Bernoulli **22**(1), 2016, 302–324
DOI: 10.3150/14-BEJ659

On empirical distribution function of high-dimensional Gaussian vector components with an application to multiple testing

SYLVAIN DELATTRE[1] and ETIENNE ROQUAIN[2]

[1]*LPMA, UMR 7599, Univ. Paris Diderot, Sorbonne Paris Cité, F-75205, Paris, France.*
E-mail: sylvain.delattre@univ-paris-diderot.fr
[2]*LPMA, UMR 7599, Sorbonne Universités, UPMC, Univ. Paris 6, F-75005, Paris, France.*
E-mail: etienne.roquain@upmc.fr

This paper presents a study of the asymptotical behavior of the empirical distribution function (e.d.f.) of Gaussian vector components, whose correlation matrix $\Gamma^{(m)}$ is dimension-dependent. By contrast with the existing literature, the vector is not assumed to be stationary. Rather, we make a "vanishing second order" assumption ensuring the covariance matrix $\Gamma^{(m)}$ is not too far from the identity matrix, while the behavior of the e.d.f. is affected by $\Gamma^{(m)}$ only through the sequence $\gamma_m = m^{-2}\sum_{i\neq j}\Gamma^{(m)}_{i,j}$, as m grows to infinity. This result recovers some of the previous results for stationary long-range dependencies while it also applies to various, high-dimensional, non-stationary frameworks, for which the most correlated variables are not necessarily close to each other. Finally, we present an application of this work to the multiple testing problem, which was the initial statistical motivation for developing such a methodology.

Keywords: empirical distribution function; factor model; false discovery rate; functional central limit theorem; functional delta method; Gaussian triangular arrays; Hermite polynomials; sample correlation matrix

1. Introduction

1.1. Motivation and background

Pertaining to the florishing field of statistics for high-dimensional data, the Benjamini–Hochberg (BH) procedure has become a well accepted and commonly used method when testing a large number of null hypotheses simultaneously. Its quality is measured via the false discovery proportion (FDP), the proportion of errors among the rejected null hypotheses, whose expectation is the celebrated false discovery rate (FDR), see Benjamini and Hochberg [4]. The methodology of Neuvial [26] shows that the FDP of BH procedure is an (Hadamard differentiable) functional of empirical cumulative distribution functions (e.d.f. in short). Via the functional delta method (see, e.g., van der Vaart [33]), this rises the problem of obtaining functional central limit theorems for e.d.f. in a setting which is suitable for high-dimensional data.

1350-7265

A colossal number of work aimed at extending Donsker's theorem (Doob [12], Donsker [11] and Dudley [15]) to a more relaxed setup. Among them, a particularly prospering research field deals with the introduction of weak dependence between the original variables, mainly by using mixing conditions. Here, we do not attempt to provide an exhaustive list for such results and we refer the reader to, for example, Arcones [1], Dedecker and Prieur [7], [14] for reviews. When restricted to the Gaussian subordinated setting, asymptotics for the e.d.f. are described in the two well-known papers of Dehling and Taqqu [8] (long-range) and Csörgő and Mielniczuk [6] (short-range). Both studies make a *stationarity* assumption: the covariance matrix between the variables is assumed to be of the form

$$\Gamma^{(m)}_{i,j} = r\big(|i-j|\big), \qquad 1 \le i, j \le m,$$

for some function $r(\cdot)$ vanishing at infinity and not depending on m.

However, in high-dimensional data, while the dimension m can be very large (typically, several thousands), the matrix $\Gamma^{(m)}$ is generally complex and not-necessarily locally structured. This is typically the case when latent variables (factors) have a simultaneous impact on all the variables (see, e.g., Friguet, Kloareg and Causeur [20], Sun, Zhang and Owen [32] and Fan, Han and Gu [17] and references therein), which leads to "spiked" correlation matrices (as referred to by Johnstone [24]). In a more general view, the larger the dimension, the more stringent the stationary assumption.

1.2. Presentation of the main result

Let us consider $\{Y^{(m)}, m \ge 1\}$ a triangular array for which each vector $Y^{(m)} = (Y^{(m)}_1, \dots, Y^{(m)}_m)$ is a m-dimensional Gaussian vector, defined on some probability space $(\Omega_m, \mathcal{F}_m, \mathbb{P}_m)$, with zero mean and covariance matrix $\Gamma^{(m)}$. For the sake of simplicity, assume that each $Y^{(m)}_i$ is of variance 1, that is, $\Gamma^{(m)}_{i,i} = 1$ for all i. Denote $\bar\Phi(z) = \mathbb{P}(Z \ge z)$, for $z \in \mathbb{R}$, $Z \sim \mathcal{N}(0,1)$, the upper tail distribution function of a standard Gaussian variable, and consider the empirical cumulative distribution function:

$$\widehat{\mathbb{F}}_m(t) = m^{-1} \sum_{i=1}^{m} \mathbf{1}\big\{\bar\Phi\big(Y^{(m)}_i\big) \le t\big\}, \qquad t \in [0,1]. \tag{1}$$

Here, we consider the e.d.f. of the $\bar\Phi(Y^{(m)}_i)$'s rather than the one of the $Y^{(m)}_i$'s to get uniformly distributed variables. The variables can therefore be interpreted as p-values, which is convenient for multiple testing, see Section 4. To study (1), let us introduce the following quantities:

$$\gamma_m = m^{-2} \sum_{i \ne j} \Gamma^{(m)}_{i,j}; \tag{2}$$

$$r_m = \big(m^{-1} + |\gamma_m|\big)^{-1/2}. \tag{3}$$

In a nutshell, our main result is as follows: by assuming, when $m \to \infty$,

$$\frac{r_m^2}{m^2} \sum_{i \neq j} \left(\Gamma_{i,j}^{(m)}\right)^2 \to 0; \qquad \text{(vanish-secondorder)}$$

$$\frac{r_m^{4+\varepsilon_0}}{m^2} \sum_{i \neq j} \left(\Gamma_{i,j}^{(m)}\right)^4 \to 0, \qquad \text{for some } \varepsilon_0 > 0; \qquad (H_1)$$

$$m\gamma_m \to \theta, \qquad \text{for some } \theta \in [-1, +\infty]; \qquad (H_2)$$

the following weak convergence holds (in the Skorokhod topology):

$$r_m(\widehat{\mathbb{F}}_m - I) \rightsquigarrow \mathbb{Z}, \qquad \text{as } m \to \infty, \tag{4}$$

where $I(t) = t$ and $\mathbb{Z}$ is some continuous Gaussian process on $[0, 1]$ with a distribution only function of θ. Specifically, denoting ϕ the standard Gaussian density:

(i) if $m\gamma_m \to \theta < +\infty$, we have $r_m \propto m^{1/2}$ and the process $m^{1/2}(\widehat{\mathbb{F}}_m - I)$ converges to a (continuous Gaussian) process with covariance function given by $(t, s) \mapsto t \wedge s - ts + \theta\phi(\bar{\Phi}^{-1}(t))\phi(\bar{\Phi}^{-1}(s))$. Hence, the limit process is a standard Brownian bridge when $\theta = 0$, but has a covariance function smaller (resp. larger) if $\theta < 0$ (resp. $\theta > 0$).

(ii) if $m\gamma_m \to \theta = +\infty$, we have $r_m \sim (\gamma_m)^{-1/2} \ll m^{1/2}$ and $(\gamma_m)^{-1/2}(\widehat{\mathbb{F}}_m - I)$ converge to the process $\phi(\bar{\Phi}^{-1}(\cdot))Z$ for $Z \sim \mathcal{N}(0, 1)$. Hence, the "Brownian" part asymptotically disappears.

The regimes (i) and (ii) are illustrated in Figure 1: as $m\gamma_m$ grows, the influence of the "Brownian" part decreases while that of the (randomly rescaled) function $\phi(\bar{\Phi}^{-1}(\cdot))$ increases. Also, the scale of the Y-axis indicates that the $m^{1/2}$ is not a suitable rate for large values of $m\gamma_m$.

Let us briefly discuss our novel conditions. Condition (vanish-secondorder) is the starting point of our study: it corresponds to assume that the expansion of the covariance function of $r_m(\widehat{\mathbb{F}}_m - I)$ asymptotically stops at order 1. This is a crucial L^2-type tool to elaborate our proofs in a possibly non-stationary regime. However, the price to pay is that it does not cover regimes where (some of) the greater orders matter asymptotically, as in the case of short range dependence (tridiagonal 1/2–1–1/2 for instance). As for condition (H_1), it is only used to prove that $r_m(\widehat{\mathbb{F}}_m - I)$ is C-tight and we suspect it to be unnecessary, although we did not manage to remove it formally from our assumption set. Condition (H_2) is not restrictive because it holds up to consider a subsequence.

Finally, in the regime (ii), we show that the convergence (4) is maintained when replacing the set of assumptions (vanish-secondorder), (H_1) and (H_2) by the two following conditions:

$$\frac{r_m^{2+\varepsilon_0}}{m^2} \sum_{i \neq j} \left(\Gamma_{i,j}^{(m)}\right)^2 = \mathrm{o}(1), \qquad \text{with } \varepsilon_0 > 0; \qquad (H_3)$$

$$m\gamma_m^{1+\varepsilon_0} \to \theta = +\infty, \qquad \text{with } \varepsilon_0 > 0. \qquad (H_4)$$

Roughly speaking, it shows that, up to add some "safety margin" ε_0 in the convergences, assumption (H_1) can be removed in regime (ii).

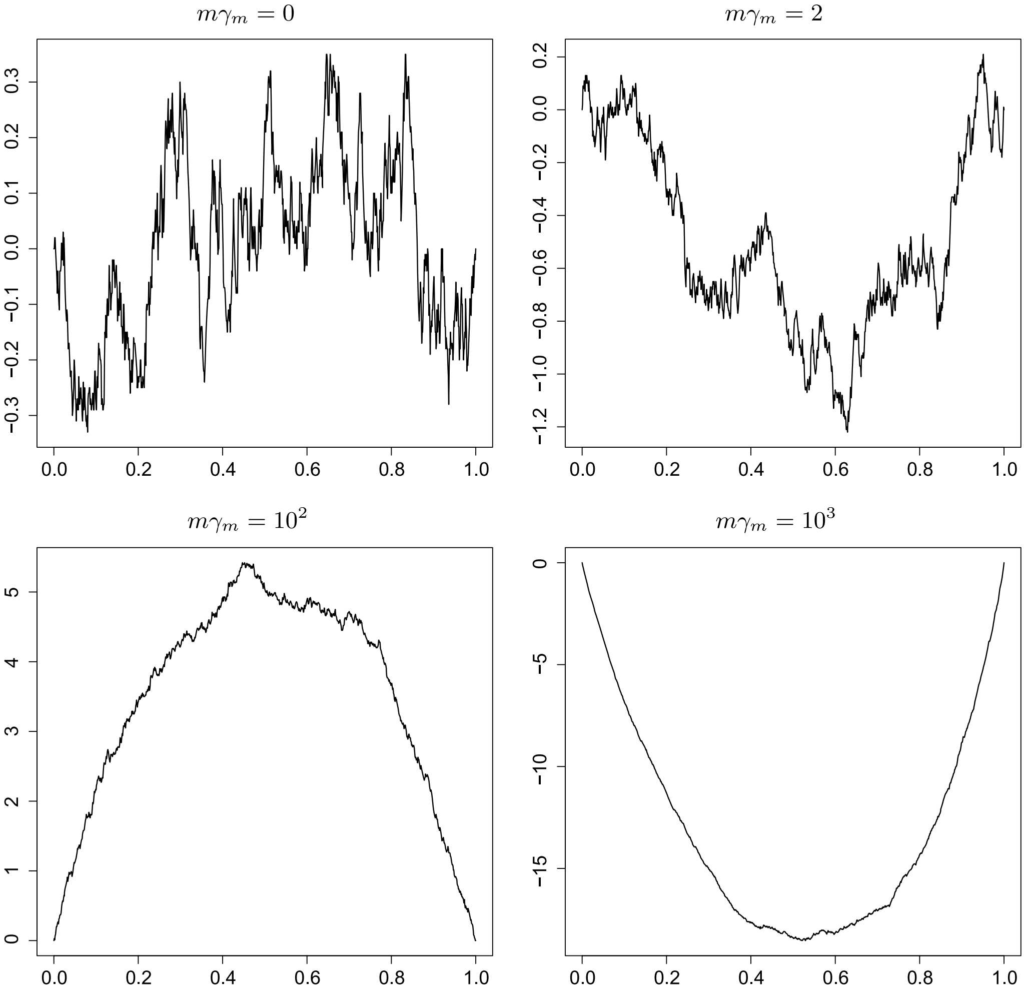

Figure 1. Plot of $t \mapsto m^{1/2}(\widehat{\mathbb{F}}_m(t) - t)$ for some observed $Y(\omega)$. These realizations have been generated in the equi-correlated model $\Gamma_{i,j}^{(m)} = \rho_m$, $i \neq j$ (see (14)) and for $m = 10^4$.

1.3. Relation to existing literature

Compared to previous studies using the stationary paradigm, our assumptions are markedly different: first, the covariance matrix $\Gamma^{(m)}$ is allowed to depend on m, that is, the $Y^{(m)}$'s form a triangular array of Gaussian variables. Second, $\Gamma^{(m)}$ needs not be locally structured, that is, $\Gamma_{i,j}^{(m)}$ is not necessarily related to the distance between i and j. Instead, our conditions are permutation invariant, that is, are unchanged when permuting the columns of the triangular array. This is quite natural because the e.d.f. is itself permutation invariant. Third, our approach shows that the negative correlations can decrease the asymptotic covariance or even increase the convergence rate.

As a counterpart, when restricted to the stationary setting, our assumptions are admittedly not optimal: it includes long-range of Dehling and Taqqu [8] (for an Hermite rank equal to 1), but excludes short range of Csörgő and Mielniczuk [6]. As explained above, this restriction comes from (vanish-secondorder), which implicitly truncates the covariance expansion in the limit. Nevertheless, our result opens a window for other dependence models as factor models or sample correlation matrices for instance.

Next, let us mention the interesting studies of Soulier [30] and Bardet and Surgailis [3] in which the stationarity assumption has also been removed. In the first study, Soulier [30] proves a central limit theorems (CLT) for variables of the form $T(Y_i^{(m)})$, by assuming that $\max_{i\neq j}\{\Gamma_{i,j}^{(m)}\}$ becomes sufficiently small as m grows. On the one hand, this work deals with a function T with arbitrary Hermite rank, while our work is restricted to the function $T(y)=\mathbf{1}\{y\le t\}-t$ (of Hermite rank 1). On the other hand, we obtain a convergence which is functional w.r.t. parameter t and our conditions on the $\Gamma_{i,j}^{(m)}$ involve *averages* (and not suprema). In the second work, Bardet and Surgailis [3] establishes CLT's for Gaussian subordinated arrays. There are two major differences with our approach: first, they deal with a CLT for the partial-sum process and not with a functional CLT for the e.d.f. Second, their assumptions are not of the same nature, because they require that $|\Gamma_{i,j}^{(m)}|\le r(|i-j|)$ for all i,j, for some function $r(\cdot)$, independent of m, and vanishing at infinity. So the latter still relies on a stationary structure.

Finally, Delattre and Roquain [9] studied in a previous work the (non-stationary) equi-correlated case where $\Gamma_{i,j}^{(m)}=\rho_m$, $i\neq j$, for some correlation ρ_m tending to zero (at some arbitrary rate). In a nutshell, our approach generalizes this result, by finding general conditions on $\Gamma^{(m)}$ such that, in our general $\Gamma^{(m)}$-dependent framework, the asymptotical distribution of the e.d.f. is the same as for the ρ_m-equi-correlated with $\rho_m=(m(m-1))^{-1}\sum_{i\neq j}\Gamma_{i,j}^{(m)}$.

1.4. Organization of the paper

In Section 2, we study the covariance function of $\widehat{\mathbb{F}}_m$ under (vanish-secondorder). The main theorem is formally stated in Section 3 together with many illustrative examples. This new methodology is then applied to the multiple testing problem in Section 4. The proof of the main result is presented in Section 5; it mainly relies on central limit theorems for martingale arrays and on a suitable tightness criterion. To make the proof as clear as possible, some technical and auxiliary results are deferred to a supplementary file, see Supplementary Material in [10], whose sections are denoted by adding a “S-” in the reference number (writing, e.g., Section S-1).

2. Covariance of $\widehat{\mathbb{F}}_m$ under (vanish-secondorder)

Throughout the paper, to alleviate the notation, we will often denote $\mathbb{P}_m$ by $\mathbb{P}$, $Y^{(m)}$ by Y and $\Gamma^{(m)}$ by Γ when not ambiguous.

Let us consider the sequence of Hermite polynomials $H_\ell(x)$, $\ell\ge 0$, $x\in\mathbb{R}$ (see Appendix S-4).

By using Melher's formula, the covariance function of the process $\widehat{\mathbb{F}}_m(\cdot)$ can be described as a function of the correlation matrix Γ of Y.

Proposition 2.1. *Consider $\widehat{\mathbb{F}}_m(\cdot)$ the process defined by (1) and the function family $\{c_\ell(\cdot), \ell \geq 1\}$ defined by*

$$c_\ell(t) = H_{\ell-1}\big(\bar{\Phi}^{-1}(t)\big)\phi\big(\bar{\Phi}^{-1}(t)\big), \qquad t \in [0,1], \ell = 1, 2, \dots. \tag{5}$$

Then for all $t, s \in [0,1]$, we have

$$\mathrm{Cov}\big(\widehat{\mathbb{F}}_m(t), \widehat{\mathbb{F}}_m(s)\big) = \sum_{\ell \geq 1} \frac{c_\ell(t)c_\ell(s)}{\ell!}\bigg(m^{-2}\sum_{i,j}(\Gamma_{i,j})^\ell\bigg). \tag{6}$$

This result is well known and can be found, for example, in Theorem 2 of Schwartzman and Lin [29], see also Lemma 10.1, chapter IV in Rozanov [28] (we provide a proof in Appendix S-4 for completeness). While (6) is an exact expression, we can try to approximate the covariance $\mathrm{Cov}(\widehat{\mathbb{F}}_m(t), \widehat{\mathbb{F}}_m(s))$ when m grows to infinity, while making some assumption on the matrix $\Gamma = \Gamma^{(m)}$.

First, let us note the following: since $m^{-2}\sum_{i,j}(\Gamma_{i,j})^\ell = (\ell!)^{-1}\,\mathrm{Var}(m^{-1}\sum_{i=1}^m H_\ell(Y_i)) \geq 0$ (by using (S-31) in Appendix S-4), expression (6) shows that the following conditions are equivalent as m tends to infinity,

$$\forall t \in [0,1], \qquad \mathrm{Var}\big(\widehat{\mathbb{F}}_m(t)\big) = \mathrm{o}(1), \tag{7}$$

$$\forall \ell \geq 1, \qquad m^{-2}\sum_{i,j}(\Gamma_{i,j})^\ell = \mathrm{o}(1), \tag{8}$$

$$m^{-2}\sum_{i,j}(\Gamma_{i,j})^2 = \mathrm{o}(1). \tag{LLN-dep}$$

Note that in (8), the case $\ell = 1$ follows from $\ell = 2$ by Cauchy–Schwarz's inequality. A consequence is that condition (LLN-dep) is required in order to have $\forall t \in [0,1], \widehat{\mathbb{F}}_m(t) \xrightarrow{P} t$. Moreover, the rate r_m defined by (3) satisfies $1 \leq r_m \leq \sqrt{m}$ and $(m^{-2}\sum_{i,j}(\Gamma_{i,j})^2)^{-1/4} \leq r_m$, the latter coming from $m^{-1} + |m^{-2}\sum_{i \neq j}\Gamma_{i,j}| \leq m^{-2}\sum_{i,j}|\Gamma_{i,j}| \leq (m^{-2}\sum_{i,j}(\Gamma_{i,j})^2)^{1/2}$. Hence, r_m tends to infinity under (LLN-dep) but not faster than $\sqrt{m}$.

Second, let us rewrite (6) as follows:

$$\begin{aligned}\mathrm{Cov}\big(\widehat{\mathbb{F}}_m(t), \widehat{\mathbb{F}}_m(s)\big) &= m^{-1}(t \wedge s - ts) + \gamma_m c_1(t)c_1(s) \\ &\quad + \sum_{\ell \geq 2}\bigg(m^{-2}\sum_{i \neq j}(\Gamma_{i,j})^\ell\bigg)c_\ell(t)c_\ell(s)(\ell!)^{-1},\end{aligned} \tag{9}$$

where γ_m is defined by (2). The latter holds because, for two independent $\mathcal{N}(0,1)$ variables U and V, we have $m^{-1}\sum_{\ell \geq 1} c_\ell(t)c_\ell(s)(\ell!)^{-1} = \mathrm{Cov}(\mathbf{1}\{\bar{\Phi}(U) \leq t\}, \mathbf{1}\{\bar{\Phi}(V) \leq s\})$. In expansion (9), the second order term (i.e., the sum over $\ell \geq 2$) is negligible w.r.t. the other terms if (vanish-secondorder) holds. Hence, assuming now (vanish-secondorder), we obtain that the rescaled covariance $\mathrm{Cov}(r_m\widehat{\mathbb{F}}_m(t), r_m\widehat{\mathbb{F}}_m(s))$ of $r_m\widehat{\mathbb{F}}_m$ converges to the following covariance function

$$K(t,s) = \frac{1}{1+|\theta|}(t \wedge s - ts) + \frac{\theta}{1+|\theta|}c_1(t)c_1(s), \tag{10}$$

where θ is defined in (H_2) and where we use the conventions $\theta/(1+|\theta|) = 1$ and $1/(1+|\theta|) = 0$ when $\theta = +\infty$. Note that (H_2) always holds up to consider a subsequence, because $m\gamma_m \geq -1$ from the non-negativeness of $\Gamma^{(m)}$.

Remark 2.2. In the RHS of expression (10), the second term is not necessarily a covariance function because θ can be negative. Nevertheless, K can be written as $K(t,s) = \frac{1}{1+|\theta|}\widetilde{K}(t,s) + \frac{1+\theta}{1+|\theta|}c_1(t)c_1(s)$, where

$$\widetilde{K}(t,s) = t \wedge s - ts - c_1(t)c_1(s) \tag{11}$$

turns out to be a covariance function; considering a Wiener process $(W_t)_{t\in[0,1]}$, $\widetilde{K}$ is the covariance function of the process $W_t - tW_1 - c_1(t)\int_0^1 \bar{\Phi}^{-1}(s)\,\mathrm{d}W_s$, which is the orthogonal projection in L^2 of W_t onto the orthogonal of the linear space spanned by W_1 and $\int_0^1 \bar{\Phi}^{-1}(s)\,\mathrm{d}W_s$. Interestingly, the latter also shows that the original covariance K given by (10) can be seen as the covariance function of $\mathbb{Z}_t = (1+|\theta|)^{-1/2}(W_t - tW_1) + (1+|\theta|)^{-1/2}((1+\theta)^{1/2} - 1)c_1(t)\int_0^1 \bar{\Phi}^{-1}(s)\,\mathrm{d}W_s$.

3. Main result

3.1. Statement

Our main result establishes that the convergence of the covariance functions investigated in Section 2 can be extended to the case of a weak convergence of process. For this, we should consider the other technical assumptions described in Section 1.2.

Theorem 3.1. *Let us consider the empirical distribution function $\widehat{\mathbb{F}}_m$ defined by* (1). *Assume that the covariance matrix $\Gamma^{(m)}$ depends on m in such a way that* (vanish-secondorder) *and* (H_1) *hold with r_m defined by* (3) *and assume* (H_2). *Consider $(\mathbb{Z}_t)_{t\in[0,1]}$ the Gaussian continuous process with covariance function K defined by* (10). *Then we have the convergence* (*in the Skorokhod topology*)

$$r_m(\widehat{\mathbb{F}}_m - I) \rightsquigarrow \mathbb{Z}, \qquad \text{as } m \to \infty, \tag{12}$$

where $I(t) = t$ denotes the identity function. Moreover, the result holds by replacing the set of assumptions {(vanish-secondorder), (H_1) *and* (H_2)} *by* {(H_3) *and* (H_4)}.

Theorem 3.1 is illustrated in the next section, which provides several examples.

3.2. Examples

Let us first note that assumptions (vanish-secondorder) and (H_1) always hold under the following condition

$$|\Gamma^{(m)}_{i,j}| \le a_m \qquad \text{for all } i \neq j \quad \text{and} \quad a_m \text{ satisfies } m^{1+\delta} a_m^2 \to 0 \qquad \text{for some } \delta > 0. \quad (13)$$

Also remember that, as mentioned in Section 1.2, regime (i) (resp. (ii)) referred to the case where $\theta < \infty$ (resp. $\theta = \infty$). We now give several types of matrix Γ^m for which Theorem 3.1 can be applied.

Equi-correlation

Let us start with the following simple example:

$$\Gamma^{(m)} = \begin{pmatrix} 1 & \rho_m & \dots & \rho_m \\ \rho_m & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho_m \\ \rho_m & \dots & \rho_m & 1 \end{pmatrix} = (1-\rho_m) I_m + \rho_m \begin{pmatrix} 1 \\ \vdots \\ \vdots \\ 1 \end{pmatrix} \begin{pmatrix} 1 \\ \vdots \\ \vdots \\ 1 \end{pmatrix}^T, \quad (14)$$

where $\rho_m \in [-(m-1)^{-1}, 1]$ is some parameter. We easily check that γ_m defined by (2) is given by $m\gamma_m = (m-1)\rho_m$ and that the assumptions of Theorem 3.1 are all satisfied if $\rho_m \to 0$ and $m\rho_m$ converges to some $\theta \in [-1, +\infty]$, which yields convergence (12). This is in accordance with Lemma 3.3 of Delattre and Roquain [9].

This simple example already shows that, following the choice of the sequence $(\rho_m)_m$, the empirical distribution function can have various asymptotic behaviors. For instance, taking $\rho_m = -(m-1)^{-1}$ gives a process in regime (i) with a minimal asymptotic covariance function ($\theta = -1$, see (11)), while taking $\rho_m \sim m^{-2/3}$ leads to a rate $r_m \sim m^{1/3} \ll m^{1/2}$ and thus a process converging in regime (ii).

Alternate equi-correlation

Let us consider the covariance matrix:

$$\Gamma^{(m)} = \begin{pmatrix} 1 & -\rho_m & \rho_m & \dots & \\ -\rho_m & 1 & -\rho_m & \ddots & \vdots \\ \rho_m & \ddots & \ddots & \ddots & \rho_m \\ \vdots & \ddots & -\rho_m & 1 & -\rho_m \\ & \dots & \rho_m & -\rho_m & 1 \end{pmatrix} = (1-\rho_m) I_m + \rho_m \begin{pmatrix} 1 \\ -1 \\ 1 \\ \vdots \end{pmatrix} \begin{pmatrix} 1 \\ -1 \\ 1 \\ \vdots \end{pmatrix}^T, \quad (15)$$

where $\rho_m \in [-(m-1)^{-1}, 1]$ is a given parameter. Clearly, γ_m is such that

$$m\gamma_m = 2\rho_m m^{-1} \sum_{i=1}^{m-1} \sum_{k=1}^{i} (-1)^k = -\rho_m \lfloor m/2 \rfloor / (m/2).$$

Hence the rate r_m defined by (3) is $r_m \sim \sqrt{m}$ and assumptions of Theorem 3.1 are fulfilled (with $\theta = 0$) by assuming that $m^{1+\delta}\rho_m^2 \to 0$, with $\delta > 0$ (because (13) holds). Hence, under that assumption, $\sqrt{m}(\widehat{\mathbb{F}}_m - I)$ converges to a standard Brownian bridge.

Maybe surprisingly, this example shows that, even if the correlations are "strong" (e.g., $\rho_m \sim m^{-2/3}$, to be compared with the equi-correlated case), positive and negative correlations can exactly compensate each other to provide the same convergence result as under independence.

Long-range stationary correlations

Let us consider the correlation matrix of the following form:

$$\Gamma_{i,j}^{(m)} = r\big(|j-i|\big), \qquad \text{for } r(0) = 1, \qquad r(k) = k^{-D}L(k), \qquad 0 < D < 1, \tag{16}$$

where $L : (0, +\infty) \to (0, +\infty)$ is slowly varying at infinity ($\forall t > 0$, $L(tx) \sim L(x)$ as $x \to +\infty$). This framework is often referred to as "long-range dependence" in literature dealing with a stationary setup (see, e.g., Dehling and Taqqu [8] and Doukhan, Lang and Surgailis [13]). We can prove than assumptions (H_3) and (H_4) are satisfied (see Section S-2.1). Hence, by using Theorem 3.1 under (13), we derive

$$\gamma_m^{-1/2}(\widehat{\mathbb{F}}_m - I) \rightsquigarrow c_1(\cdot)Z, \qquad \text{as } m \to \infty,$$

for $Z \sim \mathcal{N}(0, 1)$. This is in accordance with Theorem 1.1 of Dehling and Taqqu [8] (see in particular Example 1 therein).

Finally, let us note that assumption (vanish-secondorder) of Theorem 3.1 is not satisfied for a covariance matrix of the type (16) taken with $D \geq 1$ (short-range) (the other terms in the covariance expansion (9) are required in the limit, see Csörgő and Mielniczuk [6]).

"Vanishing" short/long range correlations

Let us modify slightly the matrix (16), by letting:

$$\Gamma_{i,j}^{(m)} = \rho_m r\big(|j-i|\big), \qquad \text{for } r(0) = 1, \qquad r(k) = k^{-D}, \qquad D > 0, \tag{17}$$

where ρ_m is some non-negative parameter (we removed the slowly varying function for the sake of simplicity). When ρ_m varies in function of m, note that the latter is not of the stationary type. We have

$$m\gamma_m \sim \begin{cases} 2\rho_m \dfrac{m^{1-D}}{(1-D)(2-D)} & \text{if } D \in [0, 1), \\ 2\rho_m \log m & \text{if } D = 1, \\ 2\rho_m \displaystyle\sum_{k \geq 1} k^{-D} & \text{if } D > 1, \end{cases} \tag{18}$$

see (S-11) in the supplement [10] for a proof. Assuming that the quantity (18) as a limit (denoted θ), (vanish-secondorder) and (H_1) hold if $m^{1+\delta}\rho_m^2 \to 0$ with $\delta > 0$ (because (13) holds). The

Table 1. Rate r_m defined by (3) in function of $D \geq 0$ and ρ_m such that $\rho_m = \mathrm{o}(m^{-(1/2+\delta)})$ for some $\delta > 0$, for the particular covariance (17)

	$D \in [0, 1)$	$D \geq 1$
$\theta < \infty$	$\rho_m m^{1-D} = \mathrm{O}(1)$ $r_m \sim \sqrt{m}$	$\theta = 0$ $r_m \sim \sqrt{m}$
$\theta = \infty$	$\rho_m m^{1-D} \to \infty$ $r_m \sim \rho_m^{-1/2} m^{D/2}$	Not possible

resulting rate of convergence r_m is given as a function of D and ρ_m in Table 1. Markedly, weak short-range correlations ($D > 1$) always yields $r_m \sim m^{1/2}$ while weak long-range correlations ($D < 1$) can give both regimes. For instance, taking $\rho_m \sim m^{-2/3}$ yields $r_m \sim m^{D/2+1/3}$ for $D < 1/3$ and $r_m \sim m^{1/2}$ otherwise. Overall, the convergence rate increases with D.

"Vanishing" factor model

This model is also referred to as "spiked" covariance matrix, see Johnstone [24]. It assumes that the k-first eigenvalues of the covariance matrix are greater than 1 (for some fixed value of k) while the other are all equal to 1. In our setting where we consider only *correlation* matrices, we assume that the sequence of eigenvalues is constant after some fixed rank k. Precisely, let us consider a matrix $\Gamma^{(m)}$ of the following form:

$$\Gamma^{(m)} = (1 - \rho_m) I_m + \rho_m P H P^T, \tag{19}$$

where H is a $k \times k$ diagonal matrix with diagonal entries $h_1^{(m)}, \dots, h_k^{(m)} \in (1, \infty)$, where $P = (p_{i,r}^{(m)})_{1 \leq i \leq m, 1 \leq r \leq k}$ is an $m \times k$ matrix such that $P^T P = I_k$ and where $\rho_m \in [-1, 1]$ is some parameter. Importantly, k is taken fixed with m. The k first eigenvalues of $\Gamma^{(m)}$ are thus given by $1 - \rho_m + \rho_m h_r^{(m)}$, $r = 1, \dots, k$, while the remaining eigenvalues are all equal to $1 - \rho_m$. Hence, to ensure that $\Gamma^{(m)}$ given by (19) is a well defined correlation matrix, we should additionally assume that for all $r = 1, \dots, k$, $1 - \rho_m + \rho_m h_r^{(m)} \geq 0$, and that $P H P^T$ has diagonal entries equal to 1, that is, for all $i = 1, \dots, m$, $\sum_{r=1}^k h_r^{(m)} (p_{i,r}^{(m)})^2 = 1$. Note that the latter requires $h_1^{(m)} + \cdots + h_k^{(m)} = m$ and thus $\max_r \{h_r^{(m)}\} \geq m/k$.

In Section S-2.2, we prove that (vanish-secondorder) and (H_1) hold provided that

$$r_m^{2+\delta} \rho_m^2 \to 0 \qquad \text{with } \delta > 0. \tag{20}$$

Hence, Theorem 3.1 applies (up to consider a subsequence such that the convergence (H_2) holds). In (20), the rate r_m can be computed by using the definition, see (3), or the following expression.

$$m\gamma_m = \rho_m \sum_{r=1}^k h_r^{(m)} \left(m^{-1/2} \sum_{i=1}^m p_{i,r}^{(m)} \right)^2 - \rho_m. \tag{21}$$

The rate of convergence thus intrinsically depends on the asymptotic behavior of the coordinate-wise mean of each eigenvector $(p_{i,r}^{(m)})_{1\leq i\leq m}$.

To further illustrate this example, we can focus on the particular case where $k=1$. In that case, the model can be equivalently written as

$$\Gamma^{(m)} = (1-\rho_m)I_m + \rho_m \xi\xi^T, \tag{22}$$

where $\xi = \xi^{(m)}$ is a $m \times 1$ vector in $\{-1,1\}^m$ and where $\rho_m \in [-(m-1)^{-1}, 1]$. The model (22) contains as particular instances the equicorrelated matrix ($\xi^{(m)} = (1\ 1 \cdots 1)^T$) and the alternate equicorrelated matrix ($\xi^{(m)} = (1\ -1\ 1 \cdots)^T$) that we have studied above. We easily check that condition (20) recovers the conditions that we obtained in each of theses particular cases. In general, for an arbitrary $\xi^{(m)} \in \{-1,1\}^m$, since the quantity in (21) is equal to

$$\rho_m \left(m^{-1/2} \sum_{i=1}^m \xi_i^{(m)} \right)^2 - \rho_m, \tag{23}$$

the rate r_m is directly related to the number of -1 and $+1$ into $\xi^{(m)}$. For instance, if $\xi^{(m)} = (U_1,\ldots,U_m)$ where $U_1, U_2, \ldots$ are i.i.d. centered random signs, we have by the central limit theorem that the quantity (23) tends to 0 (in probability) whenever $\rho_m \to 0$, which gives a rate $r_m \sim \sqrt{m}$ (in probability). Hence, we obtain the convergence (12) with the same rate and asymptotic variance as in the independent case whenever $m^{1+\delta}\rho_m^2 \to 0$ with $\delta > 0$.

Sample correlation matrix

We consider the model where the correlation matrix is generated *a priori* as a Gaussian empirical correlation matrix. Namely, let us assume that

$$\Gamma^{(m)} = D^{-1}SD^{-1}, \qquad \text{for } S = n_m^{-1}X^TX \quad \text{and} \quad D = \operatorname{diag}(S_{1,1},\ldots,S_{m,m})^{1/2}, \tag{24}$$

where X is a $n_m \times m$ matrix with i.i.d. $\mathcal{N}(0,1)$ entries. Assume $m/n_m \to 0$ as m tends to infinity, which, in a statistical setup, corresponds to assume that the number m of variables (columns of X) is of smaller order than the sample size n_m.

A by-product of Theorem 2 in Bai and Yin [2] (adding a number of variables which is a vanishing small proportion of the sample size) is that,

$$\|S - I_m\|_2 \xrightarrow{P} 0,$$

where $\|\cdot\|_2$ denotes the Euclidian-operator norm, that is, $\|S - I_m\|_2 = \max_{1\leq i\leq m} |\lambda_i^{(m)} - 1|$ and $\lambda_1^{(m)},\ldots,\lambda_m^{(m)}$ denote the eigenvalues of S. Hence, $\max_{1\leq i\leq m}|S_{i,i} - 1| \xrightarrow{P} 0$, which in turn implies $\|\Gamma^{(m)} - I_m\|_2 \xrightarrow{P} 0$. Next, simple arguments entail the following inequalities:

$$\left| m^{-1} \sum_{i\neq j} \Gamma_{i,j}^{(m)} \right| = m^{-1} \left| \left\langle (1\cdots 1)^T, \left(\Gamma^{(m)} - I_m\right)(1\cdots 1)^T \right\rangle \right| \leq \left\| \Gamma^{(m)} - I_m \right\|_2;$$

$$r_m^2 m^{-2} \sum_{i \neq j} \big(\Gamma_{i,j}^{(m)}\big)^2 \leq m^{-1} \sum_{i=1}^{m} \big(\lambda_i^{(m)} - 1\big)^2 \leq \big\| \Gamma^{(m)} - I_m \big\|_2^2;$$

$$r_m^{4+\varepsilon_0} m^{-2} \sum_{i \neq j} \big(\Gamma_{i,j}^{(m)}\big)^4 \leq \Big\{ \min_{1 \leq i \leq n_m} |S_{i,i}| \Big\}^{-4} m^{\varepsilon_0/2} \sum_{i \neq j} (S_{i,j})^4.$$

Moreover, we easily check that $\mathbb{E}(n_m^{1/2} S_{i,j})^4 = \mathbb{E}(n_m^{-1/2} \sum_{k=1}^{n_m} X_{k,i} X_{k,j})^4$ is upper bounded by some positive constant. Hence, by assuming that the sequence n_m satisfies

$$m^{1+\delta}/n_m \to 0 \qquad \text{for some } \delta > 0,$$

the above inequalities implies that the rate is $r_m \sim \sqrt{m}$, that (H_2) holds with $\theta = 0$ and that (vanish-secondorder) and (H_1) are satisfied (all these convergences holding in probability). Hence, Theorem 3.1 can be applied and this shows that the asymptotic of the empirical distribution function is the same as under independence.

4. Application to multiple testing

4.1. Context

The so-called "Benjamini and Hochberg procedure" (BH procedure), widely popularized after the celebrated paper Benjamini and Hochberg [4], is often given as the default procedure to provide a false discovery proportion (FDP) close to some pre-specified error level α. More specifically, the BH procedure provides that the *expectation* of the FDP, called the false discovery rate (FDR), is bounded by the nominal level α under independence of the tests (and also for some other types of dependence, see Benjamini and Yekutieli [5], Farcomeni [18] and Kim and van de Wiel [25]). The present work shows that, under our assumptions, the distribution of the FDP of BH procedure is widening around its expectation, as the quantity γ_m defined by (2) grows.

The formal link between the FDP, the BH procedure and e.d.f.'s has been delineated in Genovese and Wasserman [21] and Farcomeni [19] (FDP at a fixed threshold) and consolidated later in Neuvial [26] (FDP at BH threshold). Here, we follow the approach of Neuvial [26], by using that the FDP of BH procedure is a Hadamard differentiable function of (rescaled) empirical distribution functions. Convergence results are thus derived from Theorem 3.1 by applying the (partial) functional delta method, see Proposition S-1.1.

4.2. Multiple testing setting

Let us add to the original vector $Y \sim \mathcal{N}(0, \Gamma)$ an unknown vector $H = (H_i)_{1 \leq i \leq m} \in \{0, 1\}^m$ as follows: for $1 \leq i \leq m$,

$$X_i = \delta H_i + Y_i, \tag{25}$$

for some positive number δ (assumed to be fixed with m). Hence, $X \sim \mathcal{N}(\delta H, \Gamma)$. Now consider the statistical problem of finding H from the observation of $X = (X_i)_{1\le i\le m}$. From an intuitive point of view, H is the "signal" (unknown parameter of interest), Y is the "noise" (unobserved) while Γ and δ are "nuisance" parameters. An important remark is that Γ can be sometimes known, for instance in genome-wide association study, see Section 2 in Fan, Han and Gu [17]. If Γ is unknown, the quantities γ_m can be often estimated by a standard sample variance estimator, see Remark 4.1.

Let us define the following e.d.f.'s: for $t \in [0, 1]$,

$$\widehat{\mathbb{F}}_{0,m}(t) = m_0^{-1}\sum_{i=1}^{m}(1 - H_i)\mathbf{1}\{\bar{\Phi}(X_i) \le t\}; \tag{26}$$

$$\widehat{\mathbb{F}}_{1,m}(t) = m_1^{-1}\sum_{i=1}^{m} H_i\mathbf{1}\{\bar{\Phi}(X_i) \le t\}; \tag{27}$$

$$\widehat{\mathbb{G}}_m(t) = m^{-1}\sum_{i=1}^{m}\mathbf{1}\{\bar{\Phi}(X_i) \le t\} = \frac{m_0}{m}\widehat{\mathbb{F}}_{0,m}(t) + \frac{m_1}{m}\widehat{\mathbb{F}}_{1,m}(t), \tag{28}$$

where $m_0 = \sum_{i=1}^m (1 - H_i)$ and $m_1 = \sum_{i=1}^m H_i$. Here, the quality of a procedure that rejects each null hypothesis "$H_i = 0$" whenever $\bar{\Phi}(X_i) \le t$ is given by

$$\mathrm{FDP}_m(t) = \frac{(m_0/m)\widehat{\mathbb{F}}_{0,m}(t)}{\widehat{\mathbb{G}}_m(t)}, \tag{29}$$

where we used the convention $0/0 = 0$. Now, define the following functional: for $\alpha \in (0, 1)$, $\mathcal{T}(H) = \sup\{t \in [0, 1] : H(t) \ge t/\alpha\}$, for $H \in D(0, 1)$, with the convention $\sup\{\varnothing\} = 0$. Classically, the BH procedure (at level α) corresponds the thresholding $\mathcal{T}(\widehat{\mathbb{G}}_m)$, see Genovese and Wasserman [21]. In the sequel, the random variable $\mathrm{FDP}_m(\mathcal{T}(\widehat{\mathbb{G}}_m))$ is denoted by FDP_m for short.

As proved in the supplemental file (see Section S-1 in [10]), in this model, the asymptotic properties of FDP_m generally depends on which null hypotheses are true or not, which can be considered as a limitation. Nevertheless, this fact is inherent to the multiple testing setting considered here, because the dependencies accounting in the FDP of the BH procedure are related to the sub-matrix $(\Gamma_{i,j})_{i,j:H_i=H_j=0}$ and thus are linked to the location of the true null hypotheses. A convenient way to circumvent this problem is to add *random effects*, by assuming that, previously and independently to the model (25), we have drawn $H = (H_1, \ldots, H_m)$ for $H_1, H_2, \ldots$ i.i.d. Bernoulli variables of parameter $\pi_1 = 1 - \pi_0$, for some $\pi_0 \in (0, 1)$. Thus X follows the distribution $\mathcal{N}(\delta H, \Gamma)$ *conditionally* on H. The corresponding global (unconditional) model, often referred to as the *two-group mixture model* has been widely used in the multiple testing literature, see, for example, Efron *et al.* [16], Storey [31], Genovese and Wasserman [21] and Roquain and Villers [27].

Remark 4.1. It is common that the X_i's are obtained via an averaging on several i.i.d. replications, that is, $X_i = n^{-1/2} \sum_{j=1}^{n} Z_{i,j}$ where $(Z_{i,1})_{1\le i\le m}, \dots, (Z_{i,n})_{1\le i\le m}$ are n i.i.d. Gaussian vectors $\mathcal{N}(\delta n^{-1/2} H, \Gamma)$. In that case, the following holds:

- $\gamma_m = \mathrm{Var}(\bar{X}_m)$ can be estimated by the sample variance $\widehat{\gamma}_m$ of $\bar{Z}_j = m^{-1} \sum_{i=1}^{m} Z_{i,j}$, $1 \le j \le n$. For instance, a confidence bound can be obtained for γ_m by noting that $(n-1)\widehat{\gamma}_m/\gamma_m \sim \chi^2(n-1)$.
- To check whether (vanish-secondorder) holds, the quantity $(2/m^2) \sum_{i,j=1}^{m} \Gamma_{i,j}^2$ can be overestimated by the empirical variance of $\bar{Z}_j^2 = m^{-1} \sum_{i=1}^{m} Z_{i,j}^2$, $1 \le j \le n$. This holds because $H^T \Gamma H \ge 0$ and $\mathrm{Var}(\bar{Z}_1^2) = 2m^{-2} \sum_{i,i'=1}^{m} \Gamma_{i,i'}^2 + 4\delta^2 n^{-1} H^T \Gamma H$.

4.3. Result

Denote $G(t) = \pi_0 F_0(t) + \pi_1 F_1(t)$, where $F_0(t) = t$, $F_1(t) = \bar{\Phi}(\bar{\Phi}^{-1}(t) - \delta)$. The following result is proved in Section S-1.

Corollary 4.2. *Consider the two-group mixture model defined above, generated from parameters* $\delta > 0$, $\pi_0 \in (0,1)$ *and a correlation matrix* $\Gamma = \Gamma^{(m)}$. *Assume that* Γ *satisfies either* {(vanish-secondorder) *and* (H_1)} *or* {(H_3) *and* (H_4)}. *Let* $\alpha \in (0,1)$ *and* $t^\star = t^\star(\delta, \alpha)$ *be the unique* $t \in (0,1)$ *such that* $G(t) = t/\alpha$. *Let* $h(t^\star) = (\phi(\bar{\Phi}^{-1}(t^\star))/t^\star)^2$. *Then the sequence of r.v.* FDP_m *defined by* (29) *enjoys the following convergence*:

$$\frac{FDP_m - \pi_0 \alpha}{\pi_0 \alpha \{(1/t^\star - \pi_0)/(\pi_0 m) + h(t^\star)\gamma_m\}^{1/2}} \rightsquigarrow \mathcal{N}(0,1), \tag{30}$$

where γ_m *is defined by* (2).

Corollary 4.2 provides a theoretical support for the following fact: as m grows to infinity, the concentration of FDP_m around $\pi_0\alpha$ deteriorates when γ_m increases, so when positive correlations appear between the individual statistical tests. However, notice that, as for the e.d.f. convergence, negative correlations help to decrease γ_m and can entail a concentration even better than under independence when γ_m is negative (although this phenomenon is necessary of limited amplitude because $\gamma_m \ge -1/m$).

To illustrate further Corollary 4.2, Figure 2 displays the true distribution of FDP_m, together with the Gaussian approximation obtained by Corollary 4.2. The two-group mixture model chosen to generate the X_i's uses a factor model (19) for Γ with the following parameters: $k = 3$, $m\rho_m \in \{0, 10, 10^2, 10^3\}$, $h_1/m = 0.4$, $h_2/m = 0.3$, $h_3/m = 0.6$, and $p_1 = (1,1,\dots)/m^{1/2}$, $p_2 = (1,-1,1,\dots)/m^{1/2}$, $p_3 = (1,\dots,1,-1,\dots,-1)/m^{1/2}$ (m is assumed to be even). The parameters of the mixture are $\pi_0 = 0.9$ and $\delta = 3$. The BH procedure is taken at level $\alpha = 0.25$.

This experiment shows that, even for a relatively small values for ρ_m ($\rho_m = 0.002$ or $\rho_m = 0.02$), the FDP distribution can be largely affected by the dependencies. Also, for $m = 5000$ (left picture), while the Gaussian approximation looks accurate for $m\rho_m \in \{0, 10, 100\}$, this seems more questionable when $m\rho_m = 1000$. This non-Gaussian phenomenon, whose amplitude increases with ρ_m (for a fixed m), shows the limit of the proposed methodology. As a matter of

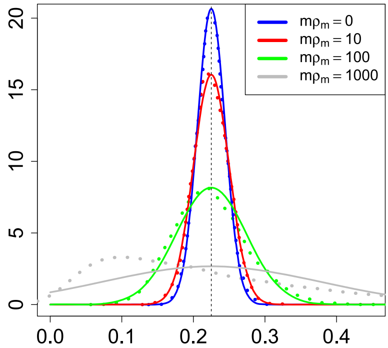

Figure 2. Distribution of FDP_m (29) under a 3-factor model, see text; the dotted lines corresponds to the true distribution computed over 5000 simulations, the solid lines display the Gaussian approximation given by Corollary 4.2 (whose mean, $\pi_0\alpha = 0.225$, is displayed by the dashed vertical line).

fact, additional experiments show that the approximation induced by Theorem 3.1 is still valid for $m = 5000$ and $m\rho_m = 1000$. Hence, we suspect that, for this case, the FDP cannot be approximated by a linear function of the e.d.f.'s which results in a poor accuracy when using the delta method. Finally, the right display in Figure 2 shows that, as one can expect, this phenomenon disappears by increasing the value of m.

To conclude, this study reinforces the fact that it is desirable to incorporate the dependence (e.g., γ_m) in a multiple testing procedure in order to "stabilize" the behavior of the FDP. This is an exciting direction for a future work.

5. Proof of Theorem 3.1

5.1. A related result and additional notation

Let us define the "modified" empirical distribution function $\widetilde{\mathbb{F}}_m$ by the following relation: for $t \in [0, 1]$,

$$r_m\big(\widehat{\mathbb{F}}_m(t) - t\big) = r_m\big(\widetilde{\mathbb{F}}_m(t) - t\big) + c_1(t) r_m \overline{Y}_m. \tag{31}$$

The convergence of the two processes $r_m(\widehat{\mathbb{F}}_m - I)$ and $r_m(\widetilde{\mathbb{F}}_m - I)$ are strongly related by (31). The main idea of our proof is to deduce the convergence of $r_m(\widehat{\mathbb{F}}_m - I)$ from the one of $r_m(\widetilde{\mathbb{F}}_m - I)$. Precisely, the following result will be proved together with Theorem 3.1 in the sequel.

Proposition 5.1. *Under one of the two sets of assumptions of Theorem 3.1, let us consider the corrected empirical distribution function $\widetilde{\mathbb{F}}_m$ defined by (31) and a continuous process $(\widetilde{\mathbb{Z}}_t)_{t\in[0,1]}$ with covariance function $\widetilde{K}$ defined by (11). Then we have the convergence (in the Skorokhod topology)*

$$r_m(\widetilde{\mathbb{F}}_m - I) \rightsquigarrow \widetilde{\mathbb{Z}}/\big(1 + |\theta|\big)^{1/2}, \qquad \text{as } m \to \infty, \tag{32}$$

where $I(t) = t$ denotes the identity function.

Additionally, throughout the section, we use the following notation

$$h_t(x) = \mathbf{1}\{\bar{\Phi}(x) \le t\} - t - c_1(t)x, \tag{33}$$

so that $\widetilde{\mathbb{F}}_m(t) - t = m^{-1} \sum_{i=1}^m h_t(Y_i)$. Finally, we will sometimes use the following assumption:

$$\begin{aligned}&\text{there exists } \eta > 0 \text{ (independent on } m) \\ &\text{lower bounding the } m \text{ eigenvalues of } \Gamma^{(m)}.\end{aligned} \tag{eigenvalues-away0}$$

5.2. Convergence of finite dimensional laws for $\widetilde{\mathbb{F}}_m$

Let us prove the following result.

Proposition 5.2. *Assume that the covariance matrix Γ depends on m in such a way that (vanish-secondorder) holds with r_m defined by (3) and assume (H_2). Consider a continuous process $(\widetilde{\mathbb{Z}}_t)_{t\in[0,1]}$ with covariance function $\widetilde{K}$ defined by (11). Then, the process $(r_m(\widetilde{\mathbb{F}}_m - I), Y_1^{(m)})$ (jointly) converges to $\mathcal{L}(\widetilde{\mathbb{Z}}/(1 + |\theta|)^{1/2}) \otimes \mathcal{N}(0, 1)$ in the sense of the finite dimensional convergence. In particular, the convergence (32) holds in the sense of the finite dimensional convergence.*

Proof. The proof is based on central limit theorems for martingale arrays as presented, for example, in Chapter 3 of Hall and Heyde [22].

First, since we aim at obtaining a convergence jointly with $Y_1^{(m)}$, a (somewhat technical but useful) task is to define the array of random variables $(Y_i^{(m)}, 1 \le i \le m, m \ge 1)$ is such a way that $Y_1^{(m)}$ is fixed with m. This is possible by first considering some variable $Z \sim \mathcal{N}(0, 1)$, by letting $Y_1^{(m)} = Z$ for all $m \ge 1$, and then by choosing for each $m \ge 2$, the variables $Y_i^{(m)}, 2 \le i \le m$, such that

- $(Z, Y_i^{(m)}, 2 \le i \le m) \sim \mathcal{N}(0, \Gamma^{(m)})$;
- $\{(Y_i^{(m)})_{2\le i\le m}, m \ge 2\}$ is a family of mutually independent vectors conditionally on Z.

This also define a common underlying space $(\Omega, \mathcal{F}, \mathbb{P})$ for the array of random variables.

Now, define the following nested array of σ-field: for $m \ge 1$, $\mathcal{G}_{m,0} = \sigma(\varnothing)$ and for $1 \le i \le m$,

$$\mathcal{G}_{m,i} = \sigma\big(Y_j^{(\ell)}, 1 \le j \le i \wedge \ell, 1 \le \ell \le m\big).$$

Next, let us consider for each $t \in [0, 1]$, the martingale array $(M_{m,i}(t), \mathcal{G}_{m,i}, 1 \le i \le m, m \ge 1)$ defined as follows:

$$M_{m,i}(t) = \sum_{j=1}^{i} X_{m,j}(t) \qquad \text{for } X_{m,j}(t) = \frac{r_m}{m}\big(h_t\big(Y_j^{(m)}\big) - \mathbb{E}\big(h_t\big(Y_j^{(m)}\big)|\mathcal{G}_{m,j-1}\big)\big). \tag{34}$$

Clearly,

$$r_m\big(\widetilde{\mathbb{F}}_m(t)-t\big)=M_{m,m}(t)+\frac{r_m}{m}\sum_{i=1}^m \mathbb{E}\big(h_t\big(Y_i^{(m)}\big)|\mathcal{G}_{m,i-1}\big). \tag{35}$$

Also note that we can replace each $\mathcal{G}_{m,i}$ by $\mathcal{F}_{m,i}=\sigma(Y_1^{(m)},\dots,Y_i^{(m)})$ ($\mathcal{F}_{m,0}=\sigma(\varnothing)$) in the above expression, because $(Y_i^{(m)}, 2\le i\le m)$ is independent of $(Y_j^{(\ell)}, 2\le j\le i\wedge \ell, 2\le \ell<m)$, conditionally on $Y_1^{(m)}$.

Case 1: (eigenvalues-away0) *is assumed*. We show in Lemma S-3.1 expression (S-16) that the second term in the RHS of (35) has a vanishing variance as m tends to infinity. Therefore, it remains to show that the conclusion of Proposition 5.2 holds for the process $M_{m,m}$, which we prove by using Lindeberg's theorem. We use Corollary 3.1 page 58 in Hall and Heyde [22] (or more precisely its generalization to the multidimensional case). The conditions are as follows:

(i) for all $t\in[0,1]$, for all $\varepsilon>0$, $\sum_{i=1}^m \mathbb{E}((X_{m,i}(t))^2\mathbf{1}\{|X_{m,i}(t)|>\varepsilon\}|\mathcal{F}_{m,i-1})\xrightarrow{P}0$;
(ii) for all $t,s\in[0,1]$, $\sum_{i=1}^m \mathbb{E}(X_{m,i}(t)X_{m,i}(s)|\mathcal{F}_{m,i-1})\xrightarrow{P}\widetilde{K}(t,s)$.

To check (i), let us fix $t\in[0,1]$ and prove $\sum_{i=1}^m \mathbb{E}(X_{m,i}(t))^4=\mathrm{o}(1)$. By definition, we have

$$\begin{aligned}\sum_{i=1}^m \mathbb{E}\big(X_{m,i}(t)\big)^4 &= \frac{r_m^4}{m^4}\sum_{i=1}^m \mathbb{E}\big(h_t\big(Y_i^{(m)}\big)-\mathbb{E}\big(h_t\big(Y_i^{(m)}\big)|\mathcal{F}_{m,i-1}\big)\big)^4\\ &\le 2^4\left(\frac{r_m^4}{m^3}m^{-1}\sum_{i=1}^m \mathbb{E}\big(h_t\big(Y_i^{(m)}\big)\big)^4+\frac{r_m^4}{m^4}\sum_{i=1}^m \mathbb{E}\big(\mathbb{E}\big(h_t\big(Y_i^{(m)}\big)|\mathcal{F}_{m,i-1}\big)\big)^4\right)\\ &\le 2^5\frac{r_m^4}{m^3}m^{-1}\sum_{i=1}^m \mathbb{E}\big(h_t\big(Y_i^{(m)}\big)\big)^4.\end{aligned}$$

Now, the RHS of the previous display converges to zero because $r_m\le\sqrt{m}$ and $\mathbb{E}(h_t(Y_i^{(m)}))^4<\infty$. This proves condition (i) of Lindeberg's theorem.

Let us now turn to condition (ii). For $t,s\in[0,1]$, we obviously obtain

$$\begin{aligned}&\sum_{i=1}^m \mathbb{E}\big(X_{m,i}(t)X_{m,i}(s)|\mathcal{F}_{m,i-1}\big)\\ &\quad=\frac{r_m^2}{m^2}\sum_{i=1}^m \mathbb{E}\big(h_t\big(Y_i^{(m)}\big)h_s\big(Y_i^{(m)}\big)|\mathcal{F}_{m,i-1}\big)\\ &\qquad-\frac{r_m^2}{m^2}\sum_{i=1}^m \mathbb{E}\big(h_t\big(Y_i^{(m)}\big)|\mathcal{F}_{m,i-1}\big)\mathbb{E}\big(h_s\big(Y_i^{(m)}\big)|\mathcal{F}_{m,i-1}\big).\end{aligned} \tag{36}$$

Next, by using $ab \le 2(a^2 + b^2)$ for all $a, b \in \mathbb{R}$ together with (S-15), the second term in the RHS of (36) tends to zero in probability. Moreover, we have

$$\begin{aligned}
&\operatorname{Var}\biggl(\frac{r_m^2}{m^2}\sum_{i=1}^{m}\bigl(h_t\bigl(Y_i^{(m)}\bigr)h_s\bigl(Y_i^{(m)}\bigr) - \mathbb{E}\bigl(h_t\bigl(Y_i^{(m)}\bigr)h_s\bigl(Y_i^{(m)}\bigr)|\mathcal{F}_{m,i-1}\bigr)\bigr)\biggr) \\
&\quad = \frac{r_m^4}{m^4}\sum_{i=1}^{m}\operatorname{Var}\bigl(h_t\bigl(Y_i^{(m)}\bigr)h_s\bigl(Y_i^{(m)}\bigr) - \mathbb{E}\bigl(h_t\bigl(Y_i^{(m)}\bigr)h_s\bigl(Y_i^{(m)}\bigr)|\mathcal{F}_{m,i-1}\bigr)\bigr),
\end{aligned}$$

because the elements inside the sum are martingale increments. Hence, the quantity inside the above display tends to zero. Combining the latter with (36) establishes condition (ii) of Lindeberg's theorem provided that the following holds:

$$\frac{r_m^2}{m^2}\sum_{i=1}^{m} h_t\bigl(Y_i^{(m)}\bigr)h_s\bigl(Y_i^{(m)}\bigr) \xrightarrow{P} \bigl(1+|\theta|\bigr)^{-1}\widetilde{K}(t,s).$$

This comes directly from the law of large number stated in Lemma S-5.2, because $r_m^2/m \to (1+|\theta|)^{-1}$ by (3) and (H_2).

Applying Lindeberg's theorem (in the underlying space described above), for any $t_1, \dots, t_k \in [0, 1]$, the random vector

$$Z_m = \bigl(M_{m,m}(t_1), \dots, M_{m,m}(t_k)\bigr)$$

converges stably in the following sense (see, e.g., Jacod and Shiryaev [23], Definition 5.28): for all (fixed) bounded random variable U and continuous bounded function f in $\mathbb{R}^k$,

$$\mathbb{E}\bigl(Uf(Z_m)\bigr) \to \mathbb{E}(U)\mathbb{E}\bigl(f(Z)\bigr) \qquad \text{as } m \to \infty,$$

where Z is a centered multivariate Gaussian vector with covariance $(1+|\theta|)^{-1}(\widetilde{K}(t_i, t_j))_{1\le i,j\le k}$. This implies that (Z_m, Y_1) converges (jointly) in distribution to $\mathcal{L}(Z) \otimes \mathcal{N}(0, 1)$. This finishes the proof of Proposition 5.2 in the case where (eigenvalues-away0) is assumed to hold.

Case 2: (eigenvalues-away0) *is not assumed*. The strategy is to apply Lemma S-5.3 in order to reduce the study to “Case 1” above. For any $\varepsilon > 0$, let

$$Y_i^{\varepsilon} = \frac{Y_i + \varepsilon\xi_i}{(1+\varepsilon^2)^{1/2}},$$

where $\xi_1, \xi_2, \dots$ are i.i.d. $\mathcal{N}(0, 1)$ variables, independent of all the Y_i's. The covariance matrix of $(Y_1^{\varepsilon}, \dots, Y_m^{\varepsilon})$ is obviously

$$\Gamma^{\varepsilon} = \frac{\varepsilon^2}{1+\varepsilon^2}I_m + \frac{1}{1+\varepsilon^2}\Gamma.$$

Clearly, the corresponding rate (3) is $r_m^{\varepsilon} = (m^{-1} + (1+\varepsilon^2)^{-1}|\gamma_m|)^{-1/2}$. It is related to r_m via the following inequalities: $r_m \le r_m^{\varepsilon} \le (1+\varepsilon^2)^{1/2} r_m$. Hence, Γ^{ε} satisfies (vanish-secondorder) and (H_2) with θ replaced by $\theta^{\varepsilon} = \frac{1}{1+\varepsilon^2}\theta$. Since it also satisfies (eigenvalues-away0), by using Proposition 5.2 in the “Case 1” above, it satisfies for any $t_1, \dots, t_k \in [0, 1]$,

(a) $(r_m^\varepsilon(\widetilde{\mathbb{F}}_m^\varepsilon(t_1)-t_1),\dots,r_m^\varepsilon(\widetilde{\mathbb{F}}_m^\varepsilon(t_k)-t_k),Y_1^\varepsilon)\rightsquigarrow\mathcal{L}(\frac{(\widetilde{\mathbb{Z}}(t_1),\dots,\widetilde{\mathbb{Z}}(t_k))}{(1+|\theta^\varepsilon|)^{1/2}})\otimes\mathcal{N}(0,1)$, where $\widetilde{\mathbb{F}}_m^\varepsilon(t)-t=m^{-1}\sum_{i=1}^m h_t(Y_i^\varepsilon)$ for all t. Next, we clearly have,

(b) $\frac{(\widetilde{\mathbb{Z}}(t_1),\dots,\widetilde{\mathbb{Z}}(t_k))}{(1+|\theta^\varepsilon|)^{1/2}}\rightsquigarrow\frac{(\widetilde{\mathbb{Z}}(t_1),\dots,\widetilde{\mathbb{Z}}(t_k))}{(1+|\theta|)^{1/2}}$ as $\varepsilon\to 0$.

Let us now prove that for any $t\in[0,1]$,

$$\limsup_m\bigl\{\mathbb{E}\bigl|r_m\bigl(\widetilde{\mathbb{F}}_m(t)-t\bigr)-r_m^\varepsilon\bigl(\widetilde{\mathbb{F}}_m^\varepsilon(t)-t\bigr)\bigr|\bigr\}\to 0\qquad\text{as }\varepsilon\to 0.\tag{37}$$

This will conclude the proof by applying Lemma S-5.3. First, we write

$$\begin{aligned}
&\mathbb{E}\bigl|r_m\bigl(\widetilde{\mathbb{F}}_m(t)-t\bigr)-r_m^\varepsilon\bigl(\widetilde{\mathbb{F}}_m^\varepsilon(t)-t\bigr)\bigr|\\
&\quad\le\mathbb{E}\left|r_m/m\sum_{i=1}^m\bigl(h_t(Y_i)-h_t\bigl(Y_i^\varepsilon\bigr)\bigr)\right|+\bigl(r_m^\varepsilon-r_m\bigr)\mathbb{E}\left|m^{-1}\sum_{i=1}^m h_t\bigl(Y_i^\varepsilon\bigr)\right|\\
&\quad\le\left\{(r_m/m)^2\mathbb{E}\left(\sum_{i=1}^m\bigl(h_t(Y_i)-h_t\bigl(Y_i^\varepsilon\bigr)\bigr)\right)^2\right\}^{1/2}+\bigl(\bigl(1+\varepsilon^2\bigr)^{1/2}-1\bigr)\mathbb{E}\bigl|r_m^\varepsilon\bigl(\widetilde{\mathbb{F}}_m^\varepsilon(t)-t\bigr)\bigr|.
\end{aligned}$$

By taking the lim sup in the above display, it only remains to show

$$\limsup_m\left\{(r_m/m)^2\mathbb{E}\left(\sum_{i=1}^m\bigl(h_t(Y_i)-h_t\bigl(Y_i^\varepsilon\bigr)\bigr)\right)^2\right\}\to 0\qquad\text{as }\varepsilon\to 0.\tag{38}$$

This can be proved by using Lemma S-4.3 (S-39) as follows:

$$\begin{aligned}
&(r_m/m)^2\mathbb{E}\left(\sum_{i=1}^m\bigl(h_t(Y_i)-h_t\bigl(Y_i^\varepsilon\bigr)\bigr)\right)^2\\
&\quad=(r_m/m)^2\sum_{i,j=1}^m\mathbb{E}\bigl(\bigl(h_t(Y_i)-h_t\bigl(Y_i^\varepsilon\bigr)\bigr)\bigl(h_t(Y_j)-h_t\bigl(Y_j^\varepsilon\bigr)\bigr)\bigr)\\
&\quad=(r_m/m)^2\sum_{i,j=1}^m\bigl(\mathbb{E}\bigl(h_t(Y_i)h_t(Y_j)\bigr)-\mathbb{E}\bigl(h_t(Y_i)h_t\bigl(Y_j^\varepsilon\bigr)\bigr)\\
&\qquad\qquad-\mathbb{E}\bigl(h_t\bigl(Y_i^\varepsilon\bigr)h_t(Y_j)\bigr)+\mathbb{E}\bigl(h_t\bigl(Y_i^\varepsilon\bigr)h_t\bigl(Y_j^\varepsilon\bigr)\bigr)\bigr)\\
&\quad=(r_m/m)^2\sum_{i,j=1}^m\sum_{\ell\ge 2}\frac{(c_\ell(t))^2}{\ell!}(\Gamma_{i,j})^\ell\bigl(1+\bigl(1+\varepsilon^2\bigr)^{-\ell}-2\bigl(1+\varepsilon^2\bigr)^{-\ell/2}\bigr)
\end{aligned}$$

because $\mathrm{Cov}(Y_i,Y_j)=\Gamma_{i,j}$, $\mathrm{Cov}(Y_i^\varepsilon,Y_j)=\mathrm{Cov}(Y_i,Y_j^\varepsilon)=\Gamma_{i,j}/(1+\varepsilon^2)^{1/2}$ and $\mathrm{Cov}(Y_i^\varepsilon,Y_j^\varepsilon)=\Gamma_{i,j}/(1+\varepsilon^2)$. Next, by separating the case $i=j$ and $i\neq j$, the previous display can be upper bounded by

$$\sum_{\ell\geq 2}\frac{(c_\ell(t))^2}{\ell!}\left|1+\left(1+\varepsilon^2\right)^{-\ell}-2\left(1+\varepsilon^2\right)^{-\ell/2}\right|+(r_m/m)^2\sum_{i\neq j}(\Gamma_{i,j})^2\times 4\sum_{\ell\geq 2}\frac{(c_\ell(t))^2}{\ell!}.$$

While the first term above does not depend on m and converges to zero as $\varepsilon\to 0$, the second term above as a $\limsup_m$ equal to zero by (vanish-secondorder). This implies (38) and finishes the proof. □

Remark 5.3. Interestingly, Proposition 5.2 is related to Theorem 3.1 of Soulier [30], in the particular case where $\theta<\infty$ and $\sup_{i\neq j}(\Gamma_{i,j}^{(m)})^2=\mathrm{O}(m^{-2}\sum_{i\neq j}(\Gamma_{i,j}^{(m)})^2)$.

5.3. Convergence of finite dimensional laws for $\widehat{\mathbb{F}}_m$

In this section, we aim at proving the following result.

Proposition 5.4. *Consider the assumptions of Proposition 5.2. Then, (12) holds in the sense of the finite dimensional convergence.*

Proof. From expression (31), we investigate the (joint) convergence of $(r_m(\widetilde{\mathbb{F}}_m-I),r_m\overline{Y}_m)$.

Case 1: $\theta=-1$. In that case, $r_m^2\operatorname{Var}(\overline{Y}_m)\to 0$. Hence, we can directly use Proposition 5.2 to state that $(r_m(\widetilde{\mathbb{F}}_m-I),r_m\overline{Y}_m)$ converges to $\mathcal{L}(\widetilde{\mathbb{Z}}/(1+|\theta|)^{1/2})\otimes\delta_0$ in the sense of the finite dimensional convergence. This establishes Proposition 5.4 in that case.

Case 2: $\theta>-1$. Now, $r_m^2\operatorname{Var}(\overline{Y}_m)$ is converging to some positive real number, namely $(1+\theta)/(1+|\theta|)>0$. In particular, $\operatorname{Var}(\overline{Y}_m)>0$ for m large enough. Let us define the random variable

$$Y_0=\overline{Y}_m(\operatorname{Var}\overline{Y}_m)^{-1/2}.$$

We now consider the $(m+1)$-dimensional random vector $(Y_i)_{0\leq i\leq m}$, which is centered, with a covariance matrix denoted $\Lambda^{(m+1)}=(\Lambda_{i,j}^{(m+1)})_{0\leq i,j\leq m}$ and such that $\Lambda_{0,0}^{(m+1)}=1$, $\Lambda_{i,j}^{(m+1)}=\Gamma_{i,j}^{(m)}$ for $1\leq i,j\leq m$. We easily check that $\Lambda^{(m+1)}$ satisfies (vanish-secondorder) and (H_2) with the same value of θ and a rate asymptotically equivalent to the original r_m, see Lemma S-3.2. Hence, Proposition 5.2 shows that (by using notation therein),

$$\left(r_m\left((m+1)^{-1}\sum_{i=0}^m h_t(Y_i)\right),Y_0\right)\rightsquigarrow\mathcal{L}\left(\widetilde{\mathbb{Z}}/\left(1+|\theta|\right)^{1/2}\right)\otimes\mathcal{N}(0,1),$$

in the sense of the finite dimensional convergence. Since $r_m h_t(Y_0)/m$ tends to zero in probability, the last display can be rewritten as

$$\left(r_m(\widetilde{\mathbb{F}}_m-I),\overline{Y}_m(\operatorname{Var}\overline{Y}_m)^{-1/2}\right)\rightsquigarrow\mathcal{L}\left(\widetilde{\mathbb{Z}}/\left(1+|\theta|\right)^{1/2}\right)\otimes\mathcal{N}(0,1).$$

Finally, since $r_m^2\operatorname{Var}(\overline{Y}_m)\to(1+\theta)/(1+|\theta|)$, we finish the proof by applying (31). □

5.4. Tightness under (vanish-secondorder), (H_1) and (H_2)

To complete the proof of Proposition 5.1, we prove that the process $X_m = r_m(\widetilde{\mathbb{F}}_m - I)$ is tight in the Skorokhod space. This also implies tightness for $r_m(\widehat{\mathbb{F}}_m - I)$ by (31) because c_1 is a continuous function on [0, 1], itself entailing Theorem 3.1.

We consider here the set of assumptions (vanish-secondorder), (H_1) and (H_2) (the second set of assumptions is examined in Section 5.5). For proving the tightness of X_m, we use Proposition S-5.1. This is possible because $|c_1(t) - c_1(s)| \leq L|t-s|^{1/2}$, $0 \leq s, t \leq 1$ for some constant $L > 1$ (see Lemma S-4.4). In Section S-3.3, we finish the proof by showing that for large m,

$$\mathbb{E}\big|X_m(t) - X_m(s)\big|^4 \leq C\big(|t-s|^{3/2} + (r_m)^{-\varepsilon_0}|t-s|\big), \qquad \text{for all } t, s \in [0, 1], \tag{39}$$

for some constant $C > 0$ and for a constant $\varepsilon_0 > 0$ such that (H_1) holds.

5.5. Tightness under (H_3) and (H_4)

Obviously, (H_3) and (H_4) imply (vanish-secondorder), (H_2) with $\theta = +\infty$, and $r_m \sim {\gamma_m}^{-1/2}$. Hence, Proposition 5.2 entails that the finite dimensional laws of $X_m = r_m(\widetilde{\mathbb{F}}_m - I)$ converge to 0 and it only remains to prove that X_m is tight. This can be done as in the previous section, except that we use $\kappa = 2$ in Proposition S-5.1. Namely, we prove in Section S-3.3 that, for large m,

$$\mathbb{E}\big|X_m(t) - X_m(s)\big|^2 \leq C\gamma_m^{\delta_0}|t-s|, \qquad \text{for all } t, s \in [0, 1], \tag{40}$$

for some constants $C > 0$, $\delta_0 > 0$.

Acknowledgements

We would like to warmly acknowledge the Editor, Associate Editor and referee for their helpful comments. We are grateful to Jérôme Dedecker, Stephane Gaiffas, Pierre Neuvial and Mathieu Rosenbaum for interesting discussions. The second author was supported by the French Agence Nationale de la Recherche (ANR grant references: ANR-09-JCJC-0027-01, ANR-PARCIMONIE, ANR-09-JCJC-0101-01) and by the French ministry of foreign and european affairs (EGIDE - PROCOPE project number 21887 NJ).

Supplementary Material

Supplement to "On empirical distribution function of high-dimensional Gaussian vector components with an application to multiple testing." (DOI: 10.3150/14-BEJ659SUPP; .pdf). Supplement that essentially contains the auxiliary results required to prove Theorem 3.1 (main theorem) and Corollary 4.2 (multiple testing application).

References

[1] Arcones, M.A. (1994). Limit theorems for nonlinear functionals of a stationary Gaussian sequence of vectors. *Ann. Probab.* **22** 2242–2274. MR1331224

[2] Bai, Z.D. and Yin, Y.Q. (1993). Limit of the smallest eigenvalue of a large-dimensional sample covariance matrix. *Ann. Probab.* **21** 1275–1294. MR1235416

[3] Bardet, J.-M. and Surgailis, D. (2013). Moment bounds and central limit theorems for Gaussian subordinated arrays. *J. Multivariate Anal.* **114** 457–473. MR2993899

[4] Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **57** 289–300. MR1325392

[5] Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.* **29** 1165–1188. MR1869245

[6] Csörgő, S. and Mielniczuk, J. (1996). The empirical process of a short-range dependent stationary sequence under Gaussian subordination. *Probab. Theory Related Fields* **104** 15–25. MR1367664

[7] Dedecker, J. and Prieur, C. (2007). An empirical central limit theorem for dependent sequences. *Stochastic Process. Appl.* **117** 121–142. MR2287106

[8] Dehling, H. and Taqqu, M.S. (1989). The empirical process of some long-range dependent sequences with an application to U-statistics. *Ann. Statist.* **17** 1767–1783. MR1026312

[9] Delattre, S. and Roquain, E. (2011). On the false discovery proportion convergence under Gaussian equi-correlation. *Statist. Probab. Lett.* **81** 111–115. MR2740072

[10] Delattre, S. and Roquain, E. (2014). Supplement to "On empirical distribution function of high-dimensional Gaussian vector components with an application to multiple testing." DOI:10.3150/14-BEJ659SUPP.

[11] Donsker, M.D. (1952). Justification and extension of Doob's heuristic approach to the Komogorov–Smirnov theorems. *Ann. Math. Stat.* **23** 277–281. MR0047288

[12] Doob, J.L. (1949). Heuristic approach to the Kolmogorov–Smirnov theorems. *Ann. Math. Stat.* **20** 393–403. MR0030732

[13] Doukhan, P., Lang, G. and Surgailis, D. (2002). Asymptotics of weighted empirical processes of linear fields with long-range dependence. *Ann. Inst. Henri Poincaré Probab. Stat.* **38** 879–896. En l'honneur de J. Bretagnolle, D. Dacunha-Castelle, I. Ibragimov. MR1955342

[14] Doukhan, P., Lang, G., Surgailis, D. and Teyssière, G., eds. (2010). *Dependence in Probability and Statistics*. *Lecture Notes in Statistics* **200**. Berlin: Springer. MR2741808

[15] Dudley, R.M. (1966). Weak convergences of probabilities on nonseparable metric spaces and empirical measures on Euclidean spaces. *Illinois J. Math.* **10** 109–126. MR0185641

[16] Efron, B., Tibshirani, R., Storey, J.D. and Tusher, V. (2001). Empirical Bayes analysis of a microarray experiment. *J. Amer. Statist. Assoc.* **96** 1151–1160. MR1946571

[17] Fan, J., Han, X. and Gu, W. (2012). Estimating false discovery proportion under arbitrary covariance dependence. *J. Amer. Statist. Assoc.* **107** 1019–1035. MR3010887

[18] Farcomeni, A. (2006). More powerful control of the false discovery rate under dependence. *Stat. Methods Appl.* **15** 43–73. MR2281214

[19] Farcomeni, A. (2007). Some results on the control of the false discovery rate under dependence. *Scand. J. Stat.* **34** 275–297. MR2346640

[20] Friguet, C., Kloareg, M. and Causeur, D. (2009). A factor model approach to multiple testing under dependence. *J. Amer. Statist. Assoc.* **104** 1406–1415. MR2750571

[21] Genovese, C. and Wasserman, L. (2004). A stochastic process approach to false discovery control. *Ann. Statist.* **32** 1035–1061. MR2065197

[22] Hall, P. and Heyde, C.C. (1980). *Martingale Limit Theory and Its Application*. *Probability and Mathematical Statistics*. New York–London: Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers]. MR0624435

[23] Jacod, J. and Shiryaev, A.N. (2003). *Limit Theorems for Stochastic Processes*, 2nd ed. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **288**. Berlin: Springer. MR1943877

[24] Johnstone, I.M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29** 295–327. MR1863961

[25] Kim, K.I. and van de Wiel, M.A. (2008). Effects of dependence in high-dimensional multiple testing problems. *BMC Bioinformatics* **9** 114.

[26] Neuvial, P. (2008). Asymptotic properties of false discovery rate controlling procedures under independence. *Electron. J. Stat.* **2** 1065–1110. MR2460858

[27] Roquain, E. and Villers, F. (2011). Exact calculations for false discovery proportion with application to least favorable configurations. *Ann. Statist.* **39** 584–612. MR2797857

[28] Rozanov, Yu.A. (1967). *Stationary Random Processes*. San Francisco, CA: Holden-Day, Inc. Translated from the Russian by A. Feinstein. MR0214134

[29] Schwartzman, A. and Lin, X. (2011). The effect of correlation in false discovery rate estimation. *Biometrika* **98** 199–214. MR2804220

[30] Soulier, P. (2001). Moment bounds and central limit theorem for functions of Gaussian vectors. *Statist. Probab. Lett.* **54** 193–203. MR1858634

[31] Storey, J.D. (2003). The positive false discovery rate: A Bayesian interpretation and the q-value. *Ann. Statist.* **31** 2013–2035. MR2036398

[32] Sun, Y., Zhang, N.R. and Owen, A.B. (2012). Multiple hypothesis testing adjusted for latent variables, with an application to the AGEMAP gene expression data. *Ann. Appl. Stat.* **6** 1664–1688. MR3058679

[33] van der Vaart, A.W. (1998). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge: Cambridge Univ. Press. MR1652247

Received May 2013 and revised October 2013