Chapter 17
Concentration Inequalities for Randomly Permuted Sums

Mélisande Albert

Abstract Initially motivated by the study of the non-asymptotic properties of non-parametric tests based on permutation methods, concentration inequalities for uniformly permuted sums have been largely studied in the literature. Recently, Delyon et al. proved a new Bernstein-type concentration inequality based on martingale theory. This work presents a new proof of this inequality based on the fundamental inequalities for random permutations of Talagrand. The idea is to first obtain a rough inequality for the square root of the permuted sum, and then, iterate the previous analysis and plug this first inequality to obtain a general concentration of permuted sums around their median. Then, concentration inequalities around the mean are deduced. This method allows us to obtain the Bernstein-type inequality up to constants, and, in particular, to recovers the Gaussian behavior of such permuted sums under classical conditions encountered in the literature. Then, an application to the study of the second kind error rate of permutation tests of independence is presented.

Keywords Concentration inequalities · Random permutations

Mathematics Subject Classification 60E15, 60C05

17.1 Introduction and Motivation

This article presents concentration inequalities for randomly permuted sums defined by

$$Z_n = \sum_{i=1}^{n} a_{i,\Pi(i)},$$

M. Albert (✉)
Institut de Mathématiques de Toulouse; UMR5219, Université de Toulouse; CNRS, INSA, Toulouse, France
e-mail: melisande.albert@insa-toulouse.fr

where $\{a_{i,j}\}_{1\le i,j\le n}$ are real numbers, and Π is a uniformly distributed random permutation of the set $\{1,\ldots,n\}$. Initially motivated by hypothesis testing in the non-parametric framework (see [28] for instance), such sums have been largely studied from an asymptotic point of view in the literature. A first combinatorial central limit theorem is proved by Wald and Wolfowitz in [28], in the particular case when the real numbers $a_{i,j}$ are of a product form $b_i \times c_j$, under strong assumptions that have been released for instance by Noether [22]. Then, Hoeffding obtains stronger results in such product case, and generalizes those results to not necessarily product type real terms $a_{i,j}$ in [15]. More precisely, he considers

$$d_{i,j} = a_{i,j} - \frac{1}{n}\sum_{k=1}^{n} a_{k,j} - \frac{1}{n}\sum_{l=1}^{n} a_{i,l} + \frac{1}{n^2}\sum_{k,l=1}^{n} a_{k,l}. \tag{17.1}$$

In particular, $\mathrm{Var}(Z_n) = \frac{1}{n-1}\sum_{i=1}^{n} d_{i,j}^2$. Then he proves (see [15, Theorem 3]) that, if

$$\lim_{n\to+\infty} \frac{\frac{1}{n}\sum_{1\le i,j\le n} d_{i,j}^r}{\left(\frac{1}{n}\sum_{i,j=1}^{n} d_{i,j}^2\right)^{r/2}} = 0, \quad \text{for some } r > 2, \tag{17.2}$$

then the distribution of $Z_n = \sum_{i=1}^{n} a_{i,\Pi(i)}$ is asymptotically normal, that is, for all x in $\mathbb{R}$,

$$\lim_{n\to+\infty} \mathbb{P}\left(Z_n - \mathbb{E}[Z_n] \le x\sqrt{\mathrm{Var}(Z_n)}\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{y^2}{2}} dy.$$

He also considers a stronger (in the sense that it implies (17.2)), but simpler condition in [15, Theorem 3], precisely

$$\frac{\max_{1\le i,j\le n}\{|d_{i,j}|\}}{\sqrt{\frac{1}{n}\sum_{i,j=1}^{n} d_{i,j}^2}} \xrightarrow[n\to+\infty]{} 0, \tag{17.3}$$

under which such an asymptotic Gaussian limit holds. Similar results have been obtained later, for instance by Motoo [21], under the following Lindeberg-type condition that is for all $\varepsilon > 0$,

$$\lim_{n\to+\infty} \sum_{1\le i,j\le n} \left(\frac{d_{i,j}}{d}\right)^2 \mathbb{1}_{\left|\frac{d_{i,j}}{d}\right|>\varepsilon} = 0, \tag{17.4}$$

where $d^2 = n^{-1}\sum_{1\le i,j\le n} d_{i,j}^2$. In particular, Motoo proves in [21] that such Lindeberg-type condition is weaker than Hoeffding's ones in the sense that (17.4) is implied by (17.2) (and thus by (17.3)). A few years later, Hájek [13] proves in

the product case, that the condition (17.4) is in fact necessary. A simpler proof of the sufficiency of the Lindeberg-type condition is given by Schneller [25] based on Stein's method.

Afterwards, the next step was to study the convergence of the conditional distribution when the terms $a_{i,j}$ in the general case, or $b_i \times c_j$ in the product case, are random. Notably, Dwass studies in [10] the limit of the randomly permuted sum in the product case, where only the c_j's are random, and proves that the conditional distribution given the c_j's converges almost surely (a.s.) to a Gaussian distribution. Then, Shapiro and Hubert [26] generalized this study to weighted U-statistics of the form $\sum_{i \neq j} b_{i,j} h(X_i, X_j)$ where the X_i's are independent and identically distributed (i.i.d.) random variables. In a first time, they show some a.s. asymptotic normality of this statistic. In a second time, they complete Jogdeo's [17] work in the deterministic case, proving asymptotic normality of permuted statistics based on the previous weighted U-statistic. More precisely, they consider the rank statistic $\sum_{i \neq j} b_{i,j} h(X_{R_i}, X_{R_j})$, where R_i is the rank of V_i in a sample $V_1, \ldots, V_n$ of i.i.d. random variables with a continuous distribution function. In particular, notice that considering such rank statistics is equivalent to considering uniformly permuted statistics. In [2], the previous combinatorial central limit theorems is generalized to permuted sums of non-i.i.d. random variables $\sum_{i=1}^{n} Y_{i,\Pi(i)}$, for particular forms of random variables $Y_{i,j}$. The main difference with the previous results comes from the fact that the random variables $Y_{i,j}$ are not necessarily exchangeable.

Hence, the asymptotic behavior of permuted sums has been vastly investigated in the literature, allowing to deduce good properties for permutation tests based on such statistics, like the asymptotic size, or the power (see for instance [23] or [2]). Yet, such results are purely asymptotic, while, in many application fields, such as neurosciences for instance as described in [2], few exploitable data are available. Hence, such asymptotic results may not be sufficient. This is why a non-asymptotic approach is preferred here, leading to concentration inequalities. In the sequel, unless specified, we will thus drop the index n and denote $Z = Z_n$.

Concentration inequalities have been vastly investigated in the literature, and the interested reader can refer to the books of Ledoux [18], Massart [19], or the more recent one of Boucheron et al. [8] for some overall reviews. Yet in many cases, they provide precise tail bounds for well-behaved functions or sums of independent random variables. For instance, let us recall the classical Bernstein inequality stated for instance in [19, Proposition 2.9 and Corollary 2.10].

Theorem 17.1.1 (Bernstein's Inequality, Massart [19]) *Let $X_1, \ldots, X_n$ be independent real valued random variables. Assume that there exists some positive numbers v and c such that*

$$\sum_{i=1}^{n} \mathbb{E}\left[X_i^2 \right] \leq v,$$

and for all integers $k \geq 3$,

$$\sum_{i=1}^{n} \mathbb{E}\left[(X_i)_+^k\right] \leq \frac{k!}{2} v c^{k-2},$$

where $(\cdot)_+ = \max\{\cdot, 0\}$ denotes the positive part.
  Let $S = \sum_{i=1}^{n}(X_i - \mathbb{E}[X_i])$, then for every positive x,

$$\mathbb{P}\left(S \geq \sqrt{2vx} + cx\right) \leq e^{-x}. \tag{17.5}$$

Moreover, for any positive t,

$$\mathbb{P}(S \geq t) \leq \exp\left(-\frac{t^2}{2(v + ct)}\right). \tag{17.6}$$

Notice that both forms of Bernstein's inequality appear in the literature. Yet, due to its form, (17.5) is rather preferred in statistics, even though (17.6) is more classical.

The work in this article is based on the pioneering work of Talagrand (see [27] for a review) who investigates the concentration of measure phenomenon for product measures. Of main interest here, he proves the following inequality for random permutations in [27, Theorem 5.1].

Theorem 17.1.2 (Talagrand [27]) *Denote by $\mathfrak{S}_n$ the set of all permutations of $\{1, \ldots, n\}$. Define for any subset $A \subset \mathfrak{S}_n$, and permutation $\tau \in \mathfrak{S}_n$,*

$$U_A(\tau) = \left\{s \in \{0, 1\}^n \; ; \; \exists \sigma \in A \text{ such that } \forall 1 \leq i \leq n, \; s_i = 0 \implies \sigma(i) = \tau(i)\right\}.$$

Then, consider $V_A(\tau) = \text{ConvexHull}(U_A(\tau))$, and

$$f(A, \tau) = \min\left\{\sum_{i=1}^{n} v_i^2 \; ; \; v = (v_i)_{1 \leq i \leq n} \in V_A(\tau)\right\}.$$

Then, if P_n denotes the uniform distribution on $\mathfrak{S}_n$,

$$\int_{\mathfrak{S}_n} e^{\frac{1}{16} f(A, \tau)} dP_n(\tau) \leq \frac{1}{P_n(A)}.$$

Therefore, by Markov's inequality, for all $t > 0$,

$$P_n\left(\tau \; ; \; f(A, \tau) \geq t^2\right) \leq \frac{e^{-t^2/16}}{P_n(A)}. \tag{17.7}$$

This result on random permutations is fundamental, and is a key point to many other non-asymptotic works on random permutations. Among them emerges McDiarmid's article [20] in which he derives from Talagrand's inequality, exponential concentration inequalities around the median for randomly permuted functions of the observation under Lipschitz-type conditions and applied to randomized methods for graph coloring. More recently, Adamczak et al. obtained in [4] some concentration inequality under convex-Lipschitz conditions when studying the empirical spectral distribution of random matrices. In particular, they prove the following theorem (precisely [4, Theorem 3.1]).

Theorem 17.1.3 (Adamczak et al. [4]) *Consider $x_1, \ldots, x_n$ in $[0, 1]$ and let $\varphi : [0, 1]^n \to \mathbb{R}$ be an L-Lipschitz convex function. Let Π be a uniform random permutation of the set $\{1, \ldots, n\}$ and denote $Y = \varphi\big(x_{\Pi(1)}, \ldots, x_{\Pi(n)}\big)$. Then, there exists some positive absolute constant c such that, for all $t > 0$,*

$$\mathbb{P}(Y - \mathbb{E}[Y] \geq t) \leq 2 \exp\left(-\frac{ct^2}{L^2}\right).$$

Yet, the Lipschitz assumptions may be very restrictive and may not be satisfied by the functions considered in the application fields (see Sect. 17.3.1 for instance). Hence, the idea is to exploit the attractive form of a sum. Based on Stein's method, initially introduced to study the Gaussian behavior of sums of dependent random variables, Chatterjee studies permuted sums of non-negative numbers in [9]. He obtains in [9, Proposition 1.1] the following first Bernstein-type concentration inequality for non-negative terms around the mean.

Theorem 17.1.4 (Chatterjee [9]) *Let $\big\{a_{i,j}\big\}_{1 \leq i, j \leq n}$ be a collection of numbers from $[0, 1]$. Let $Z = \sum_{i=1}^{n} a_{i, \Pi(i)}$, where Π is drawn from the uniform distribution over the set of all permutations of $\{1, \ldots, n\}$. Then, for any $t \geq 0$,*

$$\mathbb{P}(|Z - \mathbb{E}[Z]| \geq t) \leq 2 \exp\left(-\frac{t^2}{4\mathbb{E}[Z] + 2t}\right). \tag{17.8}$$

Notice that because of the expectation term in the right-hand side of (17.8), the link with Hoeffding's combinatorial central limit theorem (for instance) is not so clear.

In [6, Theorem 4.3], this result is sharpened in the sense that this expectation term is replaced by a variance term, allowing us to provide a non-asymptotic version of such combinatorial central limit theorem. This result is moreover generalized to any real numbers (not necessarily non-negative). More precisely, based on martingale theory, they prove the following result.

Theorem 17.1.5 (Bercu et al. [6]) *Let* $\left\{a_{i,j}\right\}_{1 \leq i, j \leq n}$ *be an array of real numbers from* $[-M_a, M_a]$. *Let* $Z = \sum_{i=1}^{n} a_{i,\Pi(i)}$, *where* Π *is drawn from the uniform distribution over the set of all permutations of* $\{1, \ldots, n\}$. *Then, for any* $t > 0$,

$$\mathbb{P}(|Z - \mathbb{E}[Z]| \geq t) \leq 4 \exp\left(-\frac{t^2}{16(\theta \frac{1}{n} \sum_{i,j=1}^{n} a_{i,j}^2 + M_a t/3)}\right), \qquad (17.9)$$

where $\theta = \dfrac{5}{2} \ln(3) - \dfrac{2}{3}$.

In this work, we obtain a similar result (up to constants) but based on a completely different approach. Moreover, this approach provides a direct proof for a concentration inequality of a permuted sum around its median.

The present work is organized as follows. In Sect. 17.2 are formulated the main results. Section 17.2.1 is devoted to the permuted sums of non-negative numbers. Based on Talagrand's result, a first rough concentration inequality for the square root of permuted sum is obtained in Lemma 17.2.1. Then by iterating the previous analysis and plugging this first inequality, a general concentration of permuted sums around their median is obtained in Proposition 17.2.1. Finally, the concentration inequality of Proposition 17.2.2 around the mean is deduced. In Sect. 17.2.2, the previous inequalities are generalized to permuted sums of not necessarily non-negative terms in Theorem 17.2.1. Section 17.3 presents an application to the study of non-asymptotic properties of a permutation independence test in statistics. In particular, a sharp control of the critical value of the test is deduced from the main result. The proofs are detailed in Sect. 17.4. Finally, the Appendix contains technical results for the non-asymptotic control of the second kind error rate of the permutation test introduced in Sect. 17.3.

17.2   Bernstein-Type Concentration Inequalities for Permuted Sums

Let us first introduce some general notation. In the sequel, denote by $\mathfrak{S}_n$ the set of permutations of $\{1, 2, \ldots, n\}$. For all collection of real numbers $\left\{a_{i,j}\right\}_{1 \leq i, j \leq n}$, and for each τ in $\mathfrak{S}_n$, consider the permuted sum

$$Z(\tau) = \sum_{i=1}^{n} a_{i,\tau(i)}.$$

Let Π be a uniform random permutation in $\mathfrak{S}_n$, and $Z := Z(\Pi)$. Denote med (Z) its median, that is which satisfies

$$\mathbb{P}(Z \geq \text{med}(Z)) \geq 1/2 \quad \text{and} \quad \mathbb{P}(Z \leq \text{med}(Z)) \geq 1/2.$$

This study is divided in two steps. The first one is restrained to non-negative terms. The second one extends the previous results to general terms, based on a trick involving both non-negative and negative parts.

17.2.1   Concentration of Permuted Sums of Non-negative Numbers

In the present section, consider a non-negative collection of numbers $\{a_{i,j}\}_{1\le i,j\le n}$. The proof of the concentration inequality around the median in Proposition 17.2.1 needs a preliminary step which is presented in Lemma 17.2.1. It provides concentration inequality for the square root of the sum. It allows us then by iterating the same argument, and plugging the obtained inequality to the square root of the sum of the squares, namely $\sqrt{\sum_{i=1}^{n} a_{i,\Pi(i)}^2}$, to be able to sharpen Chatterjee's concentration inequality (17.8).

Lemma 17.2.1 *Let $\{a_{i,j}\}_{1\le i,j\le n}$ be a collection of non-negative numbers, and Π be a uniform random permutation in $\mathfrak{S}_n$. Consider $Z = \sum_{i=1}^{n} a_{i,\Pi(i)}$. Then, for all $t > 0$,*

$$\mathbb{P}\left(\sqrt{Z} \ge \sqrt{\mathrm{med}\,(Z)} + t\sqrt{\max_{1\le i,j\le n}\{a_{i,j}\}}\right) \le 2e^{-t^2/16}, \qquad (17.10)$$

and

$$\mathbb{P}\left(\sqrt{Z} \le \sqrt{\mathrm{med}\,(Z)} - t\sqrt{\max_{1\le i,j\le n}\{a_{i,j}\}}\right) \le 2e^{-t^2/16}. \qquad (17.11)$$

In particular, one obtains the following two-sided concentration for the square root of a randomly permuted sum of non-negative numbers,

$$\mathbb{P}\left(\left|\sqrt{Z} - \sqrt{\mathrm{med}\,(Z)}\right| > t\sqrt{\max_{1\le i,j\le n}\{a_{i,j}\}}\right) \le 4e^{-t^2/16}.$$

The idea of the proof is the same that the one of Adamczak et al. in [4, Theorem 3.1], but with a sum instead of a convex Lipschitz function. In a similar way, it is based on Talagrand's inequality for random permutations recalled in Theorem 17.1.2.

In the following are presented two concentration inequalities in the non-negative case; the first one around the median, and the second one around the mean. It is well known that both are equivalent up to constants, but here, both are detailed in order to give the order of magnitude of the constants. The transition from the median to the

mean can be obtained thanks to Ledoux' trick in the proof of [18, Proposition 1.8] allowing to reduce exponential concentration inequalities around any constant m (corresponding in our case to med (Z)) to similar inequalities around the mean. This trick consists in using the exponentially fast decrease around m to upper bound the difference between m and the mean. Yet, this approach leads to drastic multiplicative constants (of the order $8e^{16\pi}$ as shown in [1]). Better constants can be deduced from the following lemma.

Lemma 17.2.2 *For any real valued random variable X,*

$$|\mathbb{E}[X] - \text{med}(X)| \leq \sqrt{\text{Var}(X)}.$$

In particular, we obtain the following results.

Proposition 17.2.1 *Let $\{a_{i,j}\}_{1 \leq i, j \leq n}$ be a collection of non-negative numbers and Π be a uniform random permutation in $\mathfrak{S}_n$. Consider $Z = \sum_{i=1}^{n} a_{i,\Pi(i)}$. Then, for all $x > 0$,*

$$\mathbb{P}\left(|Z - \text{med}(Z)| > \sqrt{\text{med}\left(\sum_{i=1}^{n} a_{i,\Pi(i)}^2\right) x + x \max_{1 \leq i, j \leq n} \{a_{i,j}\}}\right) \leq 8 \exp\left(\frac{-x}{16}\right).$$
(17.12)

Since in many applications, the concentration around the mean is more adapted, the following proposition shows that one may obtain a similar behavior around the mean, at the cost of higher constants.

Proposition 17.2.2 *Let $\{a_{i,j}\}_{1 \leq i, j \leq n}$ be a collection of non-negative numbers, and Π be a uniform random permutation in $\mathfrak{S}_n$. Consider $Z = \sum_{i=1}^{n} a_{i,\Pi(i)}$.*
Then, for all $x > 0$,

$$\mathbb{P}\left(|Z - \mathbb{E}[Z]| \geq 2\sqrt{\left(\frac{1}{n}\sum_{i,j=1}^{n} a_{i,j}^2\right) x + \max_{1 \leq i, j \leq n} \{a_{i,j}\} x}\right) \leq 8e^{1/16} \exp\left(-\frac{x}{16}\right).$$
(17.13)

This concentration inequality is called a Bernstein-type inequality restricted to non-negative sums, due to its resemblance to the standard Bernstein inequality, as recalled in Theorem 17.1.1. The main difference here lies in the fact that the random variables in the sum are not independent. Moreover, this inequality implies a more popular form of Bernstein's inequality stated in Corollary 17.2.1.

Corollary 17.2.1 *With the same notation and assumptions as in Proposition 17.2.2, for all $t > 0$,*

$$\mathbb{P}(|Z - \mathbb{E}[Z]| \geq t) \leq 8e^{1/16} \exp\left(\frac{-t^2}{16\left(4\frac{1}{n}\sum_{i,j=1}^{n} a_{i,j}^2 + 2\max_{1\leq i,j\leq n}\{a_{i,j}\}\, t\right)} \right).$$
(17.14)

Comment Recall Chatterjee's result in [9, Proposition 2.1], quoted in Theorem 17.1.4, which can easily be rewritten with our notation, and for a collection of non-negative numbers not necessarily in [0, 1], by

$$\forall t > 0, \quad \mathbb{P}(|Z - \mathbb{E}[Z]| \geq t) \leq 2\exp\left(\frac{-t^2}{4M_a\frac{1}{n}\sum_{i,j=1}^{n} a_{i,j} + 2M_a t} \right),$$

where M_a denotes the maximum $\max_{1\leq i,j\leq n}\{a_{i,j}\}$. As mentioned in [6], the inequality in (17.14) is sharper up to constants, because of the quadratic term, since the inequality $\sum_{i,j=1}^{n} a_{i,j}^2 \leq M_a \sum_{i,j=1}^{n} a_{i,j}$ always holds.

17.2.2 Concentration of Permuted Sums in the General Case

In this section, the collection of numbers $\{a_{i,j}\}_{1\leq i,j\leq n}$ is no longer assumed to be non-negative. The following general concentration inequality for randomly permuted sums directly derives from Proposition 17.2.2.

Theorem 17.2.1 *Let $\{a_{i,j}\}_{1\leq i,j\leq n}$ be a collection of any real numbers, and Π be a uniform random permutation in $\mathfrak{S}_n$. Consider $Z = \sum_{i=1}^{n} a_{i,\Pi(i)}$. Then, for all $x > 0$,*

$$\mathbb{P}\left(|Z - \mathbb{E}[Z]| \geq 2\sqrt{2\left(\frac{1}{n}\sum_{i,j=1}^{n} a_{i,j}^2\right)x} + 2\max_{1\leq i,j\leq n}\{|a_{i,j}|\}\, x \right) \leq 16e^{1/16}\exp\left(-\frac{x}{16}\right).$$
(17.15)

Once again, the obtained inequality is a Bernstein-type inequality. Moreover, it is also possible to obtain a more popular form of Bernstein-type inequalities applying the same trick based on the non-negative and the negative parts from Corollary 17.2.1.

Corollary 17.2.2 *With the same notation as in Theorem 17.2.1, for all $t > 0$,*

$$\mathbb{P}(|Z - \mathbb{E}[Z]| \geq t) \leq 16e^{1/16}\exp\left(\frac{-t^2}{256\left(\mathrm{Var}(Z) + \max_{1\leq i,j\leq n}\{|a_{i,j}|\}\, t\right)} \right).$$

Comments One recovers a Gaussian behavior of the centered permuted sum obtained by Hoeffding in [15, Theorem 3] under the same assumptions. Indeed, in the proof of Corollary 17.2.2, one obtains the following intermediate result (see (17.41)), that is

$$\mathbb{P}(|Z - \mathbb{E}[Z]| \geq t) \leq 16e^{1/16} \exp\left(\frac{-t^2}{64\left(4\frac{1}{n}\sum_{i,j=1}^n d_{i,j}^2 + \max_{1 \leq i,j \leq n}\{|d_{i,j}|\}t\right)}\right),$$

where the $d_{i,j}$'s are defined in (17.1). Yet, $\text{Var}(Z) = (n-1)^{-1}\sum_{i,j=1}^n d_{i,j}^2$ (see [15, Theorem 2]). Hence, applying this inequality to

$$t = x\sqrt{\text{Var}(Z)} \geq x\sqrt{\frac{1}{n}\sum_{i,j=1}^n d_{i,j}^2},$$

for $x > 0$ leads to

$$\mathbb{P}\left(|Z - \mathbb{E}[Z]| \geq x\sqrt{\text{Var}(Z)}\right) \leq 16e^{1/16} \exp\left(\frac{-x^2}{256\left(1 + \frac{\max_{1 \leq i,j \leq n}\{|d_{i,j}|\}}{\sqrt{\frac{1}{n}\sum_{i,j=1}^n d_{i,j}^2}}x\right)}\right),$$

Hence, under Hoeffding's simpler condition (17.3), namely

$$\lim_{n \to +\infty} \frac{\max_{1 \leq i,j \leq n} d_{i,j}^2}{\frac{1}{n}\sum_{i,j=1}^n d_{i,j}^2} = 0,$$

one recovers, (denoting $Z = Z_n$ depending on n),

$$\lim_{n \to +\infty} \mathbb{P}\left(|Z_n - \mathbb{E}[Z_n]| \geq x\sqrt{\text{Var}(Z_n)}\right) \leq 16e^{1/16}e^{-x^2/256},$$

which is a Gaussian tail that is, up to constants, close in spirit to the one obtained by Hoeffding in [15, Theorem 3].

17.3 Application to Independence Testing

17.3.1 Statistical Motivation

Let $\mathcal{X}$ represent a separable set. Given an i.i.d. n-sample $\mathbb{X}_n = (X_1, \ldots, X_n)$, where each X_i is a couple (X_i^1, X_i^2) in $\mathcal{X}^2$ with distribution P of marginals P^1 and P^2,

we aim at testing the null hypothesis $(\mathcal{H}_0)$ "$P = (P^1 \otimes P^2)$" against the alternative $(\mathcal{H}_1)$ "$P \neq (P^1 \otimes P^2)$". The considered test statistic is defined by

$$T(\mathbb{X}_n) = \frac{1}{n-1} \left(\sum_{i=1}^{n} \varphi(X_i^1, X_i^2) - \frac{1}{n} \sum_{i,j=1}^{n} \varphi(X_i^1, X_j^2) \right), \qquad (17.16)$$

where φ is a measurable real-valued function on $\mathcal{X}^2$. Denoting for any real-valued measurable function g on $\mathcal{X}^2$,

$$\mathbb{E}_P[g] = \int_{\mathcal{X}^2} g\left(x^1, x^2\right) dP\left(x^1, x^2\right) \quad \text{and} \quad \mathbb{E}_{\perp\!\!\!\perp}[g] = \int_{\mathcal{X}^2} g\left(x^1, x^2\right) dP^1\left(x^1\right) dP^2\left(x^2\right),$$
$$(17.17)$$

one may notice that, $T(\mathbb{X}_n)$ is an unbiased estimator of

$$\mathbb{E}\left[T(\mathbb{X}_n)\right] = \mathbb{E}_P[\varphi] - \mathbb{E}_{\perp\!\!\!\perp}[\varphi],$$

which is equal to 0 under $(\mathcal{H}_0)$.

For more details on the choice of the test statistic, the interested reader can refer to [2] (motivated by synchrony detection in neuroscience for instance). The particular case where $\mathcal{X} = [0, 1]$ and φ is a two-dimensional isotropic Haar wavelet is studied in [1, Chapter 4] and recalled below. More precisely, consider a resolution scale j in $\mathbb{N}$, a translation $k = (k_1, k_2)$ in $\mathcal{K}_j := \{0, 1, \ldots, 2^j - 1\}^2$. Consider the functions defined for all (x^1, x^2) in $[0, 1]^2$ by

$$\varphi_0(x^1, x^2) = \phi(x^1)\phi(x^2), \quad \text{and} \quad \begin{cases} \varphi_{(1,j,k)}(x^1, x^2) = \phi_{j,k_1}(x^1)\psi_{j,k_2}(x^2), \\ \varphi_{(2,j,k)}(x^1, x^2) = \psi_{j,k_1}(x^1)\phi_{j,k_2}(x^2), \\ \varphi_{(3,j,k)}(x^1, x^2) = \psi_{j,k_1}(x^1)\psi_{j,k_2}(x^2), \end{cases}$$

where $\phi = \mathbb{1}_{[0,1)}$ and $\psi = \mathbb{1}_{[0,1/2)} - \mathbb{1}_{[1/2,1)}$ are respectively the one-dimensional Haar father and Haar mother wavelets and

$$\Phi_{j,k}(\cdot) = 2^{j/2}\Phi(2^j \cdot -k)$$

denotes the dilated/translated wavelet at scale j in $\mathbb{N}$ for Φ being either ϕ or ψ. Notice that $\mathcal{K}_j$ corresponds to the set of translations k such that for any $1 \leq i \leq 3$, the intersection between the supports of the wavelets $\varphi_{(i,j,k)}$ and $[0, 1)^2$ is not empty.

Then the function φ is taken out of the family $\{\varphi_\delta, \delta \in \Lambda\}$, with

$$\Lambda = \{0\} \cup \left\{(i, j, k) \in \{1, 2, 3\} \times \mathbb{N} \times \mathcal{K}_j\right\},$$

which constitutes an orthonormal basis of $\mathbb{L}_2\left([0, 1]^2\right)$. Notice that in this case, the Lipschitz assumptions of Adamczak et al. (see Theorem 17.1.3) are not satisfied, since the Haar wavelet functions are not even continuous.

The critical value of the test is obtained from the permutation approach, inspired by Hoeffding [16], and Romano [23]. Let Π be a uniformly distributed random permutation of $\{1, \ldots, n\}$ independent of $\mathbb{X}_n$ and consider the permuted sample

$$\mathbb{X}_n^{\Pi} = (X_1^{\Pi}, \ldots X_n^{\Pi}), \quad \text{where } \forall 1 \leq i \leq n, \quad X_i^{\Pi} = (X_i^1, X_{\Pi(i)}^2),$$

obtained from permuting only the second coordinates. Then, under $(\mathcal{H}_0)$, the original sample $\mathbb{X}_n$ and the permuted one $\mathbb{X}_n^{\Pi}$ have the same distribution. Hence, the critical value of the upper-tailed test, denoted by $q_{1-\alpha}(\mathbb{X}_n)$, is the $(1 - \alpha)$-quantile of the conditional distribution of the permuted statistic $T(\mathbb{X}_n^{\Pi})$ given the sample $\mathbb{X}_n$, where the permuted test statistic is equal to

$$T(\mathbb{X}_n^{\Pi}) = \frac{1}{n-1} \left(\sum_{i=1}^{n} \varphi(X_i^1, X_{\Pi(i)}^2) - \frac{1}{n} \sum_{i,j=1}^{n} \varphi(X_i^1, X_j^2) \right).$$

More precisely, given $\mathbb{X}_n$, if

$$T^{(1)}(\mathbb{X}_n) \leq T^{(2)}(\mathbb{X}_n) \leq \cdots \leq T^{(n!)}(\mathbb{X}_n)$$

denote the ordered values of all the permuted test statistic $T(\mathbb{X}_n^{\tau})$, when τ describes the set of all permutations of $\{1, \ldots, n\}$, then the critical value is equal to

$$q_{1-\alpha}(\mathbb{X}_n) = T^{(n! - \lfloor n!\alpha \rfloor)}(\mathbb{X}_n). \tag{17.18}$$

The corresponding test rejects the null hypothesis when $T(\mathbb{X}_n) > q_{1-\alpha}(\mathbb{X}_n)$, here denoted by

$$\Delta_\alpha(\mathbb{X}_n) = \mathbb{1}_{T(\mathbb{X}_n) > q_{1-\alpha}(\mathbb{X}_n)}. \tag{17.19}$$

In [2], the asymptotic properties of such test are studied. Based on a combinatorial central limit theorem in a non-i.i.d. case, the test is proved to be, under mild conditions, asymptotically of prescribed size, and power equal to one under any reasonable alternatives. Yet, as explained above, such purely asymptotic properties may be insufficient when applying these tests in neuroscience for instance. Moreover, the delicate choice of φ, generally out of a parametric family $\{\varphi_\delta\}_\delta$ (which reduces to the choice of the parameter δ), is a real question, especially, in neuroscience, where it has some biological meaning, as mentioned in [2] and [3]. A possible approach to overcome this issue is to aggregate several tests for different parameters δ, and reject independence if at least one of them does. In particular, this approach should give us information on how to choose this parameter. Yet, to do so, non-asymptotic controls are necessary.

From a non-asymptotic point of view, since the test is non-asymptotically of prescribed level by construction, remains the non-asymptotic control of the second kind error rate, that is the probability of wrongly accepting the null hypothesis. In

the spirit of [11, 12, 24], the idea is to study the uniform separation rates of testing, in order to study the optimality in the minimax sense (see [5]).

From now on, consider an alternative P satisfying $(\mathcal{H}_1)$, and an i.i.d. sample $\mathbb{X}_n$ from such distribution P. Assume moreover that the alternative satisfies $\mathbb{E}_P[\varphi] > \mathbb{E}_{\perp\!\!\!\perp}[\varphi]$, that is $\mathbb{E}[T(\mathbb{X}_n)] > 0$. The initial step is to find some condition on P guaranteeing the control of the second kind error rate, namely $\mathbb{P}(\Delta_\alpha(\mathbb{X}_n) = 0)$, by a prescribed value $\beta > 0$. Intuitively, since the expectation of the test statistic $\mathbb{E}[T(\mathbb{X}_n)]$ is equal to zero under the null hypothesis, the test should be more efficient in rejecting $(\mathcal{H}_0)$ for large values of this expectation. So, the aim is to find conditions of the form $\mathbb{E}[T(\mathbb{X}_n)] \geq s$ for some threshold s to be determined. Yet, one of the main difficulties here comes from the randomness of the critical value. The idea, as in [11], is thus to introduce $q_{1-\beta/2}^\alpha$ the $(1 - \beta/2)$-quantile of the critical value $q_{1-\alpha}(\mathbb{X}_n)$ and deduce from Chebychev's inequality (see section "A First Condition Ensuing from Chebychev's Inequality" in the Appendix), that the second kind error rate is controlled by β as soon as

$$\mathbb{E}[T(\mathbb{X}_n)] \geq q_{1-\beta/2}^\alpha + \sqrt{\frac{2}{\beta} \operatorname{Var}(T(\mathbb{X}_n))}. \qquad (17.20)$$

Usually, the goal in general minimax approaches is to express, for well-chosen functions φ, some distance between the alternative P and the null hypothesis $(\mathcal{H}_0)$ in terms of $\mathbb{E}[T(\mathbb{X}_n)]$ for which minimax lower-bounds are known (see for instance [11, 12]). The objective is then to control, up to a constant, such distance (and in particular each term in the right-hand side of (17.20)) by the minimax rate of independence testing with respect to such distance on well-chosen regularity subspaces of alternatives, in order to prove the optimality of the method from a theoretical point of view. The interested reader could refer to the thesis [1, Chapter 4] for more details about this kind of development in the density case. It is not in the scope of the present article to develop such minimax theory in the general case, but to provide some general tools providing some sharp control of each term in the right-hand side of (17.20) which consists in a very first step of this approach. Some technical computations imply that the variance term can be upper bounded, up to a multiplicative constant, by $n^{-1}(\mathbb{E}_P[\varphi^2] + \mathbb{E}_{\perp\!\!\!\perp}[\varphi^2])$ (see Lemma 17.3.1). Hence, the challenging part relies in the quantile term. At this point, several ideas have been explored.

17.3.2 Why Concentration Inequalities Are Necessary

A first idea to control the conditional quantile of the permuted test statistic is based on the non-asymptotic control of the critical value obtained in section

"Control of the Critical Value Based on Hoeffding's Approach" in the Appendix
(see Eq. (17.49)), following Hoeffding's idea (see [16, Theorem 2.1]), that leads to
the condition

$$\mathbb{E}\left[T(\mathbb{X}_n)\right] \geq \frac{4}{\sqrt{\alpha}} \sqrt{\frac{2}{\beta} \frac{\mathbb{E}_P\left[\varphi^2\right] + \mathbb{E}_{\perp\!\!\!\perp}\left[\varphi^2\right]}{n}}. \tag{17.21}$$

The proof of this result is detailed in section "A First Condition Ensuing from
Hoeffding's Approach" in the Appendix. Yet, this result may not be sharp enough,
especially in α. Indeed, as explained above, the next step consists in aggregating
several tests for different functions φ out of a parametric family $\{\varphi_\delta\}_\delta$ in a purpose
of adaptivity. Generally, when aggregating tests, as in multiple testing methods,
the multiplicity of the tests has to be taken into account. In particular, the single
prescribed level of each individual test should be corrected. Several corrections
exist, such as the Bonferroni one, which consists in dividing the global desired level
α by the number of tests M. Yet, for such correction, the lower-bound in (17.21)
comes with a cost in $\sqrt{M}$, which is too large to provide optimal rates. Even with
more sophisticated corrections than the Bonferroni one (see, e.g., [11, 12, 24]), the
control by a term of order $\sqrt{1/\alpha}$ is too large, since classically in the literature, the
dependence on α should be of the order of $\sqrt{\ln(1/\alpha)}$. Hence, the bound ensuing
from this first track being not sharp enough, the next idea was to investigate other
non-asymptotic approaches for permuted sums.

   Such approaches have also been studied in the literature. For instance, Ho and
Chen [14] obtain non-asymptotic Berry-Esseen type bounds in the $\mathbb{L}^p$-distance
between the cumulative distribution function (c.d.f.) of the standardized permuted
sum of i.i.d. random variables and the c.d.f. of the normal distribution, based
on Stein's method. In particular, they obtain the rate of convergence to a normal
distribution in $\mathbb{L}^p$-distance under Lindeberg-type conditions. Then, Bolthausen [7]
considers a different approach, also based on Stein's method allowing to extend
Ho and Chen's results in the non-identically distributed case. More precisely, he
obtains bounds in the $\mathbb{L}^\infty$-distance in the non-random case. In particular, in the
deterministic case (which can easily be generalized to random cases), considering
the notation introduced above, he obtains the following non-asymptotic bound:

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P}\left(Z - \mathbb{E}\left[Z\right] \leq x\sqrt{\mathrm{Var}(Z)}\right) - \Phi_{0,1}(x) \right| \leq \frac{C}{n\sqrt{\mathrm{Var}(Z)}^3} \sum_{i,j=1}^{n} \left|d_{i,j}\right|^3,$$

where C is an absolute constant, and $\Phi_{0,1}$ denotes the standard normal distribution
function. In particular, when applying this result to answer our motivation by
considering random variables $\varphi(X_i^1, X_j^2)$ instead of the deterministic terms $a_{i,j}$, and
working conditionally on the sample $\mathbb{X}_n$, the permuted statistic $T(\mathbb{X}_n^\Pi)$ corresponds

to $(n-1)^{-1}(Z - \mathbb{E}[Z])$. Therefore, the previous inequality implies that, for all t in $\mathbb{R}$,

$$
\mathbb{P}\big(T\big(\mathbb{X}_n^\Pi\big) > t\big|\mathbb{X}_n\big) \leq \left[1 - \Phi_{0,1}\left(\frac{t}{\sqrt{\mathrm{Var}\big(T\big(\mathbb{X}_n^\Pi\big)\big|\mathbb{X}_n\big)}}\right)\right]
$$
$$
+ \frac{C}{n(n-1)^{2/3}\sqrt{\mathrm{Var}\big(T\big(\mathbb{X}_n^\Pi\big)\big|\mathbb{X}_n\big)}^{-3}} \sum_{i,j}\big|D_{i,j}\big|^3,
$$
(17.22)

where $D_{i,j}$ is defined by

$$
D_{i,j} = \varphi(X_i^1, X_j^2) - \frac{1}{n}\sum_{l=1}^n \varphi(X_i^1, X_l^2) - \frac{1}{n}\sum_{k=1}^n \varphi(X_k^1, X_j^2) + \frac{1}{n^2}\sum_{k,l=1}^n \varphi(X_k^1, X_l^2).
$$

Yet, by definition of conditional quantiles, the critical value $q_{1-\alpha}(\mathbb{X}_n)$ is the smallest value of t such that $\mathbb{P}(T(\mathbb{X}_n) > t|\mathbb{X}_n) \leq \alpha$. Hence, considering (17.22), one can easily make the first term of the sum in the right-hand side of the inequality as small as one wants by choosing t large enough. However, the second term being fixed, nothing guarantees that the upper-bound in (17.22) can be constrained to be smaller than α. Thus, this result cannot be applied in order to control non-asymptotically the critical value. Concentration inequalities seem thus to be adequate here, as they provide sharp non-asymptotic results, with usually exponentially small controls which leads to the desired logarithmic dependency in α, as mentioned above.

17.3.3 A Sharp Control of the Conditional Quantile and a New Condition Guaranteeing a Control of the Second Kind Error Rate

Sharp controls of the quantiles are provided in the following proposition.

Proposition 17.3.1 *Consider the same notation as in Sect. 17.3.1 and let $q_{1-\beta/2}^\alpha$ be the $(1-\beta/2)$-quantile of the conditional quantile $q_{1-\alpha}(\mathbb{X}_n)$. Then, there exists two universal positive constants C' and c_0 such that*

$$
q_{1-\alpha}(\mathbb{X}_n) \leq \frac{C'}{n-1}\left\{\sqrt{\frac{1}{n}\sum_{i,j=1}^n \varphi^2(X_i^1, X_j^2)}\sqrt{\ln\left(\frac{c_0}{\alpha}\right)} + \|\varphi\|_\infty \ln\left(\frac{c_0}{\alpha}\right)\right\}.
$$
(17.23)

As a consequence, there exists a universal positive constants C such that

$$q_{1-\beta/2}^{\alpha} \leq C \left\{ \sqrt{\frac{2}{\beta} \ln\left(\frac{c_0}{\alpha}\right)} \left(\frac{\sqrt{\mathbb{E}_P[\varphi^2]}}{n} + \frac{\sqrt{\mathbb{E}_{\perp\!\!\!\perp}[\varphi^2]}}{\sqrt{n}} \right) + \frac{\|\varphi\|_\infty}{n} \ln\left(\frac{c_0}{\alpha}\right) \right\}.$$
(17.24)

Moreover, a control of the variance term is obtained in the following lemma based on the Cauchy-Schwartz inequality.

Lemma 17.3.1 *Let $n \geq 4$ and $\mathbb{X}_n$ be a sample of n i.i.d. random variables with distribution P and marginals P^1 and P^2. Let T be the test statistic defined in (17.16), and $\mathbb{E}_P[\cdot]$ and $\mathbb{E}_{\perp\!\!\!\perp}[\cdot]$ be notation introduced in (17.17). Then, if both $\mathbb{E}_P[\varphi^2] < +\infty$ and $\mathbb{E}_{\perp\!\!\!\perp}[\varphi^2] < +\infty$,*

$$\mathrm{Var}(T(\mathbb{X}_n)) \leq \frac{1}{n} \left(\sqrt{\mathbb{E}_P[\varphi^2]} + 2\sqrt{\mathbb{E}_{\perp\!\!\!\perp}[\varphi^2]} \right)^2.$$

Proposition 17.3.1 and Lemma 17.3.1 both imply that the right-hand side of (17.20) is upper bounded by

$$C'' \left\{ \sqrt{\frac{2}{\beta} \left[\ln\left(\frac{c_0}{\alpha}\right) + 1 \right] \frac{(\mathbb{E}_P[\varphi^2] + \mathbb{E}_{\perp\!\!\!\perp}[\varphi^2])}{n}} + \frac{\|\varphi\|_\infty}{n} \ln\left(\frac{c_0}{\alpha}\right) \right\}, \quad (17.25)$$

where C'' is a universal constant.

Indeed, the control of $q_{1-\beta/2}^{\alpha}$ is implied by (17.24) combined with the concavity property of the square-root function. Lemma 17.3.1 directly implies that the variance term satisfies

$$\mathrm{Var}(T(\mathbb{X}_n)) \leq \frac{8}{n} \left(\mathbb{E}_P[\varphi^2] + \mathbb{E}_{\perp\!\!\!\perp}[\varphi^2] \right),$$

Finally, if $\mathbb{E}[T(\mathbb{X}_n)]$ is larger than the quantity in (17.25), then condition (17.20) is satisfied which directly provides that $\mathbb{P}(\Delta_\alpha(\mathbb{X}_n) = 0) \leq \beta$, that is the second kind error rate of the test Δ_α is less than or equal to the prescribed value β. One may notice that this time, the dependence in α is, as expected, of the order of $\sqrt{\ln(1/\alpha)}$.

17.4 Proofs

17.4.1 Proof of Lemma 17.2.1

Sketch of Proof From now on, fix $t > 0$. Recall the notation introduced by Talagrand in Theorem 17.1.2. The main purpose of these notation is to introduce

some notion of distance between a permutation τ in $\mathfrak{S}_n$ and a subset A of $\mathfrak{S}_n$. To do so, the idea is to reduce the set of interest to a simpler one, that is $[0, 1]^n$, by considering

$$U_A(\tau) = \left\{ s \in \{0, 1\}^n \; ; \; \exists \sigma \in A \text{ such that } \forall 1 \leq i \leq n, \; s_i = 0 \implies \sigma(i) = \tau(i) \right\}.$$

One may notice that the permutation τ belongs to A if and only if 0 belongs to the set $U_A(\tau)$. Hence, the corresponding distance between the permutation τ and the set A is coded by the distance between 0 and the set $U_A(\tau)$ and thus defined by

$$f(A, \tau) = \min \left\{ \sum_{i=1}^{n} v_i^2 \; ; \; v = (v_i)_{1 \leq i \leq n} \in V_A(\tau) \right\},$$

where $V_A(\tau) = \text{ConvexHull}(U_A(\tau))$. One may notice in particular that A contains τ if and only if the distance $f(A, \tau) = 0$.

The global frame of the proof of Lemma 17.2.1 (and also Proposition 17.2.1) relies on the following steps. The first step consists in proving that

$$\mathbb{P}\left(\sqrt{Z} \geq \sqrt{C_A} + t \sqrt{\max_{1 \leq i,j \leq n} \{a_{i,j}\}} \right) \leq \frac{e^{-t^2/16}}{\mathbb{P}(Z \in A)}, \tag{17.26}$$

for some subset A of $\mathfrak{S}_n$ of the shape $A = \{\sigma \in \mathfrak{S}_n \; ; \; Z(\sigma) \leq C_A\}$ for some constant C_A to be chosen later. For this purpose, since Talagrand's inequality for random permutations (see Theorem 17.1.2) provides that

$$\mathbb{P}\left(f(A, \Pi) \geq t^2 \right) \leq \frac{e^{-t^2/16}}{\mathbb{P}(\Pi \in A)},$$

it is sufficient to prove that

$$\mathbb{P}\left(f(A, \Pi) \geq t^2 \right) \geq \mathbb{P}\left(\sqrt{Z} \geq \sqrt{C_A} + t \sqrt{\max_{1 \leq i,j \leq n} \{a_{i,j}\}} \right),$$

to obtain (17.26). To do so, the idea, as in [4], is to show that the assertion $f(A, \Pi) < t^2$ implies that $\sqrt{Z} < \sqrt{C_A} + t \sqrt{\max_{1 \leq i,j \leq n} \{a_{i,j}\}}$, and to conclude by contraposition.

Then, the two following steps consist in choosing appropriate constants C_A in (17.26) depending on the median of Z, such that both $\mathbb{P}\left(\sqrt{Z} \geq \sqrt{C_A} + t \sqrt{\max_{1 \leq i,j \leq n} \{a_{i,j}\}} \right)$ and $\mathbb{P}(Z \in A)$ are greater than $1/2$, in order to control both probabilities

$$\mathbb{P}\left(\sqrt{Z} \geq \sqrt{\text{med}(Z)} + t \sqrt{\max_{1 \leq i,j \leq n} \{a_{i,j}\}} \right) \text{ and } \mathbb{P}\left(\sqrt{Z} \leq \sqrt{\text{med}(Z)} - t \sqrt{\max_{1 \leq i,j \leq n} \{a_{i,j}\}} \right)$$

respectively in (17.10) and (17.11).

First Step: Preliminary Study Assume $f(A, \Pi) < t^2$. Then, by definition of the distance f, there exists some $s^1, \ldots, s^m$ in $U_A(\Pi)$, and some non-negative weights $p_1, \ldots, p_m$ satisfying $\sum_{j=1}^m p_j = 1$ such that

$$\sum_{i=1}^n \left[\left(\sum_{j=1}^m p_j s_i^j \right)^2 \right] < t^2.$$

For each $1 \le j \le m$, since s^j belongs to $U_A(\Pi)$, one may consider a permutation σ_j in A associated to s^j (that is satisfying $s_i^j = 0 \implies \sigma_j(i) = \Pi(i)$). Then, since the $a_{i,j}$ are non-negative, and from the Cauchy-Schwartz inequality,

$$Z - \sum_{j=1}^m p_j Z(\sigma_j) = \sum_{i=1}^n \sum_{j=1}^m p_j \left(a_{i,\Pi(i)} - a_{i,\sigma_j(i)} \right)$$

$$= \sum_{i=1}^n \sum_{j=1}^m p_j \left(a_{i,\Pi(i)} - a_{i,\sigma_j(i)} \right) s_i^j$$

$$\le \sum_{i=1}^n \left[\left(\sum_{j=1}^m p_j s_i^j \right) a_{i,\Pi(i)} \right]$$

$$\le \sqrt{\sum_{i=1}^n \left(\sum_{j=1}^m p_j s_i^j \right)^2} \sqrt{\sum_{i=1}^n a_{i,\Pi(i)}^2}$$

$$< t \sqrt{\max_{1 \le i,j \le n} \{a_{i,j}\}} \sqrt{Z}.$$

Thus, as the σ_j's all belong to $A = \{\sigma \; ; \; Z(\sigma) \le C_A\}$,

$$Z < C_A + t \sqrt{\max_{1 \le i,j \le n} \{a_{i,j}\}} \sqrt{Z}.$$

Therefore, by solving the second-order polynomial in $\sqrt{Z}$ above, one obtains

$$\sqrt{Z} < \frac{t \sqrt{\max_{1 \le i,j \le n} \{a_{i,j}\}} + \sqrt{t^2 \max_{1 \le i,j \le n} \{a_{i,j}\} + 4 C_A}}{2} \le t \sqrt{\max_{1 \le i,j \le n} \{a_{i,j}\}} + \sqrt{C_A}.$$

Finally, by contraposition,

$$\mathbb{P}\left(\sqrt{Z} \geq \sqrt{C_A} + t\sqrt{\max_{1 \leq i,j \leq n} \{a_{i,j}\}}\right) \leq \mathbb{P}\left(f(A, \Pi) \geq t^2\right),$$

which, combined with (17.7) of Theorem 17.1.2 provides (17.26).

Second Step: Proof of (17.10) Taking $C_A = \text{med}(Z)$ guarantees $\mathbb{P}(Z \in A) \geq 1/2$ and thus, (17.26) provides (17.10).

Third Step: Proof of (17.11) Taking $C_A = \left(\sqrt{\text{med}(Z)} - t\sqrt{\max_{1 \leq i,j \leq n}\{a_{i,j}\}}\right)^2$ implies

$$\mathbb{P}\left(\sqrt{Z} \geq \sqrt{C_A} + t\sqrt{\max_{1 \leq i,j \leq n}\{a_{i,j}\}}\right) = \mathbb{P}\left(\sqrt{Z} \geq \sqrt{\text{med}(Z)}\right) = \mathbb{P}(Z \geq \text{med}(Z)) \geq \frac{1}{2}.$$

So finally, again by (17.26),

$$\mathbb{P}\left(\sqrt{Z} \leq \sqrt{\text{med}(Z)} - t\sqrt{\max_{1 \leq i,j \leq n}\{a_{i,j}\}}\right) = \mathbb{P}(Z \in A)$$

$$\leq \frac{e^{-t^2/16}}{\mathbb{P}\left(\sqrt{Z} \geq \sqrt{C_A} + t\sqrt{\max_{1 \leq i,j \leq n}\{a_{i,j}\}}\right)}$$

$$\leq 2e^{-t^2/16},$$

which ends the proof of the Lemma.

17.4.2 Proof of Lemma 17.2.2

Let X be any real random variable. Recall that

$$\text{med}(X) \in \operatorname*{argmin}_{m \in \mathbb{R}} \mathbb{E}\left[|X - m|\right].$$

In particular, thanks to Jensen's inequality,

$$|\mathbb{E}[X] - \text{med}(X)| \leq \mathbb{E}\left[|X - \text{med}(X)|\right]$$

$$\leq \mathbb{E}\left[|X - \mathbb{E}[X]|\right]$$

$$\leq \sqrt{\mathbb{E}\left[(X - \mathbb{E}[X])^2\right]}$$

$$\leq \sqrt{\text{Var}(X)}. \tag{17.27}$$

17.4.3 Proof of Proposition 17.2.1

From now on, fix $x > 0$, and consider $t = x^2$. This proof is again based on Talagrand's inequality for random permutations, combined with (17.10) in Lemma 17.2.1. It follows exactly the same progression as in the proof of Lemma 17.2.1; the preliminary step consists in working with subsets $A \subset \mathfrak{S}_n$ of the form $A = \{\sigma \in \mathfrak{S}_n \; ; \; Z(\sigma) \leq C_A\}$ for some constant C_A, in order to obtain for all $v > 0$,

$$\mathbb{P}\left(Z \geq C_A + t\left(\sqrt{\text{med}\left(\sum_{i=1}^{n} a_{i,\Pi(i)}^2\right)} + v \max_{1 \leq i,j \leq n}\{a_{i,j}\}\right)\right) \leq \frac{e^{-t^2/16}}{\mathbb{P}(Z \in A)} + 2e^{-v^2/16}.$$
(17.28)

The second and third step consist in picking up a well-chosen constant C_A and a well-chosen $v > 0$ in order to obtain respectively

$$\mathbb{P}\left(Z \geq \text{med}\,(Z) + t\left(\sqrt{\text{med}\left(\sum_{i=1}^{n} a_{i,\Pi(i)}^2\right)} + (t \vee C_0) \max_{1 \leq i,j \leq n}\{a_{i,j}\}\right)\right) \leq 4e^{-t^2/16},$$
(17.29)

and

$$\mathbb{P}\left(Z \leq \text{med}\,(Z) - t\left(\sqrt{\text{med}\left(\sum_{i=1}^{n} a_{i,\Pi(i)}^2\right)} + (t \vee C_0) \max_{1 \leq i,j \leq n}\{a_{i,j}\}\right)\right) \leq 4e^{-t^2/16},$$
(17.30)

where $C_0 = 4\sqrt{\ln(8)}$. The final step combines (17.29) and (17.30) in order to prove (17.12).

First Step: Preliminary Study Let $A = \{\sigma \in \mathfrak{S}_n \; ; \; Z(\sigma) \leq C_A\}$ with C_A a general constant, and fix $v > 0$. Assume, this time, that both

$$f(A, \Pi) < t^2 \quad \text{and} \quad \sqrt{\sum_{i=1}^{n} a_{i,\Pi(i)}^2} < \sqrt{\text{med}\left(\sum_{i=1}^{n} a_{i,\Pi(i)}^2\right)} + v \max_{1 \leq i,j \leq n}\{a_{i,j}\}.$$
(17.31)

Then, as in the preliminary study of the proof of Lemma 17.2.1, from the first assumption in (17.31), there exists some $s^1, \ldots, s^m$ in $U_A(\Pi)$, and some non-negative weights $p_1, \ldots, p_m$ satisfying $\sum_{j=1}^{m} p_j = 1$ such that

$$\sum_{i=1}^{n} \left[\left(\sum_{j=1}^{m} p_j s_i^j \right)^2 \right] < t^2.$$

For each $1 \leq j \leq m$, consider σ_j in A associated to s^j, that is a permutation σ_j in A satisfying $s_i^j = 0 \implies \sigma_j(i) = \Pi(i)$. Then, combining the Cauchy-Shwartz inequality with the second assumption in (17.31) leads to

$$
\begin{aligned}
Z - \sum_{j=1}^{m} p_j Z(\sigma_j) &= \sum_{i=1}^{n} \sum_{j=1}^{m} p_j \left(a_{i,\Pi(i)} - a_{i,\sigma_j(i)} \right) s_i^j \\
&\leq \sum_{i=1}^{n} \left[\left(\sum_{j=1}^{m} p_j s_i^j \right) a_{i,\Pi(i)} \right] \\
&\leq \sqrt{\sum_{i=1}^{n} \left(\sum_{j=1}^{m} p_j s_i^j \right)^2} \sqrt{\sum_{i=1}^{n} a_{i,\Pi(i)}^2} \\
&< t \left(\sqrt{\mathrm{med} \left(\sum_{i=1}^{n} a_{i,\Pi(i)}^2 \right)} + v \max_{1 \leq i,j \leq n} \{a_{i,j}\} \right).
\end{aligned}
$$

Notice that here, the reasoning begins exactly as in the proof of Lemma 17.2.1. Yet, the second assumption in (17.31), which can be controlled using that lemma, allows us to sharpen the inequality. Thus, as the σ_j's all belong to $A = \{\sigma \; ; \; Z(\sigma) \leq C_A\}$,

$$Z < C_A + t \left(\sqrt{\mathrm{med} \left(\sum_{i=1}^{n} a_{i,\Pi(i)}^2 \right)} + v \max_{1 \leq i,j \leq n} \{a_{i,j}\} \right). \tag{17.32}$$

Hence, by contraposition of (17.31) $\implies$ (17.32), one obtains

$$
\mathbb{P} \left(Z \geq C_A + t \left(\sqrt{\mathrm{med} \left(\sum_{i=1}^{n} a_{i,\Pi(i)}^2 \right)} + v \max_{1 \leq i,j \leq n} \{a_{i,j}\} \right) \right)
$$
$$
\leq \mathbb{P} \left(f(A, \Pi) \geq t^2 \right) + \mathbb{P} \left(\sqrt{\sum_{i=1}^{n} a_{i,\Pi(i)}^2} \geq \sqrt{\mathrm{med} \left(\sum_{i=1}^{n} a_{i,\Pi(i)}^2 \right)} + v \max_{1 \leq i,j \leq n} \{a_{i,j}\} \right),
$$

and (17.28) follows from Theorem 17.1.2 and (17.10) in Lemma 17.2.1.

Second Step: Proof of (17.29) Consider $C_A = \text{med}\,(Z)$ so that $\mathbb{P}(Z \in A) \geq 1/2$. Thus, if $v = t$ in (17.28),

$$\mathbb{P}\left(Z \geq \text{med}\,(Z) + t\left(\sqrt{\text{med}\left(\sum_{i=1}^{n} a_{i,\Pi(i)}^2\right)} + (t \vee C_0) \max_{1 \leq i,j \leq n} \{a_{i,j}\}\right)\right)$$

$$\leq \mathbb{P}\left(Z \geq \text{med}\,(Z) + t\left(\sqrt{\text{med}\left(\sum_{i=1}^{n} a_{i,\Pi(i)}^2\right)} + t \max_{1 \leq i,j \leq n} \{a_{i,j}\}\right)\right)$$

$$\leq 4e^{-t^2/16}.$$

Notice that the maximum with the constant in $(t \vee C_0)$ is not necessary in the case only a control of the right-tail is wanted.

Third Step: Proof of (17.30) Consider now

$$C_A = \text{med}\,(Z) - t\left(\sqrt{\text{med}\left(\sum_{i=1}^{n} a_{i,\Pi(i)}^2\right)} + v \max_{1 \leq i,j \leq n} \{a_{i,j}\}\right),$$

so that

$$\mathbb{P}\left(Z \geq C_A + t\left(\sqrt{\text{med}\left(\sum_{i=1}^{n} a_{i,\Pi(i)}^2\right)} + v \max_{1 \leq i,j \leq n} \{a_{i,j}\}\right)\right) = \mathbb{P}(Z \geq \text{med}\,(Z)) \geq \frac{1}{2}.$$

Hence, on the one hand, from (17.28),

$$\mathbb{P}(Z \in A) \leq \frac{e^{-t^2/16}}{\left(\frac{1}{2} - 2e^{-v^2/16}\right)}.$$

Thus, if $v = C_0 = 4\sqrt{\ln(8)}$, then $\left(1/2 - 2e^{-v^2/16}\right) = 1/4$, and $\mathbb{P}(Z \in A) \leq 4e^{-t^2/16}$.

On the other hand, as $(t \vee C_0) \geq C_0 = v$,

$$\mathbb{P}(Z \in A) \geq \mathbb{P}\left(Z \leq \text{med}\,(Z) - t\left(\sqrt{\text{med}\left(\sum_{i=1}^{n} a_{i,\Pi(i)}^2\right)} + (t \vee C_0) \max_{1 \leq i,j \leq n} \{a_{i,j}\}\right)\right),$$

which ends the proof of (17.30).

Fourth Step: Proof of (17.12) Both (17.29) and (17.30) lead to

$$\mathbb{P}\left(|Z - \mathrm{med}\,(Z)| > t\left(\sqrt{\mathrm{med}\left(\sum_{i=1}^{n} a_{i,\Pi(i)}^2\right)} + (t \vee C_0)\max_{1 \le i,j \le n}\{a_{i,j}\}\right)\right) \le 8e^{-t^2/16}.$$

Thus, on the one hand, if $t \ge C_0$, that is $t \vee C_0 = t$, and (17.12) holds. On the other hand, if $t < C_0$,

$$\mathbb{P}\left(|Z - \mathrm{med}\,(Z)| > t\left(\sqrt{\mathrm{med}\left(\sum_{i=1}^{n} a_{i,\Pi(i)}^2\right)} + t\max_{1 \le i,j \le n}\{a_{i,j}\}\right)\right) \le 1$$

$$\le e^{C_0^2/16 - t^2/16} = 8e^{-t^2/16},$$

which ends the proof of the Proposition by taking $x = \sqrt{t}$.

17.4.4   Proof of Proposition 17.2.2

First, for a better readability, let

$$M = \max_{1 \le i,j \le n}\{a_{i,j}\} \quad \text{and} \quad V = \mathbb{E}\left[\sum_{i=1}^{n} a_{i,\Pi(i)}^2\right] = \frac{1}{n}\sum_{i,j=1}^{n} a_{i,j}^2.$$

Then, $\mathrm{med}\left(\sum_{i=1}^{n} a_{i,\Pi(i)}^2\right) \le 2V$ since by Markov's inequality, for all non-negative random variable X, $\mathrm{med}\,(X) \le 2\mathbb{E}\,[X]$. Indeed,

$$\frac{1}{2} \le \mathbb{P}(X \ge \mathrm{med}\,(X)) \le \frac{\mathbb{E}\,[X]}{\mathrm{med}\,(X)}.$$

Thus, by Proposition 17.2.1, one obtains that, for all $x > 0$,

$$\mathbb{P}\left(|Z - \mathrm{med}\,(Z)| \ge \sqrt{2Vx} + Mx\right) \le 8e^{-x/16}. \tag{17.33}$$

The following is based on Lemma 17.2.2, and provides an upper-bound of the difference between the expectation and the median of Z.

Lemma 17.4.1 *With the notation defined above,*

$$|\mathbb{E}\,[Z] - \mathrm{med}\,(Z)| \le \sqrt{2V}.$$

Proof (Proof of Lemma 17.4.1) Lemma 17.2.2 implies that

$$|\mathbb{E}[Z] - \text{med}(Z)| \le \sqrt{\text{Var}(Z)}.$$

Let us prove that

$$\text{Var}(Z) \le 2V. \tag{17.34}$$

Indeed,

$$\text{Var}(Z) = \mathbb{E}\left[\left(\sum_{i=1}^{n} a_{i,\Pi(i)} - \frac{1}{n}\sum_{i,j=1}^{n} a_{i,j}\right)^2\right]$$

$$= \mathbb{E}\left[\left(\sum_{i,j=1}^{n} a_{i,j}\left(\mathbb{1}_{\Pi(i)=j} - \frac{1}{n}\right)\right)^2\right]$$

$$= \sum_{i,j=1}^{n}\sum_{k,l=1}^{n} a_{i,j}a_{k,l}E_{i,j,k,l},$$

where

$$E_{i,j,k,l} = \mathbb{E}\left[\left(\mathbb{1}_{\Pi(i)=j} - \frac{1}{n}\right)\left(\mathbb{1}_{\Pi(k)=l} - \frac{1}{n}\right)\right] = \mathbb{E}\left[\mathbb{1}_{\Pi(i)=j}\mathbb{1}_{\Pi(k)=l}\right] - \frac{1}{n^2}.$$

In particular,

$$E_{i,j,k,l} = \begin{cases} \dfrac{1}{n} - \dfrac{1}{n^2} \le \dfrac{1}{n} & \text{if } i = k \text{ and } j = l, \\[2ex] \dfrac{-1}{n^2} \le 0 & \text{if } i = k \text{ and } j \ne l \text{ or } i \ne k \text{ and } j = l, \\[2ex] \dfrac{1}{n(n-1)} - \dfrac{1}{n^2} = \dfrac{1}{n^2(n-1)} & \text{if } i \ne k \text{ and } j \ne l. \end{cases}$$

Therefore, from the Cauchy-Schwarz inequality applied to the second sum below (of $n^2(n-1)^2$ terms), one obtains

$$\text{Var}(Z) \le \frac{1}{n}\sum_{i,j=1}^{n} a_{i,j}^2 + \frac{1}{n^2(n-1)}\sum_{i \ne k}\sum_{j \ne l} a_{i,j}a_{k,l}$$

$$\le V + \frac{\sqrt{n^2(n-1)^2}}{n^2(n-1)}\sqrt{\sum_{i \ne k}\sum_{j \ne l} a_{i,j}^2 a_{k,l}^2}$$

$$\text{Var}(Z) \leq V + \frac{1}{n}\sqrt{\sum_{i,j} a_{i,j}^2 \sum_{k,l} a_{k,l}^2}$$

$$= 2V.$$

Finally, combining (17.27) and (17.34) ends the proof of Lemma 17.4.1.   □

Therefore, one deduces from Lemma 17.4.1 and Eq. (17.33) that for all $x > 0$,

$$\mathbb{P}\left(|Z - \mathbb{E}[Z]| \geq \sqrt{2V} + \sqrt{2Vx} + Mx\right) \leq 8e^{-x/16}. \tag{17.35}$$

Now, as in [8, Corollary 2.11], introduce $h_1 : u \in \mathbb{R}^+ \mapsto 1 + u - \sqrt{1 + 2u}$. Then, in particular, h_1 is non-decreasing, convex, one to one function on $\mathbb{R}^+$ with inverse function $h_1^{-1} : v \in \mathbb{R}^+ \mapsto v + \sqrt{2v}$. Indeed,

$$h_1\left(h_1^{-1}(v)\right) = 1 + v + \sqrt{2v} - \sqrt{1 + 2v + 2\sqrt{2v}}$$

$$= 1 + v + \sqrt{2v} - \sqrt{\left(1 + \sqrt{2v}\right)^2} = v,$$

and

$$h_1^{-1}(h_1(u)) = 1 + u - \sqrt{1 + 2u} + \sqrt{2 + 2u - 2\sqrt{1 + 2u}}$$

$$= u + 1 - \sqrt{1 + 2u} + \sqrt{1 - 2\sqrt{1 + 2u} + 1 + 2u}$$

$$= 1 + u - \sqrt{1 + 2u} + \sqrt{\left(1 - \sqrt{1 + 2u}\right)^2} = u.$$

Consider $\mathfrak{a}$ and $\mathfrak{c}$ defined by $\mathfrak{a} = V/M$ and $\mathfrak{c} = M^2/V$, such that $\mathfrak{a}\mathfrak{c} = M$ and $\mathfrak{a}^2\mathfrak{c} = V$ and thus

$$\sqrt{2Vx} + Mx = \mathfrak{a}h_1^{-1}(\mathfrak{c}x).$$

Then, from (17.35),

$$\mathbb{P}\left(|Z - \mathbb{E}[Z]| \geq \sqrt{2\mathfrak{a}^2\mathfrak{c}} + \mathfrak{a}h_1^{-1}(\mathfrak{c}x)\right) \leq 8e^{-x/16}.$$

Let $t > 0$, and consider the two following cases.

First Case If $t \geq \sqrt{2V} = \sqrt{2\mathfrak{a}^2\mathfrak{c}}$, then define $x = \frac{1}{\mathfrak{c}} h_1 \left(\frac{t}{\mathfrak{a}} - \sqrt{2\mathfrak{c}} \right)$ such that $t = \sqrt{2\mathfrak{a}^2\mathfrak{c}} + \mathfrak{a} h_1^{-1}(\mathfrak{c}x)$. Then,

$$\mathbb{P}(|Z - \mathbb{E}[Z]| \geq t) \leq 8 \exp\left(-\frac{1}{16\mathfrak{c}} h_1 \left(\frac{t}{\mathfrak{a}} - \sqrt{2\mathfrak{c}} \right) \right).$$

Yet, by convexity of h_1,

$$h_1 \left(\frac{t}{\mathfrak{a}} - \sqrt{2\mathfrak{c}} \right) \geq 2 h_1 \left(\frac{t}{2\mathfrak{a}} \right) - h_1 \left(\sqrt{2\mathfrak{c}} \right).$$

Hence,

$$\mathbb{P}(|Z - \mathbb{E}[Z]| \geq t) \leq 8 \exp\left(\frac{1}{16\mathfrak{c}} h_1 \left(\sqrt{2\mathfrak{c}} \right) \right) \exp\left(-\frac{1}{8\mathfrak{c}} h_1 \left(\frac{t}{2\mathfrak{a}} \right) \right).$$

Moreover, $\sqrt{2\mathfrak{c}} \leq \mathfrak{c} + \sqrt{2\mathfrak{c}} = h_1^{-1}(\mathfrak{c})$, hence

$$\frac{1}{16\mathfrak{c}} h_1 \left(\sqrt{2\mathfrak{c}} \right) \leq \frac{1}{16}.$$

So finally in this case,

$$\mathbb{P}(|Z - \mathbb{E}[Z]| \geq t) \leq 8 e^{1/16} \exp\left(-\frac{1}{8\mathfrak{c}} h_1 \left(\frac{t}{2\mathfrak{a}} \right) \right). \tag{17.36}$$

Second Case If $t < \sqrt{2V} = \sqrt{2\mathfrak{a}^2\mathfrak{c}}$,

$$\mathbb{P}(|Z - \mathbb{E}[Z]| \geq t) \quad \leq \quad 1 \quad = \quad \exp\left(\frac{1}{8\mathfrak{c}} h_1 \left(\frac{t}{2\mathfrak{a}} \right) \right) \exp\left(-\frac{1}{8\mathfrak{c}} h_1 \left(\frac{t}{2\mathfrak{a}} \right) \right)$$

Moreover, in this case, since $\sqrt{2\mathfrak{c}}/2 \leq h_1^{-1}(\mathfrak{c}/4)$, hence

$$\frac{1}{8\mathfrak{c}} h_1 \left(\frac{t}{2\mathfrak{a}} \right) \leq \frac{1}{8\mathfrak{c}} h_1 \left(\frac{\sqrt{2\mathfrak{c}}}{2} \right) \leq \frac{1}{32},$$

and thus

$$\mathbb{P}(|Z - \mathbb{E}[Z]| \geq t) \leq e^{1/32} \exp\left(-\frac{1}{8\mathfrak{c}} h_1 \left(\frac{t}{2\mathfrak{a}} \right) \right). \tag{17.37}$$

Finally, combining (17.36) and (17.37) leads, in all cases, to

$$\mathbb{P}(|Z - \mathbb{E}[Z]| \geq t) \leq 8e^{1/16} \exp\left(-\frac{1}{8\mathfrak{c}} h_1\left(\frac{t}{2\mathfrak{a}}\right)\right). \qquad (17.38)$$

Now, in order to obtain the Bernstein-type inequality, let $x = \frac{2}{\mathfrak{c}} h_1\left(\frac{t}{2\mathfrak{a}}\right)$, then

$$t = 2\mathfrak{a} h_1^{-1}\left(\frac{\mathfrak{c}x}{2}\right) = \mathfrak{a}\mathfrak{c}x + 2\sqrt{\mathfrak{a}^2\mathfrak{c}x} = 2\sqrt{Vx} + Mx,$$

and thus for all $x > 0$,

$$\mathbb{P}\left(|Z - \mathbb{E}[Z]| \geq 2\sqrt{Vx} + Mx\right) \leq 8e^{1/16} \exp\left(-\frac{x}{16}\right), \qquad (17.39)$$

which ends the proof of the Proposition.

17.4.5   Proof of Corollary 17.2.1

Consider the same notation as in both Proposition 17.2.2 and its proof. This proof follows the one of [19, Corollary 2.10]. Notice that for all $u \geq 0$,

$$h_1(u) \geq \frac{u^2}{2(1+u)}.$$

Hence, from (17.38) in the proof of Proposition 17.2.2, for all $t \geq 0$,

$$\mathbb{P}(|Z - \mathbb{E}[Z]| \geq t) \leq 8e^{1/16} \exp\left(-\frac{1}{8\mathfrak{c}} h_1\left(\frac{t}{2\mathfrak{a}}\right)\right)$$

$$\leq 8e^{1/16} \exp\left(-\frac{t^2}{64\mathfrak{a}^2\mathfrak{c}\,(1 + t/2\mathfrak{a})}\right)$$

$$= 8e^{1/16} \exp\left(-\frac{t^2}{32\,(2\mathfrak{a}^2\mathfrak{c} + \mathfrak{a}\mathfrak{c}t)}\right)$$

$$= 8e^{1/16} \exp\left(-\frac{t^2}{32\,(V + Mt)}\right).$$

which ends the proof of the corollary.

17.4.6 Proof of Theorem 17.2.1

For a better readability, let us introduce $a_{i,j}^+ = a_{i,j}\mathbb{1}_{a_{i,j}\geq 0}$ (respectively $a_{i,j}^- = -a_{i,j}\mathbb{1}_{a_{i,j}<0}$), and denote $Z^+ = \sum_{i=1}^n a_{i,\Pi(i)}^+$ (respectively $Z^- = \sum_{i=1}^n a_{i,\Pi(i)}^-$). Then

$$Z = \sum_{i=1}^n a_{i,\Pi(i)} = Z^+ - Z^-.$$

Moreover, if v (respectively v^+ and v^-) denotes $\frac{1}{n}\sum_{i,j=1}^n a_{i,j}^2$ (respectively $\frac{1}{n}\sum_{i,j=1}^n (a_{i,j}^+)^2$ and $\frac{1}{n}\sum_{i,j=1}^n (a_{i,j}^-)^2$), then $v = v^+ + v^-$ and, from the concavity property of the square root function,

$$\sqrt{2v} \geq \sqrt{v^+} + \sqrt{v^-}.$$

Furthermore, if M^+ (respectively M^-) denotes $\max_{1\leq i,j\leq n}\{a_{i,j}^+\}$ (respectively $\max_{1\leq i,j\leq n}\{a_{i,j}^-\}$), then $2M = 2\max_{1\leq i,j\leq n}\{|a_{i,j}|\} \geq M^+ + M^-$.

Finally, applying Proposition 17.2.2 to Z^+ and Z^- which are both sums of non-negative numbers leads to

$$\mathbb{P}\Big(|Z-\mathbb{E}[Z]| \geq 2\sqrt{2vx} + 2Mx\Big)$$

$$\leq \mathbb{P}\Big(|Z^+ - \mathbb{E}[Z^+]| + |Z^- - \mathbb{E}[Z^-]| \geq 2\sqrt{v^+x} + M^+x + 2\sqrt{v^-x} + M^-x\Big)$$

$$\leq \mathbb{P}\Big(|Z^+ - \mathbb{E}[Z^+]| \geq 2\sqrt{v^+x} + M^+x\Big) + \mathbb{P}\Big(|Z^- - \mathbb{E}[Z^-]| \geq 2\sqrt{v^-x} + M^-x\Big)$$

$$\leq 16e^{1/16}\exp\Big(-\frac{x}{16}\Big),$$

which ends the proof of the Theorem.

17.4.7 Proof of Corollary 17.2.2

Consider the same notation as in the proof of Theorem 17.2.1, and let $t > 0$. Let M denote the maximum $\max_{1\leq i,j\leq n}\{|a_{i,j}|\}$. On the one hand, $M^+ \leq M$ and $M^- \leq M$, and on the other hand, $v^+ \leq v$ and $v^- \leq v$. Therefore, applying Corollary 17.2.1, one obtains

$$\mathbb{P}(|Z - \mathbb{E}[Z]| \geq t) \leq \mathbb{P}\big(|Z^+ - \mathbb{E}[Z^+]| + |Z^- - \mathbb{E}[Z^-]| \geq t\big)$$

$$\leq \mathbb{P}\big(|Z^+ - \mathbb{E}[Z^+]| \geq t/2\big) + \mathbb{P}\big(|Z^- - \mathbb{E}[Z^-]| \geq t/2\big)$$

$$\mathbb{P}(|Z - \mathbb{E}[Z]| \geq t) \leq 8e^{1/16} \exp\left(\frac{-(t/2)^2}{16\left(4v^+ + 2M^+t/2\right)}\right) + 8e^{1/16} \exp\left(\frac{-(t/2)^2}{16\left(4v^- + 2M^-t/2\right)}\right)$$

$$\leq 16e^{1/16} \exp\left(\frac{-t^2}{64\left(4v + Mt\right)}\right),$$

which leads to the following intermediate result

$$\mathbb{P}(|Z - \mathbb{E}[Z]| \geq t) \leq 16e^{1/16} \exp\left(\frac{-t^2}{64\left(4\frac{1}{n}\sum_{i,j=1}^n a_{i,j}^2 + \max_{1 \leq i,j \leq n}\left\{|a_{i,j}|\right\} t\right)}\right). \tag{17.40}$$

In order to make the variance appear, consider Hoeffding's centering trick recalled in (17.1) and introduce

$$d_{i,j} = a_{i,j} - \frac{1}{n}\sum_{k=1}^n a_{k,j} - \frac{1}{n}\sum_{l=1}^n a_{i,l} + \frac{1}{n^2}\sum_{k,l=1}^n a_{k,l} = \frac{1}{n^2}\sum_{k,l=1}^n \left(a_{i,j} - a_{k,j} - a_{i,l} + a_{k,l}\right).$$

One may easily verify that for all i_0 and j_0, $\sum_{i=1}^n d_{i,j_0} = \sum_{j=1}^n d_{i_0,j} = 0$. Moreover,

$$\sum_{i=1}^n d_{i,\Pi(i)} = \sum_{i=1}^n a_{i,\Pi(i)} - \frac{1}{n}\sum_{i,j=1}^n a_{i,j} = Z - \mathbb{E}[Z] \quad \text{and} \quad \mathbb{E}\left[\sum_{i=1}^n d_{i,\Pi(i)}\right] = \frac{1}{n}\sum_{i,j=1}^n d_{i,j} = 0.$$

In particular, applying Eq. (17.40) to the permuted sum of the $d_{i,j}$'s leads to

$$\mathbb{P}(|Z - \mathbb{E}[Z]| \geq t) \leq 16e^{1/16} \exp\left(\frac{-t^2}{64\left(4\frac{1}{n}\sum_{i,j=1}^n d_{i,j}^2 + \max_{1 \leq i,j \leq n}\left\{|d_{i,j}|\right\} t\right)}\right). \tag{17.41}$$

Then, it is sufficient to notice that, on the one hand, from [15, Theorem 2],

$$\text{Var}(Z) = \frac{1}{n-1}\sum_{i,j=1}^n d_{i,j}^2 \geq \frac{1}{n}\sum_{i,j=1}^n d_{i,j}^2,$$

and on the other hand,

$$\max_{1 \leq i,j \leq n}\left\{|d_{i,j}|\right\} \leq 4 \max_{1 \leq i,j \leq n}\left\{|a_{i,j}|\right\},$$

to end the proof of Corollary 17.2.2.

17.4.8 Proof of Proposition 17.3.1

The proof of Proof of Proposition 17.3.1 is divided into two steps. The first step consists in controlling the conditional quantile $q_{1-\alpha}(\mathbb{X}_n)$ and the second step provides an upper-bound for $q^{\alpha}_{1-\beta/2}$.

First Step Let us prove (17.23), that is

$$q_{1-\alpha}(\mathbb{X}_n) \leq \frac{C'}{n-1} \left\{ \sqrt{\frac{1}{n} \sum_{i,j=1}^{n} \varphi^2(X_i^1, X_j^2)} \sqrt{\ln\left(\frac{c_0}{\alpha}\right)} + \|\varphi\|_\infty \ln\left(\frac{c_0}{\alpha}\right) \right\}.$$

Introduce $\tilde{Z}(\mathbb{X}_n) = \sum_{i=1}^{n} \varphi(X_i^1, X_{\Pi(i)}^2)$. Then, notice that

$$T^{\Pi}(\mathbb{X}_n) = \frac{1}{n-1} \left(\tilde{Z}(\mathbb{X}_n) - \mathbb{E}\left[\tilde{Z}(\mathbb{X}_n) \Big| \mathbb{X}_n \right] \right). \tag{17.42}$$

Therefore, applying Theorem 17.2.1 to the conditional probability given $\mathbb{X}_n$, one obtains that there exist universal positive constants c_0 and c_1 such that, for all $x > 0$,

$$\mathbb{P}\left(\left| \tilde{Z}(\mathbb{X}_n) - \mathbb{E}\left[\tilde{Z}(\mathbb{X}_n) \Big| \mathbb{X}_n \right] \right| \geq 2 \sqrt{2 \left(\frac{1}{n} \sum_{i,j=1}^{n} \varphi^2(X_i^1, X_j^2) \right) x + 2\|\varphi\|_\infty x} \Bigg| \mathbb{X}_n \right)$$
$$\leq c_0 \exp\left(-c_1 x\right).$$

In particular, from (17.42), one obtains

$$\mathbb{P}\left(\left| T(\mathbb{X}_n^{\Pi}) \right| \geq \frac{2}{n-1} \left(\sqrt{2 \left(\frac{1}{n} \sum_{i,j=1}^{n} \varphi^2(X_i^1, X_j^2) \right) x} + \|\varphi\|_\infty x \right) \Bigg| \mathbb{X}_n \right)$$
$$\leq c_0 \exp\left(-c_1 x\right).$$

Yet, by definition of the quantile, $q_{1-\alpha}(\mathbb{X}_n)$ is the smallest u such that

$$\mathbb{P}\left(\left| T(\mathbb{X}_n^{\Pi}) \right| \geq u \Big| \mathbb{X}_n \right) \leq \alpha.$$

Thus taking x such that $c_0 \exp(-c_1 x) = \alpha$, that is $x = c_1^{-1} \ln(c_0/\alpha)$, one obtains (17.23) with $C' = 2\max\{\sqrt{2/c_1}, 1/c_1\}$ which is a universal positive constant.

Second Step Let us now control the quantile $q^{\alpha}_{1-\beta/2}$. Since (17.23) is always true, by definition of $q^{\alpha}_{1-\beta/2}$, one has that $q^{\alpha}_{1-\beta/2}$ is upper bounded by the $(1-\beta/2)$-quantile of the right-hand side of (17.23). Yet, the only randomness left in the right-

hand side of (17.23) comes from the randomness of $\frac{1}{n} \sum_{i,j=1}^{n} \varphi^2(X_i^1, X_j^2)$, and thus it is sufficient to control its $(1 - \beta/2)$-quantile.

Besides, applying Markov's inequality, one obtains for all $x > 0$,

$$\mathbb{P}\left(\frac{1}{n} \sum_{i,j=1}^{n} \varphi^2(X_i^1, X_j^2) \geq x\right) \leq \frac{\mathbb{E}\left[\frac{1}{n} \sum_{i,j=1}^{n} \varphi^2(X_i^1, X_j^2)\right]}{x},$$

with $\mathbb{E}\left[\frac{1}{n} \sum_{i,j=1}^{n} \varphi^2(X_i^1, X_j^2)\right] = \mathbb{E}_P[\varphi^2] + (n-1)\mathbb{E}_{\perp\!\!\!\perp}[\varphi^2]$, and thus, taking

$$x = \frac{2}{\beta}\left(\mathbb{E}_P[\varphi^2] + (n-1)\mathbb{E}_{\perp\!\!\!\perp}[\varphi^2]\right),$$

one has that the $(1 - \beta/2)$-quantile of $\frac{1}{n} \sum_{i,j=1}^{n} \varphi^2(X_i^1, X_j^2)$ is upper bounded by x, and thus, the $(1 - \beta/2)$-quantile of $\sqrt{\frac{1}{n} \sum_{i,j=1}^{n} \varphi^2(X_i^1, X_j^2)}$ is itself upper bounded by

$$\sqrt{\frac{2}{\beta}}\left(\sqrt{\mathbb{E}_P[\varphi^2]} + \sqrt{n}\sqrt{\mathbb{E}_{\perp\!\!\!\perp}[\varphi^2]}\right).$$

Finally,

$$q_{1-\beta/2}^{\alpha} \leq \frac{2C'}{n}\left\{\sqrt{\frac{2}{\beta}}\left(\sqrt{\mathbb{E}_P[\varphi^2]} + \sqrt{n}\sqrt{\mathbb{E}_{\perp\!\!\!\perp}[\varphi^2]}\right)\sqrt{\ln\left(\frac{c_0}{\alpha}\right)} + \|\varphi\|_\infty \ln\left(\frac{c_0}{\alpha}\right)\right\}.$$

which is exactly (17.24) for any constant $C \geq 2C'$.

17.5 Proof of Lemma 17.3.1

Let us now prove Lemma 17.3.1. Let $n \geq 4$ and $\mathbb{X}_n$ be an i.i.d. sample with distribution P. First notice that one can write

$$T(\mathbb{X}_n) = \frac{1}{n(n-1)} \sum_{i \neq j} \left(\varphi(X_i^1, X_i^2) - \varphi(X_i^1, X_j^2)\right).$$

In particular, one recovers that $\mathbb{E}[T(\mathbb{X}_n)] = \mathbb{E}_P[\varphi] - \mathbb{E}_{\perp\!\!\!\perp}[\varphi]$.

For a better readability, let us introduce for all $i \neq j$ in $\{1, 2, \ldots, n\}$,

$$Y_i = \varphi(X_i^1, X_i^2) - \mathbb{E}_P[\varphi] \quad \text{and} \quad Z_{i,j} = \varphi(X_i^1, X_j^2) - \mathbb{E}_{\perp\!\!\!\perp}[\varphi].$$

Then,

$$\mathbb{E}[Y_i] = \mathbb{E}[Z_{i,j}] = 0, \quad \text{and} \quad \begin{cases} \mathbb{E}[Y_i^2] = \mathrm{Var}_P(\varphi) \le \mathbb{E}_P[\varphi^2], \\ \mathbb{E}[Z_{i,j}^2] = \mathrm{Var}_{\perp\!\!\!\perp}(\varphi) \le \mathbb{E}_{\perp\!\!\!\perp}[\varphi^2]. \end{cases} \tag{17.43}$$

One can write

$$T(\mathbb{X}_n) - \mathbb{E}[T(\mathbb{X}_n)] = \frac{1}{n(n-1)} \sum_{i \ne j} (Y_i - Z_{i,j}),$$

and thus,

$$\mathrm{Var}(T(\mathbb{X}_n)) = \mathbb{E}\left[\left(\frac{1}{n(n-1)} \sum_{i \ne j} (Y_i - Z_{i,j})\right)^2\right]$$

$$= \frac{1}{n^2(n-1)^2} \sum_{i \ne j} \sum_{k \ne l} \mathbb{E}[(Y_i - Z_{i,j})(Y_k - Z_{k,l})]$$

$$= A_n - 2B_n + C_n,$$

with

$$A_n = \frac{1}{n^2} \sum_{i,k=1}^{n} \mathbb{E}[Y_i Y_k],$$

$$B_n = \frac{1}{n^2(n-1)} \sum_{i=1}^{n} \sum_{k \ne l} \mathbb{E}[Y_i Z_{k,l}],$$

$$C_n = \frac{1}{n^2(n-1)^2} \sum_{i \ne j} \sum_{k \ne l} \mathbb{E}[Z_{i,j} Z_{k,l}],$$

where each sum is taken for indexes contained in $\{1, 2, \ldots, n\}$. In particular, since just an upper-bound of the variance is needed, it is sufficient to write

$$\mathrm{Var}(T(\mathbb{X}_n)) \le |A_n| + 2|B_n| + |C_n|, \tag{17.44}$$

and to study each term separately.

Study of A_n Since by construction, the Y_i's are centered, and independent (as the X_i's are),

$$A_n = \frac{1}{n^2} \left(\sum_i \mathbb{E}\left[Y_i^2\right] + \sum_{i \neq k} \mathbb{E}\left[Y_i\right]\mathbb{E}\left[Y_k\right] \right)$$

$$= \frac{1}{n}\mathbb{E}\left[Y_1^2\right],$$

and in particular, from (17.43),

$$|A_n| \leq \frac{1}{n}\mathbb{E}_P\left[\varphi^2\right]. \tag{17.45}$$

Study of B_n If i, k and l are all different, using once again the independence of the X_i's and a centering argument, then $\mathbb{E}\left[Y_i Z_{k,l}\right] = \mathbb{E}\left[Y_i\right]\mathbb{E}\left[Z_{k,l}\right] = 0$. Thus

$$B_n = \frac{1}{n^2(n-1)} \sum_{i \neq k} \left(\mathbb{E}\left[Y_i Z_{i,k}\right] + \mathbb{E}\left[Y_i Z_{k,i}\right] \right)$$

$$= \frac{1}{n} \left(\mathbb{E}\left[Y_1 Z_{1,2}\right] + \mathbb{E}\left[Y_1 Z_{2,1}\right] \right).$$

In particular, applying the Cauchy-Schwartz inequality, and from (17.43), one obtains

$$|B_n| \leq \frac{2}{n}\sqrt{\mathbb{E}\left[Y_1^2\right]\mathbb{E}\left[Z_{1,2}^2\right]} \leq \frac{2}{n}\sqrt{\mathbb{E}_P\left[\varphi^2\right]\mathbb{E}_{\perp\!\!\!\perp}\left[\varphi^2\right]}. \tag{17.46}$$

Study of C_n Still by an independence and a centering argument, if i, j, k and l are all different, $\mathbb{E}\left[Z_{i,j}Z_{k,l}\right] = \mathbb{E}\left[Z_{i,j}\right]\mathbb{E}\left[Z_{k,l}\right] = 0$. Thus, if $I_n^{[3]}$ denotes the set of triplets (i, j, k) in $\{1, \ldots, n\}^3$ which are all different, one obtains

$$C_n = \frac{1}{n^2(n-1)^2} \left\{ \sum_{(i,j,k) \in I_n^{[3]}} \left(\mathbb{E}\left[Z_{i,j}Z_{i,k}\right] + 2\mathbb{E}\left[Z_{i,j}Z_{k,i}\right] + \mathbb{E}\left[Z_{j,i}Z_{k,i}\right] \right) \right.$$

$$\left. + \sum_{i \neq j} \left(\mathbb{E}\left[Z_{i,j}^2\right] + \mathbb{E}\left[Z_{i,j}Z_{j,i}\right] \right) \right\}$$

$$= \frac{n-2}{n(n-1)} \left(\mathbb{E}\left[Z_{1,2}Z_{1,3}\right] + 2\mathbb{E}\left[Z_{1,2}Z_{3,1}\right] + \mathbb{E}\left[Z_{2,1}Z_{3,1}\right] \right)$$

$$+ \frac{1}{n(n-1)} \left(\mathbb{E}\left[Z_{1,2}^2\right] + \mathbb{E}\left[Z_{1,2}Z_{2,1}\right] \right).$$

In particular, applying the Cauchy-Schwartz inequality, and using (17.43), each expectation in the previous equation satisfies $\mathbb{E}\left[Z_{i,j}Z_{k,l}\right] \leq \mathbb{E}\left[Z_{1,2}^2\right] \leq \mathbb{E}_{\perp\!\!\!\perp}\left[\varphi^2\right]$, and thus

$$|C_n| \leq \left(\frac{4(n-2)}{n(n-1)} + \frac{2}{n(n-1)}\right) \mathbb{E}_{\perp\!\!\!\perp}\left[\varphi^2\right] \leq \frac{4}{n}\mathbb{E}_{\perp\!\!\!\perp}\left[\varphi^2\right]. \tag{17.47}$$

Finally, combining (17.44), (17.45), (17.46), and (17.47) leads to

$$\mathrm{Var}(T(\mathbb{X}_n)) \leq \frac{1}{n}\left(\sqrt{\mathbb{E}_P\left[\varphi^2\right]} + 2\sqrt{\mathbb{E}_{\perp\!\!\!\perp}\left[\varphi^2\right]}\right)^2,$$

which ends the proof of the lemma.

Appendix: A Non-asymptotic Control of the Second Kind Error Rates

Consider the notation from Sect. 17.3. Since this section focuses on the study of the second kind error rate of the test, in all the sequel, the observation is assumed to satisfy the alternative ($\mathcal{H}_1$). Let thus P be an alternative, that is $P \neq (P^1 \otimes P^2)$, $n \geq 4$ and $\mathbb{X}_n = (X_i, \ldots, X_n)$ be an i.i.d. sample from distribution P. Fix α and β be two fixed values in $(0, 1)$. Consider T the test statistic introduced in (17.16), the (random) critical value $q_{1-\alpha}(\mathbb{X}_n)$ defined in (17.18), and the corresponding permutation test defined in (17.19) by

$$\Delta_\alpha(\mathbb{X}_n) = \mathbb{1}_{T(\mathbb{X}_n)>q_{1-\alpha}(\mathbb{X}_n)},$$

which precisely rejects independence when $T(\mathbb{X}_n) > q_{1-\alpha}(\mathbb{X}_n)$. Notice that this test is exactly the upper-tailed test by permutation introduced in [2].

The aim of this section is to provide different conditions on the alternative P ensuring a control of the second kind error rate by a fixed value $\beta > 0$, that is $\mathbb{P}(\Delta_\alpha(\mathbb{X}_n) = 0) \leq \beta$. The following steps constitute the first steps of a general study of the separation rates for the previous independence test, and is worked through in the specific case of continuous real-valued random variables in [1, Chapter 4].

Recall the notation introduced in (17.17) for a better readability. For all real-valued measurable function g on $\mathcal{X}^2$, denote respectively

$$\mathbb{E}_P[g] = \mathbb{E}\left[g(X_1^1, X_1^2)\right] \quad \text{and} \quad \mathbb{E}_{\perp\!\!\!\perp}[g] = \mathbb{E}\left[g(X_1^1, X_2^2)\right],$$

the expectations of $g(X)$ under the alternative P, that is if $X \sim P$, and under the null hypothesis $(\mathcal{H}_0)$, that is if $X \sim (P^1 \otimes P^2)$.

Assume the following moment assumption holds, that is

$$(\mathcal{A}_{Mmt,2}) \quad \Big| \quad \text{both } \mathbb{E}_P\Big[\varphi^2\Big] < +\infty \text{ and } \mathbb{E}_{\perp\!\!\perp}\Big[\varphi^2\Big] < +\infty,$$

so that all variance and second-order moments exist. Then, the following statements hold.

1. By Chebychev's inequality, one has $\mathbb{P}(\Delta_\alpha(\mathbb{X}_n) = 0) \leq \beta$ as soon as Condition (17.20) is satisfied, that is

$$\mathbb{E}\left[T(\mathbb{X}_n)\right] \geq q_{1-\beta/2}^\alpha + \sqrt{\frac{2}{\beta} \operatorname{Var}(T(\mathbb{X}_n))}.$$

2. On the one hand,

$$\operatorname{Var}(T(\mathbb{X}_n)) \leq \frac{8}{n} \left(\mathbb{E}_P\Big[\varphi^2\Big] + \mathbb{E}_{\perp\!\!\perp}\Big[\varphi^2\Big]\right), \tag{17.48}$$

3. On the other hand, in order to control the quantile $q_{1-\beta/2}^\alpha$, let us first upper bound the conditional quantile, following Hoeffding's approach based on the Cauchy-Schwarz inequality, by

$$q_{1-\alpha}(\mathbb{X}_n) \leq \sqrt{\frac{1-\alpha}{\alpha} \operatorname{Var}\big(T\big(\mathbb{X}_n^\Pi\big)\big|\mathbb{X}_n\big)}. \tag{17.49}$$

4. Markov's inequality allows us to deduce the following bound for the quantile:

$$q_{1-\beta/2}^\alpha \leq 2\sqrt{\frac{1-\alpha}{\alpha}} \sqrt{\frac{2}{\beta} \frac{\left(\mathbb{E}_{\perp\!\!\perp}\Big[\varphi^2\Big] + \mathbb{E}_P\Big[\varphi^2\Big]\right)}{n}}. \tag{17.50}$$

5. Finally, combining (17.20), (17.48) and (17.50) ensures that $\mathbb{P}(\Delta_\alpha(\mathbb{X}_n) = 0) \leq \beta$ as soon as Condition (17.21) is satisfied, that is

$$\mathbb{E}\left[T(\mathbb{X}_n)\right] \geq \frac{4}{\sqrt{\alpha}} \sqrt{\frac{2}{\beta} \frac{\mathbb{E}_P\Big[\varphi^2\Big] + \mathbb{E}_{\perp\!\!\perp}\Big[\varphi^2\Big]}{n}}.$$

This section is divided in five subsections, each one of them respectively proving a point stated above. The first one proves the sufficiency of Condition (17.20) in order to control the second kind error rate. The second, third and fourth ones provide respectively upper-bounds of the variance term, the critical value and the quantile $q_{1-\beta/2}^\alpha$. Finally, the fifth one provides the sufficiency of Condition (17.21).

A First Condition Ensuing from Chebychev's Inequality

In this section, we prove the sufficiency of a first simple condition, derived from Chebychev's inequality in order to control the second error rate. Assume that (17.20) is satisfied, that is

$$\mathbb{E}\left[T(\mathbb{X}_n)\right] \geq q_{1-\beta/2}^{\alpha} + \sqrt{\frac{2}{\beta}\text{Var}(T(\mathbb{X}_n))}.$$

Then,

$$
\begin{aligned}
\mathbb{P}(\Delta_\alpha(\mathbb{X}_n) = 0) &= \mathbb{P}(T(\mathbb{X}_n) \leq q_{1-\alpha}(\mathbb{X}_n)) &\text{(17.51)}\\
&= \mathbb{P}\Big(\{T(\mathbb{X}_n) \leq q_{1-\alpha}(\mathbb{X}_n)\} \cap \Big\{q_{1-\alpha}(\mathbb{X}_n) \leq q_{1-\beta/2}^{\alpha}\Big\}\Big)\\
&\quad + \mathbb{P}\Big(\{T(\mathbb{X}_n) \leq q_{1-\alpha}(\mathbb{X}_n)\} \cap \Big\{q_{1-\alpha}(\mathbb{X}_n) > q_{1-\beta/2}^{\alpha}\Big\}\Big)\\
&\leq \mathbb{P}\Big(T(\mathbb{X}_n) \leq q_{1-\beta/2}^{\alpha}\Big) + \mathbb{P}\Big(q_{1-\alpha}(\mathbb{X}_n) > q_{1-\beta/2}^{\alpha}\Big)\\
&\leq \mathbb{P}\Big(T(\mathbb{X}_n) \leq q_{1-\beta/2}^{\alpha}\Big) + \frac{\beta}{2}, &\text{(17.52)}
\end{aligned}
$$

by definition of the quantile $q_{1-\beta/2}^{\alpha}$. Yet, from (17.20) one obtains from Chebychev's inequality that

$$
\begin{aligned}
\mathbb{P}\Big(T(\mathbb{X}_n) \leq q_{1-\beta/2}^{\alpha}\Big) &\leq \mathbb{P}\left(T(\mathbb{X}_n) \leq \mathbb{E}\left[T(\mathbb{X}_n)\right] - \sqrt{\frac{2}{\beta}\text{Var}(T(\mathbb{X}_n))}\right)\\
&\leq \mathbb{P}\left(|T(\mathbb{X}_n) - \mathbb{E}\left[T(\mathbb{X}_n)\right]| \geq \sqrt{\frac{2}{\beta}\text{Var}(T(\mathbb{X}_n))}\right)\\
&\leq \frac{\beta}{2}. &\text{(17.53)}
\end{aligned}
$$

Finally, both (17.52) and (17.53) lead to the desired control $\mathbb{P}(\Delta_\alpha(\mathbb{X}_n) = 0) \leq \beta$ which ends the proof.

Control of the Variance in the General Case

To upper bound the variance term, we apply Lemma 17.3.1 which directly implies that

$$\text{Var}(T(\mathbb{X}_n)) \leq \frac{2}{n}\left(\mathbb{E}_P\left[\varphi^2\right] + 4\mathbb{E}_{\perp\!\!\!\perp}\left[\varphi^2\right]\right),$$

which directly leads to (17.48).

Control of the Critical Value Based on Hoeffding's Approach

This section is devoted to the proof the inequality (17.49), namely

$$q_{1-\alpha}(\mathbb{X}_n) \leq \sqrt{\frac{1-\alpha}{\alpha}} \operatorname{Var}\big(T\left(\mathbb{X}_n^\Pi\right)\big|\mathbb{X}_n\big).$$

The proof of this upper-bound follows Hoeffding's approach in [16], and relies on a normalizing trick, and the Cauchy-Schwarz inequality. From now on, for a better readability, denote respectively $\mathbb{E}^*[\cdot]$ and $\operatorname{Var}^*(\cdot)$ the conditional expectation and variance given the sample $\mathbb{X}_n$.

As in Hoeffding [16], the first step is to center and normalize the permuted test statistic. Yet, by construction the permuted test statistic is automatically centered, that is $\mathbb{E}^*[T\left(\mathbb{X}_n^\Pi\right)] = 0$, as one can notice that

$$T\left(\mathbb{X}_n^\Pi\right) = \frac{1}{n-1} \left(\sum_{i=1}^n \varphi\left(X_i^1, X_{\Pi(i)}^2\right) - \mathbb{E}^*\left[\sum_{i=1}^n \varphi\left(X_i^1, X_{\Pi(i)}^2\right) \right] \right).$$

Therefore, just consider the normalizing term

$$\nu(\mathbb{X}_n) = \operatorname{Var}^*(T\left(\mathbb{X}_n^\Pi\right)) = \mathbb{E}^*\left[T\left(\mathbb{X}_n^\Pi\right)^2 \right] = \frac{1}{n!} \sum_{\tau \in \mathfrak{S}_n} \left(T\left(\mathbb{X}_n^\tau\right) \right)^2.$$

Two cases appear: either $\nu(\mathbb{X}_n) = 0$ or not.

In the first case, the nullity of the conditional variance implies that all the permutations of the test statistic are equal. Hence, for all permutation τ of $\{1, \ldots, n\}$, one has $T(\mathbb{X}_n^\tau) = T(\mathbb{X}_n)$. Since the centering term $\mathbb{E}^*\left[\sum_{i=1}^n \varphi\left(X_i^1, X_{\Pi(i)}^2\right)\right] = n^{-1} \sum_{i,j=1}^n \varphi(X_i^1, X_j^2)$ is permutation invariant, one obtains the equality of the permuted sums, that is

$$\sum_{i=1}^n \varphi\left(X_i^1, X_{\tau(i)}^2\right) = \sum_{i=1}^n \varphi\left(X_i^1, X_i^2\right),$$

and this for all permutation τ. In particular, the centering term is also equal to $\sum_{i=1}^n \varphi\left(X_i^1, X_i^2\right)$. Indeed, by invariance of the sum (applied in the third equality below),

$$\frac{1}{n} \sum_{i,j=1}^n \varphi\left(X_i^1, X_j^2\right) = \frac{1}{n} \sum_{i,j=1}^n \varphi\left(X_i^1, X_j^2\right) \left[\frac{1}{(n-1)!} \sum_{\tau \in \mathfrak{S}_n} \mathbb{1}_{\tau(i)=j} \right]$$

$$= \frac{1}{n!} \sum_{\tau \in \mathfrak{S}_n} \sum_{i=1}^n \varphi\left(X_i^1, X_{\tau(i)}^2\right) \left[\sum_{j=1}^n \mathbb{1}_{\tau(i)=j} \right]$$

$$\frac{1}{n} \sum_{i,j=1}^{n} \varphi\left(X_i^1, X_j^2\right) = \frac{1}{n!} \sum_{\tau \in \mathfrak{S}_n} \left(\sum_{i=1}^{n} \varphi\left(X_i^1, X_i^2\right)\right)$$

$$= \sum_{i=1}^{n} \varphi\left(X_i^1, X_i^2\right).$$

Therefore, $T(\mathbb{X}_n)$ is equal to zero, and thus, so is $q_{1-\alpha}(\mathbb{X}_n)$. Finally, inequality (17.55) is satisfied since

$$q_{1-\alpha}(\mathbb{X}_n) = 0 \leq 0 = \sqrt{\frac{1-\alpha}{\alpha}} \, \mathrm{Var}\left(T\left(\mathbb{X}_n^{\Pi}\right) \middle| \mathbb{X}_n\right).$$

Consider now the second case, and assume $v(\mathbb{X}_n) > 0$. Let us introduce the (centered and) normalized statistic

$$T'(\mathbb{X}_n) = \frac{1}{\sqrt{v(\mathbb{X}_n)}} \left(T(\mathbb{X}_n)\right).$$

In particular, the new statistic $T'(\mathbb{X}_n)$ satisfies

$$\mathbb{E}^*\left[T'\left(\mathbb{X}_n^{\Pi}\right)\right] = 0 \quad \text{and} \quad \mathrm{Var}^*\left(T'\left(\mathbb{X}_n^{\Pi}\right)\right) \leq 1.$$

One may moreover notice that the normalizing term $v(\mathbb{X}_n)$ is permutation invariant, that is, for all permutations τ and τ' in $\mathfrak{S}_n$,

$$v\left(\mathbb{X}_n^{\tau}\right) = v(\mathbb{X}_n) = v\left(\mathbb{X}_n^{\tau'}\right).$$

In particular, since $v(\mathbb{X}_n) > 0$,

$$T\left(\mathbb{X}_n^{\tau}\right) \leq T\left(\mathbb{X}_n^{\tau'}\right) \quad \Leftrightarrow \quad T'\left(\mathbb{X}_n^{\tau}\right) \leq T'\left(\mathbb{X}_n^{\tau'}\right).$$

Therefore, as the test Δ_α depends only on the comparison of the $\left\{T\left(\mathbb{X}_n^{\tau}\right)\right\}_{\tau \in \mathfrak{S}_n}$, the test statistic T can be replaced by T', and the new critical value becomes

$$q'_{1-\alpha}(\mathbb{X}_n) = T'^{(n!-\lfloor n!\alpha \rfloor)}(\mathbb{X}_n) = \frac{T^{(n!-\lfloor n!\alpha \rfloor)}(\mathbb{X}_n)}{v(\mathbb{X}_n)} = \frac{q_{1-\alpha}(\mathbb{X}_n)}{v(\mathbb{X}_n)}. \qquad (17.54)$$

Moreover, following the proof of Theorem 2.1. of Hoeffding [16], one can show (as below) that

$$q'_{1-\alpha}(\mathbb{X}_n) \leq \sqrt{\frac{1-\alpha}{\alpha}}. \qquad (17.55)$$

Hence, combining (17.55) with (17.54) leads straightforwardly to (17.49).

Finally, remains the proof of (17.55). There are two cases:

First Case If $q'_{1-\alpha}(\mathbb{X}_n) \leq 0$, then (17.55) is satisfied.

Second Case If $q'_{1-\alpha}(\mathbb{X}_n) > 0$, then introduce $Y = q'_{1-\alpha}(\mathbb{X}_n) - T'\left(\mathbb{X}_n^{\Pi}\right)$.

First, since by construction, $\mathbb{E}^*[T'(\mathbb{X}_n^{\Pi})] = 0$, one directly obtains $\mathbb{E}^*[Y] = q'_{1-\alpha}(\mathbb{X}_n)$. Hence,

$$0 < q'_{1-\alpha}(\mathbb{X}_n) = \mathbb{E}^*[Y] \leq \mathbb{E}^*[Y \mathbb{1}_{Y>0}],$$

and by the Cauchy-Schwarz inequality,

$$\left(q'_{1-\alpha}(\mathbb{X}_n)\right)^2 \quad \leq \quad \left(\mathbb{E}^*[Y \mathbb{1}_{Y>0}]\right)^2 \quad \leq \quad \mathbb{E}^*\left[Y^2\right] \mathbb{E}^*[\mathbb{1}_{Y>0}],$$

Yet, on one hand,

$$
\begin{aligned}
\mathbb{E}^*\left[Y^2\right] &= \mathbb{E}^*\left[\left(q'_{1-\alpha}(\mathbb{X}_n) - T'\left(\mathbb{X}_n^{\Pi}\right)\right)^2\right] \\
&= \left(q'_{1-\alpha}(\mathbb{X}_n)\right)^2 + \mathbb{E}^*\left[\left(T'\left(\mathbb{X}_n^{\Pi}\right)\right)^2\right] - 2q'_{1-\alpha}(\mathbb{X}_n)\mathbb{E}^*[T'\left(\mathbb{X}_n^{\Pi}\right)] \\
&= \left(q'_{1-\alpha}(\mathbb{X}_n)\right)^2 + \mathrm{Var}^*(T'\left(\mathbb{X}_n^{\Pi}\right)) \\
&\leq \left(q'_{1-\alpha}(\mathbb{X}_n)\right)^2 + 1,
\end{aligned}
$$

since by the normalizing initial step, $\mathrm{Var}^*(T'\left(\mathbb{X}_n^{\Pi}\right)) \leq 1$.

And, on the other hand,

$$
\begin{aligned}
\mathbb{E}^*[\mathbb{1}_{Y>0}] &= \mathbb{E}^*\left[\mathbb{1}_{T'(')(\mathbb{X}_n^{\Pi}) < q'_{1-\alpha}(\mathbb{X}_n)}\right] \\
&= \frac{\#\left\{\tau \in \mathfrak{S}_n \; ; \; T'(\mathbb{X}_n^{\tau}) < T'^{(n! - \lfloor n!\alpha \rfloor)}(\mathbb{X}_n)\right\}}{n!} \\
&\leq \frac{(n! - \lfloor n!\alpha \rfloor) - 1}{n!} = 1 - \frac{\lfloor n!\alpha \rfloor + 1}{n!} \\
&< 1 - \frac{n!\alpha}{n!} = 1 - \alpha.
\end{aligned}
$$

So finally,

$$\left(q'_{1-\alpha}(\mathbb{X}_n)\right)^2 \leq (1-\alpha)\left(\left(q'_{1-\alpha}(\mathbb{X}_n)\right)^2 + 1\right),$$

which is equivalent to $\left(q'_{1-\alpha}(\mathbb{X}_n)\right)^2 \leq (1-\alpha)/\alpha$, and thus ends the proof of (17.55).

Control of the Quantile of the Critical Value

The control of the conditional quantile allows us to upper bound its own quantile $q_{1-\beta/2}^{\alpha}$ as stated in (17.50), that is

$$q_{1-\beta/2}^{\alpha} \leq 2\sqrt{\frac{1-\alpha}{\alpha}}\sqrt{\frac{2}{\beta}\frac{\left(\mathbb{E}_{\perp\!\!\!\perp}\left[\varphi^2\right]+\mathbb{E}_P\left[\varphi^2\right]\right)}{n}}.$$

Indeed, (17.49) ensures that

$$q_{1-\alpha}(\mathbb{X}_n) \leq \sqrt{\frac{1-\alpha}{\alpha}}\sqrt{\mathbb{E}\left[T\left(\mathbb{X}_n^{\Pi}\right)^2\Big|\mathbb{X}_n\right]},$$

and in particular, the $(1-\beta/2)$-quantile of $q_{1-\alpha}(\mathbb{X}_n)$ satisfies

$$q_{1-\beta/2}^{\alpha} \leq \sqrt{\frac{1-\alpha}{\alpha}}\sqrt{\zeta_{1-\beta/2}}, \tag{17.56}$$

where $\zeta_{1-\beta/2}$ is the $(1-\beta/2)$-quantile of $\mathbb{E}\left[T\left(\mathbb{X}_n^{\Pi}\right)^2\Big|\mathbb{X}_n\right]$. Yet, from Markov's inequality, for all positive x,

$$\mathbb{P}\left(\mathbb{E}\left[T\left(\mathbb{X}_n^{\Pi}\right)^2\Big|\mathbb{X}_n\right] \geq x\right) \leq \frac{\mathbb{E}\left[T\left(\mathbb{X}_n^{\Pi}\right)^2\right]}{x}.$$

In particular, the choice of $x = 2\mathbb{E}\left[T\left(\mathbb{X}_n^{\Pi}\right)^2\right]/\beta$ leads to the control of the quantile

$$\zeta_{1-\beta/2} \leq \frac{2\mathbb{E}\left[T\left(\mathbb{X}_n^{\Pi}\right)^2\right]}{\beta}. \tag{17.57}$$

Moreover, noticing that one can write

$$T\left(\mathbb{X}_n^{\Pi}\right) = \frac{1}{n-1}\sum_{i,j=1}^{n}\left(\mathbb{1}_{\Pi(i)=j}-\frac{1}{n}\right)\varphi(X_i^1, X_j^2),$$

the second-order moment in (17.57) can be rewritten

$$\mathbb{E}\left[T\left(\mathbb{X}_n^{\Pi}\right)^2\right] = \frac{1}{(n-1)^2}\mathbb{E}\left[\left(\sum_{i,j=1}^{n}\left(\mathbb{1}_{\Pi(i)=j}-\frac{1}{n}\right)\varphi(X_i^1, X_j^2)\right)^2\right]$$

$$= \frac{1}{(n-1)^2}\sum_{i,j=1}^{n}\sum_{k,l=1}^{n}E_{i,j,k,l}\times\mathbb{E}\left[\varphi(X_i^1, X_j^2)\varphi(X_k^1, X_l^2)\right],$$

by independence between Π and $\mathbb{X}_n$, where

$$E_{i,j,k,l} = \mathbb{E}\left[\left(\mathbb{1}_{\Pi(i)=j} - \frac{1}{n}\right)\left(\mathbb{1}_{\Pi(k)=l} - \frac{1}{n}\right)\right] = \mathbb{E}\left[\mathbb{1}_{\Pi(i)=j}\mathbb{1}_{\Pi(k)=l}\right] - \frac{1}{n^2}.$$

On the one hand, for all $1 \le i, j, k, l \le n$, the Cauchy-Schwarz inequality always ensures

$$\mathbb{E}\left[\varphi(X_i^1, X_j^2)\varphi(X_k^1, X_l^2)\right] \le \sqrt{\mathbb{E}\left[\varphi^2(X_i^1, X_j^2)\right]\mathbb{E}\left[\varphi^2(X_k^1, X_l^2)\right]} \le \mathbb{E}_{\perp\!\!\!\perp}\left[\varphi^2\right] + \mathbb{E}_P\left[\varphi^2\right], \tag{17.58}$$

since for all $1 \le i, j \le n$, $\mathbb{E}\left[\varphi^2(X_i^1, X_j^2)\right] \le \mathbb{E}_{\perp\!\!\!\perp}\left[\varphi^2\right] + \mathbb{E}_P\left[\varphi^2\right]$.

On the other hand, remains to control the sum $(n-1)^{-2}\sum_{i,j=1}^{n}\sum_{k,l=1}^{n}E_{i,j,k,l}$. Three cases appear.

First Case If $i \ne k$ and $j \ne l$ (occurring $[n(n-1)]^2$ times), then

$$E_{i,j,k,l} = \frac{1}{n(n-1)} - \frac{1}{n^2} = \frac{1}{n^2(n-1)}.$$

Second Case If $[i \ne k$ and $j = l]$ or $[i = k$ and $j \ne l]$, then

$$E_{i,j,k,l} = 0 - 1/n^2 \le 0.$$

Third Case If $i = k$ and $j = l$ (occurring $n(n-1)$ times), then

$$E_{i,j,k,l} = \frac{1}{n} - \frac{1}{n^2} = \frac{n-1}{n^2} \le \frac{1}{n}.$$

Therefore,

$$\frac{1}{(n-1)^2}\sum_{i,j=1}^{n}\sum_{k,l=1}^{n}E_{i,j,k,l} \le \frac{1}{(n-1)^2}\left([n(n-1)]^2 \times \frac{1}{n^2(n-1)} + n(n-1) \times \frac{1}{n}\right)$$

$$\le \frac{2}{n-1}$$

$$\le \frac{4}{n}. \tag{17.59}$$

Finally, both (17.58) and (17.59) imply that

$$\mathbb{E}\left[T\left(\mathbb{X}_n^{\Pi}\right)^2\right] \le \frac{4}{n}\left(\mathbb{E}_{\perp\!\!\!\perp}\left[\varphi^2\right] + \mathbb{E}_P\left[\varphi^2\right]\right), \tag{17.60}$$

Therefore, combining (17.56), (17.57) and (17.60) ends the proof of (17.50).

A First Condition Ensuing from Hoeffding's Approach

Back to the condition (17.20) derived from Chebychev's inequality, both (17.48) and (17.50) imply that

$$
q_{1-\beta/2}^{\alpha} + \sqrt{\frac{2}{\beta} \operatorname{Var}(T(\mathbb{X}_n))} \leq \sqrt{\frac{2}{\beta} \frac{\left(\mathbb{E}_P\left[\varphi^2\right] + \mathbb{E}_{\perp\!\!\!\perp}\left[\varphi^2\right]\right)}{n}} \left(2\sqrt{\frac{1-\alpha}{\alpha}} + \sqrt{8}\right),
$$

with $2\sqrt{(1-\alpha)/\alpha} + \sqrt{8} \leq 4/\sqrt{\alpha}$, since $\sqrt{1-\alpha} + \sqrt{\alpha} \leq \sqrt{2}$. Finally, the right-hand side of condition (17.20) being upper bounded by

$$
\frac{4}{\sqrt{\alpha}} \sqrt{\frac{2}{\beta} \frac{\left(\mathbb{E}_P\left[\varphi^2\right] + \mathbb{E}_{\perp\!\!\!\perp}\left[\varphi^2\right]\right)}{n}},
$$

which is exactly the right-hand side of (17.21), this ensures the sufficiency of condition 17.21 to control the second kind error rate by β.

References

1. M. Albert, Tests of independence by bootstrap and permutation: an asymptotic and non-asymptotic study. Application to Neurosciences. PhD thesis, Nice, 2015
2. M. Albert, Y. Bouret, M. Fromont, P. Reynaud-Bouret, Bootstrap and permutation tests of independence for point processes. Ann. Stat. **43**(6), 2537–2564 (2015)
3. M. Albert, Y. Bouret, M. Fromont, P. Reynaud-Bouret, Surrogate data methods based on a shuffling of the trials for synchrony detection: the centering issue. Neural Comput. **28**(11), 2352–2392 (2016)
4. R. Adamczak, D. Chafaï, P. Wolff, Circular law for random matrices with exchangeable entries (2014), arXiv preprint arXiv:1402.3660
5. Y. Baraud, Non-asymptotic minimax rates of testing in signal detection. Bernoulli **8**(5), 577–606 (2002)
6. B. Bercu, B. Delyon, E. Rio, *Concentration Inequalities for Sums and Martingales* (Springer, Cham, 2015)
7. E. Bolthausen, An estimate of the remainder in a combinatorial central limit theorem. Probab. Theory Relat. Fields **66**(3), 379–386 (1984)
8. S. Boucheron, G. Lugosi, P. Massart, *Concentration Inequalities: A Nonasymptotic Theory of Independence* (Oxford University Press, Oxford, 2013)
9. S. Chatterjee, Stein's method for concentration inequalities. Probab. Theory Relat. Fields **138**(1), 305–321 (2007)
10. M. Dwass, On the asymptotic normality of some statistics used in non-parametric tests. Ann. Math. Stat. **26**, 334–339 (1955)
11. M. Fromont, B. Laurent, P. Reynaud-Bouret, Adaptive tests of homogeneity for a Poisson process. Ann. Inst. Henri Poincaré Probab. Stat. **47**(1), 176–213 (2011)
12. M. Fromont, B. Laurent, P. Reynaud-Bouret, The two-sample problem for poisson processes: adaptive tests with a nonasymptotic wild bootstrap approach. Ann. Stat. **41**(3), 1431–1461 (2013)

13. J. Hájek, Some extensions of the Wald-Wolfowitz-Noether theorem. Ann. Math. Stat. **32**(2), 506–523 (1961)
14. S.T. Ho, L.H.Y. Chen, An L_p bound for the remainder in a combinatorial central limit theorem. Ann. Probab. **6**(2), 231–249 (1978)
15. W. Hoeffding, A combinatorial central limit theorem. Ann. Math. Stat. **22**(4), 558–566 (1951)
16. W. Hoeffding, The large-sample power of tests based on permutation of the observations. Ann. Math. Stat. **23**(2), 169–192 (1952)
17. K. Jogdeo, Asymptotic normality in nonparametric methods. Ann. Math. Stat. **39**, 905–922 (1968)
18. M. Ledoux, *The Concentration of Measure Phenomenon*, vol. 89 (American Mathematical Society, Providence, 2005)
19. P. Massart, J. Picard, *Concentration Inequalities and Model Selection*, vol. 1896 (Springer, Berlin, 2007)
20. C. McDiarmid, Concentration for independent permutations. Comb. Probab. Comput. **11**(02), 163–178 (2002)
21. M. Motoo, On the Hoeffding's combinatrial central limit theorem. Ann. Inst. Stat. Math. **8**(1), 145–154 (1956)
22. G.E. Noether, On a theorem by Wald and Wolfowitz. Ann. Math. Stat. **20**, 455–458 (1949)
23. J.P. Romano, Bootstrap and randomization tests of some nonparametric hypotheses. Ann. Stat. **17**(1), 141–159 (1989)
24. L. Sansonnet, C. Tuleau-Malot, A model of Poissonian interactions and detection of dependence. Stat. Comput. **25**(2), 449–470 (2015)
25. W. Schneller, A short proof of Motoo's combinatorial central limit theorem using stein's method. Probab. Theory Relat. Fields **78**(2), 249–252 (1988)
26. C.P. Shapiro, L. Hubert, Asymptotic normality of permutation statistics derived from weighted sums of bivariate functions. Ann. Stat. **7**, 788–794 (1979)
27. M. Talagrand, Concentration of measure and isoperimetric inequalities in product spaces. Publ. Math. Inst. Hautes Etudes Sci. **81**(1), 73–205 (1995)
28. A. Wald, J. Wolfowitz, Statistical tests based on permutations of the observations. Ann. Math. Stat. **15**(4), 358–372 (1944)