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

Sylvain Delattre, Etienne Roquain

▶ To cite this version:

Sylvain Delattre, Etienne Roquain. On empirical distribution function of high-dimensional Gaussian vector components with an application to multiple testing. 2013. <hal-00739749v2>

HAL Id: hal-00739749

https://hal.archives-ouvertes.fr/hal-00739749v2

Submitted on 3 May 2013 (v2), last revised 18 Jan 2016 (v3)

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

Sylvain Delattre

Université Paris Diderot, LPMA,
e-mail: sylvain.delattre@univ-paris-diderot.fr

and

Etienne Roquain

UPMC Université Paris 6, LPMA,
e-mail: etienne.roquain@upmc.fr

Abstract:

This paper introduces a new framework to study the asymptotical behavior of the empirical distribution function (e.d.f.) of Gaussian vector components, whose correlation matrix $\Gamma^{(m)}$ is dimension-dependent. Hence, by contrast with the existing literature, the vector is not assumed to be stationary. Rather, we make a "vanishing second order" assumption ensuring that 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 $\gamma_m = m^{-2} \sum_{i \neq j} \Gamma_{i,j}^{(m)}$, 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 next 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.

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 (1995). The methodology of Neuvial (2008) 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 (1998)), this rises the problem of obtaining functional central limit theorems for e.d.f. in a setting which is suitable for high-dimensional data.

A colossal number of work aimed at extending Donsker's theorem (Doob, 1949; Donsker, 1952; Dudley, 1966) 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, e.g., Dedecker and Prieur (2007); Doukhan et al. (2010) for detailed 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 (1989) (long-range) and Csörgő and Mielniczuk (1996) (short-range). Both studies make a *stationarity* assumption: the covariance matrix between the variables is assumed to be of the form

$$\Gamma_{i,j}^{(m)} = r(|i-j|),\ 1 \leq i,j \leq m,$$

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 et al., 2009; Sun et al., 2012; Fan et al., 2012 and references therein), which leads to "spiked" correlation matrices (as refered to by Johnstone, 2001). 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 $\{Y^{(m)}, m \geq 1\}$ a triangular array for which each $Y^{(m)} = \left(Y_1^{(m)}, \ldots, Y_m^{(m)}\right)$ 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 $Y_i^{(m)}$ is of variance 1, that is, $\Gamma_{i,i}^{(m)} = 1$ for all i. Denote $\Phi(z) = \mathbb{P}(Z \geq z)$, 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:

$$\widehat{\mathbb{F}}_m(t) = m^{-1} \sum_{i=1}^{m} \mathbf{1}\left\{\Phi(Y_i^{(m)}) \leq t\right\}, \quad t \in [0,1]. \tag{1}$$

Here, we consider the e.d.f. of the $\Phi(Y_i^{(m)})$'s rather than the one of the $Y_i^{(m)}$'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:

$$\gamma_m = m^{-2} \sum_{i \neq j} \Gamma_{i,j}^{(m)}; \tag{2}$$

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

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

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

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

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

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

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

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 $(t, s) \mapsto t \wedge s - ts + \theta \, \phi(\Phi^{-1}(t))\phi(\Phi^{-1}(s))$. 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 $\phi(\Phi^{-1}(\cdot))Z$ 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 $\phi(\Phi^{-1}(\cdot))$ 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, athough 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, 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:

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

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

Roughly speaking, it shows that, up to add some "safety margin" ε_0 in the convergence, Assumption (H_1) can be removed in regime (ii).
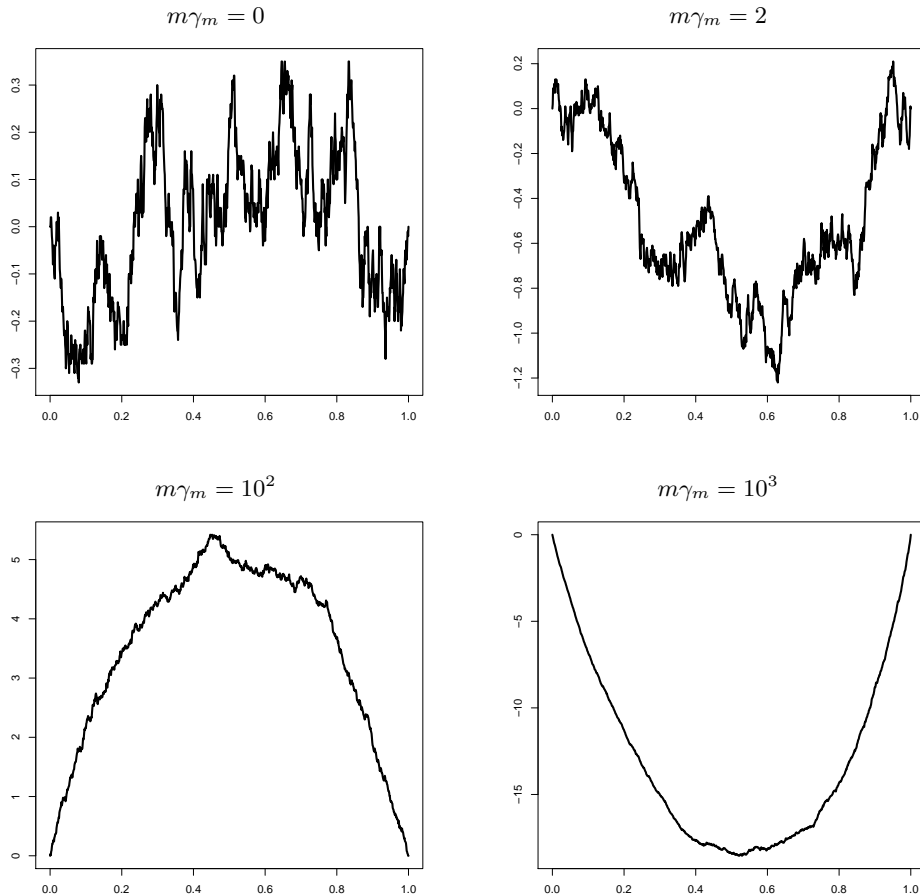
FIG 1. *Plot of $t \mapsto m^{1/2}(\widehat{\mathbb{F}}_m(t) - t)$ for some observed $Y(\omega)$. These realizations have been generated in the equi-correlated model $\Gamma_{i,j}^{(m)} = \rho_m$, $i \neq j$ (see (14)) and for $m = 10^4$.*

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 (1989) but excludes short range of Csörgő and Mielniczuk (1996). 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. In particular, it covers the result of Delattre

and Roquain (2011), obtained in the 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).

Finally, let us mention the interesting work of Bardet and Surgailis (2011) in which the stationarity assumption has also been removed, by establishing central limit theorems (CLT) 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 $|\Gamma_{i,j}^{(m)}| \leq r(|i-j|)$ for all i, j, for some function $r(\cdot)$, independent of m, and vanishing at infinity.

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 appendices.

2. Preliminaries: 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 $H_\ell(x)$, $\ell \geq 0$, $x \in \mathbb{R}$ (see Appendix B). 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*

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

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

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

This result can be found, e.g., in Theorem 2 of Schwartzman and Lin (2011) (see also Theorem 1 of Efron, 2010). We provide a proof in Appendix B for completeness. While (6) is an exact expression, we can try to approximate the covariance $\text{Cov}(\widehat{\mathbb{F}}_m(t), \widehat{\mathbb{F}}_m(s))$ when m grows to infinity, while making some assumption on the matrix $\Gamma = \Gamma^{(m)}$.

Firstly, let us note the following: since $m^{-2} \sum_{i,j} (\Gamma_{i,j})^\ell = (\ell!)^{-1} \text{Var}\left(m^{-1} \sum_{i=1}^m H_\ell(Y_i)\right) \geq 0$ (by using (70) in Appendix B), expression (6) shows that the following conditions are equivalent as m tends to infinity,

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

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

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

As a consequence, Condition (LLN-dep) is required as soon as a convergence result of the form (4) holds. Note that the rate r_m defined by (3) satisfies $1 \leq \left(m^{-2} \sum_{i,j} (\Gamma_{i,j})^2\right)^{-1/4} \leq r_m \leq \sqrt{m}$. Hence r_m tends to infinity under (LLN-dep) but not faster than $\sqrt{m}$.

Secondly, let us rewrite (6) as follows:

$$\text{Cov}(\widehat{\mathbb{F}}_m(t), \widehat{\mathbb{F}}_m(s)) = m^{-1}(t \wedge s - ts) + \gamma_m c_1(t)c_1(s)$$

$$+ \sum_{\ell \geq 2} \left(m^{-2} \sum_{i \neq j} (\Gamma_{i,j})^\ell\right) c_\ell(t)c_\ell(s)(\ell!)^{-1}, \tag{9}$$

where γ_m is defined by (2). The latter holds because, for two independent $\mathcal{N}(0,1)$ variables U and V, we have $m^{-1} \sum_{\ell \geq 1} c_\ell(t)c_\ell(s)(\ell!)^{-1} = \text{Cov}(\mathbf{1}\{\Phi(U) \leq t\}, \mathbf{1}\{\Phi(V) \leq s\})$. 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 $\text{Cov}(r_m\widehat{\mathbb{F}}_m(t), r_m\widehat{\mathbb{F}}_m(s))$ 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 nonnegativeness 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 $W_t - tW_1 - c_1(t) \int_0^1 \Phi^{-1}(s)dW_s$, which is the orthogonal projection in L^2 of W_t onto the orthogonal of the linear space spanned by W_1 and $\int_0^1 \Phi^{-1}(s)dW_s$. Interestingly, the latter also shows that the original covariance K given by (10) can be seen as the covariance function of $\mathbb{Z}_t = (1 + |\theta|)^{-1/2}(W_t - tW_1) + (1 + |\theta|)^{-1/2}((1 + \theta)^{1/2} - 1)c_1(t) \int_0^1 \Phi^{-1}(s)dW_s$.

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 $(\mathbb{Z}_t)_{t \in [0,1]}$ a continuous process*

with covariance function K defined by (10). *Then we have the convergence (in the Skorokhod topology)*

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

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 (commented) examples.

3.2. Examples

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

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

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:

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

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 (2011).

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:

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

$$\tag{15}$$

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

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

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:

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

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 (1989); Doukhan et al. (2002)). First, standard calculations easily show that for all $\nu \geq 0$,

$$m^{-1} \sum_{i \neq j} |j-i|^{-\nu} = 2m^{-1} \sum_{i=1}^{m-1} \sum_{k=1}^{i} k^{-\nu} \sim \begin{cases} 2\frac{m^{1-\nu}}{(1-\nu)(2-\nu)} & \text{if } \nu \in [0,1) \\ 2\log m & \text{if } \nu = 1 \\ 2\sum_{k \geq 1} k^{-\nu} & \text{if } \nu > 1 \end{cases}. \tag{17}$$

Thus, for any $\nu_1 \in (D, 1)$, since L is slowly varying,

$$m\gamma_m \gtrsim m^{-1} \sum_{i \neq j} |j-i|^{-\nu_1} \gtrsim m^{1-\nu_1},$$

by applying (17), where the "$u_m \lesssim v_m$" means $u_m = O(v_m)$. This entails $m\gamma_m^{1+(1-\nu_1)/(2\nu_1)} \gtrsim m^{(1-\nu_1)/2}$ and thus Assumption (H_4) holds. In particular, $r_m \sim \gamma_m^{-1/2}$. Additionally, for any $\nu_2 \in (D, 1)$ and $\nu_3 \in (0, 2D)$ such that $\nu_3/\nu_2 > 1$, by applying again (17),

$$\gamma_m^{-\delta} m^{-2} \sum_{i \neq j} (\Gamma_{i,j})^2 \lesssim m^{\nu_1 \delta} m^{-2} \sum_{i \neq j} |j-i|^{-\nu_3} \lesssim m^{\nu_2 \delta - \nu_3} \vee (m^{\nu_2 \delta - 1} \log m)$$

for any $\delta > 0$. We derive (H_3) because taking $\delta > 1$ such that $\delta < \nu_3/\nu_2$ and $\delta < 1/\nu_2$ is possible. By using Theorem 3.1 under (13), we derive

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

for $Z \sim \mathcal{N}(0,1)$. This is in accordance with Theorem 1.1 of Dehling and Taqqu (1989) (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 (1996)).

Weak short/long range correlations Let us modify slightly the matrix (16), by letting:

$$\Gamma_{i,j}^{(m)} = \rho_m \, r(|j - i|), \quad \text{for } r(0) = 1, \ r(k) = k^{-D}, \ D > 0, \tag{18}$$

where ρ_m is some nonnegative 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. From (17), we have

$$m\gamma_m \sim \begin{cases} 2\rho_m \frac{m^{1-D}}{(1-D)(2-D)} & \text{if } D \in [0,1) \\ 2\rho_m \log m & \text{if } D = 1 \\ 2\rho_m \sum_{k \geq 1} k^{-D} & \text{if } D > 1 \end{cases} \tag{19}$$

Assuming that the quantity (19) as a limit (denoted θ) and that $\rho_m \to 0$ as m grows to infinity, (vanish-secondorder) and (H_1) hold if $m^{1+\delta}\rho_m^2 \to 0$ with $\delta > 0$ (because (13) holds). The 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.

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

TABLE 1
Rate r_m defined by (3) in function of $D \geq 0$ and ρ_m such that $\rho_m = o(m^{-(1/2+\delta)})$ for some $\delta > 0$, for the particular covariance (18).

Vanishing factor model "Spiked" covariance matrix has been introduced in Johnstone (2001). 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:

$$\Gamma^{(m)} = (1 - \rho_m)I_m + \rho_m PHP^T, \tag{20}$$

where H is a $k \times k$ diagonal matrix with diagonal entries $h_1^{(m)}, \ldots, h_k^{(m)} \in (1, \infty)$, 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 $1 - \rho_m + \rho_m h_r^{(m)}$, $r = 1, \ldots, k$, while the remaining eigenvalues are all equal to $1 - \rho_m$. Hence, to ensure that $\Gamma^{(m)}$ given by (20) is a well defined correlation matrix, we should additionally assume that for all $r = 1, \ldots, k$, $1 - \rho_m + \rho_m h_r^{(m)} \geq 0$, and that PHP^T

has diagonal entries equal to 1, that is, for all $i = 1, \ldots, m$, $\sum_{r=1}^{k} h_r^{(m)} (p_{i,r}^{(m)})^2 = 1$. Note that the latter requires $h_1^{(m)} + \cdots + h_k^{(m)} = m$ and thus $\max_r \{h_r^{(m)}\} \geq m/k$.

Next, by using (20), the conditions above and some properties of the Frobenius norm, we can derive the following:

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

$$m^{-2} \sum_{i \neq j} \left(\Gamma_{i,j}^{(m)} \right)^2 = \rho_m^2 \left(\sum_{r=1}^{k} (h_r^{(m)}/m)^2 - 1/m \right). \tag{22}$$

Since the RHS of (22) is upper-bounded by $\rho_m^2(k - m^{-1})$ and lower-bounded by $\rho_m^2(k^{-2} - m^{-1})$ and since k is taken fixed with m, condition (LLN-dep) is satisfied if and only if $\rho_m \to 0$ while (vanish-secondorder) holds if and only if $r_m \rho_m \to 0$. Additionally, we have

$$m^{-2} \sum_{i \neq j} \left(\Gamma_{i,j}^{(m)} \right)^4 = \rho_m^4 \left(m^{-2} \sum_{r_1, \ldots, r_4} h_{r_1} \ldots h_{r_4} \left(\sum_{i=1}^{m} p_{i,r_1} \ldots p_{i,r_4} \right)^2 - 1/m \right)$$

$$\leq \rho_m^4 \left(m^{-2} \left(\sum_{r_1, r_2} h_{r_1} h_{r_2} \sum_{i=1}^{m} p_{i,r_1}^2 p_{i,r_2}^2 \right)^2 - 1/m \right)$$

$$= \rho_m^4 (1 - 1/m),$$

where we used the Cauchy-Schwartz inequality (we dropped the dependence in m in the notation for short). Finally, the assumption of Theorem 3.1 are all fulfilled provided that

$$r_m^{2+\delta} \rho_m^2 \to 0 \text{ with } \delta > 0 \tag{23}$$

(up to consider a subsequence making the quantity into (21) converges to some θ). In (23), the rate r_m can be computed by using the definition, see (3), or expression (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

$$\Gamma^{(m)} = (1 - \rho_m) I_m + \rho_m \xi \xi^T, \tag{24}$$

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 (24) 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 (23) 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

$$\rho_m \left(m^{-1/2} \sum_{i=1}^{m} \xi_i^{(m)} \right)^2 - \rho_m, \tag{25}$$

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. random signs, we have by the central limit theorem that the quantity (25) 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

$$\Gamma^{(m)} = D^{-1}SD^{-1}, \text{ for } S = n_m^{-1}X^TX \text{ and } D = \text{diag}(S_{1,1}, \cdots, S_{m,m})^{1/2} \qquad (26)$$

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 (1993) (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, $||S - I_m||_2 = \max_{1 \le i \le m} |\lambda_i^{(m)} - 1|$ and $\lambda_1^{(m)}, \ldots, \lambda_m^{(m)}$ denote the eigenvalues of S. Hence, $\max_{1 \le i \le m} |S_{i,i} - 1| \xrightarrow{P} 0$, which in turn implies $||\Gamma^{(m)} - I_m||_2 \xrightarrow{P} 0$. Next, simple arguments entail the following inequalities:

$$\left| m^{-1} \sum_{i \ne j} \Gamma_{i,j}^{(m)} \right| = m^{-1}| < (1 \cdots 1)^T, (\Gamma^{(m)} - I_m)(1 \cdots 1)^T > | \le ||\Gamma^{(m)} - I_m||_2;$$

$$r_m^2 m^{-2} \sum_{i \ne j} \left(\Gamma_{i,j}^{(m)} \right)^2 \le m^{-1} \sum_{i=1}^m (\lambda_i^{(m)} - 1)^2 \le ||\Gamma^{(m)} - I_m||_2^2;$$

$$r_m^{4+\varepsilon_0} m^{-2} \sum_{i \ne j} \left(\Gamma_{i,j}^{(m)} \right)^4 \le \left\{ \min_{1 \le i \le n_m} |S_{i,i}| \right\}^{-4} m^{\varepsilon_0/2} \sum_{i \ne j} (S_{i,j})^4.$$

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

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

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. The curse of dependence

The so-called "Benjamini and Hochberg procedure" (BH procedure), widely popularized after the celebrated paper Benjamini and Hochberg (1995), 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 α under independence of the tests (and also for some type of positive dependence, see Benjamini and Yekutieli (2001)). Furthermore, many authors reported that the FDR of the BH procedure is essentially unaffected by dependencies, see,

e.g., Farcomeni (2006); Kim and van de Wiel (2008). It is therefore tempting to conclude that the BH procedure works whatever the dependencies are. However, as noticed by Lehmann and Romano: "*control of the FDR does not prohibit the FDP from varying, even if its average value is bounded*", see Lehmann and Romano (2005). In addition, some authors have exhibited that the distribution of the FDP of BH can be wide spread in a particular (unrealistic) equi-correlated framework, by using simulations, see, e.g., Table 2 in Korn et al. (2004) and by using a theoretical study, see Delattre and Roquain (2011). The present work brings a broad theoretical support for this, by showing that the distribution of the FDP of BH procedure is widening 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 (2004); Farcomeni (2007) (FDP at a fixed threshold) and consolidated later in Neuvial (2008) (FDP at BH threshold). Here, we follow the approach of Neuvial (2008), 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 C.2.

4.2. Two-group model, FDP and BH procedure

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

$$X_i = \delta H_i + Y_i, \tag{27}$$

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, generally assumed to be unknown.

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

$$\widehat{\mathbb{F}}_{0,m}(t) = m_0^{-1} \sum_{i=1}^{m} (1 - H_i) \mathbf{1}\{\Phi(X_i) \le t\}; \tag{28}$$

$$\widehat{\mathbb{F}}_{1,m}(t) = m_1^{-1} \sum_{i=1}^{m} H_i \mathbf{1}\{\Phi(X_i) \le t\}; \tag{29}$$

$$\widehat{\mathbb{G}}_m(t) = m^{-1} \sum_{i=1}^{m} \mathbf{1}\{\Phi(X_i) \le t\} = \frac{m_0}{m}\widehat{\mathbb{F}}_{0,m}(t) + \frac{m_1}{m}\widehat{\mathbb{F}}_{1,m}(t), \tag{30}$$

where $m_0 = \sum_{i=1}^{m}(1 - H_i)$ and $m_1 = \sum_{i=1}^{m} H_i$. The proportions m_0/m and m_1/m are supposed to converge when m grows to infinity and the limits are denoted by $\pi_0 \in (0, 1)$ and $\pi_1 \in (0, 1)$, respectively. From Section 2, when Γ satisfies (LLN-dep), the e.c.d.f.'s $\widehat{\mathbb{F}}_{0,m}(t)$, $\widehat{\mathbb{F}}_{1,m}(t)$ and $\widehat{\mathbb{G}}_m(t)$ converge in probability and we denote in what follows the limiting c.d.f.'s by $F_0(t) = t$, $F_1(t) = \Phi(\Phi^{-1}(t) - \delta)$ and $G(t) = \pi_0 F_0(t) + \pi_1 F_1(t)$, respectively.

Here, the quality of a procedure that rejects each null hypothesis "$H_i = 0$" whenever $\Phi(X_i) \le t$ is given by

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

where we used the convention $0/0 = 0$. Now, define the following functional: for $\alpha \in (0, 1)$,

$$\mathcal{T}(H) = \sup\{t \in [0, 1] \ : \ H(t) \geq t/\alpha\} \text{ for } H \in D(0, 1),$$

with the convention $\sup\{\emptyset\} = 0$. Classically, the BH procedure (at level α) corresponds the thresholding $\mathcal{T}(\widehat{\mathbb{G}}_m)$, see Genovese and Wasserman (2004). In the sequel, we study the asymptotic behavior of $\mathrm{FDP}_m(\mathcal{T}(\widehat{\mathbb{G}}_m))$, denoted by FDP_m for short.

4.3. Partial functional delta method

Since we have $\widehat{\mathbb{G}}_m(\mathcal{T}(\widehat{\mathbb{G}}_m)) = \mathcal{T}(\widehat{\mathbb{G}}_m)/\alpha$ a.s., the FDP of BH procedure corresponds to the random variable

$$\mathrm{FDP}_m = \alpha \frac{\frac{m_0}{m}\widehat{\mathbb{F}}_{0,m}(\mathcal{T}(\widehat{\mathbb{G}}_m))}{\mathcal{T}(\widehat{\mathbb{G}}_m)} = \Psi\left(\frac{m_0}{m}\widehat{\mathbb{F}}_{0,m}, \frac{m_1}{m}\widehat{\mathbb{F}}_{1,m}\right), \tag{31}$$

where we used the following functional:

$$\Psi(H_0, H_1) = \alpha \frac{H_0(\mathcal{T}(H_0 + H_1))}{\mathcal{T}(H_0 + H_1)}, \text{ for } (H_0, H_1) \in D(0, 1)^2, \tag{32}$$

still using the conventions $\sup\{\emptyset\} = 0$ and $0/0 = 0$. By Corollary 7.12 in Neuvial (2008), $\mathcal{T}$ is Hadamard differentiable at function G, tangentially to the set $C(0, 1)$ of continuous functions on $(0, 1)$ and w.r.t. the supremum norm (we refer to Section 20.2 in van der Vaart (1998) for a formal definition of Hadamard differentiable functions). This holds because G is strictly concave and $\lim_{t\to 0} G(t)/t = +\infty$, which yields in particular $\mathcal{T}(G) \in (0, 1)$. As a consequence, standard calculations show that Ψ is Hadamard differentiable at $(\pi_0 F_0, \pi_1 F_1)$ tangentially to $C(0, 1)$, with derivative

$$\dot{\Psi}_{(\pi_0 F_0, \pi_1 F_1)}(H_0, H_1) = \alpha \frac{H_0(\mathcal{T}(G))}{\mathcal{T}(G)}, \text{ for } (H_0, H_1) \in C(0, 1)^2. \tag{33}$$

Now, by using (31), the functional delta method provides the asymptotic behavior of FDP_m from the one of $(\frac{m_0}{m}\widehat{\mathbb{F}}_{0,m}, \frac{m_1}{m}\widehat{\mathbb{F}}_{1,m})$. As a matter of fact, since the derivative $\dot{\Psi}_{(\pi_0 F_0, \pi_1 F_1)}(H_0, H_1)$ only depends on H_0 while the limit processes are (a.s.) continuous, establishing convergence results separately for $\widehat{\mathbb{F}}_{0,m}$ and $\widehat{\mathbb{F}}_{1,m}$ is sufficient (we do not need to consider the joint process $(\frac{m_0}{m}\widehat{\mathbb{F}}_{0,m}, \frac{m_1}{m}\widehat{\mathbb{F}}_{1,m})$). We have precisely formulated this argument in Proposition C.2. This is an interesting novelty w.r.t. the methodology of Neuvial (2008). Hence, applying (twice) Theorem 3.1 we are able to derive a convergence result for FDP_m.

4.4. Results

First, let us introduce the following additional quantities:

$$r_{0,m} = \left(m_0^{-1} + \left|m_0^{-2}\sum_{i\neq j}(1 - H_i)(1 - H_j)\Gamma_{i,j}\right|\right)^{-1/2}; \tag{34}$$

$$r_{1,m} = \left(m_1^{-1} + \left|m_1^{-2}\sum_{i\neq j}H_i H_j \Gamma_{i,j}\right|\right)^{-1/2}. \tag{35}$$

Corollary 4.1. *Consider the two-group model* (27), *generated from parameters* δ, $H = H^{(m)}$ *and a correlation matrix* $\Gamma = \Gamma^{(m)}$. *Assume that* m_0 *(depending on H) is such that* $\sqrt{m}(m_0/m - \pi_0) \to 0$. *Assume that* Γ *satisfies either* {(vanish-secondorder) *and* (H_1)} *or* {(H_3) *and* (H_4)}. *Assume that the rates* r_m, $r_{0,m}$ *and* $r_{1,m}$, *respectively defined by* (3), (34) *and* (35), *grow proportionally to infinity as m tends to infinity. 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 $h(t^\star) = (\phi(\Phi^{-1}(t^\star))/t^\star)^2$. Then the sequence of r.v.* FDP_m *defined by* (31) *enjoys the following convergence:*

$$\frac{FDP_m - \pi_0\alpha}{\pi_0\alpha\left\{(1/t^\star - 1)/m_0 + h(t^\star)\gamma_{0,m}\right\}^{1/2}} \rightsquigarrow \mathcal{N}(0,1), \tag{36}$$

where $\gamma_{0,m} = m_0^{-2}\sum_{i\neq j}(1-H_i)(1-H_j)\Gamma_{i,j}$.

Proof. First, classically, it is sufficient to prove that the convergence (36) holds up to consider a subsequence. Hence, we can assume that (H_2) and the convergences

$$m_0^{-1}\sum_{i\neq j}(1-H_i)(1-H_j)\Gamma_{i,j} \to \theta_0; \tag{37}$$

$$m_1^{-1}\sum_{i\neq j}H_iH_j\Gamma_{i,j} \to \theta_1;$$

hold, with θ, θ_0 and θ_1 valued in $[-1, +\infty]$. Also note that since $r_m \propto r_{0,m}$ (resp. $r_m \propto r_{1,m}$), the sub-matrices $(\Gamma_{i,j})_{i,j:H_i=H_j=0}$ and $(\Gamma_{i,j})_{i,j:H_i=H_j=1}$ satisfies the same assumption set as Γ. Now, let us write

$$r_{0,m}\left(\frac{m_0}{m}\widehat{\mathbb{F}}_{0,m}(t) - \pi_0 F_0(t)\right) = r_{0,m}m^{-1}\sum_{i=1}^m(1-H_i)(\mathbf{1}\left\{\Phi(X_i) \leq t\right\} - t) + r_{0,m}t(m_0/m - \pi_0). \tag{38}$$

In the RHS of (38), while the second term converges to 0 by assumption, a consequence of Theorem 3.1 is that the first term converges to a process with covariance function

$$\pi_0^2\left[\frac{1}{1+|\theta_0|}(t \wedge s - ts) + \frac{\theta_0}{1+|\theta_0|}c_1(t)c_1(s)\right], \quad \text{for all } t, s \in [0,1].$$

Obviously, a similar result holds for the process $r_{1,m}(\frac{m_1}{m}\widehat{\mathbb{F}}_{1,m} - \pi_1 F_1)$.

Applying the (partial) functional delta method as explained in Proposition C.2 (by using $r_{0,m} \propto r_{1,m}$ and (33)), we obtain

$$r_{0,m}(\text{FDP}_m - \pi_0\alpha) \rightsquigarrow \mathcal{N}\left(0, (\alpha\pi_0)^2\left[\frac{1/t^\star - 1}{1+|\theta_0|} + \frac{\theta_0}{1+|\theta_0|}(c_1(t^\star)/t^\star)^2\right]\right). \tag{39}$$

Finally, we easily derive (36) by separating the cases $\theta_0 < +\infty$ and $\theta_0 = +\infty$. $\square$

As an illustration, Corollary 4.1 can be used in the independent case ($\gamma_{0,m} = 0$) or ρ_m-equi-correlated case ($\gamma_{0,m} = \rho_m$), so recovering the previous results of Neuvial (2008, 2009) (in the Gaussian case) and Delattre and Roquain (2011), respectively. This holds for any H satisfying $\sqrt{m}(m_0/m - \pi_0) \to 0$. Note that, in general, the quantity $\gamma_{0,m}$ depends on the unknown H and not only on Γ. Hence, the asymptotic properties of FDP_m potentially 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 BH's 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 *prior random effects*, by assuming that, previously and independently to the model (27), 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, e.g. Efron et al. (2001); Storey (2003); Genovese and Wasserman (2004); Roquain and Villers (2011). By contrast with the previous model, H is random. In particular, $m_0 = \sum_{i=1}^m (1 - H_i) \sim \mathcal{B}(m, \pi_0)$ and $\sqrt{m}(m_0/m - \pi_0)$ does not degenerate at the limit, which adds some extra variance in the FDP convergence result. The counterpart is that the statement is substantially simplified, as we can see below.

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 $h(t^\star) = (\phi(\Phi^{-1}(t^\star))/t^\star)^2$. Then the sequence of r.v. FDP_m defined by (31) enjoys the following convergence:*

$$\frac{FDP_m - \pi_0 \alpha}{\pi_0 \alpha \left\{ (1/t^\star - \pi_0)/(\pi_0 m) + h(t^\star)\gamma_m \right\}^{1/2}} \rightsquigarrow \mathcal{N}(0,1), \tag{40}$$

where γ_m is defined by (2).

Proof. Again, it is sufficient to state the result up to consider a subsequence. Thus (H_2) holds without loss of generality. First check that (vanish-secondorder) entails

$$\frac{r_m^2}{m^2} \left(\sum_{i \neq j} (1 - H_i)(1 - H_j)\Gamma_{i,j} - \pi_0^2 \sum_{i \neq j} \Gamma_{i,j} \right) = o_P(1), \tag{41}$$

(computing, e.g., the variance of the latter) and this convergence can be made a.s. by taking a suitable subsequence. A consequence of (41) is that $\gamma_{0,m} \sim \gamma_m$ a.s. (in particular, θ_0 defined by (37) equals $\pi_0\theta$.) This implies $r_m \propto r_{0,m}$ (a.s.) and thus the adequate assumption set for the sub-matrices $(\Gamma_{i,j})_{i,j:H_i=H_j=0}$ and $(\Gamma_{i,j})_{i,j:H_i=H_j=1}$. Now, by using (38), we obtain that $r_{0,m} \left(\frac{m_0}{m} \widehat{\mathbb{F}}_{0,m}(t) - \pi_0 F_0(t) \right)$ converges (unconditionally) to a process with covariance function defined by: for all $t, s \in [0,1]$,

$$\pi_0^2 \left[\frac{1}{1 + \pi_0|\theta|}(t \wedge s - ts) + \frac{\pi_0\theta}{1 + \pi_0|\theta|}c_1(t)c_1(s) \right] + \frac{\pi_0(1 - \pi_0)}{1/\pi_0 + |\theta|}ts$$
$$= \pi_0^2 \left[\frac{1}{1 + \pi_0|\theta|}(t \wedge s - \pi_0 ts) + \frac{\pi_0\theta}{1 + \pi_0|\theta|}c_1(t)c_1(s) \right]$$

Obviously, a similar result holds for the process $r_{1,m}(\frac{m_1}{m}\widehat{\mathbb{F}}_{1,m} - \pi_1 F_1)$. We finish the proof by applying the (partial) functional delta method, see Proposition C.2. $\square$
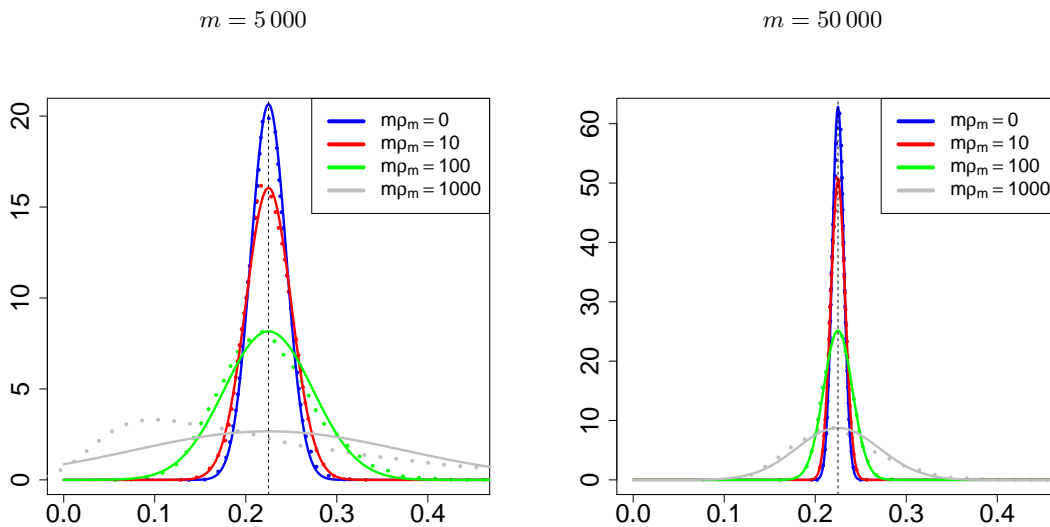
FIG 2. *The curse of dependence for BH procedure: distribution of FDP_m (31); the dotted lines corresponds to the true distribution computed over $5\,000$ simulations, the solid lines display the Gaussian approximation given by Corollary 4.2 (whose mean, $\pi_0\alpha = 0.225$, is displayed by the dashed vertical line). Simulations made in a 3-factor model, see text.*

4.5. Discussion

Corollary 4.2 provides a theoretical support for the "curse of dependence" of BH procedure: 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, perhaps surprisingly, negative correlations help to decrease γ_m and can yields to a concentration even better than under independence when γ_m is negative (although this phenomenon is necessary of limited amplitude because $\gamma_m \geq -1/m$).

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 (20) 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 $p_1 = (1, 1, \ldots, 1)/m^{1/2}$, $p_2 = (1, -1, 1, -1, \ldots, 1, -1)/m^{1/2}$, $p_3 = (1, 1, \ldots, 1, -1, -1, \ldots, -1)/m^{1/2}$. 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 $m = 5\,000$ (left picture), while the Gaussian approximation looks accurate for $m\rho_m \in \{0, 10, 100\}$, this seems more questionable when $m\rho_m = 1\,000$. 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 fact, additional experiments show that the approximation induced by Theorem 3.1 is still valid for $m = 5\,000$ and $m\rho_m = 1\,000$. As a consequence, we believe that the observed bias comes from the functional delta method, because the functional Ψ (32) cannot be considered as linear in that case. Finally, the right display in Figure 2 shows that, as one can expect, this phenomenon disappears by increasing the value of m.

This study reinforces the idea that the BH procedure should be used very carefully when there are dependencies between the individual tests. Following the work of Romano and Wolf (2007), an interesting task would be to correct the BH procedure by taking into account these

dependencies while still providing a valid control 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 $t \in [0, 1]$,

$$r_m(\widehat{\mathbb{F}}_m(t) - t) = r_m(\widetilde{\mathbb{F}}_m(t) - t) + c_1(t)r_m\overline{Y}_m. \tag{42}$$

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 (42). 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 (42) 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)*

$$r_m(\widetilde{\mathbb{F}}_m - I) \rightsquigarrow \widetilde{\mathbb{Z}}/(1 + |\theta|)^{1/2}, \text{ as } m \to \infty, \tag{43}$$

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

Additionally, throughout the section, we use the following notation

$$h_t(x) = \mathbf{1}\left\{\Phi(x) \leq t\right\} - t - c_1(t)x, \tag{44}$$

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:

there exists $\eta > 0$ (independent on m) lower bounding the m eigenvalues of $\Gamma^{(m)}$.

(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 (43) holds in the sense of the finite dimensional convergence.*

Proof. The proof is based on central limit theorems for martingale arrays as presented, e.g., in Chapter 3 of Hall and Heyde (1980).

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 $(Y_i^{(m)}, 1 \leq i \leq m, m \geq 1)$ 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 $m \geq 1$, and then by choosing for each $m \geq 2$, the variables $Y_i^{(m)}, 2 \leq i \leq m$, such that

- $(Z, Y_i^{(m)}, 2 \leq i \leq m) \sim \mathcal{N}(0, \Gamma^{(m)})$;
- $\{(Y_i^{(m)})_{2 \leq i \leq m}, m \geq 2\}$ 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 $m \geq 1$, $\mathcal{G}_{m,0} = \sigma(\emptyset)$ and for $1 \leq i \leq m$,

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

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

$$M_{m,i}(t) = \sum_{j=1}^{i} X_{m,j}(t) \quad \text{for } X_{m,j}(t) = \frac{r_m}{m} \left(h_t(Y_j^{(m)}) - \mathbb{E}\left(h_t(Y_j^{(m)}) \mid \mathcal{G}_{m,j-1} \right) \right). \tag{45}$$

Clearly,

$$r_m(\widetilde{\mathbb{F}}_m(t) - t) = M_{m,m}(t) + \frac{r_m}{m} \sum_{i=1}^{m} \mathbb{E}\left(h_t(Y_i^{(m)}) \mid \mathcal{G}_{m,i-1} \right). \tag{46}$$

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

Case 1: (eigenvalues-away0) **is assumed** We show in Lemma A.1 expression (66) that the second term in the RHS of (46) 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 (1980) (or more precisely its generalization to the multidimensional case). The conditions are as follows:

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

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

$$\sum_{i=1}^{m} \mathbb{E}(X_{m,i}(t))^4 = \frac{r_m^4}{m^4} \sum_{i=1}^{m} \mathbb{E}\left(h_t(Y_i^{(m)}) - \mathbb{E}\left(h_t(Y_i^{(m)}) \mid \mathcal{F}_{m,i-1} \right) \right)^4$$

$$\leq 2^4 \left(\frac{r_m^4}{m^3} m^{-1} \sum_{i=1}^{m} \mathbb{E}\left(h_t(Y_i^{(m)}) \right)^4 + \frac{r_m^4}{m^4} \sum_{i=1}^{m} \mathbb{E}\left(\mathbb{E}\left(h_t(Y_i^{(m)}) \mid \mathcal{F}_{m,i-1} \right) \right)^4 \right)$$

$$\leq 2^5 \frac{r_m^4}{m^3} m^{-1} \sum_{i=1}^{m} \mathbb{E}\left(h_t(Y_i^{(m)}) \right)^4.$$

Now, the RHS of the previous display converges to zero because $r_m \leq \sqrt{m}$ 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

$$\sum_{i=1}^{m} \mathbb{E}(X_{m,i}(t)X_{m,i}(s) \mid \mathcal{F}_{m,i-1}) = \frac{r_m^2}{m^2} \sum_{i=1}^{m} \mathbb{E}(h_t(Y_i^{(m)})h_s(Y_i^{(m)}) \mid \mathcal{F}_{m,i-1})$$

$$- \frac{r_m^2}{m^2} \sum_{i=1}^{m} \mathbb{E}(h_t(Y_i^{(m)}) \mid \mathcal{F}_{m,i-1})\mathbb{E}(h_s(Y_i^{(m)}) \mid \mathcal{F}_{m,i-1}). \quad (47)$$

Next, by using $ab \leq 2(a^2 + b^2)$ for all $a, b \in \mathbb{R}$ together with (65), the second term in the RHS of (47) tends to zero in probability. Moreover, we have

$$\text{Var}\left(\frac{r_m^2}{m^2} \sum_{i=1}^{m} \left((h_t(Y_i^{(m)})h_s(Y_i^{(m)}) - \mathbb{E}(h_t(Y_i^{(m)})h_s(Y_i^{(m)}) \mid \mathcal{F}_{m,i-1}))\right)\right)$$

$$= \frac{r_m^4}{m^4} \sum_{i=1}^{m} \text{Var}\left((h_t(Y_i^{(m)})h_s(Y_i^{(m)}) - \mathbb{E}(h_t(Y_i^{(m)})h_s(Y_i^{(m)}) \mid \mathcal{F}_{m,i-1}))\right),$$

because the elements inside the sum are martingale increments. Hence, the quantity inside the above display tends to zero. Combining the latter with (47) establishes condition (ii) of Lindeberg's theorem provided that the following holds:

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

This comes directly from the law of large number stated in Lemma C.3, 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 $t_1, \ldots, t_k \in [0, 1]$, the random vector

$$Z_m = (M_{m,m}(t_1), \ldots, M_{m,m}(t_k))$$

converges stably in the following sense (see, e.g., Jacod and Shiryaev (2003) Definition 5.28): for all (fixed) bounded random variable U and continuous bounded function f in $\mathbb{R}^k$,

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

where Z is a centered multivariate Gaussian vector with covariance $(1+|\theta|)^{-1}(\widetilde{K}(t_i, t_j))_{1 \leq i,j \leq k}$. 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 C.4 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 $\xi_1, \xi_2, \ldots$ are i.i.d. $\mathcal{N}(0, 1)$ variables, independent of all the Y_i's. The covariance matrix of $(Y_1^\varepsilon, \ldots, Y_m^\varepsilon)$ is obviously

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

Clearly, the corresponding rate (3) is $r_m^\varepsilon = \left(m^{-1} + (1+\varepsilon^2)^{-1}\,|\gamma_m|\right)^{-1/2}$. It is related to r_m via the following inequalities: $r_m \leq r_m^\varepsilon \leq (1+\varepsilon^2)^{1/2} r_m$. 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 $t_1, \ldots, t_k \in [0, 1]$,

(a) $\left(r_m^\varepsilon(\widetilde{\mathbb{F}}_m^\varepsilon(t_1) - t_1), \ldots, r_m^\varepsilon(\widetilde{\mathbb{F}}_m^\varepsilon(t_k) - t_k), Y_1^\varepsilon\right) \rightsquigarrow \mathcal{L}\left(\frac{(\widetilde{\mathbb{Z}}(t_1),\ldots,\widetilde{\mathbb{Z}}(t_k))}{(1+|\theta^\varepsilon|)^{1/2}}\right) \otimes \mathcal{N}(0,1),$

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

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

$$\limsup_m \left\{ \mathbb{E}\left| r_m(\widetilde{\mathbb{F}}_m(t) - t) - r_m^\varepsilon(\widetilde{\mathbb{F}}_m^\varepsilon(t) - t)\right| \right\} \to 0 \text{ as } \varepsilon \to 0. \tag{48}$$

This will conclude the proof by applying Lemma C.4. First, we write

$$\mathbb{E}\left| r_m(\widetilde{\mathbb{F}}_m(t) - t) - r_m^\varepsilon(\widetilde{\mathbb{F}}_m^\varepsilon(t) - t)\right|$$

$$\leq \mathbb{E}\left| r_m/m \sum_{i=1}^m (h_t(Y_i) - h_t(Y_i^\varepsilon))\right| + (r_m^\varepsilon - r_m)\mathbb{E}\left| m^{-1}\sum_{i=1}^m h_t(Y_i^\varepsilon)\right|$$

$$\leq \left\{ (r_m/m)^2\, \mathbb{E}\left(\sum_{i=1}^m (h_t(Y_i) - h_t(Y_i^\varepsilon))\right)^2 \right\}^{1/2} + ((1+\varepsilon^2)^{1/2} - 1)\mathbb{E}\left| r_m^\varepsilon(\widetilde{\mathbb{F}}_m^\varepsilon(t) - t)\right|.$$

By taking the $\limsup$ in the above display, it only remains to show

$$\limsup_m \left\{ (r_m/m)^2\, \mathbb{E}\left(\sum_{i=1}^m (h_t(Y_i) - h_t(Y_i^\varepsilon))\right)^2 \right\} \to 0 \text{ as } \varepsilon \to 0. \tag{49}$$

This can be proved by using Lemma B.3 (78) as follows:

$$(r_m/m)^2\, \mathbb{E}\left(\sum_{i=1}^m (h_t(Y_i) - h_t(Y_i^\varepsilon))\right)^2$$

$$= (r_m/m)^2 \sum_{i,j=1}^m \mathbb{E}\left((h_t(Y_i) - h_t(Y_i^\varepsilon))(h_t(Y_j) - h_t(Y_j^\varepsilon))\right)$$

$$= (r_m/m)^2 \sum_{i,j=1}^m \left(\mathbb{E}\left(h_t(Y_i)h_t(Y_j)\right) - \mathbb{E}\left(h_t(Y_i)h_t(Y_j^\varepsilon)\right) - \mathbb{E}\left(h_t(Y_i^\varepsilon)h_t(Y_j)\right) + \mathbb{E}\left(h_t(Y_i^\varepsilon)h_t(Y_j^\varepsilon)\right)\right)$$

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

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

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

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 (49) and finishes the proof. $\qquad\square$

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

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

Proposition 5.3. *Consider the assumptions of Proposition 5.2. Then, (12) holds in the sense of the finite dimensional convergence.*

Proof. From expression (42), 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 5.3 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 $(Y_i)_{0 \le i \le m}$, which is centered, with a covariance matrix denoted $\Lambda^{(m+1)} = (\Lambda_{i,j}^{(m+1)})_{0 \le i,j \le m}$ and such that $\Lambda_{0,0}^{(m+1)} = 1$, $\Lambda_{i,j}^{(m+1)} = \Gamma_{i,j}^{(m)}$ for $1 \le i,j \le m$. 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 A.2. Hence, Proposition 5.2 shows that (by using notation therein),

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

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

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

Finally, since $r_m^2 \operatorname{Var}(\overline{Y}_m) \to (1+\theta)/(1+|\theta|)$, we finish the proof by applying (42). $\qquad\square$

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 (42) 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 C.1. This is possible because $|c_1(t) - c_1(s)| \le L|t-s|^{1/2}$, $0 \le s, t \le 1$ for some constant $L > 1$ (see Lemma B.4). Below, we prove that (80) holds in the following way: for large m,

$$\mathbb{E}\big|X_m(t) - X_m(s)\big|^4 \le C\big(|t-s|^{3/2} + (r_m)^{-\varepsilon_0}|t-s|\big), \text{ for all } t, s \in [0,1], \qquad (50)$$

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

To establish (50), fix $t, s \in [0, 1]$, $s \leq t$ and write

$$\mathbb{E}\big|X_m(t) - X_m(s)\big|^4 = \frac{r_m^4}{m^4} \sum_{i,j,k,\ell} \mathbb{E}\big(\overline{h}(Y_i)\overline{h}(Y_j)\overline{h}(Y_k)\overline{h}(Y_\ell)\big), \tag{51}$$

where we let $\overline{h}(x) = \mathbf{1}\{s < \Phi(x) \leq t\} - (t - s) - (c_1(t) - c_1(s))x = h_t(x) - h_s(x)$. Now, we split the sum in the RHS of (51) following the value of the cardinal of $\{i, j, k, \ell\}$.

Sum over $\#\{i, j, k, \ell\} = 1$ The corresponding summation is $\frac{r_m^4}{m^4} \sum_{i=1}^m \mathbb{E}\big((\overline{h}(Y_i))^4\big)$. We have

$$\mathbb{E}\big((\overline{h}(Y_i))^4\big) \leq 3^4 \left(|t - s| + |t - s|^4 + \mathbb{E}(|Y_1|^4)L^4|t - s|^{4/2}\right) \leq C_1|t - s|,, \tag{52}$$

for $C_1 = 5\,3^4 L^4 > 0$. Since $r_m^2 \leq m$, we obtain

$$\frac{r_m^4}{m^4} \sum_{i=1}^m \mathbb{E}\big((\overline{h}(Y_i))^4\big) \leq \frac{C_1}{m}|t - s|. \tag{53}$$

Sum over $\#\{i, j, k, \ell\} = 2$ Up to a multiplicative constant, we should consider the sum

$$\frac{r_m^4}{m^4} \sum_{i \neq j} \mathbb{E}\big((\overline{h}(Y_i))^2(\overline{h}(Y_j))^2\big) = T_1^{(1)} + T_2^{(1)},$$

where, for an arbitrary $\eta_1 > 0$, $T_1^{(1)}$ and $T_2^{(1)}$ are defined by

$$T_1^{(1)} = \frac{r_m^4}{m^4} \sum_{i \neq j} \mathbf{1}\{|\Gamma_{i,j}| > \eta_1\}\mathbb{E}\big((\overline{h}(Y_i))^2(\overline{h}(Y_j))^2\big); \tag{54}$$

$$T_2^{(1)} = \frac{r_m^4}{m^4} \sum_{i \neq j} \mathbf{1}\{|\Gamma_{i,j}| \leq \eta_1\}\mathbb{E}\big((\overline{h}(Y_i))^2(\overline{h}(Y_j))^2\big). \tag{55}$$

On the one hand, by using (52),

$$T_1^{(1)} \leq \frac{r_m^4}{\eta_1^2 m^4} \sum_{i \neq j} |\Gamma_{i,j}|^2 \mathbb{E}\big((\overline{h}(Y_i))^2(\overline{h}(Y_j))^2\big) \leq \frac{C_1}{\eta_1^2 m} \left(\frac{r_m^2}{m^2} \sum_{i \neq j} |\Gamma_{i,j}|^2\right)|t - s|. \tag{56}$$

On the other hand, by using (74) in Proposition B.1 (with $g_1 = g_2 = (\overline{h})^2$ and $d = 2$), we obtain that for any $i \neq j$ such that $|\Gamma_{i,j}| \leq \eta_1$ (choosing $\eta_1 > 0$ such that $2\sqrt{3\eta_1} < 1$),

$$\mathbb{E}\big((\overline{h}(Y_i))^2(\overline{h}(Y_j))^2\big) \leq \frac{1}{(1 - 2\sqrt{3\eta_1})^2} \left(\mathbb{E}\big(|\overline{h}(Z)|^{8/3}\big)\right)^{3/2} \leq \frac{C_2}{(1 - 2\sqrt{3\eta_1})^2}|t - s|^{3/2},$$

for $C_2 = 3^4 L^4 \left(\mathbb{E}\big(|Z|^{8/3}\big)\right)^{3/2} \in (0, \infty)$. Hence, we get

$$T_2^{(1)} \leq \frac{C_2}{(1 - 2\sqrt{3\eta_1})^2}|t - s|^{3/2}. \tag{57}$$

Sum over $\#\{i,j,k,\ell\} = 3$ Up to a multiplicative constant, we should consider the sum

$$\frac{r_m^4}{m^4} \sum_{i,j,k \neq} \mathbb{E}\left(\overline{h}(Y_i)\overline{h}(Y_j)(\overline{h}(Y_k))^2\right) = T_1^{(2)} + T_2^{(2)},$$

where, for an arbitrary $\eta_2 > 0$, $T_1^{(2)}$ and $T_2^{(2)}$ are defined similarly to (54) and (55), by separating the case where $\max_{e_1 \neq e_2 \in \{i,j,k\}} |\Gamma_{e_1,e_2}|$ is above or below η_2.
  On the one hand, by using (52), we have

$$T_1^{(2)} \leq \frac{C_1 r_m^4}{m^4} \sum_{i,j,k \neq} \mathbf{1}\left\{\max_{e_1 \neq e_2 \in \{i,j,k\}} |\Gamma_{e_1,e_2}| > \eta_2\right\}|t - s|$$

$$\leq 3\frac{C_1/\eta_2^4}{m}\left(\frac{r_m^4}{m^2}\sum_{i \neq j} |\Gamma_{i,j}|^4\right)|t - s|. \tag{58}$$

On the other hand, by using (75) in Proposition B.1 (with $g_1 = g_2 = \overline{h}$, $g_3 = (\overline{h})^2$, $f_1 = f_2 = \mathbf{1}\{s < \Phi(\cdot) \leq t\}$, $d = 3$ and $d' = 2$), we obtain that for any distinct i, j, k such that $\max_{e_1 \neq e_2 \in \{i,j,k\}} |\Gamma_{e_1,e_2}| \leq \eta_2$ (choosing $\eta_2 > 0$ such that $3\sqrt{3\eta_2} < 1$),

$$\mathbb{E}\left(\overline{h}(Y_i)\overline{h}(Y_j)(\overline{h}(Y_k))^2\right) \leq \max_{e_1 \neq e_2 \in \{i,j,k\}} |\Gamma_{e_1,e_2}|^2 \frac{27^2}{(1 - 3\sqrt{3\eta_2})^3}|t - s|^{3/2} \times \sqrt{C_2}.$$

This yields

$$T_2^{(2)} \leq 3\frac{\sqrt{C_2}\, 27^2}{(1 - 3\sqrt{3\eta_2})^3}\left(\frac{r_m^2}{m^2}\sum_{i \neq j} |\Gamma_{i,j}|^2\right)|t - s|^{3/2}. \tag{59}$$

Sum over $\#\{i,j,k,\ell\} = 4$ The last sum to be considered is

$$\frac{r_m^4}{m^4} \sum_{i,j,k,\ell \neq} \mathbb{E}\left(\overline{h}(Y_i)\overline{h}(Y_j)\overline{h}(Y_k)\overline{h}(Y_\ell)\right) = T_1^{(3)} + T_2^{(3)},$$

where, for an arbitrary $\eta_3 > 0$, $T_1^{(3)}$ and $T_2^{(3)}$ are defined similarly to (54) and (55), by separating the case where $\max_{e_1 \neq e_2 \in \{i,j,k,\ell\}} |\Gamma_{e_1,e_2}|$ is above or below η_3. As before,

$$T_1^{(3)} \leq \frac{C_1 r_m^4}{m^4} \sum_{i,j,k,\ell \neq} \mathbf{1}\left\{\max_{e_1 \neq e_2 \in \{i,j,k,\ell\}} |\Gamma_{e_1,e_2}| > \eta_3\right\}|t - s|$$

$$\leq 6\frac{C_1}{\eta_3^4}\left(\frac{r_m^4}{m^2}\sum_{i \neq j} |\Gamma_{i,j}|^4\right)|t - s|. \tag{60}$$

Next, by using (75) in Proposition B.1 (with $g_i = \overline{h}$, $f_i = \mathbf{1}\{s < \Phi(\cdot) \leq t\}$ and $d' = d = 4$), we obtain that (choosing $\eta_3 > 0$ such that $4\sqrt{3\eta_3} < 1$),

$$T_2^{(3)} \leq \frac{r_m^4}{m^4} \sum_{i,j,k,\ell \neq} \max_{e_1 \neq e_2 \in \{i,j,k,\ell\}} |\Gamma_{e_1,e_2}|^4 \frac{48^4}{(1 - 4\sqrt{3\eta_3})^4}|t - s|^3$$

$$\leq 6\frac{48^4}{(1 - 4\sqrt{3\eta_3})^4}\left(\frac{r_m^4}{m^2}\sum_{i \neq j} |\Gamma_{i,j}|^4\right)|t - s|^3. \tag{61}$$

Finally, we obtain (50) by combining the bounds (53),(56),(57),(58),(59),(60),(61) and by using the assumptions (vanish-secondorder) and (H_1).

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 C.1. Namely, we prove that, for large m,

$$\mathbb{E}\big|X_m(t) - X_m(s)\big|^2 \leq C\gamma_m^{\delta_0}|t - s|, \text{ for all } t, s \in [0, 1], \tag{62}$$

for some constants $C > 0$, $\delta_0 > 0$. To prove (62), we write (by using the same notation as in the previous section)

$$
\begin{aligned}
\mathbb{E}\big|X_m(t) - X_m(s)\big|^2 =& \frac{r_m^2}{m^2} \sum_{i,j} \mathbb{E}\big(\overline{h}(Y_i)\overline{h}(Y_j)\big) \\
\leq& C_3 \frac{r_m^2}{m}|t - s| + C_3|t - s|\frac{r_m^2}{\eta^2 m^2} \sum_{i \neq j} |\Gamma_{i,j}|^2 \\
&+ \frac{r_m^2}{m^2} \sum_{i \neq j} \mathbf{1}\left\{|\Gamma_{i,j}| \leq \eta\right\}\mathbb{E}\big(\overline{h}(Y_i)\overline{h}(Y_j)\big)
\end{aligned}
$$

for some $\eta > 0$ and by letting $C_3 = 4\,3^2 L^2 > 0$. Applying now (75) in Proposition B.1 (with $g_i = \overline{h}$, $f_i = \mathbf{1}\{s < \Phi(\cdot) \leq t\}$ for $i = 1, 2$ and $d' = d = 2$), we obtain that (choosing $\eta > 0$ such that $2\sqrt{3\eta} < 1$),

$$\frac{r_m^2}{m^2} \sum_{i \neq j} \mathbf{1}\left\{|\Gamma_{i,j}| \leq \eta\right\}\mathbb{E}\big(\overline{h}(Y_i)\overline{h}(Y_j)\big) \leq \frac{(12)^2}{(1 - 2\sqrt{3\eta})^2}|t - s|^{3/2}\left(\frac{r_m^2}{m^2} \sum_{i \neq j} |\Gamma_{i,j}|^2\right).$$

Finally, since (H_3) and (H_4) provide $\frac{r_m^2}{m^2}\sum_{i \neq j}|\Gamma_{i,j}|^2 = O(\gamma_m^{\varepsilon_1})$ and $r_m^2/m = O(\gamma_m^{\varepsilon_2})$ for some $\varepsilon_1, \varepsilon_2 > 0$, the criterion (62) is proved with $\delta_0 = \varepsilon_1 \wedge \varepsilon_2$ and the proof is finished.

Appendix A: Technical results for proving the main theorem

Lemma A.1. *Assume that $\Gamma^{(m)}$ satisfies (vanish-secondorder) and (eigenvalues-away0). For $1 \leq i \leq m$, let us consider the filtration $\{\mathcal{F}_i\}_{0 \leq i \leq m}$ defined by $\mathcal{F}_0 = \sigma(\emptyset)$ and $\mathcal{F}_i = \sigma(Y_1, \ldots, Y_i)$, and denote $\sigma_i^2 = \text{Var}\left[\mathbb{E}\left(Y_i \mid \mathcal{F}_{i-1}\right)\right]$. Consider the function $h_t(\cdot)$ defined by (44), the Hermite polynomials $H_\ell(\cdot)$ defined by (69) and the coordinates $c_\ell(\cdot)$ defined by (5). Then*

the following holds:

$$\frac{r_m^2}{m^2} \sum_{i=1}^m \sigma_i^2 \to 0; \tag{63}$$

$$\frac{r_m^2}{m^2} \sum_{i,j} \left(\mathbb{E}\left[\mathbb{E}\left(Y_i \mid \mathcal{F}_{i-1} \right) \mathbb{E}\left(Y_j \mid \mathcal{F}_{j-1} \right) \right] \right)^2 \to 0; \tag{64}$$

$$\frac{r_m^2}{m^2} \sum_{i=1}^m \mathbb{E}\left[\left(\mathbb{E}\left(h_t(Y_i) \mid \mathcal{F}_{i-1} \right) \right)^2 \right] \to 0, \quad \textit{for any } t \in [0,1]; \tag{65}$$

$$\mathbb{E}\left[\left(\frac{r_m}{m} \sum_{i=1}^m \mathbb{E}\left(h_t(Y_i) \mid \mathcal{F}_{i-1} \right) \right)^2 \right] \to 0, \quad \textit{for any } t \in [0,1]. \tag{66}$$

Proof. By using Cholesky's decomposition, we can write $\Gamma = RR^T$ where R is $m \times m$ a lower triangular matrix. Hence, denoting by $R_{1,.}, \dots R_{m,.}$ the lines of R, we have $< R_{i,.}, R_{j,.} >= \Gamma_{i,j}$ for all i, j. Moreover, since we can write $Y_i = \sum_{j=1}^i R_{i,j} Z_j$ for some $Z_1, \dots, Z_m$ i.i.d. $\mathcal{N}(0,1)$, we have $R_{i,i}^2 = \text{Var}(Y_i \mid \mathcal{F}_{i-1}) = 1 - \sigma_i^2$ and $\sigma_i^2 = \sum_{j=1}^{i-1} R_{i,j}^2$ for all i.

Let us now prove (63). From (eigenvalues-away0), we have for all $x \in \mathbb{R}^m$, $||R^T x||^2 = x^T \Gamma x \geq \eta ||x||^2$. Hence for all $x \in \mathbb{R}^m$, $||Rx||^2 \geq \eta ||x||^2$. Thus, we have

$$\sum_{i<j} \Gamma_{i,j}^2 = \sum_{j=1}^m \sum_{i=1}^{j-1} \left([RR_{j.}^T]_i \right)^2 \geq \eta \sum_{j=1}^m \sum_{i=1}^{j-1} R_{j,i}^2 = \eta \sum_{i=1}^m \sigma_i^2,$$

which proves (63) by (vanish-secondorder). As for (64), we have for $i < j$,

$$\mathbb{E}\left[\mathbb{E}\left(Y_i \mid \mathcal{F}_{i-1} \right) \mathbb{E}\left(Y_j \mid \mathcal{F}_{j-1} \right) \right] = \mathbb{E}\left[\sum_{k=1}^{i-1} R_{i,k} Z_k \sum_{\ell=1}^{j-1} R_{j,\ell} Z_\ell \right] = \sum_{k=1}^{i-1} R_{i,k} R_{j,k} = \Gamma_{i,j} - R_{i,i} R_{j,i}.$$

Hence, we obtain

$$\sum_{i<j} \left(\mathbb{E}\left[\mathbb{E}\left(Y_i \mid \mathcal{F}_{i-1} \right) \mathbb{E}\left(Y_j \mid \mathcal{F}_{j-1} \right) \right] \right)^2 = \sum_{i<j} (\Gamma_{i,j} - R_{i,i} R_{j,i})^2 \leq 2 \left(\sum_{i<j} \Gamma_{i,j}^2 + \sum_{i<j} R_{j,i}^2 \right),$$

which establishes (64) by (63) and (vanish-secondorder).

Next, let us establish the following equality in $L^2(\mathbb{P}_m)$: for any $i = 1, \dots, m$ and $t \in [0,1]$,

$$\mathbb{E}\left(h_t(Y_i) \mid \mathcal{F}_{i-1} \right) = \sum_{\ell \geq 2} \frac{c_\ell(t)}{\ell!} \sigma_i^\ell H_\ell \left(\frac{\mathbb{E}\left(Y_i \mid \mathcal{F}_{i-1} \right)}{\sigma_i} \right), \tag{67}$$

where the RHS of (67) is 0 if $\sigma_i = 0$. For this, consider some $1 \leq i \leq m$ and assume $\sigma_i > 0$ (otherwise the result is obvious). Let $\widetilde{Y}_i = \frac{\mathbb{E}(Y_i \mid \mathcal{F}_{i-1})}{\sigma_i} \sim \mathcal{N}(0,1)$. By using the multivariate Gaussian structure of Y, the distribution of Y_i conditionally on $\mathcal{F}_{i-1}$ only depends on $\widetilde{Y}_i$. Hence, we can write $\mathbb{E}\left(h_t(Y_i) \mid \mathcal{F}_{i-1} \right) = g(\widetilde{Y}_i)$ for a (unique) function g in $L^2(\mathbb{R}, \mathcal{N}(0,1))$. We now consider the expansion of g w.r.t. the Hermite polynomials in that space:

$$g(\cdot) = \sum_{\ell \geq 0} \frac{\mathbb{E}(g(\widetilde{Y}_i) H_\ell(\widetilde{Y}_i))}{\ell!} H_\ell(\cdot),$$

and we can compute each coordinate $\mathbb{E}(g(\widetilde{Y}_i)H_\ell(\widetilde{Y}_i))$ in the following way: for any $\ell \geq 0$,

$$
\begin{aligned}
\mathbb{E}\left[H_\ell(\widetilde{Y}_i)\mathbb{E}\left(h_t(Y_i) \mid \mathcal{F}_{i-1}\right)\right] &= \mathbb{E}\left[H_\ell(\widetilde{Y}_i)h_t(Y_i)\right] \\
&= \sum_{\ell' \geq 2} \frac{c_{\ell'}(t)}{(\ell')!}\mathbb{E}\left[H_\ell(\widetilde{Y}_i)H_{\ell'}(Y_i)\right] \\
&= \frac{c_\ell(t)}{\ell!}\sigma_i^\ell \, \ell! \, \mathbf{1}\left\{\ell \geq 2\right\},
\end{aligned}
$$

by using Fubini's theorem (because $\sum_{\ell' \geq 2} \frac{|c_{\ell'}(t)|}{(\ell')!}\mathbb{E}\left[|H_\ell(\widetilde{Y}_i)H_{\ell'}(Y_i)|\right] \leq (\ell!)^{1/2} \sum_{\ell' \geq 2} \frac{|c_{\ell'}(t)|}{(\ell'!)^{1/2}} < \infty$), and by applying (70) with $\mathrm{Cov}(Y_i, \widetilde{Y}_i) = \sigma_i$. This proves (67).

Finally, by using (67), (70) and notation above, we have

$$
\begin{aligned}
\mathbb{E}\left[\left(\frac{r_m}{m}\sum_{i=1}^m \mathbb{E}\left(h_t(Y_i) \mid \mathcal{F}_{i-1}\right)\right)^2\right] &= \frac{r_m^2}{m^2}\sum_{i,j}\mathbb{E}\left[\mathbb{E}\left(h_t(Y_i) \mid \mathcal{F}_{i-1}\right)\mathbb{E}\left(h_t(Y_j) \mid \mathcal{F}_{j-1}\right)\right] \\
&= \frac{r_m^2}{m^2}\sum_{i,j}\sum_{\ell \geq 2}\sum_{\ell' \geq 2}\frac{c_\ell(t)}{\ell!}\frac{c_{\ell'}(t)}{(\ell'!)}\sigma_i^\ell \sigma_j^{\ell'}\mathbb{E}\left[H_\ell(\widetilde{Y}_i)H_{\ell'}(\widetilde{Y}_j)\right] \\
&= \frac{r_m^2}{m^2}\sum_{i,j}\sum_{\ell \geq 2}\frac{c_\ell(t)^2}{\ell!}\sigma_i^\ell \sigma_j^\ell \left(\mathbb{E}\left[\widetilde{Y}_i\widetilde{Y}_j\right]\right)^\ell \\
&\leq \left(\sum_{\ell \geq 2}\frac{c_\ell(t)^2}{\ell!}\right)\left(\frac{r_m^2}{m^2}\sum_{i,j}\left(\mathbb{E}\left[\mathbb{E}\left(Y_i \mid \mathcal{F}_{i-1}\right)\mathbb{E}\left(Y_j \mid \mathcal{F}_{j-1}\right)\right]\right)^2\right),
\end{aligned}
$$

which proves (66) by using (64). Exactly the same calculation with "$i = j$" shows (65) from (63). $\qquad\square$

Lemma A.2. *Assume that $\Gamma^{(m)}$ satisfies (vanish-secondorder) and that $r_m^2 \mathrm{Var}(\overline{Y}_m)$ converges to some positive real number. Consider the $(m+1) \times (m+1)$ covariance matrix $\Lambda^{(m+1)}$ of $(Y_i)_{0 \leq i \leq m}$ defined in Section 5.3. Then the rate*

$$
r_{m+1}(\Lambda^{(m+1)}) = \left((m+1)^{-1} + \left|(m+1)^{-2}\sum_{0 \leq i \neq j \leq m}\Lambda_{i,j}^{(m+1)}\right|\right)^{-1/2}
$$

satisfies $r_{m+1}(\Lambda^{(m+1)}) \sim r_m$ and moreover

$$
(m+1)^{-2}r_m^2 \sum_{0 \leq i \neq j \leq m}\left(\Lambda_{i,j}^{(m+1)}\right)^2 = o(1). \tag{68}
$$

In particular, $\Lambda^{(m+1)}$ satisfies (vanish-secondorder). Finally, when (H_2) holds for $\Gamma^{(m)}$, it also holds for $\Lambda^{(m+1)}$, with the same value of θ.

Proof. By definition,

$$
m^{-2}\sum_{0 \leq i \neq j \leq m}\Lambda_{i,j} = m^{-2}\sum_{1 \leq i \neq j \leq m}\Gamma_{i,j} + 2m^{-2}\sum_{1 \leq j \leq m}\Lambda_{0,j}.
$$

Since $\Lambda_{0,j} = (\operatorname{Var}\overline{Y}_m)^{-1/2}m^{-1}\sum_{i=1}^{m}\Gamma_{i,j}$, we have

$$m^{-2}\sum_{1\le j\le m}\Lambda_{0,j} = (\operatorname{Var}\overline{Y}_m)^{-1/2}m^{-1}m^{-2}\sum_{1\le i,j\le m}\Gamma_{i,j} = m^{-1}\left(m^{-2}\sum_{1\le i,j\le m}\Gamma_{i,j}\right)^{1/2},$$

which is $o(1/m)$ because Γ satisfies (vanish-secondorder) and thus (LLN-dep). This implies $r_{m+1}(\Lambda) \sim r_m$. Next, we establish (68). Let us write

$$(m+1)^{-2}r_m^2\sum_{0\le i\ne j\le m}(\Lambda_{i,j})^2 = (m+1)^{-2}r_m^2\left(\sum_{1\le i\ne j\le m}(\Gamma_{i,j})^2 + 2\sum_{1\le j\le m}(\Lambda_{0,j})^2\right).$$

Furthermore, we have

$$\begin{aligned}
\sum_{1\le j\le m}(\Lambda_{0,j})^2 &= (\operatorname{Var}\overline{Y}_m)^{-1}\sum_{1\le j\le m}\left(m^{-1}\sum_{i=1}^{m}\Gamma_{i,j}\right)^2\\
&\le (\operatorname{Var}\overline{Y}_m)^{-1}m^{-2}\sum_{1\le i,i'\le m}\left(2\Gamma_{i,i'} + \sum_{j\notin\{i,i'\}}\Gamma_{i,j}\Gamma_{i',j}\right)\\
&\le 2 + (m\operatorname{Var}\overline{Y}_m)^{-1}\sum_{1\le i\ne j\le m}(\Gamma_{i,j})^2.
\end{aligned}$$

This implies the result, because $m\operatorname{Var}\overline{Y}_m \ge r_m^2\operatorname{Var}\overline{Y}_m$, which is bounded away from 0 by assumption. $\square$

Appendix B: Results related to Hermite polynomials

Let us first recall that the sequence of Hermite polynomials $H_\ell(x)$, $\ell \ge 0$, $x \in \mathbb{R}$, is defined by the expression: for all $\ell \ge 0$,

$$\forall x \in \mathbb{R}, \ \phi^{(\ell)}(x) = (-1)^\ell H_\ell(x)\phi(x), \tag{69}$$

where $\phi(x) = (2\pi)^{-1/2}\exp(-x^2/2)$ is the density of a Gaussian standard variable and $\phi^{(\ell)}$ denotes its ℓ-th derivative (by convention, $\phi^{(0)} = \phi$). For instance, we have $H_0(x) = 1$, $H_1(x) = x$ and $H_2(x) = x^2 - 1$.

A well known fact is that $\{H_\ell(\cdot)/(\ell!)^{1/2}, \ell \ge 0\}$ is an Hilbert basis in $L^2(\mathbb{R}, \mathcal{N}(0,1))$, the Hilbert space composed by square integrable functions w.r.t. the standard Gaussian measure. Moreover, the following property holds: for any centered 2-dimensional Gaussian vector (U,V) with $\mathbb{E}U^2 = \mathbb{E}V^2 = 1$,

$$\forall \ell, \ell' \ge 0, \ell \ne \ell', \ \mathbb{E}(H_\ell(U)H_{\ell'}(V)) = (\operatorname{Cov}(U,V))^\ell\, \ell!\, \delta_{\ell,\ell'}. \tag{70}$$

The latter can be seen as a consequence of Mehler's formula, itself being nicely presented in Foata (1981) (1.4) (see also references therein).

Proof of Proposition 2.1 Let us start by expanding, for any $t \in [0,1]$, the function $\mathbf{1}\{\Phi(\cdot) \leq t\}$ w.r.t. the Hermite polynomial basis in $L^2(\mathbb{R}, \mathcal{N}(0,1))$:

$$\mathbf{1}\{\Phi(\cdot) \leq t\} = \sum_{\ell \geq 0} c_\ell(t) H_\ell(\cdot)/(\ell!). \tag{71}$$

By applying (71) at Y_i, we obtain the following expansion in $L^2(\mathbb{P}_m)$: for all $i = 1, \ldots, m$,

$$\mathbf{1}\{\Phi(Y_i) \leq t\} = \sum_{\ell \geq 0} c_\ell(t) H_\ell(Y_i)/(\ell!). \tag{72}$$

By averaging w.r.t. i, we obtain

$$\widehat{\mathbb{F}}_m(t) - t = \sum_{\ell \geq 1} \frac{c_\ell(t)}{\ell!} m^{-1} \sum_{i=1}^m H_\ell(Y_i). \tag{73}$$

where the series in the RHS of (73) converges in $L^2(\mathbb{P}_m)$ (by using the triangle inequality). The proof is finished by combining (73) with (70).

Next, the following proposition shares some similarities with Lemma 4.5 of Taqqu (1977) and Lemma 3 of Csörgő and Mielniczuk (1996).

Proposition B.1. *Consider an integer $d \geq 2$, a positive number ρ such that $\sqrt{3}\rho d < 1$ and $Z \sim \mathcal{N}(0,1)$. Let $g_1, \ldots, g_d$ be d measurable real functions defined on $\mathbb{R}$ such that $\mathbb{E}\left(|g_i(Z)|^{4/3}\right) < +\infty$, $1 \leq i \leq d$. Let $(U_1, \ldots, U_d)$ be d-dimensional centered Gaussian vector with $\mathbb{E}U_i^2 = 1$, $1 \leq i \leq d$, and $|\mathbb{E}(U_iU_j)| \leq \rho$, $1 \leq i \neq j \leq d$. Then the following holds:*

$$\mathbb{E}\left[\prod_{i=1}^d |g_i(U_i)|\right] \leq \frac{1}{(1 - \sqrt{3}\rho d)^d} \prod_{i=1}^d \left(\mathbb{E}\left(|g_i(Z)|^{4/3}\right)\right)^{3/4}; \tag{74}$$

Furthermore, if $\mathbb{E}(g_i(Z)) = 0$ and $\mathbb{E}(Zg_i(Z)) = 0$ for $1 \leq i \leq d'$ for an integer d', $1 \leq d' \leq d$, we have

$$\left|\mathbb{E}\left[\prod_{i=1}^d g_i(U_i)\right]\right| \leq \rho^{d'} \frac{(3d^2)^{d'}}{(1 - \sqrt{3}\rho d)^d} \prod_{i=1}^d \left(\mathbb{E}\left(|f_i(Z)|^{4/3}\right)\right)^{3/4}, \tag{75}$$

where f_i is any function such that $f_i(x) = g_i(x) - \alpha_i - \beta_i x$, $x \in \mathbb{R}$, $\alpha_i, \beta_i \in \mathbb{R}$, for $1 \leq i \leq d'$ and $f_i = g_i$ otherwise.

Proof. The Kibble-Slepian formula Kibble (1945); Slepian (1972) (given, e.g., in expression (2.2) of Foata, 1981) provides that

$$\mathbb{E}\left[\prod_{i=1}^d g_i(U_i)\right] = \mathbb{E}\left(\sum_\nu \prod_{i<j} \frac{(\mathbb{E}(U_iU_j))^{\nu_{ij}}}{\nu_{ij}!} \cdot \prod_{i=1}^d g_i(Z)H_{\nu_i.}(Z)\right)$$

$$= \sum_\nu \prod_{i<j} \frac{(\mathbb{E}(U_iU_j))^{\nu_{ij}}}{\nu_{ij}!} \cdot \prod_{i=1}^d \mathbb{E}(g_i(Z)H_{\nu_i.}(Z)), \tag{76}$$

where the summation is over all the $d \times d$ symmetric matrix $\nu = (\nu_{ij})_{1 \leq i,j \leq d}$ with nonnegative integral entries and with diagonal entries equal to zero, while $\nu_{i.}$ denotes $\nu_{i1} + \cdots + \nu_{id}$. Above, we have implicitly used Fubini's theorem (the summation over ν is infinite). The next calculations show that this is indeed valid: by using the assumptions, we have

$$
\sum_{\nu} \prod_{i<j} \frac{|\mathbb{E}(U_i U_j)|^{\nu_{ij}}}{\nu_{ij}!} \cdot \prod_{i=1}^{d} \mathbb{E}|g_i(Z) H_{\nu_{i.}}(Z)|
$$

$$
\leq \sum_{\nu} \prod_{i=1}^{d} \left(\frac{\rho^{\nu_{i.}}}{\prod_j \nu_{ij}!} \right)^{1/2} \mathbb{E}|g_i(Z) H_{\nu_{i.}}(Z)|
$$

$$
\leq \sum_{x_1,\ldots,x_d \in \mathbb{N}^d} \prod_{i=1}^{d} \left(\frac{\rho^{x_{i.}}}{\prod_j x_{ij}!} \right)^{1/2} \mathbb{E}|g_i(Z) H_{x_{i.}}(Z)|
$$

$$
= \prod_{i=1}^{d} \left[\sum_{y \in \mathbb{N}^d} \left(\frac{\rho^{y_1+\cdots+y_d}}{\prod_j y_j!} \right)^{1/2} \mathbb{E}|g_i(Z) H_{y_1+\cdots+y_d}(Z)| \right]
$$

$$
= \prod_{i=1}^{d} \left[\sum_{\ell \geq 0} \rho^{\ell/2} \mathbb{E}\left| g_i(Z) H_\ell(Z)/(\ell!)^{1/2} \right| \sum_{\substack{y \in \mathbb{N}^d \\ y_1+\cdots+y_d=\ell}} \left(\frac{\ell!}{\prod_j y_j!} \right)^{1/2} \right]. \tag{77}
$$

Now, in the latter display, the sum over y is upper bounded by d^ℓ, which gives that the RHS of (77) is upper bounded by

$$
\sum_{\ell \geq 0} (\rho d^2)^{\ell/2} \mathbb{E}\left| g_i(Z) H_\ell(Z)/(\ell!)^{1/2} \right| \leq \left(\sum_{\ell \geq 0} (3\rho d^2)^{\ell/2} \right) \left(\mathbb{E}\left(|g_i(Z)|^{4/3} \right) \right)^{3/4},
$$

where the latter combines Hölder's inequality with Lemma B.2 (used with $p = 4$). This proves (74) and shows that Fubini's theorem can be applied to get (76).

Finally, we prove (75) by using (76) and the same calculations as above, except that the absolute values should be kept outside the expectations. As a result, for $1 \leq i \leq d'$, since $\mathbb{E}(g_i(Z) H_\ell(Z)) = 0$ for $\ell = 0, 1$ by assumption, the corresponding sums over ℓ start at $\ell = 2$. This establishes (75), because for all $\ell \geq 2$ and $1 \leq i \leq d'$, $\mathbb{E}(g_i(Z) H_\ell(Z)) = \mathbb{E}(f_i(Z) H_\ell(Z))$. □

The following result was obtained in the proof of Lemma 3.1 in Taqqu (1977). We provide an elementary proof below. Also, let us mention that there are more accurate such results when ℓ grows to infinity, see Theorem 2.1 in Larsson-Cohn (2002).

Lemma B.2. *For all even integer $p \geq 2$ and $\ell \geq 0$, we have $\left[\mathbb{E}\left(H_\ell(Z)/\sqrt{\ell!} \right)^p \right]^{1/p} \leq (p-1)^{\ell/2}$, for $Z \sim \mathcal{N}(0,1)$.*

Proof. For some $\ell \geq 1$, by using $H_\ell' = \ell H_{\ell-1}$ and (69), we obtain

$$
\int [H_\ell(x)]^p \phi(x) dx = (-1)^\ell \int [H_\ell(x)]^{p-1} \phi^{(\ell)}(x) \, dx,
$$

$$
= \ell(p-1) \int [H_\ell(x)]^{p-2} [H_{\ell-1}(x)]^2 \phi(x) \, dx.
$$

Next, by using Hölder's inequality, we get $\left(\int [H_\ell(x)]^p \phi(x)dx\right)^{2/p} \leq \ell(p-1)\left(\int |H_{\ell-1}(x)|^p \phi(x)dx\right)^{2/p}$, and the result is obtained by induction on ℓ. $\qquad\square$

Lemma B.3. *Consider the function $h_t(\cdot)$ defined by* (44) *and $c_\ell(\cdot)$ defined by* (5). *Let us consider a two-dimensional centered Gaussian vector (U, V) with $\mathbb{E}U^2 = \mathbb{E}V^2 = 1$. Then for any $t, s \in [0, 1]$, the following holds:*

$$\mathbb{E}(h_t(U)h_s(V)) = \sum_{\ell \geq 2} \frac{c_\ell(t)c_\ell(s)}{\ell!}(\text{Cov}(U, V))^\ell. \tag{78}$$

Proof. Expression (78) is a direct consequence of (70) and of Fubini's theorem. $\qquad\square$

Lemma B.4. *The function $c_1(\cdot) = \phi(\Phi^{-1}(\cdot))$ satisfies the following: for all $\nu \in (0, 1)$, there exists some constant $C_\nu > 0$ such that for all $s, t \in [0, 1]$,*

$$|c_1(t) - c_1(s)| \leq C_\nu |t - s|^{1-\nu}. \tag{79}$$

Proof. First note that the derivative of c_1 on $(0, 1)$ is Φ^{-1}. Classically (see, e.g., Lemma 12.3 of Abramovich et al. (2006)), there is some $x_0 \in (0, 1/2)$ such that for any $u \in (0, x_0)$, $\Phi^{-1}(u) \leq \sqrt{2\log(1/u)}$. Also, obviously, for some fixed $\nu > 0$, there is some $C'_\nu > 0$ such that for any $u \in (0, x_0)$, $\sqrt{2\log(1/u)} \leq C'_\nu u^{-\nu}$. As a consequence, since $|\Phi^{-1}|$ is bounded on $[x_0, 1 - x_0]$, there exists some constant $C''_\nu > 0$ such that for all $u \in (0, 1)$, $|\Phi^{-1}(u)| \leq C''_\nu u^{-\nu}$. This entails that for all $0 < s \leq t < 1$,

$$|c_1(t) - c_1(s)| \leq \int_s^t |\Phi^{-1}(u)|du \leq \frac{C''_\nu}{1-\nu}(t^{1-\nu} - s^{1-\nu}) \leq C_\nu(t-s)^{1-\nu}$$

by letting $C_\nu = C''_\nu/(1 - \nu) > 0$ and because $(x + y)^\delta \leq x^\delta + y^\delta$ for any $x, y \geq 0$ and any $\delta \in (0, 1)$. $\qquad\square$

Appendix C: Useful auxiliary results

The following result can certainly be considered as well known, although we failed to find a precise reference for it. It can be seen as a reformulation in our framework of classical tightness results as given, e.g., in Lemma 2 of Csörgő and Mielniczuk (1996), in Remark 2.1 of Shao and Yu (1996) and Proposition 6 of Dedecker and Prieur (2007).

Proposition C.1 (Tightness criterion for empirical distribution function with non-standard scaling parameters)**.** *Consider $\xi_1, \ldots, \xi_m$ real random variables (that need not to be independent or identically distributed) such that $\overline{\xi}_m \xrightarrow{P} 0$ as m tends to infinity, for $\overline{\xi}_m = m^{-1}\sum_{i=1}^m \xi_m$, and consider the process*

$$Z_m(t) = (a_m/m)\sum_{i=1}^m g_t(\xi_i), \text{ for } t \in [0, 1],$$

where $(a_m)_m$ is some positive sequence tending to infinity as m tends to infinity and where $g_t(x) = \mathbf{1}\{\Phi(x) \leq t\} - f_0(t) - f_1(t)x$ for functions f_0, f_1 on $[0, 1]$ such that $|f_0(t) - f_0(s)| \vee |f_1(t) - f_1(s)| \leq L|t - s|^q$, $0 \leq s, t \leq 1$, for some $q \in (0, 1]$ and $L > 0$. Assume that the following holds: for large m,

$$\mathbb{E}\big|Z_m(t) - Z_m(s)\big|^\kappa \leq C\big(|t - s|^{\delta_1} + (a_m)^{-\delta_2/q}|t - s|^{q'}\big), \text{ for all } t, s \in [0, 1], \tag{80}$$

for constants $\kappa > 0$, $C > 0$, $\delta_1 > 1$, $q' \in (0,1]$ and $\delta_2 > 1 - q'$. Then, as m grows to infinity, the sequence of processes $(Z_m)_m$ is tight in $D(0,1)$ (endowed with the Skorokhod topology and the corresponding Borel σ-field) and any limit is a.s. a continuous process.

Proof. The proof is based on standard arguments and is similar to the proof of Theorem 22.1 in Billingsley (1968). Fix $\varepsilon \in (0,1)$ and $\eta > 0$. Following Theorem 15.5 in Billingsley (1968), it is sufficient to prove that there exists a $\delta \in (0,1)$ such that for large m,

$$\mathbb{P}\left(\sup_{\substack{0 \le s,t \le 1 \\ |s-t| \le \delta}} |Z_m(t) - Z_m(s)| > \varepsilon \right) < \eta.$$

We merely check (see, e.g., the proof of Theorem 8.3 in Billingsley (1968)) that the latter holds if there exists $\delta \in (0,1)$ such that for large m,

$$\forall s \in [0,1], \ \mathbb{P}\left(\sup_{t:s \le t \le (s+\delta) \wedge 1} |Z_m(t) - Z_m(s)| > \varepsilon \right) < \eta\delta. \tag{81}$$

Let us now prove (81). Fix $s \in [0,1]$. Assumption (80) entails that for all $u, v \in [0,1]$ such that $(v-u)^q \ge \varepsilon/a_m$, we have

$$\mathbb{E}\big|Z_m(v) - Z_m(u)\big|^\kappa \le \frac{2C}{\varepsilon^{\delta_2/q}} |v-u|^{\delta_3}$$

for $\delta_3 = \delta_1 \wedge (q' + \delta_2) > 1$. Hence, if $p > 0$ is such that $p^q \ge \varepsilon/a_m$, applying Theorem 12.2 of Billingsley (1968) we have for all integer M such that $s + Mp \le 1$ and for all $\lambda > 0$,

$$\mathbb{P}\left(\max_{1 \le i \le M} |Z_m(s+ip) - Z_m(s)| > \lambda \right) \le \frac{K}{\lambda^\kappa \varepsilon^{\delta_2/q}} (Mp)^{\delta_3} \tag{82}$$

for some positive constant $K > 0$ (only depending on δ_3, κ and C). Next, we use the following inequality: for all $0 \le u, v \le 1$, $u \le v \le u+p$,

$$|Z_m(v) - Z_m(u)| \le |Z_m(u+p) - Z_m(u)| + 2La_m p^q (1 + |\bar{\xi}_m|). \tag{83}$$

The latter holds because we have

$$Z_m(v) - Z_m(u) = (a_m/m) \sum_{i=1}^m \mathbf{1}\{u < \xi_i \le v\} - a_m(f_0(v) - f_0(u)) - a_m(f_1(v) - f_1(u))\bar{\xi}_m$$

$$\le Z_m(u+p) - Z_m(u) + 2La_m p^q (1 + |\bar{\xi}_m|)$$

and $Z_m(u) - Z_m(v) \le a_m(f_0(v) - f_0(u)) + a_m(f_1(v) - f_1(u))\bar{\xi}_m \le La_m p^q (1 + |\bar{\xi}_m|)$.
   Now, by using (83), we obtain

$$\sup_{t:s \le t \le s+Mp} |Z_m(t) - Z_m(s)| \le 3 \max_{1 \le i \le M} |Z_m(s+ip) - Z_m(s)| + 2La_m p^q (1 + |\bar{\xi}_m|). \tag{84}$$

Furthermore, provided that $a_m p^q \le 2\varepsilon$, we have $\mathbb{P}(2La_m p^q(1+|\bar{\xi}_m|) > 5L\varepsilon) \le \mathbb{P}(|\bar{\xi}_m| > 1/4)$. Hence, combining (82) and (84), by taking $\delta \in (0,1)$ such that $K\delta^{\delta_3-1}/\varepsilon^{\kappa+\delta_2/q} < \eta/2$, we will obtain that for all $s \in [0, 1-\delta]$, for large m,

$$\mathbb{P}\left(\sup_{t:s \le t \le s+\delta} |Z_m(t) - Z_m(s)| > (3+5L)\varepsilon \right) \le \frac{K}{\varepsilon^{\kappa+\delta_2/q}} \delta^{\delta_3} + \mathbb{P}(|\bar{\xi}_m| > 1/4) < \eta\delta,$$

as soon as we can choose $p > 0$ and an integer M such that $Mp = \delta$ and $\varepsilon/a_m \leq p^q \leq 2\varepsilon/a_m$. This holds if there exists an integer into the interval $[\delta(a_m/\varepsilon)^{1/q}, \delta(a_m/(2\varepsilon))^{1/q}]$, which is true for large m because a_m tends to infinity. This entails (81) with ε replaced by $(3 + 5L)\varepsilon$ and the proof is finished. $\qquad\square$

Proposition C.2 (Partial functional delta method on $D(0,1)$). *Consider the linear space $D(0,1)$ of càd-làg function on $[0,1]$ and the linear space $C(0,1)$ of continuous functions on $[0,1]$. Let $\theta = (\theta_0, \theta_1) \in D(0,1)^2$. Let $\phi : D(0,1)^2 \mapsto \mathbb{R}$ be Hadamard differentiable at θ tangentially to $C(0,1)$, w.r.t. the supremum norm, and such that the derivative is of the form*

$$\dot{\phi}_\theta(H_0, H_1) = g_\theta(H_0), \text{ for any } (H_0, H_1) \in C(0,1)^2,$$

for a continuous linear mapping $g_\theta : C(0,1) \mapsto \mathbb{R}$. Consider $\mathbb{Z}_{0,m}$, $\mathbb{Z}_{1,m}$, $m \geq 1$, processes valued in $D(0,1)$ and $\mathbb{Z}_0, \mathbb{Z}_1$ two processes valued a.s. in $C(0,1)$. Assume that the two following distribution convergences hold (w.r.t. the Skorokhod topology and the corresponding Borel σ-field), for some positive sequence $(a_m)_m$ tending to infinity:

$$a_m(\mathbb{Z}_{0,m} - \theta_0) \rightsquigarrow \mathbb{Z}_0;$$
$$a_m(\mathbb{Z}_{1,m} - \theta_1) \rightsquigarrow \mathbb{Z}_1.$$

Then we have

$$a_m(\phi(\mathbb{Z}_{0,m}, \mathbb{Z}_{1,m}) - \phi(\theta)) \rightsquigarrow g_\theta(\mathbb{Z}_0). \tag{85}$$

Proof. Classically, let us show that for any subsequence $\{n\}$ there exists a further subsequence $\{\ell\}$ such that (85) holds along this subsequence. For any $\{n\}$, since both processes $a_n(\mathbb{Z}_{0,n} - \theta_0)$ and $a_n(\mathbb{Z}_{1,n} - \theta_1)$ are (Skorokhod-)tight, the joint process $(a_n(\mathbb{Z}_{0,n} - \theta_0), a_n(\mathbb{Z}_{1,n} - \theta_1))$ also is. Hence, by Prohorov's theorem, there exists a further subsequence $\{\ell\}$ such that $(a_\ell(\mathbb{Z}_{0,\ell} - \theta_0), a_\ell(\mathbb{Z}_{1,\ell} - \theta_1))$ converges in distribution. Now applying the Skorokhod's representation theorem (see, e.g., Theorem 6.7 page 70 in Billingsley (1999)), there exists random elements $T_\ell = (T_{0,\ell}, T_{1,\ell})$, $\ell \geq 1$, $T = (T_0, T_1)$, defined on a common probability space, such that $\mathcal{L}(T_\ell) = \mathcal{L}(a_\ell(\mathbb{Z}_{0,\ell} - \theta_0), a_\ell(\mathbb{Z}_{1,\ell} - \theta_1))$, $\mathcal{L}(T_0) = \mathcal{L}(\mathbb{Z}_0)$, $\mathcal{L}(T_1) = \mathcal{L}(\mathbb{Z}_1)$ and T_ℓ converges a.s. to T. Since both T_0 and T_1 belong to $C(0,1)$ (a.s.) and since any sequence of càd-làg functions converging (w.r.t. to the Skorokhod distance) to a continuous function also converges uniformly, we obtain

$$||T_{0,\ell} - T_0||_\infty + ||T_{1,\ell} - T_1||_\infty \to 0 \quad \text{a.s.}$$

Hence, the Hadamard differentiability of ϕ entails:

$$\frac{\phi(\theta + t_\ell T_\ell) - \phi(\theta)}{t_\ell} \to g_\theta(T_0) \quad \text{a.s.} \quad,$$

for any sequence $t_\ell \to 0$. By taking $t_\ell = 1/a_\ell$, we derive (85) along the subsequence $\{\ell\}$, which proves the result. $\qquad\square$

Lemma C.3. *Assume that Γ satisfies (LLN-dep). Then for any $h : \mathbb{R} \to \mathbb{R}$ measurable such that $\mathbb{E}|h(Z)| < \infty$, we have*

$$m^{-1} \sum_{i=1}^{m} h(Y_i) \xrightarrow{P} \mathbb{E}[h(Z)], \quad \text{for } Z \sim \mathcal{N}(0,1). \tag{86}$$

Proof. By Section [2], Assumption (LLN-dep) implies that $\forall t \in [0,1]$, $\widehat{\mathbb{F}}_m(t) \xrightarrow{P} t$. Since $h \in L^1(\mathbb{R}, \mathcal{N}(0,1))$, for any $\varepsilon > 0$, there is a continuous bounded function h_ε such that $\mathbb{E}|h(Z) - h_\varepsilon(Z)| \leq \varepsilon$. Moreover, by definition of the weak convergence, (86) holds for $h = h_\varepsilon$ (for instance, the convergence in probability can be seen as an a.s. convergence up to consider subsequence). Since we have

$$\sup_{m \geq 1} \left\{ \mathbb{E} \left| m^{-1} \sum_{i=1}^m (h(Y_i) - h_\varepsilon(Y_i)) \right| \right\} \leq \sup_{m \geq 1} \left\{ m^{-1} \sum_{i=1}^m \mathbb{E} \left| h(Y_i) - h_\varepsilon(Y_i) \right| \right\} \leq \varepsilon,$$

we can conclude by using Lemma C.4. $\qquad\square$

The following lemma is classical, see, e.g., Theorem 4.2 in Billingsley (1968).

Lemma C.4. *For $n \geq 1$ and $\varepsilon > 0$, let X_n^ε, X_n, X^ε, X be real random variables (X_n and X_n^ε being defined on the same probability space) and such that*

(a) $\forall \varepsilon > 0$, $X_n^\varepsilon \rightsquigarrow X^\varepsilon$ *as $n \to \infty$;*
(b) $X^\varepsilon \rightsquigarrow X$ *as $\varepsilon \to 0$;*
(c) $\limsup_{n \to \infty} \{\mathbb{E}|X_n^\varepsilon - X_n|\} \to 0$ *as $\varepsilon \to 0$.*

Then $X_n \rightsquigarrow X$.

Acknowledgments

References

Abramovich, F., Benjamini, Y., Donoho, D. L., and Johnstone, I. M. (2006). Adapting to unknown sparsity by controlling the false discovery rate. *Ann. Statist.*, 34(2):584–653.

Bai, Z. D. and Yin, Y. Q. (1993). Limit of the smallest eigenvalue of a large-dimensional sample covariance matrix. *Ann. Probab.*, 21(3):1275–1294.

Bardet, J.-M. and Surgailis, D. (2011). Moment bounds and central limit theorems for Gaussian subordinated arrays. *ArXiv e-prints*.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B*, 57(1):289–300.

Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.*, 29(4):1165–1188.

Billingsley, P. (1968). *Convergence of probability measures.* John Wiley & Sons Inc., New York.

Billingsley, P. (1999). *Convergence of probability measures.* Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, second edition. A Wiley-Interscience Publication.

Csörgő, S. and Mielniczuk, J. (1996). The empirical process of a short-range dependent stationary sequence under Gaussian subordination. *Probab. Theory Related Fields*, 104(1):15–25.

Dedecker, J. and Prieur, C. (2007). An empirical central limit theorem for dependent sequences. *Stochastic Process. Appl.*, 117(1):121–142.

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(4):1767–1783.

Delattre, S. and Roquain, E. (2011). On the false discovery proportion convergence under Gaussian equi-correlation. *Statist. Probab. Lett.*, 81(1):111–115.

Donsker, M. D. (1952). Justification and extension of Doob's heuristic approach to the Komogorov-Smirnov theorems. *Ann. Math. Statistics*, 23:277–281.

Doob, J. L. (1949). Heuristic approach to the Kolmogorov-Smirnov theorems. *Ann. Math. Statistics*, 20:393–403.

Doukhan, P., Lang, G., and Surgailis, D. (2002). Asymptotics of weighted empirical processes of linear fields with long-range dependence. *Ann. Inst. H. Poincaré Probab. Statist.*, 38(6):879–896. En l'honneur de J. Bretagnolle, D. Dacunha-Castelle, I. Ibragimov.

Doukhan, P., Lang, G., Surgailis, D., and Teyssière, G., editors (2010). *Dependence in probability and statistics*, volume 200 of *Lecture Notes in Statistics*. Springer-Verlag, Berlin.

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

Efron, B. (2010). Correlated z -values and the accuracy of large-scale statistical estimates. *J. Amer. Statist. Assoc.*, 105(491):1042–1055.

Efron, B., Tibshirani, R., Storey, J. D., and Tusher, V. (2001). Empirical Bayes analysis of a microarray experiment. *J. Amer. Statist. Assoc.*, 96(456):1151–1160.

Fan, J., Han, X., and Gu, W. (2012). Estimating false discovery proportion under arbitrary covariance dependence. *Journal of the American Statistical Association*, 107(499):1019–1035.

Farcomeni, A. (2006). More powerful control of the false discovery rate under dependence. *Stat. Methods Appl.*, 15(1):43–73.

Farcomeni, A. (2007). Some results on the control of the false discovery rate under dependence. *Scand. J. Statist.*, 34(2):275–297.

Foata, D. (1981). Some Hermite polynomial identities and their combinatorics. *Adv. in Appl. Math.*, 2(3):250–259.

Friguet, C., Kloareg, M., and Causeur, D. (2009). A factor model approach to multiple testing under dependence. *J. Amer. Statist. Assoc.*, 104(488):1406–1415.

Genovese, C. and Wasserman, L. (2004). A stochastic process approach to false discovery control. *Ann. Statist.*, 32(3):1035–1061.

Hall, P. and Heyde, C. C. (1980). *Martingale limit theory and its application.* Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York. Probability and Mathematical Statistics.

Jacod, J. and Shiryaev, A. N. (2003). *Limit theorems for stochastic processes*, volume 288 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, second edition.

Johnstone, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.*, 29(2):295–327.

Kibble, W. F. (1945). An extension of a theorem of Mehler's on Hermite polynomials. *Proc. Cambridge Philos. Soc.*, 41:12–15.

Kim, K. I. and van de Wiel, M. (2008). Effects of dependence in high-dimensional multiple testing problems. *BMC Bioinformatics*, 9(1):114.

Korn, E. L., Troendle, J. F., McShane, L. M., and Simon, R. (2004). Controlling the number of

false discoveries: application to high-dimensional genomic data. *J. Statist. Plann. Inference*, 124(2):379–398.

Larsson-Cohn, L. (2002). L^p-norms of Hermite polynomials and an extremal problem on Wiener chaos. *Ark. Mat.*, 40(1):133–144.

Lehmann, E. L. and Romano, J. P. (2005). Generalizations of the familywise error rate. *Ann. Statist.*, 33:1138–1154.

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

Neuvial, P. (2009). Corrigendum to "Asymptotic properties of false discovery rate controlling procedures under independence" [mr2460858]. *Electron. J. Stat.*, 3:1083.

Romano, J. P. and Wolf, M. (2007). Control of generalized error rates in multiple testing. *Ann. Statist.*, 35(4):1378–1408.

Roquain, E. and Villers, F. (2011). Exact calculations for false discovery proportion with application to least favorable configurations. *Ann. Statist.*, 39(1):584–612.

Schwartzman, A. and Lin, X. (2011). The effect of correlation in false discovery rate estimation. *Biometrika*, 98(1):199–214.

Shao, Q.-M. and Yu, H. (1996). Weak convergence for weighted empirical processes of dependent sequences. *Ann. Probab.*, 24(4):2098–2127.

Slepian, D. (1972). On the symmetrized Kronecker power of a matrix and extensions of Mehler's formula for Hermite polynomials. *SIAM J. Math. Anal.*, 3:606–616.

Storey, J. D. (2003). The positive false discovery rate: a Bayesian interpretation and the q-value. *Ann. Statist.*, 31(6):2013–2035.

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. *The Annals of Applied Statistics*, 6(4):1664–1688.

Taqqu, M. S. (1977). Law of the iterated logarithm for sums of non-linear functions of Gaussian variables that exhibit a long range dependence. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 40(3):203–238.

van der Vaart, A. W. (1998). *Asymptotic statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge.