \mathbb{R} , which metrizes weak convergence.

Assertion (13) can be deduced from Chen (1978, Theorem 5.1). An alternative proof is given below.

Before proving Theorems 1 and 2 let us describe the symmetrization used here: First of all one may assume that

$$\text{all numbers } n = n, \text{ are even.} \tag{14}$$

For let k be uniformly distributed on $\{1, 2, \dots, n + 1\}$ such that ϕ, Π, k are independent. Then

$$\Pi'(i) := \begin{cases} \Pi(i) & \text{if } k \neq i \leq n, \\ \Pi(k) & \text{if } k < n + 1 = i, \\ n + 1 & \text{if } k = i, \end{cases}$$

defines a uniformly distributed random permutation Π' of $\{1, 2, \dots, n + 1\}$. With $\phi_{ij} := 0$ for $(i, j) \in \{1, 2, \dots, n + 1\}^2 \setminus \{1, 2, \dots, n\}^2$ one may replace S with

$$S' := \sum_{i=1}^{n+1} \phi_{i\Pi'(i)} = S - \sum_{i=1}^n \{i = k\} \phi_{i\Pi(i)},$$

because $d(\cdot, \cdot | \phi_{i\Pi(i)}; 1 \leq i \leq n + 1) \leq d$ for $d = d_1, d_2^*, d_2$, and

$$\mathbb{P} \|S - S'\| \leq n^{-2} \sum_{ij} \mathbb{P} \|\phi_{ij}\| \rightarrow 0 \quad \text{under (5),}$$

$$\mathbb{P} \|S - S'\|^2 \leq n^{-2} \sum_{ij} \mathbb{P} \|\phi_{ij}\|^2 \rightarrow 0 \quad \text{under (8).}$$

In order to bound $\|S - \mathbb{P}S\|$, we write $S = S_I + S_I^c$, where $I := \{1, 2, \dots, n/2\}$, and

$$S_M := \sum_{i \in M} \phi_{i\Pi(i)}, \quad M^c := \{1, 2, \dots, n\} \setminus M,$$

for any set $M \subset \{1, 2, \dots, n\}$. Since one may rearrange the processes ϕ_{ij} with respect to the first index i , it suffices to bound $\|S_I - \mathbb{P}S_I\|$. This process $S_I - \mathbb{P}S_I$ is symmetrized by subtracting a dependent copy rather than an independent one. Namely, we consider $S_I - S_I^*$ in place of $S_I - \mathbb{P}S_I$, where

$$S_I^* := \sum_{i \in I} \phi_{i\Pi(i+n/2)}.$$

Note that S_I and S_I^* are dependent and identically distributed, while conditional on the random set $\Pi(I)$ they are independent and have different distributions (in general). Let $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)$ be a Rademacher sequence, independent from ϕ and Π . Then

$$\mathcal{L}(S_I - S_I^*) = \mathcal{L}\left(\sum_{i \in I} \varepsilon_i (\phi_{i\Pi(i)} - \phi_{i\Pi(i+n/2)})\right).$$

This can be seen by conditioning on the sets $\{\Pi(i), \Pi(i + n/2)\}, i \in I$. Consequently, it suffices to bound the supremum norm of

$$S_I^\varepsilon := \sum_{i \in I} \varepsilon_i \phi_{i\Pi(i)}.$$

The next two lemmata are needed for justifying this symmetrization. The first one is about the conditional expectation and variance of S_I, S_I^* given $\Pi(I)$, while the second one is a modification of the symmetrization lemma in Pollard (1984).

Lemma 1. Under (14),

$$\mathbb{P}((S_I^* - \mathbb{P}S_I)(t) | \Pi(I)) = -\mathbb{P}((S_I - \mathbb{P}S_I)(t) | \Pi(I)),$$

$$\text{Var}(S_I(t) + S_I^*(t) | \Pi(I)) \leq 2(n - 2)^{-1} (n - 1) \text{Var}(S(t)) \quad \forall t \in \mathcal{T}.$$

Lemma 2. Let Z, Z^* be two independent processes on a countable set \mathcal{T} . Further let f be a fixed function on \mathcal{T} and $\delta > 0$ such that

$$\mathbb{P}\{|(Z^* + f)(t)| \leq \delta\} \wedge \mathbb{P}\{|(Z - f)(t) - (Z^* + f)(t)| \leq \delta\} \geq q > 0 \quad \forall t \in \mathcal{T}.$$

Then

$$\mathbb{P}\{\|Z\| > (3/2)(\eta + \delta)\} \leq q^{-1} \mathbb{P}\{\|Z - Z^*\| > \eta\} \quad \forall \eta.$$

Proof of Lemma 1. Note first that $\mathcal{L}(S_I | \Pi(I) = J)$ is essentially of the same form as $\mathcal{L}(S)$ for any fixed set $J \subset \{1, 2, \dots, n\}$ with $\#J = n/2$; just replace ϕ with $(\phi_{ij})_{i \in I, j \in J}$ and Π with a uniformly distributed random bijection $\Pi': I \rightarrow J$. Furthermore, $\mathcal{L}(S_I^* | \Pi(I) = J) = \mathcal{L}(S_I | \Pi(I) = J^c)$. Now the equality for the conditional means follows from routine calculations. In analogy to (2) one obtains

$$\begin{aligned} \text{Var}(S_I(t) + S_I^*(t) | \Pi(I) = J) &= 2n^{-1}(\gamma_{IJ}(t, t) + \gamma_{IJ^c}(t, t)) \\ &+ (n/2 - 1)^{-1} \langle G^{(IJ)} \mu(t), G^{(IJ)} \mu(t) \rangle \\ &+ \langle G^{(IJ^c)} \mu(t), G^{(IJ^c)} \mu(t) \rangle, \end{aligned} \tag{15}$$

where $r_{IJ} := \sum_{i \in I} r_{ij}$, $r_{iM} := \sum_{j \in M} r_{ij}$, $r_{IM} := \sum_{i \in I, j \in M} r_{ij}$, and

$$G^{(IM)} r := (\{i \in I\} \{j \in M\} (r_{ij} - 2n^{-1}r_{IJ} - 2n^{-1}r_{iM} + 4n^{-2}r_{IM}))_{i \leq i, j \leq n}$$

for $r \in \mathbb{R}^{n \times n}$ and $M = J, J^c$. The mapping G defined in the introduction is the orthogonal projection from $\mathbb{R}^{n \times n}$ onto its subspace

$$V := \{(\alpha_i + \beta_j)_{1 \leq i, j \leq n} : \alpha, \beta \in \mathbb{R}^n\}^\perp,$$

and $G^{(IM)}$ is the orthogonal projection from $\mathbb{R}^{n \times n}$ onto

$$V^{(IM)} := V \cap \{r \in \mathbb{R}^{n \times n} : r_{ij} = 0 \forall (i, j) \notin I \times M\}.$$

Since $V^{(IJ)}$ and $V^{(IJ^c)}$ are orthogonal subspaces of V , the right-hand side of (15) is not greater than

$$2n^{-1}\gamma_{++}(t, t) + (n/2 - 1)^{-1} \langle G\mu(t), G\mu(t) \rangle,$$

and the desired inequality follows from (2). \square

Proof of Lemma 2. Suppose that $|f(t)| > (\eta + \delta)/2$ for some $t \in \mathcal{T}$. Then

$$\begin{aligned} \mathbb{P}\{\|Z - Z^*\| > \eta\} &\geq \mathbb{P}\{|Z(t) - Z^*(t)| > \eta\} \\ &\geq \mathbb{P}\{2|f(t)| - |(Z - f)(t) - (Z^* + f)(t)| > \eta\} \\ &\geq \mathbb{P}\{|(Z - f)(t) - (Z^* + f)(t)| \leq \delta\} \\ &\geq q. \end{aligned}$$

On the other hand, suppose that $\|f\| \leq (\eta + \delta)/2$. Let $\tau = \tau(Z)$ be a \mathcal{F} -random variable such that $|Z(\tau)| > (3/2)(\eta + \delta)$ whenever $\|Z\| > (3/2)(\eta + \delta)$. Then

$$\begin{aligned} \mathbb{P}\{\|Z - Z^*\| > \eta\} &\geq \mathbb{P}\{|Z(\tau) - Z^*(\tau)| > \eta \text{ and } |Z(\tau)| > (3/2)(\eta + \delta)\} \\ &\geq \mathbb{P}\{|Z(\tau)| - |f(\tau)| - |(Z^* + f)(\tau)| > \eta \\ &\quad \text{and } |Z(\tau)| > (3/2)(\eta + \delta)\} \\ &\geq \mathbb{P}\{|(Z^* + f)(\tau)| \leq \delta \text{ and } |Z(\tau)| > (3/2)(\eta + \delta)\} \\ &\geq q\mathbb{P}\{|Z(\tau)| > (3/2)(\eta + \delta)\} \\ &= q\mathbb{P}\{\|Z\| > (3/2)(\eta + \delta)\}. \quad \square \end{aligned}$$

Proof of Theorem 1. According to (5) there exist numbers $u_v > 0$ such that

$$u_v \downarrow 0 \quad \text{and} \quad n^{-1} \sum_{ij} \mathbb{P} \|\phi_{ij}\| \{\|\phi_{ij}\| > u_v\} \rightarrow 0.$$

Conditions (4) and (6) remain valid if ϕ_{ij} is replaced with $\tilde{\phi}_{ij} := \phi_{ij} \{\|\phi_{ij}\| \leq u_v\}$, and the mean of $\|S - \sum_i \tilde{\phi}_{i\Pi(i)}\|$ is not greater than $n^{-1} \sum_{ij} \mathbb{P} \|\phi_{ij}\| \{\|\phi_{ij}\| > u_v\}$. Therefore, one may assume that

$$\|\phi_{ij}\| \leq u_v \quad \forall i, j \quad \forall v.$$

As mentioned earlier, it suffices to show that $\|S_I - \mathbb{P}S_I\| \rightarrow_p 0$, where we assume (14). One may symmetrize $S_I - \mathbb{P}S_I$ as described earlier. For

$$\text{Var}(S(t)) \leq (n - 1)^{-1} \sum_{ij} \mathbb{P} \phi_{ij}(t)^2 \leq u_v(n - 1)^{-1} \mu_{++}(t),$$

and the right-hand side tends to 0 uniformly in $t \in \mathcal{T}$, by (4). Hence, Lemma 1 implies that

$$\sup_{t \in \mathcal{T}, J \subset \{1, 2, \dots, n\}: \#J = n/2} \text{Var}(S_I(t) + S_I^*(t) | \Pi(I) = J) = o(1).$$

Moreover, the processes $Z := S_I - \mathbb{P}S_I$ and $Z^* := S_I^* - \mathbb{P}S_I^*$ are independent conditional on $\Pi(I)$, and $f := \mathbb{P}(Z | \Pi(I))$ equals $-\mathbb{P}(Z^* | \Pi(I))$. Thus, one can deduce from Lemma 2 that $\|S_I - \mathbb{P}S_I\| \rightarrow_p 0$, if $\|S_I^*\| \rightarrow_p 0$.

Clearly, $|S_I^s(s) - S_I^s(t)| \leq d_1(s, t)$ for arbitrary $s, t \in \mathcal{T}$. Since all ϕ_{ij} are nonnegative, one may write

$$\begin{aligned} d_1(s, t) &= \sum_i (\phi_{i\Pi(i)}(s)^{1/2} + \phi_{i\Pi(i)}(t)^{1/2}) |\phi_{i\Pi(i)}(s)^{1/2} - \phi_{i\Pi(i)}(t)^{1/2}| \\ &\leq (S(s)^{1/2} + S(t)^{1/2}) d_2^*(s, t) \\ &\leq 2 \|S\|^{1/2} d_2^*(s, t), \end{aligned}$$

where the first inequality follows from the Cauchy–Schwarz and triangle inequalities. Hence $\mathbb{P}(\|S_t^\varepsilon\| > 2\eta|\Pi, \phi)$ is not greater than

$$N(\eta/(2\|S\|^{1/2}), \mathcal{F}, d_2^*) \sup_{t \in \mathcal{F}} \mathbb{P}(|S_t^\varepsilon(t)| > \eta|\Pi, \phi),$$

and for arbitrary $t \in \mathcal{F}$ and $\eta > 0$,

$$\mathbb{P}(|S_t^\varepsilon(t)| > \eta|\Pi, \phi) \leq 2 \exp\left(-\eta^2 / \left(2 \sum_{i \in I} \phi_{i\Pi(i)}(t)^2\right)\right) \leq 2 \exp(-\eta^2/(2u_v\|S\|)),$$

according to Hoeffding’s (1963) inequality. Consequently, it suffices to show that $\|S\|$ is bounded in probability.

We now apply LeCam’s (1983) square root trick: Let $M := \{i: \varepsilon_i = 1\}$. Since $\|S\| \leq \|S_M\| + \|S_{M^c}\|$, and since S_M, S_{M^c} are identically distributed, it suffices to show that

$$\|S_M\|^{1/2} = \|S_M^{1/2}\| = O_p(1).$$

Conditional on $(\varepsilon, \Pi(M))$ the two processes $S_M^{1/2}, S_{M^c}^{1/2}$ are independent with

$$\begin{aligned} \sup_{t \in \mathcal{F}} \mathbb{P}(S_{M^c}(t)|\varepsilon, \Pi(M)) &= (\# M^c)^{-1} \sup_{t \in \mathcal{F}} \sum_{i \in M^c, j \in \Pi(M^c)} \mu_{ij}(t) \\ &\leq (\# M^c)^{-1} \sup_{t \in \mathcal{F}} \mu_{++}(t) \\ &= O_p(1), \end{aligned}$$

because of (4) and $\# M^c/n \rightarrow_p 1/2$. Hence, one can apply the symmetrization lemma in Pollard (1984), and it suffices to show that

$$\|S_M^{1/2} - S_{M^c}^{1/2}\| = O_p(1).$$

However,

$$\begin{aligned} |(S_M^{1/2} - S_{M^c}^{1/2})(s) - (S_M^{1/2} - S_{M^c}^{1/2})(t)| \\ \leq |S_M^{1/2}(s) - S_M^{1/2}(t)| + |S_{M^c}^{1/2}(s) - S_{M^c}^{1/2}(t)| \\ \leq d_2^*(s, t|\phi_{i\Pi(i)}; i \in M) + d_2^*(s, t|\phi_{i\Pi(i)}; i \in M^c) \\ \leq 2^{1/2} d_2^*(s, t). \end{aligned}$$

Thus, $\mathbb{P}(\|S_M^{1/2} - S_{M^c}^{1/2}\| > 2\eta|\Pi, \phi)$ is not greater than

$$\begin{aligned} N(2^{-1/2}\eta, \mathcal{F}, d_2^*) \sup_{t \in \mathcal{F}} \mathbb{P}(|(S_M^{1/2} - S_{M^c}^{1/2})(t)| > \eta|\Pi, \phi) \\ = O_p(1) \sup_{t \in \mathcal{F}} \mathbb{P}(|(S_M - S_{M^c})(t)| > (S_M^{1/2} + S_{M^c}^{1/2})(t)\eta|\Pi, \phi) \\ \leq O_p(1) \sup_{t \in \mathcal{F}} \mathbb{P}(|(S_M - S_{M^c})(t)| > S(t)^{1/2} \eta|\Pi, \phi), \end{aligned}$$

and for all $t \in \mathcal{T}$ and $\eta > 0$,

$$\begin{aligned} & \mathbb{P}(|(S_M - S_{M^c})(t)| > S(t)^{1/2} \eta | \Pi, \phi) \\ & \leq 2 \exp\left(-\eta^2 S(t) / \left(2 \sum_i \phi_{i \Pi(i)}(t)^2\right)\right) \leq 2 \exp(-\eta^2 / (2u_v)) \rightarrow 0, \end{aligned}$$

by Hoeffding’s (1963) inequality. \square

Before proving Theorem 2 let us mention an auxiliary result, which follows fromoeffding (1951, Theorem 3), by a simple truncation argument.

Lemma 3. *Let $r = r_v \in \mathbb{R}^{n \times n}$ such that*

$$\langle r, r \rangle = O(n) \quad \text{and} \quad \sum_{ij} r_{ij}^2 \{r_{ij}^2 > u\} = o(n) \quad \forall u > 0.$$

Then

$$D\left(\mathcal{L}\left(\sum_i Gr_{i \Pi(i)}\right), \mathcal{N}(0, (n-1)^{-1} \langle Gr, Gr \rangle)\right) \rightarrow 0.$$

Proof of Theorem 2. The arguments of Alexander (1987) carry over with only few modifications: We define processes ψ_{ij} on $\mathcal{T} \times \mathcal{T}$ via

$$\psi_{ij}(s, t) := \phi_{ij}(s) - \phi_{ij}(t).$$

Then $d_2^2(s, t) = \sum_i \psi_{i \Pi(i)}^2(s, t)$ and $\rho^2 = \mathbb{P}d_2^2$. Note that

$$N(u, \mathcal{T} \times \mathcal{T}, d_2^*(\cdot, \cdot | \psi_{i \Pi(i)}^2; 1 \leq i \leq n)) \leq N(u/2, \mathcal{T}, d_2)^2 = O_\rho(1) \quad \forall u > 0,$$

because of (9) and

$$d_2^*((s, t), (s', t') | \psi_{i \Pi(i)}^2; 1 \leq i \leq n) \leq d_2(s, s') + d_2(t, t').$$

Moreover,

$$\sup_{s, t \in \mathcal{T}} \sum_{ij} \mathbb{P} \psi_{ij}(s, t)^2 \leq 4 \sup_{t \in \mathcal{T}} \sum_{ij} \mathbb{P} \phi_{ij}(t)^2 = O(n),$$

$$\sum_{ij} \mathbb{P} \|\psi_{ij}^2\|_{\mathcal{T} \times \mathcal{T}} \{ \|\psi_{ij}^2\|_{\mathcal{T} \times \mathcal{T}} > u \} \leq 4 \sum_{ij} \mathbb{P} \|\phi_{ij}\|^2 \{ \|\phi_{ij}\|^2 > u/4 \} = o(n) \quad \forall u > 0.$$

Consequently, the first half of (10) follows from Theorem 1 applied to ψ_{ij}^2 in place of ϕ_{ij} . This, together with (9), yields the second half of (10).

In order to prove stochastic equicontinuity of $S - \mathbb{P}S$, assertion (11), one may assume (14) and symmetrize as in the proof of Theorem 1. Thus, it suffices to show that

$$\sup_{s, t \in \mathcal{T} : \rho(s, t) \leq \alpha(v)} |S_I^c(s) - S_I^c(t)| \rightarrow_p 0 \quad \text{whenever } \alpha(v) \downarrow 0.$$

Since d_2^2 is a uniformly consistent estimator of ρ^2 , one may replace ρ with d_2 . But then the assertion follows from Hoeffding’s (1963) inequality and the Chaining lemma in Pollard (1984) applied to the conditional distribution of S_i^ε given ϕ and Π .

Let $\mathcal{T}(u)$ be a subset of \mathcal{T} such that $\#\mathcal{T}(u) = N(u, \mathcal{T}, \rho)$ and $\inf_{s \in \mathcal{T}(u)} \rho(s, t) \leq u$ for all $t \in \mathcal{T}$. By (10), $\#\mathcal{T}(u) = O(1)$, and thus $\|S - \mathbb{P}S\|_{\mathcal{T}(u)} = O_p(1)$, according to (7). But (11) implies that for any $\varepsilon > 0$ there is a $u > 0$ such that $\|S - \mathbb{P}S\| > \|S - \mathbb{P}S\|_{\mathcal{T}(u)} + 1$ with asymptotic probability at most ε . This yields (12).

As for assertion (13), one can easily show that (7) and (8) remain valid, if \mathcal{T} is replaced with the singleton $\{a_v\}$ and $\phi_{ij}(a_v) := \sum_{t \in \mathcal{T}} a_v(t) \phi_{ij}(t)$. Elementary calculations show that these conditions can be reformulated as follows:

$$\langle \mu(a_v), \mu(a_v) \rangle = O(n) \quad \text{and} \quad \sum_{ij} \mu_{ij}(a_v)^2 \{ \mu_{ij}(a_v)^2 > u \} = o(n) \quad \forall u > 0, \tag{16}$$

$$\gamma_{++}(a_v) = O(n) \quad \text{and} \quad \gamma_{++}(a_v|u) = o(n) \quad \forall u > 0, \tag{17}$$

where $\gamma_{ij}(a_v) := \text{Var}(\phi_{ij}(a_v))$ and $\gamma_{ij}(a_v|u)$ stands for $\mathbb{P}(\phi_{ij} - \mu_{ij})(a_v)^2 \{ (\phi_{ij} - \mu_{ij})(a_v)^2 > u \}$. In fact, the second half of (16) follows from the following simple inequality for integrable real random variables Y :

$$\mathbb{P}Y\{Y > u\} \geq \mathbb{P}Y\{\mathbb{P}Y > 2u\}/2 \quad \forall u > 0. \tag{18}$$

Now write $(S - \mathbb{P}S)(a_v)$ as the sum of $S^0(a_v) := \sum_i (\phi_{i\Pi(i)} - \mu_{i\Pi(i)})(a_v)$ and $M(a_v) := \sum_i (G\mu(a_v))_{i\Pi(i)}$. It follows from Lemma 3 and (16) that

$$D(\mathcal{L}(M(a_v)), \mathcal{N}(0, (n-1)^{-1} \langle G\mu(a_v), G\mu(a_v) \rangle)) \rightarrow 0.$$

Since $\text{Var}(S(a_v))$ equals $n^{-1}\gamma_{++}(a_v) + (n-1)^{-1} \langle G\mu(a_v), G\mu(a_v) \rangle$ and $M(a_v)$ depends only on Π , it suffices to show that

$$D(\mathcal{L}(S^0(a_v)|\Pi), \mathcal{N}(0, n^{-1}\gamma_{++}(a_v))) \rightarrow_p 0.$$

But this is a consequence of Lindeberg’s central limit theorem. For

$$\sum_i \gamma_{i\Pi(i)}(a_v) = n^{-1}\gamma_{++}(a_v) + o_p(1),$$

$$\sum_i \gamma_{i\Pi(i)}(a_v|u) = o_p(1) \quad \forall u > 0.$$

The latter two claims follow from (17), because $\mathbb{P}\sum_i \gamma_{i\Pi(i)}(a_v|u)$ equals $n^{-1}\gamma_{++}(a_v|u)$, while $\text{Var}(\sum_i \gamma_{i\Pi(i)}(a_v) - \sum_i \phi_{i\Pi(i)}(a_v|u))$ is not greater than $(n-1)^{-1}u\gamma_{++}(a_v)$. \square

3. Permutation processes

Let $\xi = (\xi_{ij})_{1 \leq i, j \leq n}$ be a collection of fixed points ξ_{ij} in some set X , and let \mathcal{T} be a family of real-valued functions on X such that its envelope $T(x) := \sup_{t \in \mathcal{T}} |t(x)|$ is

finite for all $x \in X$. For some constant $c > 0$ we define

$$\phi_{ij}(t) := ct(\xi_{ij}),$$

i.e. each ϕ_{ij} is a nonrandom function on \mathcal{F} . We now consider the asymptotic behavior of $S = \sum_i \phi_{i/n(i)}$, where n, ξ, c depend on v with $n \rightarrow \infty$ as $v \rightarrow \infty$, while \mathcal{F} is fixed. S is regarded as a random variable with values in the Banach space $(l_\infty(\mathcal{F}), \|\cdot\|)$ of bounded functions on \mathcal{F} . Let

$$N(u, \mathcal{F}) := \sup N(u, Q(T), \mathcal{F}, d_Q),$$

where the supremum is over all discrete measures Q on X such that $Q(T) < \infty$, and $d_Q(s, t) := Q(|s - t|)$. A useful inequality is

$$N(u, \mathcal{F}) \leq Au^{-B} \quad \forall u \in (0, 1]$$

for Vapnik–Cervonenkis subgraph classes \mathcal{F} ($A, B > 0$); see Pollard (1984, Lemma II.25).

Theorem 3. *Suppose that the following conditions are satisfied:*

$$c^2 \sum_{ij} T(\xi_{ij})^2 = O(n); \tag{19}$$

$$c^2 \sum_{ij} T(\xi_{ij})^2 \{c^2 T(\xi_{ij})^2 > u\} = o(n) \quad \forall u > 0; \tag{20}$$

$$\int_0^1 (\log N(u^2, \mathcal{F}))^{1/2} du < \infty; \tag{21}$$

for some function \tilde{K} on $\mathcal{F} \times \mathcal{F}$,

$$\|\text{Cov}(S) - \tilde{K}\|_{\mathcal{F} \times \mathcal{F}} \rightarrow 0. \tag{22}$$

Then \mathcal{F} is totally bounded with respect to the pseudodistance

$$\tilde{\rho}(s, t) := (\tilde{K}(s, s) + \tilde{K}(t, t) - 2\tilde{K}(s, t))^{1/2}.$$

Moreover, $S - \mathbb{P}S$ converges in distribution in $l_\infty(\mathcal{F})$ to a centered Gaussian process $\tilde{W} = (\tilde{W}(t))_{t \in \mathcal{F}}$ having uniformly continuous paths with respect to $\tilde{\rho}$, and $\text{Cov}(\tilde{W}) = \tilde{K}$.

An important special case is that $\xi_{ij} = (y_i, z_j) \in Y \times Z$ and $\mathcal{F} = \mathcal{F} \times \mathcal{G}$ with families \mathcal{F} and \mathcal{G} of functions on Y and Z , respectively; we identify a pair $(f, g) \in \mathcal{F}$ with the function $f \times g(y, z) := f(y)g(z)$, so that $\phi_{ij}(f, g) = cf(y_i)g(z_j)$. Processes of this type appear in various statistical applications. In the context of generalized bootstrap methods, Praestgaard and Wellner (1993) consider the case that \mathcal{G} consists of one indicator function. Applications to the changepoint problem are given by Romano (1989), Lausen and Schumacher (1992) and Dümbgen (1992). There $z_j = j/n$ and $\mathcal{G} = \{\{\cdot \leq r\} : r \in [0, 1]\}$. If one wants to extend these methods to generalized

change-point models such as in Carlstein and Krishnamoorthy (1992), one has to consider $z_j \in \mathbb{R}^d$ and more general classes \mathcal{G} .

Condition (21) is satisfied if

$$\int_0^1 (\log N(u^2, \mathcal{F}))^{1/2} du + \int_0^1 (\log N(u^2, \mathcal{G}))^{1/2} du < \infty.$$

For let F and G be the envelopes of \mathcal{F} and \mathcal{G} , respectively, and let Q be any discrete measure on $Y \times Z$ such that $Q(F \times G) < \infty$. Then

$$Q(|f \times g - f' \times g'|) \leq Q'(|f - f'|) + Q''(|g - g'|) \quad \forall f, f' \in \mathcal{F} \quad \forall g, g' \in \mathcal{G},$$

where $Q'(\cdot) := Q(\cdot \times G)$ and $Q''(\cdot) := Q(F \times \cdot)$. Since $Q'(F) = Q''(G) = Q(F \times G)$, one can easily deduce that $N(u, \mathcal{F}) \leq N(u/2, \mathcal{F})N(u/2, \mathcal{G})$.

To give a specific example, let $Y = Z = \mathbb{R}$ and $z_j := j/m$ with $1 \leq m \leq n$. Further, suppose that the points $y_i = y_{i,v}$ and $m = m_v$ satisfy

$$\sum_i y_i = 0 \quad \text{and} \quad \sum_i y_i^2 = n \quad \forall v,$$

$$\sum_i y_i^2 \{y_i^2 > um\} = o(n) \quad \forall u > 0.$$

Then the process $(m^{-1/2} \sum_i y_i \{z_{\Pi(i)} \leq r\})_{r \in [0, 1]}$ converges in distribution to the standard Brownian bridge if $m = n$ and to the standard Brownian motion on $[0, 1]$ if $m \rightarrow \infty$ and $m/n \rightarrow 0$. This is an extension of Billingsley (1968, Theorem 24.1), which follows from Theorem 3 and the subsequent remarks, when $\mathcal{F} = \{y \mapsto y\}$ and $\mathcal{G} = \{\cdot \leq r\} : r \in [0, 1]\}$.

Proof of Theorem 3. One can write $S - \mathbb{P}S$ as $\sum_i \psi_{i\Pi(i)}$, where $\psi_{ij}(t) := (G\phi(t))_{ij}$. We now show that the assumptions of Theorem 2 hold with these functions ψ_{ij} in place of ϕ_{ij} . This implies Theorem 3, because

$$\rho(s, t)^2 = (1 - n^{-1})n \text{Var}(S(s) - S(t))$$

converges to $\tilde{\rho}(s, t)^2$ uniformly in $s, t \in \mathcal{T}$, according to (22).

With the measure $q_{ij} := \delta_{\xi_{ij}} + n^{-1} \sum_k (\delta_{\xi_{ik}} + \delta_{\xi_{kj}}) + n^{-2} \sum_{kl} \delta_{\xi_{kl}}$ one can write $\psi_{ij}(t) = q_{ij}(f_{ij}t)$ with a density $f_{ij} \in [-c, c]$. Then the Cauchy-Schwarz inequality yields

$$\|\psi_{ij}\|^2 \leq 4c^2 q_{ij}(T^2).$$

Together with (18) one can deduce from (19) and (20) that

$$c^2 q_{++}(T^2) = O(n) \quad \text{and} \quad c^2 \sum_{ij} q_{ij}(T^2) \{c^2 q_{ij}(T^2) > u\} = o(n) \quad \forall u > 0, \quad (23)$$

which certainly implies (7) and (8). The random distance $d_2(s, t)^2$ is not greater than $Q(|s - t|)$, where

$$Q(\cdot) := 8c^2 \sum_i q_{i\Pi(i)}(T \cdot).$$

Thus $N(u, \mathcal{F}, d_2) \leq N(u^2/Q(T), \mathcal{F})$, and (9) follows from (21) and $Q(T) = O_p(1)$. The latter fact is a consequence of (23). \square

4. Randomized trials

Suppose that one is interested in the influence of different treatments (e.g. medical drugs) $b = 1, 2, \dots, B$ on a group of subjects that are indexed by $i = 1, 2, \dots, n$. Often it is not possible to expose one subject to several treatments, and one picks disjoint subsets D_1, D_2, \dots, D_B of $\{1, 2, \dots, n\}$ randomly, where all subjects in D_b receive treatment b . The assignment $D := (D_1, \dots, D_B)$ is uniformly distributed under the restriction that $\#D_b = n_b$ for all $b = 1, 2, \dots, B$ with fixed integers $n_b \geq 2$. Then one observes the response Y_i of subject i for all $i \in \bigcup_b D_b$.

We assume that for each subject i there are distributions $P_{i1}, P_{i2}, \dots, P_{iB}$ on a measurable space Y such that for given assignment D the Y_i are independent random variables with

$$\mathcal{L}(Y_i|D) = P_{ib} \quad \forall i \in D_b.$$

A natural estimator for the average distribution $P_b := n^{-1} \sum_i P_{ib}$ is given by

$$\hat{P}_b := n_b^{-1} \sum_{i \in D_b} \delta_{Y_i},$$

and one easily verifies that it is unbiased. The question is, how good these estimators are. More precisely, let \mathcal{F} be a family of measurable functions $f: Y \rightarrow \mathbb{R}$. Then we are interested in the behavior of the process $S - \mathbb{P}S$ on $\mathcal{F} := \mathcal{F} \times \{1, 2, \dots, B\}$, where

$$S(f, b) := n_b^{1/2} \hat{P}_b(f).$$

This example fits into our general framework as follows: Let $Y = (Y_{ij})_{1 \leq i, j \leq n}$ consist of independent Y -random variables Y_{ij} such that

$$\mathcal{L}(Y_{ij}) = P_{ib} \quad \forall j \in J_b := \left\{ j: \sum_{1 \leq c < b} n_c < j \leq \sum_{1 \leq c \leq b} n_c \right\}.$$

Further, let Y and Π be independent. Then S can be represented as $\sum_i \phi_{i\Pi(i)}$, where

$$\phi_{ij}(f, b) := n_b^{-1/2} f(Y_{ij}) \{j \in J_b\}.$$

In other words, we define $Y_i := Y_{i\Pi(i)}$ and $D_b := \Pi^{-1}(J_b)$.

An often made assumption is that

$$(iid) \quad P_{ib} = P_b \quad \forall i \quad \forall b.$$

This can be justified if n is huge compared to $n_1 + n_2 + \dots + n_B$. Under (iid) the empirical processes $((S - \mathbb{P}S)(f, b))_{f \in \mathcal{F}}, 1 \leq b \leq B$, are independent, whereas in general they are dependent. However, it is shown below that under (iid) the process S is maximally dispersed in a certain sense.

Here $(G\mu(f, b))_{ij}$ equals $n_b^{-1/2}(P_{ib} - P_b)(f)(\{j \in J_b\} - n_b/n)$, and one can derive from (2) that

$$\text{Cov}(S) = V - \Delta,$$

where

$$\begin{aligned} V((f, b), (g, c)) &:= \{b = c\} (P_b(fg) - P_b(f)P_b(g)) \\ &\quad + \{b = c\} (n(n - 1))^{-1} \sum_i (P_{ib} - P_b)(f)(P_{ib} - P_b)(g) \\ &= \text{Cov}_{(iid)}(S(f, b), S(g, c)) + \{b = c\} n_b^{-1} \Delta((f, b), (g, b)), \end{aligned}$$

$$\Delta((f, b), (g, c)) := (n(n - 1))^{-1} (n_b n_c)^{1/2} \sum_i (P_{ib} - P_b)(f)(P_{ic} - P_c)(g).$$

The function Δ is nonnegative definite, i.e. $\sum_{s, t \in \mathcal{F}} a(s)a(t)\Delta(s, t)$ is nonnegative for all $a \in \mathbb{R}^{\mathcal{F}}$ such that $a(t) = 0$ for all but finitely many $t \in \mathcal{F}$. Since each subject gets only one treatment, one cannot expect to find a consistent estimator for Δ without special assumptions. However, let

$$\hat{V}((f, b), (g, c)) := \{b = c\} (n_b - 1)^{-1} n_b (\hat{P}_b(fg) - \hat{P}_b(f)\hat{P}_b(g)).$$

Under (iid), \hat{V} is an unbiased estimator for the covariance function of S . The remarkable thing is that in general,

$$\mathbb{P} \hat{V} = V. \tag{24}$$

We now analyze the asymptotic behavior of S and \hat{V} under the assumption that the number B of treatments and the (countable) family \mathcal{F} with envelope F are fixed, while $n_b = n_{b,v} \rightarrow \infty$ for all b , and the averages P_b may also depend on v . Further, we assume that

$$\sum_b P_b(F^2) = O(1), \tag{25}$$

$$\sum_b P_b(F^2 \{F^2 > n_b u\}) \rightarrow 0 \quad \forall u > 0, \tag{26}$$

$$\int_0^1 (\log N(u^2, \mathcal{F}))^{1/2} du < \infty. \tag{27}$$

Under these conditions,

$$\|V - \text{Cov}_{(\text{iid})}(S)\|_{\mathcal{F} \times \mathcal{F}} = O(n^{-1}) \quad \text{and} \quad \|\Delta\|_{\mathcal{F} \times \mathcal{F}} = O\left(n^{-1} \max_b n_b\right).$$

Hence $\text{Cov}(S)$ can be approximated by $\text{Cov}_{(\text{iid})}(S) - \Delta$, where Δ is negligible, if $n_b/n \rightarrow 0$ for all b . In what follows let \mathbb{P}_* and \mathbb{P}^* denote inner and outer probability, respectively. For $A \subset l_\infty(\mathcal{F})$ let $B(A, \varepsilon)$ be the set of all $x \in l_\infty(\mathcal{F})$ such that $\inf_{a \in A} \|x - a\| < \varepsilon$.

Theorem 4. *Suppose that conditions (25)–(27) hold. Then the conditions of Theorem 2 are satisfied, and*

$$\|\hat{V} - V\|_{\mathcal{F} \times \mathcal{F}} \rightarrow_p 0. \tag{28}$$

Further, for arbitrary convex and symmetric sets $C_v \subset l_\infty(\mathcal{F})$,

$$\mathbb{P}_* \{S - \mathbb{P}S \in B(C_v, \varepsilon)\} \geq \mathbb{P}_{(\text{iid})}^* \{S - \mathbb{P}S \in C_v\} + o(1) \quad \forall \varepsilon > 0. \tag{29}$$

For instance, if $Y = \mathbb{R}$, one can compute simultaneous Kolmogorov–Smirnov confidence bands for the distribution functions of P_1, P_2, \dots, P_B . Namely, let $\alpha \in (0, 1)$ and $R(\alpha) > 0$ such that

$$\mathbb{P}\{\|W^0\|_{[0,1]} \leq R(\alpha)\} = (1 - \alpha)^{1/B},$$

where W^0 is a standard Brownian bridge on $[0, 1]$. Then Theorem 4 and classical results for empirical distribution functions together imply that

$$\limsup \mathbb{P}\{\|S - \mathbb{P}S\| \geq R(\alpha)\} \leq \alpha,$$

where $\mathcal{F} := \{\{\cdot \leq r\} : r \in \mathbb{R}\}$. This is a new example for the phenomenon that standard procedures, which are motivated by assumption (iid), are still reliable and often conservative in the general case. Some other references for this topic are Neyman (1923, 1990), Copas (1973), Robins (1988) or Freedman et al. (1991, Chapter 27.3).

For general Y and \mathcal{F} one can estimate $\mathcal{L}_{(\text{iid})}(S - \mathbb{P})$ consistently by the bootstrap estimator $\mathcal{L}(\hat{S} - S | D, Y)$, which is constructed as follows:

$$\hat{S}(f, b) := n_b^{-1/2} \sum_{i \in D_b} f(\hat{Y}_i),$$

where the random variables $\hat{Y}_1, \hat{Y}_2, \dots, \hat{Y}_n$ are conditionally independent given (D, Y) , and

$$\mathcal{L}(\hat{Y}_i | D, Y) = \hat{P}_b \quad \forall i \in D_b.$$

Corollary 2. *Suppose that (25)–(27) are satisfied. Then for arbitrary sets $A_v \subset l_\infty(\mathcal{F})$,*

$$\mathbb{P}_{(\text{iid})}^* \{S - \mathbb{P}S \in A_v\} \leq \mathbb{P}(\hat{S} - S \in B(A_v, \varepsilon) | D, Y) + o_p(1),$$

$$\mathbb{P}(\hat{S} - S \in A_v | D, Y) \leq \mathbb{P}_{*(\text{iid})} \{S - \mathbb{P}S \in B(A_v, \varepsilon)\} + o_p(1) \quad \forall \varepsilon > 0.$$

Theorem 4 and Corollary 2 together extend recent results of Liu and Singh (1992), who demonstrated that standard bootstrap methods typically yield conservative confidence sets, if applied to independent, but not identically distributed observations.

Proof of Theorem 4. We first check the conditions of Theorem 2. One can easily show that (25) and (26) imply (7) and (8). Moreover, $N(u, \mathcal{F}, d_2) \leq \sum_b N(u, \mathcal{F} \times \{b\}, d_2)$, and for any fixed b ,

$$d_2((f, b), (g, b))^2 = \hat{P}_b((f - g)^2) \leq 2\hat{P}_b(F|f - g|) \quad \forall f, g \in \mathcal{F}.$$

Hence $N(u, \mathcal{F} \times \{b\}, d_2) \leq N(u^2/(2\hat{P}_b(F^2)), \mathcal{F})$, and condition (9) follows from (25) and (27).

One may assume without loss of generality that $0 \in \mathcal{F}$. Then one can deduce (28) from (10), (12) and (25). For $(\hat{P}_b - P_b)(fg)$ can be written as

$$((d_2^2 - \rho^2)((f, b), (0, b)) + (d_2^2 - \rho^2)((g, b), (0, b)) - (d_2^2 - \rho^2)((f, b), (g, b)))/2,$$

while

$$|\hat{P}_b(f)\hat{P}_b(g) - P_b(f)P_b(g)| \leq n_b^{-1} \|S - \mathbb{P}S\|^2 + 2P_b(F) n_b^{-1/2} \|S - \mathbb{P}S\|.$$

It remains to prove inequality (29). Note that (25)–(27) involve only \mathcal{F} and the averages P_b . Likewise $\rho((f, b), (g, c))^2 = P_b(f^2) + P_c(g^2) - 2\{b = c\}P_b(fg)$ depends only on P_1, P_2, \dots, P_B . Thus (11) implies that for any $\delta > 0$ there is a sequence of sets $\mathcal{T}_0 = \mathcal{T}_{0,v} \subset \mathcal{F}$ and mappings $\gamma = \gamma_v$ from \mathcal{F} onto \mathcal{T}_0 such that $\#\mathcal{T}_0 = O(1)$ and

$$\limsup \mathbb{P}^* \{ \|S' - S' \circ \gamma\| \geq \delta \} < \delta,$$

$$\limsup \mathbb{P}_{(\text{iid})}^* \{ \|S' - S' \circ \gamma\| \geq \delta \} < \delta,$$

where $S' := S - \mathbb{P}S$. But (13) implies that $\mathcal{L}(S'|_{\mathcal{T}_0})$ and $\mathcal{L}_{(\text{iid})}(S'|_{\mathcal{T}_0})$ can be approximated by centered Gaussian distributions $Q = Q_v$ and $Q_{(\text{iid})} = Q_{(\text{iid}),v}$ on $l_\infty(\mathcal{T}_0)$ with covariance functions $(V - A)|_{\mathcal{T}_0 \times \mathcal{T}_0}$ and $V|_{\mathcal{T}_0 \times \mathcal{T}_0}$, respectively. Further, the mapping $l_\infty(\mathcal{T}_0) \ni x_0 \mapsto \Gamma x_0 := x_0 \circ \gamma \in l_\infty(\mathcal{F})$ is linear and isomorphic. Hence,

$$\begin{aligned} \mathbb{P}_{(\text{iid})}^* \{ S' \in C_v \} &\leq \mathbb{P}_{(\text{iid})} \{ S' \circ \gamma \in B(C_v, \delta) \} + \delta \\ &= \mathbb{P}_{(\text{iid})} \{ S'|_{\mathcal{T}_0} \in \Gamma^{-1} B(C_v, \delta) \} + \delta \\ &\leq Q_{(\text{iid})}(\Gamma^{-1} B(C_v, 2\delta)) + 2\delta \end{aligned}$$

for sufficiently large v . Since $\Gamma^{-1} B(C_v, 2\delta)$ is also convex and symmetric, and since $A|_{\mathcal{T}_0 \times \mathcal{T}_0}$ is nonnegative definite, Corollary 3 of Anderson (1955) implies that

$$\begin{aligned} Q_{(\text{iid})}(\Gamma^{-1} B(C_v, 2\delta)) &\leq Q(\Gamma^{-1} B(C_v, 2\delta)) \\ &\leq \mathbb{P} \{ S'|_{\mathcal{T}_0} \in \Gamma^{-1} B(C_v, 3\delta) \} + \delta \\ &\leq \mathbb{P}_* \{ S' \in B(C_v, 4\delta) \} + 2\delta \end{aligned}$$

for sufficiently large v . Letting $\delta \downarrow 0$ yields (29). \square

Proof of Corollary 2. Note that $\mathcal{L}(\hat{S}|D, Y)$ is the same as $\mathcal{L}_{(\text{iid})}(S)$ with \hat{P}_b in place of P_b . Moreover, the pseudodistance ρ corresponding to S has to be replaced with d_2 , and

$$\text{Cov}(\hat{S}(f, b), \hat{S}(g, c)|D, Y) = (n_b - 1)n_b^{-1} \hat{V}((f, b), (g, c)).$$

Now conditions (25) and (26) hold in probability, because $\mathbb{P} \sum_b \hat{P}_b(F^2) = \sum_b P_b(F^2) = O(1)$ and $\mathbb{P} \sum_b \hat{P}_b(F^2 \{F^2 > n_b u\}) = o(1)$ for all $u > 0$. Since $\|d_2^2 - \rho^2\|_{\mathcal{F} \times \mathcal{F}} \rightarrow_p 0$ and $\|\hat{V} - \text{Cov}_{(\text{iid})}(S)\|_{\mathcal{F} \times \mathcal{F}} \rightarrow_p 0$, Theorem 2 implies that

$$\mathbb{P} \left(\sup_{s, t \in \mathcal{F} : \rho(s, t) \leq \alpha} |(\hat{S} - S)(s) - (\hat{S} - S)(t)| > \varepsilon \middle| D, Y \right) \rightarrow_p 0 \quad (v \rightarrow \infty, \alpha \downarrow 0) \quad \forall \varepsilon > 0,$$

and

$$D \left(\mathcal{L} \left(\sum_{t \in \mathcal{F}} a_v(t) (\hat{S} - S)(t) \middle| D, Y \right), \mathcal{N} \left(0, \text{Var}_{(\text{iid})} \left(\sum_{t \in \mathcal{F}} a_v(t) S(t) \right) \right) \right) \rightarrow_p 0$$

for arbitrary functions $a_v : \mathcal{F} \rightarrow \mathbb{R}$ such that $\sum_{t \in \mathcal{F}} |a_v(t)| = O(1)$. Now the assertion of Corollary 2 follows with similar arguments as in the proof of (29). \square

Acknowledgements

I am indebted to David Freedman and Terry Speed for stimulating discussions about randomized trials and useful references. I want to thank Andreas Thumfart and the referees for various helpful comments.

References

K.S. Alexander, Central limit theorems for stochastic processes under random entropy conditions, *Probab. Theory Related Fields* 75 (1987) 351–378.
 T.W. Anderson, The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities, *Proc. Amer. Math. Soc.* 6 (1955) 170–176.
 P. Billingsley, *Convergence of Probability Measures* (Wiley, New York, 1968).
 E. Carlstein and C. Krishnamoorthy, Boundary estimation, *J. Amer. Statist. Assoc.* 87 (1992) 430–438.
 L.H.Y. Chen, Two central limit theorems for dependent random variables, *Z. Wahrscheinlichkeitstheorie verw. Geb.* 43 (1978) 223–243.
 J.B. Copas, Randomization models for the matched and unmatched 2×2 tables, *Biometrika* 60 (1973) 467–476.
 L. Dümbgen, Confidence sets for a changepoint via randomization methods, Technical report 364, University of California at Berkeley, 1992.
 D. Freedman, R. Pisani, R. Purves and A. Adhikari, *Statistics* (Norton, New York, London, 2nd ed., 1991).
 W. Hoeffding, A combinatorial central limit theorem, *Ann. Math. Statist.* 22 (1951) 558–566.
 W. Hoeffding, Probability inequalities for sums of bounded random variables, *J. Amer. Statist. Assoc.* 58 (1963) 13–30.
 B. Lausen and M. Schumacher, Maximally selected rank statistics, *Biometrics* 48 (1992) 73–85.

- L. LeCam, A remark on empirical measures, in: P. Bickel, K. Doksum and J. Hodges, eds., *A Festschrift for Erich Lehmann in honor of his sixty-fifth birthday* (Wadsworth, Belmont, 1983) pp. 305–327.
- R.Y. Liu and K. Singh, Using i.i.d. bootstrap inference for general non-i.i.d. models, *J. Statist. Plan. Inf.*, to appear.
- J. Neyman, On the application of probability theory to agricultural experiments, reprinted in *Statist. Sci.* 5 (1923, 1990) 465–480.
- D. Pollard, *Convergence of Stochastic Processes* (Springer, New York, 1984).
- D. Pollard, Empirical processes: Theory and Applications, NSF-CBMS Regional Conf. Ser. Prob. Statist. 2 (1990).
- J. Praestgaard and J.A. Wellner, Exchangeably weighted bootstraps of the empirical process, *Ann. Prob.* 21 (1993) 2053–2086.
- J. Robins, Confidence intervals for causal parameters, *Statistics in Medicine* 7 (1988) 773–785.
- J. Romano, Bootstrap and randomization tests of some nonparametric hypotheses, *Ann. Statist.* 17 (1989) 141–159.