Published in final edited form as:

J Am Stat Assoc. 2015 September 1; 110(511): 1217–1228. doi:10.1080/01621459.2014.958156.

The Empirical Distribution of a Large Number of Correlated Normal Variables

David Azriel and

Lecturer at the Faculty of Industrial Engineering and Management, Technion - Israel Institute of Technology, Haifa 32000, Israel, and Postdoctoral Research Associate in the Department of Statistics of the Wharton School of the University of Pennsylvania, 3730 Walnut Street, Philadelphia, PA 19104

Armin Schwartzman

Associate Professor at the Department of Statistics, North Carolina State University, Raleigh, NC 27695

David Azriel: davidazr@ie.technion.ac.il; Armin Schwartzman: armin.schwartzman@ncsu.edu

Abstract

Motivated by the advent of high dimensional highly correlated data, this work studies the limit behavior of the empirical cumulative distribution function (ecdf) of standard normal random variables under arbitrary correlation. First, we provide a necessary and sufficient condition for convergence of the ecdf to the standard normal distribution. Next, under general correlation, we show that the ecdf limit is a random, possibly infinite, mixture of normal distribution functions that depends on a number of latent variables and can serve as an asymptotic approximation to the ecdf in high dimensions. We provide conditions under which the dimension of the ecdf limit, defined as the smallest number of effective latent variables, is finite. Estimates of the latent variables are provided and their consistency proved. We demonstrate these methods in a real high-dimensional data example from brain imaging where it is shown that, while the study exhibits apparently strongly significant results, they can be entirely explained by correlation, as captured by the asymptotic approximation developed here.

Keywords

empirical null; dependent random variables; high dimensional data; factor analysis; asymptotic approximation; strong correlation

1 Introduction

The empirical cumulative distribution function (ecdf) and its large sample properties have a long and rich history in probability and statistics. However, most of this vast literature assumes that the variables used to construct the ecdf are independent (e.g., Wasserman, 2006, Chapter 2, and references therein). Under independence, the ecdf is a consistent estimator of the true cumulative distribution function (cdf). The consistency property continues to hold under various forms of weak dependence (e.g. Dedecker and Merlevéde (2007); Wu (2006)).

Motivated by modern problems in high dimensional data, where a large number of correlated variables are measured, it is of interest to study the asymptotic behavior of the ecdf when the variables involved are arbitrarily correlated. In particular, the present paper is motivated by large-scale multiple testing problems, where each of p tests produces a z-score Z_i , $i = 1, \dots, p$. It has been pointed out by Bradley Efron that in large-scale multiple testing problems, the observed distribution of the z-scores often does not match the theoretical null distribution $N(0, 1)$ (Efron, 2004, 2007a, b, 2008). Efron (2007a) conjectured that, even when the theoretical model is correct, the observed distribution of the test statistics can look different from the theoretical null distribution simply because of correlation between them. This interesting observation suggests that the ecdf may not always be consistent and calls for a detailed study of the ecdf of a large number of dependent variables.

Assuming that the variables $Z_1, Z_2, \dots, Z_p$ are marginally standard normal simplifies the problem and allows obtaining results for their ecdf under arbitrary dependence because all the dependence is expressed through correlation. In this situation, Efron (2007a) proposed the so-called empirical null as an approximation to the observed distribution of the z-scores, parametrized as a normal distribution with mean and variance other than 0 and 1. To further understand the effect of correlation, Efron (2010) derived the covariance function of the ecdf and applied it to estimating the variance of functions of the ecdf relevant in large-scale multiple testing, such as the local and tail false discovery rate. Schwartzman (2010) proposed to approximate the ecdf by a Gram-Charlier expansion and used its coefficients to establish some constraints to the extent of the departure of the ecdf from the marginal $N(0, 1)$. However, these approaches have not succeeded to fully characterize the behavior of the ecdf under correlation.

In this article we describe the asymptotic behavior of the ecdf of a large number of correlated standard normal variables. First, we show that in general, the ecdf need not converge to Φ , the standard normal cdf. A necessary and sufficient condition for convergence, which we call weak correlation, is that the average of the absolute pairwise correlations between the z-scores (or the average of the squares of the pairwise correlations) tends to zero with increasing dimension p . However, we show that in a wide range of strong correlation situations, the ecdf converges instead to a random distribution function. This random function can be written as a (possibly infinite) normal mixture parametrized by latent independent standard normal variables. It can be thought of as an analytic asymptotic approximation to the ecdf, where the latent variables can be consistently estimated, under some regularity conditions, from the observed data sequence. Further, it can be thought of as a dimension reduction of the p -dimensional ecdf in the sense that its inherent dimension is the number of the latent variables. We give a lower bound for this inherent dimension and show that, under certain regularity conditions, it can be achieved by a particular parametrization obtained via an eigendecomposition of the correlation matrix. This parametrization is based on the factor analysis model of Fan et al. (2012), who use it to calculate the false discovery proportion in large scale multiple testing under arbitrary correlation. Here we consider a more general framework where not only eigendecompositions are investigated.

As an illustration of the behavior of the ecdf as a random function as described in this paper, Figure 1 presents the histograms of two realizations of normal random variables under two correlation structures:

- a. One block: $\{Z_i\}_{i=1}^{1000}$ is a sequence of exchangeable random variables with correlation $\rho = 0.9$, i.e., for any $i \neq j$, $\text{cor}(Z_i, Z_j) = \rho$
- b. Two independent blocks: $\{Z_i\}_{i=1}^{1000}$ consists of two intercalated independent sequences of exchangeable random variables, i.e., for any $i \neq j$ such that $|i - j|$ is even, $\text{cor}(Z_i, Z_j) = \rho$ (and 0 otherwise).

It can be seen that for both structures the empirical distribution differs from the standard normal density (red line). In case (a) the histogram is shifted and narrower than the standard normal density, while in case (b) it looks like a mixture of two normal distributions. Notice that the histograms change between the two realizations, suggesting that the distribution may not be converging to a deterministic limit. We will show that in these cases the empirical distribution indeed does not converge to a deterministic limit, but can be asymptotically approximated by a normal mixture estimated from the data, represented here by the dashed blue line. Not surprisingly, in case (a) the asymptotic approximation is of dimension 1, containing one latent standard normal variable, and in case (b) is of dimension 2, containing two independent latent standard normal variables. The fitted density in case (a) can be interpreted as Efron's empirical null model of a shifted and scaled normal, but clearly that model cannot capture the behavior of the distribution in case (b).

When considering the entire range of possible correlation structures, we show that there are essentially three regimes. The first is what we call *weak correlation*; while the random variables need not to be independent, the ecdf converges to Φ . The second regime is what we call *finite dimensional correlation* and is the main interest of this paper. Under this type of correlation, the ecdf can be approximated by a random function that depends on a finite number of latent independent standard normal random variables. Furthermore, under some regularity conditions, a representation using the smallest number of latent variables can be achieved and estimated consistently. The examples of Figure 1 belong to this regime. Finally, in the third regime, the limiting random function depends on infinite number of independent standard normal random variables, like the ecdf itself.

As an illustration of the usefulness of the results presented in this article, we present an analysis of brain imaging data obtained from a study of cortical thickness of adults who had a diagnosis of attention deficit/hyperactivity disorder (ADHD) as children (Proal et al., 2011). In this study, it had been noticed before that, when searching for brain locations whose cortical thickness is related to clinical diagnosis, the histogram of the z-scores did not follow the theoretical standard normal distribution (Reiss et al., 2012). Here we do a slightly different analysis (where the correlation structure is known) and show that, while the study exhibits apparently strongly significant results, they can be entirely explained by correlation, as captured by the asymptotic approximation mentioned above.

The rest of the paper is organized as follows. After a brief treatment of weak correlation in Section 2, the main results of the paper for general correlation are given in Section 3.

Section 4 discusses how to consistently estimate the latent variables. In Section 4 we also briefly discuss the case where the correlation matrix is unknown. Several concrete examples, including those of Figure 1, are presented in detail in Section 5. A data example is analyzed in Section 6. Section 7 considers some possible extensions of this work. All the proofs are given in a supplementary material document.

2 Weak correlation

2.1 Definition

To define weak correlation some notation is needed. For a given $p \times p$ matrix $R_p = (r_{ij})$

define the following average norms: $\|R_p\|_1^{(p)} := \frac{1}{p^2} \sum_{i,j} |r_{ij}|$ and

$\|R_p\|_2^{(p)} := \left(\sum_{i,j} r_{ij}^2 \right)^{1/2} = \frac{1}{p} \left(\sum_i \lambda_i^2 \right)^{1/2}$, where the λ_i 's are the eigenvalues of R_p . The latter is simply a scaled version of the Frobenius norm, and in both cases we use the superscript (p) to denote that the norm itself, not just the matrix, changes with p . If $\{R_p\}_{p=1}^\infty$ is a sequence of correlation matrices then, as $p \rightarrow \infty$,

$$\|R_p\|_1^{(p)} \rightarrow 0 \iff \|R_p\|_2^{(p)} \rightarrow 0.$$

This is because, by Jensen's inequality, $\|R_p\|_1^{(p)} \leq \|R_p\|_2^{(p)}$, while on the other hand, $|r_{ij}| \leq 1$ and $(r_{ij})^2 \leq |r_{ij}|$, and so $\|R_p\|_1^{(p)} \geq \left\{ \|R_p\|_2^{(p)} \right\}^2$. A similar argument holds if R_p is a covariance matrix and the diagonal entries of R_p , i.e., the variances, are bounded.

Definition 1—Let $\{\xi_i\}_{i=1}^\infty$ be a sequence of standard normal variables with joint normal distribution and denote the correlation matrix of $(\xi_1, \dots, \xi_p)$ by R_p . If $\|R_p\|_1^{(p)} \rightarrow 0$, or equivalently, $\|R_p\|_2^{(p)} \rightarrow 0$, then $\{\xi_i\}_{i=1}^\infty$ is called *weakly correlated*. Otherwise it is called *strongly correlated*.

2.2 Convergence of the ecdf

Let $\{Z_i\}_{i=1}^\infty$ be a sequence of standard normal variables with joint normal distribution. The ecdf is

$$\hat{F}_p(z) := \frac{1}{p} \sum_{i=1}^p I(Z_i \leq z), \quad (1)$$

where $I(\cdot)$ denotes the indicator function. Correlation does not affect the expectation $E[\hat{F}_p(z)] = \Phi(z)$ of the ecdf, but it affects its covariance. The covariance function is given in Proposition 1 of Schwartzman (2010).

The following theorem establishes that a necessary and sufficient condition for consistency of the ecdf in $\mathcal{L}_2$ is that $\{Z_i\}_{i=1}^\infty$ is weakly correlated.

Theorem 1—Let $Z_1, Z_2, \dots, Z_p, \dots$ be $N(0, 1)$ variables with ecdf (1) and let R_p denote the correlation matrix of $(Z_1, \dots, Z_p)$.

- i. (Sufficiency) If $\{Z_i\}_{i=1}^\infty$ is weakly correlated, then $F_p(\hat{z})$ converges to $\Phi(z)$ in $\mathcal{L}_2$ uniformly:

$$\sup_z E \left\{ \hat{F}_p(z) - \Phi(z) \right\}^2 \leq \frac{1}{4p} + C \|R_p\|_1^{(p)} \rightarrow 0,$$

where C is a universal constant.

- ii. (Necessity) If $\{Z_i\}_{i=1}^\infty$ is not weakly correlated, i.e., $\|R_p\|_2^{(p)} \not\rightarrow 0$, then $F_p(\hat{z})$ does not converge to $\Phi(z)$ in $\mathcal{L}_2$ for any $z \neq 0$, i.e.:

$$E \left\{ \hat{F}_p(z) - \Phi(z) \right\}^2 \not\rightarrow 0, \quad \forall z \neq 0$$

Notice that according to Theorem 1 the convergence is either uniform or none at all; if the correlation is strong, i.e., not weak, then for any $z \neq 0$ there is no convergence to $\Phi(z)$. Under a general correlation structure the ecdf may converge to a random function. In Section 3 below we aim at identifying and estimating this function.

2.3 Examples

It is easy to check that every Gaussian autoregressive moving average (ARMA) process is weakly correlated. So is every m -dependent Gaussian sequence with banded correlation matrix so that $r_{ij} = 0$ for $i - j > m$ and fixed finite m . This includes correlation in fixed finite blocks. In all these cases the ecdf converges to the standard normal distribution.

More generally, all Gaussian stationary ergodic processes are weakly correlated. This is because ergodicity requires that the autocorrelation function $\rho(\ell) = r_{i,i+\ell}$ satisfies $|\rho(\ell)| \rightarrow 0$ as $\ell \rightarrow \infty$, and therefore

$$\|R_p\|_1^{(p)} = \frac{1}{p^2} \sum_{i,j} |r_{ij}| = \frac{1}{p^2} \left[p + 2 \sum_{\ell=1}^{p-1} (p-\ell) |\rho(\ell)| \right] \leq \frac{1}{p} + \frac{2}{p} \sum_{\ell=1}^{p-1} |\rho(\ell)| \rightarrow 0.$$

It is not hard to check that even long-range correlation, defined by $\sum_{\ell=1}^\infty |\rho(\ell)| = \infty$, also implies weak correlation except in the extreme case where $\sum_{\ell=1}^{p-1} |\rho(\ell)|$ is of order p (the largest possible).

3 General correlation

3.1 An asymptotic approximation

To describe the asymptotic behavior of the ecdf for a general correlation structure, the main idea is to decompose the correlation into a strong correlation component and a weak

correlation component. Then, the asymptotic behavior of the ecdf as a random function will be captured by the strong correlation component, while the weak component will converge as in Theorem 1.

Specifically, suppose that for every p we have the decomposition

$$R_p = A_p + B_p, \quad (2)$$

where A_p and B_p are symmetric positive semi-definite matrices. Write A_p as

$$A_p = L_p (L_p)^T, \quad (3)$$

where L_p is of dimension $p \times k(p)$; let $\ell_i^{(p)}$ be the i 'th row of L_p . If the matrix A_p is the zero matrix then we define $k(p) := 1$, and $L_p := (0, \dots, 0)^T$.

For a matrix B , we define the matrix $\text{Cor}(B)$ by

$$\{\text{Cor}(B)\}_{ij} = \begin{cases} \frac{B_{ij}}{\sqrt{B_{ii}B_{jj}}} & B_{ii}B_{jj} \neq 0 \\ 0 & B_{ii}B_{jj} = 0 \end{cases}.$$

Notice that if B is a positive semidefinite matrix, then $|\{\text{Cor}(B)\}_{ij}| \leq 1$. Obviously, if B is a correlation matrix, then $\text{Cor}(B) = B$. The following theorem states the main result.

Theorem 2—Under the previous setting and notation,

- i. For every p there exists a (non-unique) random vector $\mathbf{W}_{k(p)}^{(p)} \sim N(0, I_{k(p)})$ such that

$$\sup_z E \left\{ \hat{F}_p(z) - \bar{F}_p(z) \right\}^2 \leq \frac{1}{4p} + C \|\text{Cor}(B_p)\|_1^{(p)}, \quad (4)$$

where

$$\bar{F}_p(z) := E[\hat{F}_p(z) | \mathbf{W}_{k(p)}^{(p)}] = \frac{1}{p} \sum_{i=1}^p \Phi \left(\frac{z - \mu_i^{(p)}}{\sigma_i^{(p)}} \right) \quad (5)$$

with $\mu_i^{(p)} := \ell_i^{(p)} \mathbf{W}_{k(p)}^{(p)}$, $\sigma_i^{(p)} := \{(B_p)_{ii}\}^{1/2}$, and C is a universal constant. If $\sigma_i^{(p)} = 0$

we define $\Phi \left(\frac{z - \mu_i^{(p)}}{\sigma_i^{(p)}} \right) := I(\mu_i^{(p)} \leq z)$.

- ii. Therefore, if $\|\text{Cor}(B_p)\|_1^{(p)} \xrightarrow{p \rightarrow \infty} 0$ then $\sup_z E \left\{ \hat{F}_p(z) - \bar{F}_p(z) \right\}^2 \xrightarrow{p \rightarrow \infty} 0$.

The key idea of the proof is the following: due to decomposition (2) and equality (3), there exists a random vector $\mathbf{W}_k \sim N(0, I_k)$ such that

$$Z_i = \ell_i^{(p)} \mathbf{W}_{k(p)}^{(p)} + \xi_i, \quad i=1, \dots, p, \quad (6)$$

where $(\xi_1, \dots, \xi_p) \sim N(0, B_p)$. Therefore, the conditional random vector

$$(\tilde{Z}_1, \dots, \tilde{Z}_p) := (Z_1, \dots, Z_p) | \mathbf{W}_{k(p)}^{(p)} = \mathbf{w}_k$$

is normal with conditional mean $\mu = L_p \mathbf{w}_k$ and conditional covariance matrix B_p . The $\mathcal{L}_2$ distance between $F_p(\hat{z})$ and $\bar{F}_p(z) := E[\hat{F}_p(z) | \mathbf{W}_{k(p)}^{(p)}]$ can be essentially bounded uniformly by $\|\text{Cor}(B_p)\|_1^{(p)}$.

It is important to note that the random vector $\mathbf{W}_{k(p)}^{(p)}$ is not unique. Because $\mathbf{W}_{k(p)}^{(p)}$ has a spherically symmetric distribution, Equation (4) will hold if $\mathbf{W}_{k(p)}^{(p)}$ is replaced by $Q\mathbf{W}_{k(p)}^{(p)}$ where Q is any $p \times p$ orthonormal matrix.

Definition 2—The sequence of random functions $\{G_p(z)\}_{p=1}^\infty$ is said to converge uniformly in $\mathcal{L}_2$ to the (random) function $G(z)$ if $\sup_z E\{G_p(z) - G(z)\}^2 \xrightarrow{p \rightarrow \infty} 0$.

Corollary 1—Suppose $F_p(\hat{z})$ satisfies the conditions of Theorem 2(ii). If there exists a (random) function $F(z)$ such that $F_p(z)$ converges to $F(z)$ uniformly in $\mathcal{L}_2$, then $F_p(\hat{z})$ also converges to $F_p(\bar{z})$ uniformly in $\mathcal{L}_2$.

Theorem 2 holds for all decompositions of the form (2). We call every corresponding $F_p(\bar{z})$ an *asymptotic approximation* of $F_p(\hat{z})$. However, some decompositions may need a smaller number of latent variables $k(p)$ than others. In the next section we characterize the best decomposition in the sense of giving the asymptotic approximation with the smallest number of latent variables. If $F_p(\bar{z})$ converges to some $F(\bar{z})$, then we call the latter the *asymptotic representation* of $F_p(\hat{z})$.

3.2 Dimension reduction

Theorem 2 approximates $F_p(\hat{z})$ by $F_p(\bar{z})$, which is the projection of $F_p(\hat{z})$ onto a space generated by $\mathbf{W}_{k(p)}^{(p)}$. Thus, as explained below, $F_p(\hat{z})$, which has dimension p , is approximated by $F_p(\bar{z})$ with dimension rank (A_p) . Hence, Theorem 2 can be regarded as a dimension reduction of the empirical distribution function as stated in the following proposition.

Proposition 1—Let $\mathcal{F}$ be the collection of all one-dimensional distribution functions, where convergence is defined in the sense of weak convergence of the corresponding random variables.

- i. $F_p(\hat{z})$ has dimension p : define the mapping $\mathcal{A}(Z_1, \dots, Z_p) = F_p(\hat{z})$, from $\mathbb{R}^p$ to $\mathcal{F}$. Then $\mathcal{A}(\mathbb{R}^p) \subseteq \mathcal{F}$ is homeomorphic to $\mathbb{R}^p$.

- ii. $F_p(z)$ has dimension $\text{rank}(A_p)$: define the mapping $\mathcal{B}(\mathbf{W}_{k(p)}^{(p)}) = \overline{F}_p(z)$, from $\mathbb{R}^{k(p)}$ to $\mathcal{F}$. Then $\mathcal{B}(\mathbb{R}^{k(p)}) \subseteq \mathcal{F}$ is homeomorphic to $\mathbb{R}^{\text{RANK}(A_p)}$.

In order to achieve dimension reduction, we are interested in knowing whether there exist decompositions of the form (2) such that the approximation of Theorem 2 holds and the dimension of the approximation $\text{rank}(A_p)$ is finite. Consider the collection $\mathcal{D}$ of decompositions of R_p that satisfy the conditions of Theorem 2 Part (ii): $D \in \mathcal{D}$ if

$D = \{(A_p, B_p)\}_{p=1}^\infty$ such that A_p, B_p satisfy (2) and $\|\text{Cor}(B_p)\|_1^{(p)} \xrightarrow{p \rightarrow \infty} 0$. Clearly, $\mathcal{D}$ is a large collection and some notion of optimality is required in order to decide which $D \in \mathcal{D}$ to choose. Given Proposition 1 (ii), we are interested in decompositions where $\text{rank}(A_p)$ is the smallest. Theorem 3 below states a lower bound on the limiting rank of A_p and presents a decomposition that achieves it under certain conditions.

To state the result, we need to define the eigendecompositions of R_p as a special case of the decompositions in $\mathcal{D}$. Suppose that the eigenvalues of R_p are $\lambda_1^{(p)} \geq \dots \geq \lambda_p^{(p)}$ and the corresponding eigenvectors are $\gamma_1^{(p)}, \dots, \gamma_p^{(p)}$ (notice that everything depends on p). The eigendecomposition of R_p is $\sum_{i=1}^p \lambda_i^{(p)} \gamma_i^{(p)} \{\gamma_i^{(p)}\}^T$. For $k < p$, define

$$A_{k,p} = \sum_{i=1}^k \lambda_i^{(p)} \gamma_i^{(p)} \{\gamma_i^{(p)}\}^T, \quad B_{k,p} = \sum_{i=k+1}^p \lambda_i^{(p)} \gamma_i^{(p)} \{\gamma_i^{(p)}\}^T. \quad (7)$$

Then, the covariance matrix R_p can be expressed as $R_p = A_{k,p} + B_{k,p}$ and $A_{k,p} = L_{k,p}(L_{k,p})^T$ where

$$L_{k,p} := \left(\sqrt{\lambda_1^{(p)}} \gamma_1^{(p)}, \dots, \sqrt{\lambda_k^{(p)}} \gamma_k^{(p)} \right). \quad (8)$$

A critical quantity is the number of “big” eigenvalues, i.e., with size of order p . Define

$$\underline{K} = \sum_{i=1}^\infty \underline{K}_i \quad \text{with} \quad \underline{K}_i = \begin{cases} 1 & \text{if } \liminf_p \frac{\lambda_i^{(p)}}{p} > 0 \\ 0 & \text{if } \liminf_p \frac{\lambda_i^{(p)}}{p} = 0 \end{cases}, i=1, 2, \dots$$

and

$$\overline{K} = \sum_{i=1}^\infty \overline{K}_i \quad \text{with} \quad \overline{K}_i = \begin{cases} 1 & \text{if } \limsup_p \frac{\lambda_i^{(p)}}{p} > 0 \\ 0 & \text{if } \limsup_p \frac{\lambda_i^{(p)}}{p} = 0 \end{cases}, i=1, 2, \dots$$

By definition we have that $\underline{K} \leq \overline{K}$. Notice that $\underline{K}$ could be ∞ and that if $\underline{K} < \infty$ then $\underline{K}_i = 1$ for $i \leq \underline{K}$ and $\underline{K}_i = 0$ otherwise; the same for $\overline{K}$. Since $\underline{K}$ and $\overline{K}$ are sums of indicators, then

they are either an integer or infinity. The following theorem states that $\underline{K}$ and $\bar{K}$ give lower bounds for the limiting rank of A_p and presents a decomposition of the correlation matrix that achieves it under certain conditions.

Theorem 3—Under the previous setting and notation:

- i. If $\{(A_p, B_p)\}_{p=1}^\infty \in \mathcal{D}$, then $\liminf_p \text{rank}(A_p) \geq \underline{K}$ and $\limsup_p \text{rank}(A_p) \leq \bar{K}$ (both $\underline{K}$ and $\bar{K}$ could be ∞).
- ii. For $\bar{K} < \infty$ define the decomposition $D_{\bar{K}} = \{(A_{\bar{K},p}, B_{\bar{K},p})\}_{p=1}^\infty$ according to (7) with $k = \bar{K}$. If the nonzero diagonal terms of $B_{\bar{K},p}$ are bounded from below, i.e.

$$\liminf_p \min_{1 \leq i \leq p; \{B_{\bar{K},p}\}_{ii} \neq 0} \{B_{\bar{K},p}\}_{ii} = \varepsilon_B > 0 \quad (9)$$

then $D_{\bar{K}} \in \mathcal{D}$ and obviously $\text{rank}(A_{\bar{K},p}) = \bar{K}$ for all $p > \bar{K}$.

The regularity condition (9) guarantees that the norm of the residual covariance in $B_{\bar{K},p}$ goes to zero because the correlation in it goes to zero, not because the variance in it goes to zero.

To better appreciate the result given by Theorem 3, it is useful to define the following concepts.

Definition 3

- i. If there exists $\{(A_p, B_p)\}_{p=1}^\infty \in \mathcal{D}$ such that $\limsup_p \text{rank}(A_p) < \infty$ we say that $\{Z_i\}_{i=1}^\infty$ has finite dimensional correlation.
- ii. If further $\bar{K} := \bar{K} = \bar{K} < \infty$ then we say that $\{Z_i\}_{i=1}^\infty$ has asymptotic dimension $\bar{K}$.

With these definitions in mind, Theorem 3 implies that if $\{Z_i\}_{i=1}^\infty$ has finite dimensional correlation with asymptotic dimension $\bar{K}$ and the regularity condition (9) holds, then the decomposition $\{(A_{\bar{K},p}, B_{\bar{K},p})\}_{p=1}^\infty$ is “optimal” in the sense that $\lim_p \text{rank}(A_{\bar{K},p}) = \bar{K}$, so that it achieves the lowest dimension among all asymptotic approximations $F_p(z)$ of $F_p(z)$. We shall see in Section 5 that the regularity condition (9) holds for most typical correlation structures.

It is now possible to see that the number $\hat{K}$ determines three different convergence regimes. If $\hat{K} = \infty$ then $\{Z_i\}_{i=1}^\infty$ has no finite dimensional correlation. If $\bar{K} < \infty$ and (9) holds then $\{Z_i\}_{i=1}^\infty$ has finite dimensional correlation. When $\bar{K} = 0$ then (9) holds trivially since $\{B_{0,p}\}_{ii} = \{R_p\}_{ii} = 1$ and therefore a necessary and sufficient conditions for weak correlation, as in Theorem 1, is $\bar{K} = 0$. We summarize the results in the following corollary and in Figure 2.

Corollary 2

- i. If $\bar{K} = \infty$ then $\{Z_i\}_{i=1}^\infty$ has no finite dimensional correlation.
- ii. If $\bar{K} < \infty$ and (9) holds then $\{Z_i\}_{i=1}^\infty$ has finite dimensional correlation.

- iii. If $K := \underline{K} = K < \infty$ and (9) holds then $\{Z_i\}_{i=1}^\infty$ has asymptotic dimension K and the decomposition $\{(A_{K,p}, B_{K,p})\}_{p=1}^\infty$ achieves it in the sense that $\lim_p \text{rank}(A_{K,p}) = K$.
- iv. $K = 0$ if and only if $\{Z_i\}_{i=1}^\infty$ is weakly correlated.

The regularity condition (9) does not hold when the diagonal of $B_{\bar{K},p}$ contains elements that are arbitrarily small. In this case, captured by the label “unknown” in Figure 2, $\|B_{\bar{K},p}\|_1^{(p)}$ may be small but not necessarily $\|\text{Cor}(B_{\bar{K},p})\|_1^{(p)}$. Since our bound (4) is based on $\|\text{Cor}(B_{\bar{K},p})\|_1^{(p)}$, rather than on $\|B_{\bar{K},p}\|_1^{(p)}$, we cannot establish the asymptotic dimension in this case. However, we could not find an example of R_p , where (9) is not satisfied. A hand-waving argument that (9) typically holds is:

$$\{B_{\bar{K},p}\}_{ii} = \sum_{j=\bar{K}+1}^p \lambda_j^{(p)} \{\gamma_j^{(p)}\}_i^2 \geq \lambda_{\bar{K}+1}^{(p)} \sum_{j=\bar{K}+1}^p \{\gamma_j^{(p)}\}_i^2 \approx \lambda_{\bar{K}+1}^{(p)} > 0,$$

where $\approx$ is typically true because $\gamma_j^{(p)}$ is a normalized vector and therefore $\{\gamma_j^{(p)}\}_i^2 \approx 1/p$.

4 Estimating the asymptotic representation from the data

4.1 Estimating the latent variables

We now discuss how to estimate the underlying latent variables $\mathbf{W}_K^{(p)}$ when the sequence $\{Z_i\}_{i=1}^\infty$ has asymptotic dimension $K := \underline{K} = K$. For ease of presentation we write everything in a matrix/vector form; thus, we write $\mathbf{Z}_p := (Z_1, \dots, Z_p)^T$, $\boldsymbol{\xi}_p = (\xi_1, \dots, \xi_p)^T$. In this notation, (6) becomes the linear regression equation $\mathbf{Z}_p = L_{K,p} \mathbf{W}_K^{(p)} + \boldsymbol{\xi}_p$ and the least square estimate of $\mathbf{W}_K^{(p)}$ is $\hat{\mathbf{W}}_K^{(p)} := (L_{K,p}^T L_{K,p})^{-1} L_{K,p}^T \mathbf{Z}_p$.

When the eigendecomposition (7) is used, then the columns of $L_k(p)$ are orthogonal and $(L_{K,p}^T L_{K,p})^{-1}$ is a diagonal matrix whose i -th diagonal element is $1/\lambda_i^{(p)}$. Thus, in this case,

$$\hat{\mathbf{W}}_K^{(p)} = \begin{pmatrix} \frac{1}{\sqrt{\lambda_1^{(p)}}} \{\gamma_1^{(p)}\}^T \mathbf{Z}_p \\ \vdots \\ \frac{1}{\sqrt{\lambda_K^{(p)}}} \{\gamma_K^{(p)}\}^T \mathbf{Z}_p \end{pmatrix}. \quad (10)$$

Fan et al. (2012) and Fan and Han (2014) consider a related framework where some of the variables Z_i may have a non-zero mean, and therefore use an estimate that minimizes the $\mathcal{L}_1$ distance under sparsity assumptions.

Since $\{Z_i\}_{i=1}^p$ has strong correlation, then $\text{Cov}(\hat{\mathbf{W}}_K^{(p)})$ does not converge to 0, as p goes to infinity. However, the following proposition states that $\hat{\mathbf{W}}_K^{(p)}$ is still consistent in an $\mathcal{L}_2$ sense.

Proposition 2—Suppose that $\{Z_i\}_{i=1}^\infty$ has asymptotic dimension K . Then

$$E \left[\left\{ \hat{\mathbf{W}}_K^{(p)} - \mathbf{W}_K^{(p)} \right\}^T \left\{ \hat{\mathbf{W}}_K^{(p)} - \mathbf{W}_K^{(p)} \right\} \right] \leq K \frac{p}{\lambda_K^{(p)}} \|B_{K,p}\|_2^{(p)} \xrightarrow{p \rightarrow \infty} 0.$$

4.2 Estimating the asymptotic representation

In real data problems $\mathbf{W}_K^{(p)}$ and, thus, $F_p(\cdot)$ are unknown. Proposition 2 suggests that if we plug-in $\hat{\mathbf{W}}_K^{(p)}$ to $F_p(\cdot)$ instead of $\mathbf{W}_K^{(p)}$, then the $\mathcal{L}_2$ distance between $F_p(\hat{\cdot})$ and the plugged-in $F_p(\cdot)$ converges to zero. Indeed, this can be proved under some additional regularity conditions.

Theorem 4

i. Define the plug-in ecdf estimate

$$\hat{F}_p(z) := \frac{1}{p} \sum_{i=1}^p \Phi \left(\frac{z - \hat{\mu}_i^{(p)}}{\sigma_i^{(p)}} \right) \quad (11)$$

with $\hat{\mu}_i^{(p)} := \ell_i^{(p)} \hat{\mathbf{W}}_K^{(p)}$ under the convention $\Phi \left(\frac{z - \hat{\mu}_i^{(p)}}{0} \right) := I \left(\mu_i^{(p)} \leq z \right)$. We have that

$$\begin{aligned} \sup_z E \left\{ \hat{F}_p(z) - \bar{F}_p(z) \right\}^2 &\leq 3 \sup_z E \left\{ \hat{F}_p(z) - F_p(z) \right\}^2 + \frac{3(p - |J_p|)}{p} + \\ &+ \frac{3}{2\pi} \max_{i \in J_p} \frac{1}{\{\sigma_i^{(p)}\}^2} E \left[\left\{ \hat{\mathbf{W}}_K^{(p)} - \mathbf{W}_K^{(p)} \right\}^T \left\{ \hat{\mathbf{W}}_K^{(p)} - \mathbf{W}_K^{(p)} \right\} \right], \end{aligned}$$

where $J_p := \{1 \leq i \leq p : \sigma_i^{(p)} > 0\}$ is the set of indexes for which $\sigma_i^{(p)} > 0$ and $|J_p|$ is the cardinality of J_p .

ii. Therefore, if $\{Z_i\}_{i=1}^\infty$ has asymptotic dimension K , (9) holds and also

$$\frac{|J_p|}{p} \xrightarrow{p \rightarrow \infty} 1 \quad (12)$$

then $\sup_z E \left\{ \hat{F}_p(z) - \bar{F}_p(z) \right\}^2 \xrightarrow{p \rightarrow \infty} 0$.

4.3 Approximated eigendecomposition

In this section we study the case in which the largest K eigenvalues and eigenvectors of the correlation matrix are not known exactly and an approximation is used. We show that if the distance between the approximated eigenvalues and eigenvectors and the true ones goes to zero as p goes to infinity, then the result of the previous section still holds.

Let $\tilde{\lambda}_1^{(p)} \geq \dots \geq \tilde{\lambda}_K^{(p)}$ be an approximation to the K biggest eigenvalues, and let $\tilde{\gamma}_1^{(p)}, \dots, \tilde{\gamma}_K^{(p)}$ be norm 1 vectors that are an approximation to the corresponding eigenvectors. Define the $p \times K$ matrix $\tilde{L}_p := \left(\sqrt{\tilde{\lambda}_1^{(p)}} \tilde{\gamma}_1^{(p)}, \dots, \sqrt{\tilde{\lambda}_K^{(p)}} \tilde{\gamma}_K^{(p)} \right)$, and let $\tilde{\ell}_i^{(p)}$ be the i -th row of the matrix. Define,

$$\tilde{\mathbf{W}}_K^{(p)} := \begin{pmatrix} \frac{1}{\sqrt{\tilde{\lambda}_1^{(p)}}} \left\{ \tilde{\gamma}_1^{(p)} \right\}^T \mathbf{Z}_p \\ \vdots \\ \frac{1}{\sqrt{\tilde{\lambda}_K^{(p)}}} \left\{ \tilde{\gamma}_K^{(p)} \right\}^T \mathbf{Z}_p \end{pmatrix};$$

and let $\tilde{\mu}_i^{(p)} := \tilde{\ell}_i^{(p)} \tilde{\mathbf{W}}_K^{(p)}$ and $\tilde{\sigma}_i^{(p)} := \left\{ 1 - \sum_{j=1}^K \lambda_j^{(p)} \left\{ \gamma_j^{(p)} \right\}_i^2 \right\}^{1/2}$, where $\{\mathbf{x}\}_i$ is the i -th element of the vector $\mathbf{x}$. Finally, define $\tilde{F}_p(z) := \frac{1}{p} \sum_{i=1}^p \Phi \left(\frac{z - \tilde{\mu}_i^{(p)}}{\tilde{\sigma}_i^{(p)}} \right)$. The following proposition bounds the $\mathcal{L}_2$ distance between $F_p(z)$ and $\hat{F}_p(z)$.

Proposition 3—Suppose that $\sigma_i^{(p)}, \tilde{\sigma}_i^{(p)} \geq \varepsilon_B$, $i = 1, \dots, p$, for some $\varepsilon_B > 0$.

i. The following inequality holds:

$$\begin{aligned} & E \left\{ \tilde{F}_p(z) - \hat{F}_p(z) \right\}^2 \\ & \leq \frac{C(\varepsilon_B)}{\lambda_K^{(p)/p} \sum_{i=1}^K} \left[\left\{ \frac{1}{\sqrt{\tilde{\lambda}_i^{(p)}/p}} - \frac{1}{\sqrt{\lambda_i^{(p)}/p}} \right\}^2 + \left\{ \sqrt{\frac{\tilde{\lambda}_i^{(p)}}{p}} - \sqrt{\frac{\lambda_i^{(p)}}{p}} \right\}^2 + 2 \left\{ \tilde{\gamma}_i^{(p)} - \gamma_i^{(p)} \right\}^T \left\{ \tilde{\gamma}_i^{(p)} - \gamma_i^{(p)} \right\} \right] \\ & \quad + \frac{C(\varepsilon_B)}{\sqrt{p}} \left[\sum_{i=1}^K \left\{ \sqrt{\frac{\tilde{\lambda}_i^{(p)}}{p}} - \sqrt{\frac{\lambda_i^{(p)}}{p}} \right\}^2 + \sum_{i=1}^K \left\{ \tilde{\gamma}_i^{(p)} - \gamma_i^{(p)} \right\}^T \left\{ \tilde{\gamma}_i^{(p)} - \gamma_i^{(p)} \right\} \right]^{1/2}, \end{aligned}$$

where $C(\varepsilon_B)$ is a constant that depends on ε_B .

ii. Therefore, if $\frac{\tilde{\lambda}_i^{(p)} - \lambda_i^{(p)}}{p} \xrightarrow{p \rightarrow \infty} 0$, $\left\{ \tilde{\gamma}_i^{(p)} - \gamma_i^{(p)} \right\}^T \left\{ \tilde{\gamma}_i^{(p)} - \gamma_i^{(p)} \right\} \xrightarrow{p \rightarrow \infty} 0$, $i = 1, \dots, K$, and the conditions of Theorem 4 (ii) hold, then $\sup_z E \left\{ \tilde{F}_p(z) - \hat{F}_p(z) \right\}^2 \xrightarrow{p \rightarrow \infty} 0$.

Proposition 3 and its proof imply that if there exist consistent estimates for the first K eigenvalues and eigenvectors, then the result of Theorem 4 (ii) still holds; see also Fan and

Han (2014). A systematic study of the case that the correlation matrix is unknown is left to future research.

5 Examples

We now consider some examples of correlation structures. In Examples 1–3, $\underline{K} = \bar{K}$ is finite, conditions (9,12) are met, and therefore the decomposition $\{(A_{\underline{K},p}, B_{\underline{K},p})\}_{p=1}^{\infty}$ is optimal and $\hat{\bar{F}}_p$ is consistent, but in Example 3, $F_p(z)$ does not have a limit. In Example 4, $\underline{K} = 1$ but $\bar{K} > 1$ and in Example 5, $\underline{K} = \infty$.

5.1 Exchangeable correlation: asymptotic dimension 1

Suppose that $\{Z_i\}_{i=1}^{\infty}$ is a sequence of exchangeable random variables with correlation $\rho > 0$, i.e., for any $i \neq j$, $\text{cor}(Z_i, Z_j) = \rho$ and

$$R_p = \begin{pmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & \dots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \dots & \rho & 1 \end{pmatrix}. \quad (13)$$

The eigenvalues of R_p are $\lambda_1^{(p)} = \rho p + 1 - \rho$ (with corresponding eigenvector equal to $(1, \dots, 1)^T / \sqrt{p}$) and $\lambda_2^{(p)} = \dots = \lambda_p^{(p)} = 1 - \rho$, hence, $\underline{K} = \bar{K} = 1$. According to (7), we have that $A_{1,p} = [\rho + (1 - \rho)/p] \mathbb{1}_p$ and $B_{1,p} = (1 - \rho) [I_p - (1/p) \mathbb{1}_p]$, where I_p is the $p \times p$ identity matrix and $\mathbb{1}_p$ is a $p \times p$ matrix with 1 in every entry. It is easy to check that for every i , $\lim_p \{B_{1,p}\}_{ii} = 1 - \rho > 0$, and (9) holds. Also, $|J_p| = p$ and (12) holds. Thus, in this case $\{Z_i\}_{i=1}^{\infty}$ has finite dimensional correlation with asymptotic dimension $\underline{K} = 1$ and the decomposition $\{(A_{1,p}, B_{1,p})\}_{p=1}^{\infty}$ is optimal.

To write the asymptotic representation $F(z)$ in this case, it is easier to work with the asymptotically equivalent decomposition $A_p = \rho \mathbb{1}_p$, $B_p = (1 - \rho)I_p$, which is in D and $\text{rank}(A_p) = \underline{K} = 1$. For this decomposition $\mu_i^{(p)} = \sqrt{\rho}W_1$ and $\sigma_i^{(p)} = \sqrt{1 - \rho}$ do not depend on i or p and

$$\bar{F}_p(z) = \frac{1}{p} \sum_{i=1}^p \Phi \left(\frac{z - \mu_i^{(p)}}{\sigma_i^{(p)}} \right) = \frac{1}{p} \sum_{i=1}^p \Phi \left(\frac{z - \sqrt{\rho}W_1}{\sqrt{1 - \rho}} \right) = \Phi \left(\frac{z - \sqrt{\rho}W_1}{\sqrt{1 - \rho}} \right) = \bar{F}(z). \quad (14)$$

That is, $\hat{F}_p(z)$, which has dimension p , is approximated by $\bar{F}_p(z)$ with dimension 1. Moreover, $\bar{F}_p(z) = \bar{F}(z)$ above does not depend on p and is therefore the asymptotic representation: $\hat{F}_p(z)$ converges to $\bar{F}(z)$ in the sense of Definition 2.

Given the sequence $\{Z_i\}_{i=1}^p$, the regression estimate (10) of W_1 is $\hat{W}_1^{(p)} = \bar{Z}_p / \sqrt{\rho}$, where $\bar{Z}_p := \frac{1}{p} \sum_{i=1}^p Z_i$. An illustration of the estimated asymptotic representation $\hat{F}_p(z)$ in (11) is shown in Figure 1(a).

Notice that the distribution specified by (14) is a shifted and scaled normal distribution, corresponding to Efron's empirical null model (Efron, 2004, 2007a,b, 2008). Fitting an empirical null in this case would amount to estimating W_1 and ρ . This example shows that Efron's empirical null model is justified under exchangeable correlation. However, as it can be seen from the above theory and the following examples, it does not capture the effect of many other correlation structures.

When $\rho = 0$ in (13), positive semi-definiteness of the correlation matrix requires $\rho \geq -1/(p-1)$. Therefore, in the high-dimensional setting, too much negative correlation is not possible. Consequently, the histograms of the Z 's are typically narrower than $N(0, 1)$ (see Figure 1).

5.2 Two exchangeable correlation blocks: asymptotic dimension 2

Suppose that R_p consists of two blocks of the form (13), the first one of size $n_1(p) \times n_1(p)$ and the second of size $n_2(p) \times n_2(p)$ with $n_1(p) + n_2(p) = p$ and $\lim_{p \rightarrow \infty} n_1(p)/p = \pi$. We assume that there is a constant correlation between the two blocks ρ_B (which can be zero or negative). Thus,

$$R_p = \begin{pmatrix} \underbrace{\begin{pmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & \ddots & \rho_1 \\ \rho_1 & \rho_1 & 1 \end{pmatrix}}_{n_1(p)} & \begin{pmatrix} \rho_B \\ \rho_B \\ \rho_B \end{pmatrix} \\ \begin{pmatrix} \rho_B \\ \rho_B \\ \rho_B \end{pmatrix} & \underbrace{\begin{pmatrix} 1 & \rho_2 & \rho_2 \\ \rho_2 & \ddots & \rho_2 \\ \rho_2 & \rho_2 & 1 \end{pmatrix}}_{n_2(p)} \end{pmatrix}.$$

Let $I_1(p)$ and $I_2(p)$ denote the set of indexes that belong to Blocks 1 and 2. If

$0 \leq \rho_B < \sqrt{\rho_1 \rho_2}$, then a sequence $\{Z_i\}_{i=1}^\infty$ with the above covariance can be generated by the model

$$Z_i = \begin{cases} \sqrt{\rho_1}(\sqrt{\tilde{\rho}_B}W_1 + \sqrt{1-\tilde{\rho}_B}W_2) + \sqrt{1-\rho_1}\varepsilon_i & i \in I_1(p) \\ \sqrt{\rho_2}(\sqrt{\tilde{\rho}_B}W_1 + \sqrt{1-\tilde{\rho}_B}W_3) + \sqrt{1-\rho_2}\varepsilon_i & i \in I_2(p) \end{cases},$$

where $\tilde{\rho}_B = \rho_B / \sqrt{\rho_1 \rho_2}$, and $W_1, W_2, W_3, \{\varepsilon_i\}_{i=1}^p$ are i.i.d $N(0, 1)$. Comparing with (6), we can write

$$\begin{aligned}\bar{F}_p(z) &= \frac{n_1(p)}{p} \Phi \left(\frac{z - \sqrt{\rho_1}(\sqrt{\tilde{\rho}_B} W_1 + \sqrt{1 - \tilde{\rho}_B} W_2)}{\sqrt{1 - \rho_1}} \right) + \frac{n_2(p)}{p} \Phi \left(\frac{z - \sqrt{\rho_2}(\sqrt{\tilde{\rho}_B} W_1 + \sqrt{1 - \tilde{\rho}_B} W_3)}{\sqrt{1 - \rho_2}} \right) \\ &\rightarrow \bar{F}(z) = \pi \Phi \left(\frac{z - \sqrt{\rho_1}(\sqrt{\tilde{\rho}_B} W_1 + \sqrt{1 - \tilde{\rho}_B} W_2)}{\sqrt{1 - \rho_1}} \right) + (1 - \pi) \Phi \left(\frac{z - \sqrt{\rho_2}(\sqrt{\tilde{\rho}_B} W_1 + \sqrt{1 - \tilde{\rho}_B} W_3)}{\sqrt{1 - \rho_2}} \right).\end{aligned}$$

If $\rho_B = 0$, then the asymptotic approximation $\bar{F}(z)$ depends only on W_2, W_3 and has dimension $\underline{K} = K = 2$. In this case R_p has two uncorrelated blocks of the type presented in Section 5.1; the generalization of the formulas there is straightforward (see also Figure 1(b)).

For $\rho_B \neq 0$, however, the above $\bar{F}(z)$ depends on W_1, W_2, W_3 and, thus, has dimension three; it is suboptimal. To obtain the optimal asymptotic representation in more generality, we work with the eigendecomposition. The matrix R_p has two eigenvalues

$$\lambda_{1,2}^{(\rho)} := \frac{\eta_1^{(p)} + \eta_2^{(p)}}{2} \pm \left[\left(\frac{\eta_1^{(p)} - \eta_2^{(p)}}{2} \right)^2 + n_1(p)n_2(p)\rho_B^2 \right]^{1/2}$$

where $\eta_j^{(p)} = 1 + \rho_j \{n_j(p) - 1\}$, $j = 1, 2$, and the other $p - 2$ eigenvalues are either $1 - \rho_1$ or $1 - \rho_2$. For large p , positive semi-definiteness requires that $\rho_B^2 \leq \rho_1 \rho_2$. On the boundary, when $\rho_B^2 = \rho_1 \rho_2$, $\lambda_2^{(p)}$ converges to a positive constant and $\underline{K} = K = 1$.

We now concentrate on the case $\rho_B^2 < \rho_1 \rho_2$. Then, $\lambda_{1,2}^{(p)}$ are of order p and $\underline{K} = K = 2$. The corresponding eigenvectors are

$$\underbrace{(x_j^{(p)}, x_j^{(p)}, \dots, x_j^{(p)})}_{n_1(p) \text{ times}}, \underbrace{(y_j^{(p)}, y_j^{(p)}, \dots, y_j^{(p)})}_{n_2(p) \text{ times}}^T, \quad j=1, 2,$$

where $x_j^{(p)}$ and $y_j^{(p)}$ are given by

$$\begin{aligned}x_1^{(p)} &= \frac{-\sqrt{n_2(p)\rho_B}}{\sqrt{(\eta_1^{(p)} - \lambda_1^{(p)})^2 + n_1(p)n_2(p)\rho_B^2}}, & y_1^{(p)} &= \frac{(\eta_1^{(p)} - \lambda_1^{(p)})/\sqrt{n_2(p)}}{\sqrt{(\eta_1^{(p)} - \lambda_1^{(p)})^2 + n_1(p)n_2(p)\rho_B^2}}, \\ x_2^{(p)} &= \frac{(\eta_2^{(p)} - \lambda_2^{(p)})/\sqrt{n_1(p)}}{\sqrt{(\eta_2^{(p)} - \lambda_2^{(p)})^2 + n_1(p)n_2(p)\rho_B^2}}, & y_2^{(p)} &= \frac{-\sqrt{n_1(p)\rho_B}}{\sqrt{(\eta_2^{(p)} - \lambda_2^{(p)})^2 + n_1(p)n_2(p)\rho_B^2}}.\end{aligned}$$

We have that

$$\mu_i^{(p)} = \begin{cases} \sqrt{\lambda_1^{(p)}} x_1^{(p)} W_1 + \sqrt{\lambda_2^{(p)}} x_2^{(p)} W_2, & i \in I_1(p) \\ \sqrt{\lambda_1^{(p)}} y_1^{(p)} W_1 + \sqrt{\lambda_2^{(p)}} y_2^{(p)} W_2, & i \in I_2(p), \end{cases}$$

and

$$\sigma_i^{(p)} = \begin{cases} \sqrt{1 - [\lambda_1^{(p)} \{x_1^{(p)}\}^2 + \lambda_2^{(p)} \{x_2^{(p)}\}^2]}, & i \in I_1(p) \\ \sqrt{1 - [\lambda_1^{(p)} \{y_1^{(p)}\}^2 + \lambda_2^{(p)} \{y_2^{(p)}\}^2]}, & i \in I_2(p). \end{cases}$$

Moreover, $\{B_{2,p}\}_{jii} \xrightarrow{p \rightarrow \infty} 1 - \rho_1$ for $i \in I_1(p)$ and $\{B_{2,p}\}_{jii} \xrightarrow{p \rightarrow \infty} 1 - \rho_2$ for $i \in I_2(p)$. Thus, (9) holds with $K = \underline{K} = 2$. Also, $|J_p| = p$ and (12) holds. The corresponding asymptotic representation can be obtained by writing $F_p^-(z)$ in terms of the $\mu_i^{(p)}$ and $\sigma_i^{(p)}$ above and taking the limit as $p \rightarrow \infty$.

The regression estimate of $W_j^{(p)}$ is

$$\hat{W}_j^{(p)} = \left\{ x_j(p) \sum_{i \in I_1(p)} Z_i + y_j(p) \sum_{i \in I_2(p)} Z_i \right\} / \sqrt{\lambda_j^{(p)}} \quad j=1, 2.$$

5.3 Finite dimensional correlation with no asymptotic representation

In this example, correlation is finite with $K = 2$ and the $\mathcal{L}_2$ distance between F_p^+ and F_p^- goes to zero, but F_p^- itself has no limit. Let $0 < \rho < 1$ and $0 < \{\alpha_i\}_{i=1}^\infty < 1$ be any sequence, $W_1, W_2, \{\varepsilon_i\}_{i=1}^\infty$ independent standard random variables. Define

$$Z_i = \sqrt{\rho\alpha_i} W_1 + \sqrt{\rho(1-\alpha_i)} W_2 + \sqrt{1-\rho} \varepsilon_i;$$

then, $Z_i \sim N(0, 1)$ and $\text{Cov}(Z_i, Z_j) = \rho \{ \sqrt{\alpha_i \alpha_j} + \sqrt{(1-\alpha_i)(1-\alpha_j)} \}$ for $i \neq j$. We can write $R_p = A_p + B_p$, where $A_p = L_p(L_p)^T$,

$$L_p = \begin{pmatrix} \sqrt{\rho\alpha_1} & \sqrt{\rho(1-\alpha_1)} \\ \sqrt{\rho\alpha_2} & \sqrt{\rho(1-\alpha_2)} \\ \vdots & \vdots \\ \sqrt{\rho\alpha_p} & \sqrt{\rho(1-\alpha_p)} \end{pmatrix}, \text{ and } B_p = \begin{pmatrix} 1-\rho & 0 & \dots & 0 \\ 0 & 1-\rho & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & 1-\rho \end{pmatrix}.$$

Then, $\bar{F}_p(z) = \Phi\left(\frac{z - \mu_i}{\sqrt{1-\rho}}\right)$, where $\mu_i := \sqrt{\rho\alpha_i}W_1 + \sqrt{\rho(1-\alpha_i)}W_2$ and it is clear that for certain choices of $0 < \{\alpha_i\}_{i=1}^\infty < 1$, $F_p(z)$ has no limit for any z .

5.4 Independent exchangeable correlation blocks with no asymptotic dimension

In this example there are two independent blocks similar to the example in Section 5.2 (with $\rho_B = 0$) and with the same notation, but here $\frac{n_1(p)}{p}, \frac{n_2(p)}{p}$ change infinitely often. Consequently, there are infinitely many p 's with two "big" eigenvalues and infinitely many p 's with only one "big" eigenvalue. Therefore, in this example, $\underline{K} = 1$ while $\bar{K} = 2$. This means that correlation is finite dimensional and (9) holds, but there is no asymptotic representation and no well defined asymptotic dimension.

Consider the subsequence $m_1 = 2, m_{k+1} = m_k^2$ for $k \geq 2$; i.e., $m_k = 2^{(2^{k-1})}$. Suppose that Z_i belongs to the first block if $i \in \{m_{k-1} + 1, \dots, m_k\}$ for even k , otherwise it belongs to the second block. That is, at $p = m_k$ for even k ,

$$n_1(p) = n_1(m_{k-1}) + m_k - m_{k-1} = n_1(\sqrt{p}) + p - \sqrt{p}, \quad \text{and } n_2(p) = n_2(m_{k-1})$$

and for odd k

$$n_1(p) = n_1(m_{k-1}), \quad \text{and } n_2(p) = n_2(m_{k-1}) + m_k - m_{k-1} = n_2(\sqrt{p}) + p - \sqrt{p}.$$

When $p = m_k$ for even k , $m_k - \sqrt{m_k} \leq n_1(m_k) \leq m_k$, and therefore the eigenvalue associated with this block satisfies $\lim_{k \text{ even}} \frac{\rho n_1(m_k) + 1 - \rho}{m_k} = \rho$ but for odd k , $n_1(m_k) \leq m_{k-1} = \sqrt{m_k}$ and $\lim_{k \text{ odd}} \frac{\rho n_1(m_k) + 1 - \rho}{m_k} = 0$; and vice versa for the other block. The largest eigenvalue satisfies

$$\frac{\lambda_1^{(p)}}{p} = \max\left\{\frac{\rho n_1(p) + 1 - \rho}{p}, \frac{\rho n_2(p) + 1 - \rho}{p}\right\} \geq \frac{\rho p/2 + 1 - \rho}{p}$$

and therefore $\liminf_p \frac{\lambda_1^{(p)}}{p} \geq \rho/2 > 0$ and $\underline{K}_1 = 1$. For the second eigenvalue

$$\frac{\lambda_2^{(p)}}{p} = \min\left\{\frac{\rho n_1(p) + 1 - \rho}{p}, \frac{\rho n_2(p) + 1 - \rho}{p}\right\};$$

since $\liminf_p$ of both terms is 0, then $\liminf_p \frac{\lambda_2^{(p)}}{p} = 0$ and $\underline{K}_2 = 0$. Thus, $\underline{K} = 1$.

We next show that $\bar{K} = 2$. When $p = m_k$ one block is larger than the other block and when $p = m_{k+1}$ the other block is larger. Therefore, for each k there exists p_k between m_k and m_{k+1} such that the two blocks are equal, i.e., $n_1(p_k) = n_2(p_k) = p_k/2$. Thus,

$$\frac{\lambda_2^{(p_k)}}{p_k} = \min \left\{ \frac{\rho n_1(p_k) + 1 - \rho}{p_k}, \frac{\rho n_2(p_k) + 1 - \rho}{p_k} \right\} = \rho/2 + \frac{1 - \rho}{p_k};$$

Therefore, $\limsup_p \frac{\lambda_2^{(p)}}{p} \geq \lim_k \frac{\lambda_2^{(p_k)}}{p_k} = \rho/2 > 0$. Since the other eigenvalues are constant $(1 - \rho)$ we have that $\bar{K} = 2$.

Moreover, in this case $\|B_{1,p}\|_2^{(p)}$ does not converge to 0 and therefore $\|\text{Cor}(B_{1,p})\|_2^{(p)} \not\rightarrow 0$; hence, the decomposition $\{(A_{1,p}, B_{1,p})\}$ is not in $\mathcal{D}$. This is because

$$\|B_{1,p_k}\|_2^{(p_k)} = \frac{\left[\{\lambda_2^{(p_k)}\}^2 + \{\lambda_3^{(p_k)}\}^2 + \dots + \{\lambda_{p_k}^{(p_k)}\}^2 \right]^{1/2}}{p_k} \geq \frac{\lambda_2^{(p_k)}}{p_k} = \rho/2 + \frac{1 - \rho}{p_k},$$

so that $\liminf_k \|\text{Cor} B_{1,p_k}\|_2^{(p_k)} \geq \liminf_k \|B_{1,p_k}\|_2^{(p_k)} \geq \rho/2$.

This example can be easily generalized such that $\bar{K} = 3, 4, \dots$ or $\bar{K} = \infty$. Given $M < \infty$, to obtain $\bar{K} = M$, for even k one can divide the observations $Z_{m_k}, \dots, Z_{m_{k+1}}$ into $M - 1$ independent blocks with correlation ρ within blocks. To obtain $\bar{K} = \infty$, for even k one can divide the observations $Z_{m_k}, \dots, Z_{m_{k+1}}$ into k blocks.

5.5 No finite dimensional approximation

In this example, $\bar{K} = K = \infty$, and therefore, according to Theorem 3, for every decomposition in $\mathcal{D}$, $\lim_p \text{rank}(A_p) = \infty$. Suppose that at $p = 2^n$, R_p consists of n independent blocks of sizes of $2^{n-1}, 2^{n-2}, \dots, 2$ and an additional block of size 2, each block of the form (13), all with the same correlation parameter ρ . At $p = 2^n$, there are $n - 1$ “big” eigenvalues, $\rho(p/2^i - 1) + 1$ for $i = 1, \dots, n - 1$, one eigenvalue $1 + \rho$ (from the last block of size 2) and the rest are equal to $1 - \rho$. For any fixed $i \in \mathbb{N}$, for large enough $p \in \{2^n\}_{n \in \mathbb{N}}$ the eigenvalue $\lambda_i^{(p)}$ is equal to $\rho(p/2^i - 1) + 1$ and therefore for large enough $p \in \{2^n\}_{n \in \mathbb{N}}$,

$$\frac{\lambda_i^{(p)}}{p} = \frac{\rho(p/2^i - 1) + 1}{p} \xrightarrow{p \rightarrow \infty} \rho/2^i > 0.$$

Therefore, for any fixed $i \in \mathbb{N}$, $\underline{K}_i = 1$ and $\underline{K} = \sum_{i=1}^{\infty} \underline{K}_i = \infty$.

6 Data example

As a practical application, we use the methods developed above to analyze a high-dimensional data set obtained from brain imaging. The data belongs to a study of cortical thickness of adults who had a diagnosis of attention deficit/hyperactivity disorder (ADHD) as children (Proal et al., 2011). The data set consists of cortical thickness measurements for about 80000 cortical voxels, obtained from magnetic resonance imaging (MRI) scans, as well as demographic and behavioral measurements, for each of $n = 139$ individuals. In this study, it had been noticed by Reiss et al. (2012) that z-scores corresponding to the voxelwise relationship between cortical thickness and ADHD diagnosis did not follow the theoretical standard normal distribution. Instead, the distribution of z-scores exhibited a substantial shift away from zero, indicating a possible widespread cortical thinning over the brain for individuals with ADHD. It is unclear, however, whether those results could have been caused by correlation between voxels rather than by a real relationship with clinical diagnosis.

In order to apply the methods developed in this paper, here we perform a slightly different analysis where the correlation structure can be taken as known. Specifically, we follow the approach of Owen (2005) of performing a regression of the observed trait for the subjects on the high-dimensional predictors, one dimension at a time. Owen (2005) and also Fan et al. (2012) used this approach in the context of genomic data. In our case, the trait is a global assessment of behavior, while the predictors are the cortical thickness measurements. For ease of computation, in the following analysis we use a random sample of $p = 1000$ voxels, which is enough to show the effect we want to highlight.

6.1 Regression analysis

Let Y_j denote the global assessment of the j -th subject and let $\mathbf{X}_j = \{X_j^{(i)}\}_{i=1}^p$ be the cortical thickness in the p voxels of the j -th subject. For each voxel $i \in \{1, \dots, p\}$, consider the simple regression model

$$Y_j = \alpha^{(i)} + \beta^{(i)} X_j^{(i)} + \varepsilon_j \quad j=1, \dots, n,$$

where the ε 's are i.i.d with mean 0 and variance σ^2 . Let $\tilde{Y}_j = Y_j - \frac{1}{n} \sum_{j=1}^n Y_j$ and $\tilde{X}_j^{(i)} = X_j^{(i)} - \frac{1}{n} \sum_{j=1}^n X_j^{(i)}$ be the centered variables and define $s_{i\ell} = \sum_{j=1}^n \tilde{X}_j^{(i)} \tilde{X}_j^{(\ell)}$. The least squares estimate of $\beta^{(i)}$ is $\hat{\beta}^{(i)} = \sum_{j=1}^n \tilde{X}_j^{(i)} \tilde{Y}_j / s_{ii}$. According to the model, we can write $\hat{\beta}^{(i)} = \beta^{(i)} + \sum_{j=1}^n \tilde{X}_j^{(i)} \varepsilon_j / s_{ii}$ and so $\text{Cov}(\hat{\beta}^{(i)}, \hat{\beta}^{(\ell)}) = \sigma^2 s_{i\ell} / [s_{ii} s_{\ell\ell}]$.

Consider the p hypotheses $H_0^{(i)}: \beta^{(i)} = 0$ versus $H_1^{(i)}: \beta^{(i)} \neq 0$, $i = 1, \dots, p$. The z-score for the i -th test is $Z_i = \hat{\beta}^{(i)} / \text{se}[\hat{\beta}^{(i)}] = \hat{\beta}^{(i)} / \sqrt{s_{ii}} / \sigma$. Thus, under the global null hypothesis, i.e. that all $H_0^{(i)}$'s are true, we have that $Z_1, \dots, Z_p$ are each approximately $N(0, 1)$ with correlation

matrix $R := \{r_{i\ell}\}_{i,\ell=1}^p$ given by the pairwise correlations $r_{i\ell} := \text{Cor}(Z_i, Z_\ell) = s_{i\ell} / \sqrt{s_{ii}s_{\ell\ell}}$. Notice that the pairwise correlations between the z-scores are precisely the pairwise correlations between the cortical thickness measurements at each voxel. Because the regression is conditional on the voxelwise measurements, we take the pairwise correlations as fixed. To compute the z-scores, we use the variance estimate

$\hat{\sigma}^2 = \frac{1}{p(n-2)} \sum_{i=1}^p \sum_{j=1}^n [\tilde{Y}_j - \hat{\beta}^{(i)} \tilde{X}_j^{(i)}]^2$ and ignore its negligible contribution to their variability.

6.2 The distribution of the z-scores

Consider the decomposition $R_p = A_{k,p} + B_{k,p}$, where $A_{k,p}, B_{k,p}$ are of the form (7). The asymptotic dimension, if it exists, is unknown. In the next subsection we discuss the choice of k , for now we work with $k = 2$. After k is set, $\mathbf{W}_k^{(p)}$ is estimated via (10). Define

$$\hat{\mu}_i := \ell_i \hat{\mathbf{W}}_k^{(p)}, \quad \text{and } \sigma_i := \{(B_{k,p})_{ii}\}^{1/2} \quad i=1, \dots, p,$$

where ℓ_i is the i -th row of $L_{k,p}$ defined in (8). The empirical distribution is approximated by $\hat{F}(z)$ given by (11), and the density is approximated by $\bar{f}(z) := \frac{1}{p} \sum_{i=1}^p \frac{1}{\sigma_i} \varphi\left(\frac{z - \hat{\mu}_i}{\sigma_i}\right)$, where φ is the standard normal density. Figure 3(a) plots the histogram of the Z 's, f and $\bar{f}$. The approximation $\bar{f}$ beautifully captures the shape of the empirical distribution, even though it is based on only two latent variables.

Recall that the approximation is done under the complete null hypothesis, i.e. that there is no effect. The validity of the complete null hypothesis is confirmed by a false discovery rate (FDR) analysis which shows no significant voxels after applying the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995) at an FDR level of 0.2. Nevertheless, the z-scores exhibit a strong shift toward positive values, indicating a possible positive correlation between cortical thickness and behavior. Our approximation indicates that the reason that many Z 's are large may be because of the correlation between them and not because of a true effect. To illustrate this point we simulated $Y_1, \dots, Y_n \sim N(0, 1)$ i.i.d and independent of the X 's and repeated the same procedure as before. One such simulated instance is given in Figure 3(b). Without correlation, we would expect the histogram of the Y 's to follow the theoretical null density (red). However, the correlation is causing the impression of a strong positive effect not unlike the one seen in panel (a).

6.3 The number of latent variables

We now discuss the choice of k , the number of latent variables. Recall that the asymptotic dimension K , if it exists, is the optimal choice as discussed in Section 3.2 and is equal the number of eigenvalues of order p . A scree plot of the 50 largest eigenvalues of R_p is shown in Figure 4(a). Catell's graphical test indicates an elbow at $k = 3$, and after 10 ~ 20 eigenvalues, the value is almost constant.

Theorem 2 states that under the decomposition $R_p = A_{k,p} + B_{k,p}$, the $\mathcal{L}_2$ distance between $F_p^\wedge$ and F_p^- is bounded by $\frac{1}{4p} + C\|\text{cor}(B_{k,p})\|_2^{(p)}$. As k increases the bound decreases and $\|\text{cor}(B_{k,p})\|_2^{(p)} = 0$ when $k = p$. However, as k increases the dimension of the representation also increases since the dimension is $\text{rank}(A_{k,p})$ (Proposition 1). Furthermore, for large k the empirical distribution and the approximation, F , are very close and the overfitting problem arises. To illustrate this point, we plot different choices of k in Figure 4(b). It can be seen that when $k = 100$ the approximation is much closer to the histogram, but such high dimension is unnecessary to capture the global behavior of the histogram. The differences between $k = 2$ and $k = 10$ are rather small suggesting that $k = 2$ suffices.

7 Summary and extensions

In this work we have studied the limit of the ecdf of marginally standard normal variables when strong correlation is present. As predicted by Efron (2007a), we have shown that the limit is indeed not standard normal. Specifically, we have shown that under a regime that we call finite dimensional correlation (and some regularity conditions), the limit is a finite mixture of scaled normals with random means, which reduces to Efron's empirical null model when the correlation structure is exchangeable. Moreover, we have shown that if the correlation is not finite then the limit can still be approximated by mixture of normals but an infinite number of them may be required.

The main technique for achieving these results has been a decomposition of the correlation matrix R into two matrices A and B , where A captures the strong correlation and B captures the weakly correlated residual noise. The form of the limiting distribution of the ecdf is determined by A , while the residual noise represented by B goes to zero asymptotically. The key to achieving the asymptotic representation of the ecdf with the smallest dimension is to choose B so that it contains the largest amount of variance while remaining weakly correlated.

For future work, we consider the following extensions. First, we assumed that all random variables have variance 1. If the variances are not 1, but are bounded above and below, then the random variables could be standardized and all our results follow. If they are not bounded, then still for each finite p one can standardize and obtain inequality (4) of Theorem 2.

In Theorem 2 we proved that the $\mathcal{L}_2$ distance between $F_p^\wedge$ and F_p^- converges to zero if the residual correlation not captured by F_p^- is weak. Could the same be said of the $\mathcal{L}_2$ distance between the moments of $F_p^\wedge$ and F_p^- ? Let $\hat{m}_{n,p} := \frac{1}{p} \sum_{i=1}^p Z_i^n$ and $m_{n,p}^- := \int z^n dF_p^-(z)$ be the n -th moment of $F_p^\wedge$ and F_p^- . For the first moment and an appropriate decomposition $R_p = A_p + B_p$,

$$\overline{m}_{1,p} = \frac{1}{p} \sum_{i=1}^p \int z \frac{1}{\sigma_i^{(p)}} \varphi\left(\frac{z - \mu_i^{(p)}}{\sigma_i^{(p)}}\right) dz = \frac{1}{p} \sum_{i=1}^p \mu_i^{(p)}.$$

By the definition of $\mu_i^{(p)}$ and equation (6), we have that $Z_i - \mu_i^{(p)} = \xi_i$. Therefore,

$$E(\hat{m}_{1,p} - \bar{m}_{1,p})^2 = \frac{1}{p^2} E \left\{ \sum_{i=1}^p (Z_i - \mu_i^{(p)}) \right\}^2 = \frac{1}{p^2} E \left(\sum_{i=1}^p \xi_i \right)^2 = \|B_p\|_2^{(p)},$$

and the distance converges to zero if and only if $\|B_p\|_2^{(p)} \rightarrow 0$. This condition is weaker than that of Theorem 2 and thus convergence of the ecdf implies convergence of the first moment. Proving convergence of higher moments is more difficult and is left for future work.

About the choice of number of components k in practice, the problem is not unlike that of determining the number of components in factor analysis or the effective dimension in principal components analysis. The connection with these techniques is worth exploring.

Finally, about the normality assumption, it is used in the proof of Theorem 2 by applying Mehler's expansion to the joint density. Similar expansions are available for other distributions such as chi square and gamma (e.g., Koudou (1998); Schwartzman (2011)). Thus it may be possible to extend our results to those distributions as well.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors are grateful to Philip Reiss from the Department of Child and Adolescent Psychiatry, New York University School of Medicine, for providing the brain imaging data. This work was partially supported by NIH grant R01-CA157528.

References

- Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B*. 1995; 57:289–300.
- Dedecker J, Merlevéde F. The empirical distribution function for dependent variables: asymptotic and nonasymptotic results in L^p . *ESAIM: Probability and Statistics*. 2007; 11:102–114.
- Efron B. Large-scale simultaneous hypothesis testing: The choice of a null hypothesis. *Journal of the American Statistical Association*. 2004; 99:96–104.
- Efron B. Correlation and large-scale simultaneous hypothesis testing. *Journal of the American Statistical Association*. 2007a; 102:93–103.
- Efron B. Size, power and false discovery rates. *The Annals of Statistics*. 2007b; 35:1351–1377.
- Efron, Bradley. Simultaneous inference: When should hypothesis testing problems be combined? *The Annals of Applied Statistics*. 2008; 2:197–223.
- Efron, Bradley. Correlated z-values and the accuracy of large-scale statistical estimates. *Journal of the American Statistical Association*. 2010; 105:1042–1055. [PubMed: 21052523]
- Fan, J.; Han, X. Estimation of False Discovery Proportion with Unknown Dependence. 2014. submitted <http://arxiv.org/abs/1305.7007>
- Fan J, Han X, Gu W. Estimating False Discovery Proportion Under Arbitrary Covariance Dependence. *Journal of the American Statistical Association*. 2012; 107:1019–1035. [PubMed: 24729644]

- Koudou AE. Lancaster bivariate probability distributions with Poisson, negative binomial and gamma margins. *Test*. 1998; 7:95–110.
- Owen AB. Variance of the number of false discoveries. *Journal of the Royal Statistical Society: Series B*. 2005; 67:411–426.
- Proal E, Reiss PT, Klein RG, Mannuzza S, Gotimer K, Ramos-Olazagasti MA, Lerch JP, He Y, Zijdenbos A, Kelly C, Milham MP, Castellanos FX. Brain gray matter deficits at 33-year follow-up in adults with attention-deficit/hyperactivity disorder established in childhood. *Archives of general psychiatry*. 2011; 68:1122–1134. [PubMed: 22065528]
- Reiss PT, Schwartzman A, Lu F, Huang L, Proal E. Paradoxical results of adaptive false discovery rate procedures in neuroimaging studies. *Neuroimage*. 2012; 63:1833–1840. [PubMed: 22842214]
- Schwartzman A. Comment on “Correlated z-values and the accuracy of large-scale statistical estimates” by Bradley Efron. *Journal of the American Statistical Association*. 2010; 105(491): 1059–1063. [PubMed: 22021943]
- Schwartzman A, Lin X. The effect of correlation in false discovery rate estimation. *Biometrika*. 2011; 98:199–214. [PubMed: 23049127]
- Wasserman, L. *All of Nonparametric Statistics: A Concise Course in Nonparametric Statistical Inference*. New York: Springer; 2006.
- Wu WB. Oscillations of Empirical Distribution Functions under Dependence. *IMS lecture notes-monograph series, High dimensional Probabilities*. 2006; 51:53–61.

Figure 1.

Histograms of various instances of 1000 standard normal variables with a one-block correlation structure (a) and a two-independent-blocks correlation structure (b). The red line is the standard normal density and the blue dashed line is the asymptotic approximation we use.

Figure 2.

Different cases of asymptotic approximation to the ecdf. The corresponding examples from Section 5 below appear in parentheses.

Figure 3.

Histogram of $Z_1, Z_2, \dots, Z_{1000}$ of the real data (a) and the simulation results (b). The red line illustrates ϕ , the standard normal density, and the blue dashed line is the approximation we use, f .

Figure 4.

(a) The 50 largest eigenvalues of R_P ordered according to their size. (b) Histogram of $Z_1, Z_2, \dots, Z_{1000}$ and the approximation, f , for $k = 2$ (blue dashed), 10 (red dotted), 100 (solid green).