



**HAL**  
open science

## Detection of sparse positive dependence

Ery Arias-Castro, Rong Huang, Nicolas Verzelen

► **To cite this version:**

Ery Arias-Castro, Rong Huang, Nicolas Verzelen. Detection of sparse positive dependence. *Electronic Journal of Statistics*, 2020, 14 (1), pp.702-730. 10.1214/19-EJS1675 . hal-02922761

**HAL Id: hal-02922761**

**<https://hal.inrae.fr/hal-02922761>**

Submitted on 26 Aug 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Detection of sparse positive dependence

Ery Arias-Castro and Rong Huang

*Department of Mathematics, University of California, San Diego*  
*e-mail: [eariascastro@ucsd.edu](mailto:eariascastro@ucsd.edu); [roh019@ucsd.edu](mailto:roh019@ucsd.edu)*

and

Nicolas Verzelen

*INRAE, Montpellier SupAgro, MISTEA, Univ. Montpellier*  
*e-mail: [nicolas.verzelen@inra.fr](mailto:nicolas.verzelen@inra.fr)*

**Abstract:** In a bivariate setting, we consider the problem of detecting a sparse contamination or mixture component, where the effect manifests itself as a positive dependence between the variables, which are otherwise independent in the main component. We first look at this problem in the context of a normal mixture model. In essence, the situation reduces to a univariate setting where the effect is a decrease in variance. In particular, a higher criticism test based on the pairwise differences is shown to achieve the detection boundary defined by the (oracle) likelihood ratio test. We then turn to a Gaussian copula model where the marginal distributions are unknown. Standard invariance considerations lead us to consider rank tests. In fact, a higher criticism test based on the pairwise rank differences achieves the detection boundary in the normal mixture model, although not in the very sparse regime. We do not know of any rank test that has any power in that regime.

**Keywords and phrases:** Rare effects, positive dependence, higher criticism, Gaussian mixture Copula model, rank tests.

Received November 2018.

## Contents

1	Introduction . . . . .	703
1.1	Gaussian mixture model . . . . .	703
1.2	Gaussian mixture copula model . . . . .	705
2	Gaussian mixture model . . . . .	705
2.1	The likelihood ratio test . . . . .	706
2.2	The covariance test . . . . .	709
2.3	The higher criticism test and the extremes test . . . . .	711
2.4	Numerical experiments . . . . .	711
3	Gaussian mixture copula model . . . . .	712
3.1	The covariance rank test . . . . .	712
3.2	The higher criticism rank test . . . . .	716
3.3	Numerical experiments . . . . .	727
4	Discussion . . . . .	727

Acknowledgements . . . . . 729  
 References . . . . . 729

**1. Introduction**

The detection of rare effects has been an important problem for years in settings, and may be particularly relevant today, for example, with the search for personalized care in the health industry, where a small fraction of a population may respond particularly well, or particularly poorly, to some given treatment [20].

Following a theoretical investigation initiated in large part by Ingster [16] and broadened by Donoho and Jin [10], we are interested in studying two-component mixture models, also known as contamination models, in various asymptotic regimes defined by how the small mixture weight converges to zero. Most of the existing work in the setting of univariate data has focused on models where the contamination manifests itself as a shift in mean [12, 11, 14, 7, 19] with a few exceptions where the effect is a change in variance [1], or a change in both mean and variance [6].

In the present paper, we are interested in bivariate data, instead, and more specifically in a situation where the effect felt in the dependence between the two variables being measured. This setting has been recently considered in the literature in the context of assessing the reproducibility of studies. For example, [18] aims to identify significant features from separate studies using an expectation-maximization (EM) algorithm. They applied a copula mixture model and assumed that changes in the mean and covariance matrix differentiate the contaminated component from the null component. [23] studies another model where variables from the contamination are stochastically larger marginally. In both models, the marginal distributions have some non-null effects. Similar settings have been considered within a multiple testing framework [5, 22].

While existing work has focused on models motivated by questions of reproducibility, in the present work we come back to basics and directly address the problem of detecting a bivariate mixture with a component where the variables are independent and a component where the variables are positively dependent.

*1.1. Gaussian mixture model*

Ingster [16] and Donoho and Jin [10] started with a mixture of Gaussians, and we do the same, and in our setting, this means we consider the following mixture model

$$(X, Y) \sim (1 - \varepsilon)\mathcal{N}(0, \mathbf{I}) + \varepsilon\mathcal{N}(0, \Sigma_\rho), \quad \Sigma_\rho := \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \quad (1)$$

where  $\varepsilon \in [0, 1/2)$  is the contamination proportion and  $0 \leq \rho \leq 1$  is the correlation between the two variables under contamination. We consider the following hypothesis testing problem: based on  $(X_1, Y_1), \dots, (X_n, Y_n)$  drawn iid from (1), decide

$$\mathcal{H}_0 : \varepsilon = 0 \quad \text{versus} \quad \mathcal{H}_1 : \varepsilon > 0, \rho > 0. \quad (2)$$

Note that under the null hypothesis,  $(X, Y)$  is from the bivariate standard normal. Under the alternative,  $X$  and  $Y$  remain standard normal marginally. Following the literature on the detection of sparse mixtures [16, 10], we are most interested in a situation, asymptotic as  $n \rightarrow \infty$ , where  $\varepsilon = \varepsilon_n \rightarrow 0$ , and the central question is how large  $\rho = \rho_n$  needs to be in order to reliably distinguish these hypotheses.

The formulation (1) suggests that the alternative hypothesis is composite, but if we assume that  $(\varepsilon, \rho)$  are known under the alternative, then the likelihood ratio test (LRT) is optimal by Neyman–Pearson lemma. We start with characterizing the behavior of the LRT, which provides a benchmark. We then study some other testing procedures that do not require knowledge of the model parameters:<sup>1</sup>

- The *covariance test* rejects for large values of  $\sum_i X_i Y_i$ , and coincides with Rao’s score test in the present context. This is the classical test for independence, specifically designed for the case where  $\varepsilon = 1$  and  $\rho > 0$  under the alternative. We shall see that it is suboptimal in some regimes.
- The *extremes test* rejects for small values of  $\min_i |X_i - Y_i|$ . This test exploits the fact that, because  $\rho$  is assumed positive, the variables in the contaminated component are closer to each other than in the null component.
- The *higher criticism test* was suggested by John Tukey and deployed by [10] for the testing of sparse mixtures. We propose a version of that test based on the pairwise differences,  $U_i := (X_i - Y_i)/\sqrt{2}$ . In detail, the test rejects for large values of

$$\sup_{u \geq 0} \frac{\sqrt{n} (\hat{F}(u) - \Psi(u))}{\sqrt{\Psi(u)(1 - \Psi(u))}}, \quad (3)$$

where  $\Psi(u) := 2\Phi(u) - 1$ , with  $\Phi$  denotes the standard normal distribution function, and  $\hat{F}(u) := \frac{1}{n} \sum_{i=1}^n \mathbb{I}\{|U_i| \leq u\}$ , the empirical distribution function of  $|U_1|, \dots, |U_n|$ .

As is common practice in this line of work [16, 10], under  $\mathcal{H}_1$  we set

$$\varepsilon = n^{-\beta}, \quad \beta \in (0, 1) \text{ fixed.} \quad (4)$$

The setting where  $\beta \leq 1/2$  is often called the dense regime and the setting where  $\beta > 1/2$  is often called the sparse regime. Our analysis reveals the following:

- Dense regime.* The dense regime is most interesting when  $\rho \rightarrow 0$ . In that case, we find that the covariance test and the higher criticism test match the asymptotic performance of the likelihood ratio test to first-order, while the extremes test has no power.
- Sparse regime.* The sparse regime is most interesting when  $\rho \rightarrow 1$ . In that case, we find that the higher criticism test still performs as well as the likelihood ratio test to first order, while the covariance test is powerless, and the extremes test is suboptimal.

---

<sup>1</sup>Such procedures are said to be *adaptive*.

### 1.2. Gaussian mixture copula model

From a practical point of view, the assumption that both  $X$  and  $Y$  are normally distributed is quite stringent. Hence, we would like to know if there are nonparametric procedures that do not require such a condition but can still achieve the same performance as the likelihood ratio test. In the univariate setting where the effect arises as a shift in mean, this was investigated in [2]. In the bivariate setting, in a model for reproducibility, [23] proposes a nonparametric test based on a weighted version of Hoeffding’s test for independence.

Here, instead of model (1), we suppose  $(X, Y)$  follows a Gaussian mixture copula model (GMCM) [4], meaning that there is a latent random vector  $(Z^1, Z^2)$  such that

$$F(X) = \Phi(Z^1), \quad G(Y) = \Phi(Z^2), \tag{5}$$

$$(Z^1, Z^2) \sim (1 - \varepsilon)\mathcal{N}(0, I) + \varepsilon\mathcal{N}(0, \Sigma_\rho), \quad \Sigma_\rho := \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix},$$

where  $F$  and  $G$  are unknown distribution functions on the real line, and  $\Phi$  is the standard normal distribution function, while  $\varepsilon \in [0, 1/2)$  is the contamination proportion and  $0 \leq \rho \leq 1$  is the correlation between  $Z^1$  and  $Z^2$  in the contaminated component, as before in model (1). [18] also uses a copula mixture model, but they placed emphasis on the mean while we focus on the dependence.

We still consider the testing problem (2), but now in the context of Model (5). The setting is nonparametric in that both  $F$  and  $G$  are unknown. Model (5) is crafted in such a way that the marginal distributions of  $X$  and  $Y$  contain absolutely no information that is pertinent to the testing problem under consideration.

The model is also attractive because of an invariance under all increasing marginal transformations of the variables. This is the same invariance that leads to considering rank based methods such as the Spearman correlation test [17, Chp. 6]. In fact, we analyze the Spearman correlation test, which is the nonparametric analog to the covariance test, showing that it is first-order asymptotically optimal in the dense regime. We also propose and analyze a nonparametric version of the higher criticism based on ranks which we show is first-order asymptotically optimal in the moderately sparse regime where  $1/2 < \beta < 3/4$ . In the very sparse regime, where  $\beta > 3/4$ , we do not know of any rank-based test that has any power.

## 2. Gaussian mixture model

In this section, we focus on the Gaussian mixture model (1). We start by deriving a lower bound on the performance of the likelihood ratio test, which provides a benchmark for the other (adaptive) tests, which we subsequently analyze.

We distinguish between the dense and sparse regimes:

$$\text{dense regime} \quad \rho = n^{-\gamma}, \quad \gamma > 0 \text{ fixed}; \tag{6}$$

$$\text{sparse regime } \rho = 1 - n^{-\gamma}, \quad \gamma > 0 \text{ fixed.} \quad (7)$$

We say that a testing procedure is asymptotically powerful (resp. powerless) if the sum of its probabilities of Type I and Type II errors (its risk) has limit 0 (resp. limit inferior at least 1) in the large sample asymptote.

### 2.1. The likelihood ratio test

**Theorem 1.** *Consider the testing problem (2) with  $\varepsilon$  parameterized as in (4). In the dense regime, with  $\rho$  parameterized as in (6), the likelihood ratio test is asymptotically powerless when  $\gamma > 1/2 - \beta$ . In the sparse regime, with  $\rho$  parameterized as in (7), the likelihood ratio test is asymptotically powerless when  $\gamma < 4(\beta - 1/2)$ .*

This only provides a lower bound on what can be achieved, but it will turn out that to be sharp once we establish the performance of the higher criticism test in Proposition 2 below.

*Proof.* The proof techniques are standard and already present in [12, 16], and many of the subsequent works.

Defining  $U := (X - Y)/\sqrt{2}$  and  $V := (X + Y)/\sqrt{2}$ , the model (1) is equivalently expressed in terms of  $(U, V)$ , which has distribution

$$(U, V) \sim (1 - \varepsilon)\mathcal{N}(0, \mathbf{I}) + \varepsilon\mathcal{N}(0, \Delta_\rho), \quad \Delta_\rho := \text{diag}(1 - \rho, 1 + \rho). \quad (8)$$

Note that  $U$  and  $V$  are independent only conditional on knowing what distribution they were sampled from. In terms of the  $(U, V)$ 's, the likelihood ratio is

$$L := \prod_{i=1}^n L_i,$$

where  $L_i$  is the likelihood ratio for observation  $(U_i, V_i)$ , which in the present case takes the following expression

$$\begin{aligned} L_i &= \frac{\frac{1-\varepsilon}{2\pi} \exp(-\frac{1}{2}U_i^2 - \frac{1}{2}V_i^2) + \frac{\varepsilon}{2\pi\sqrt{1-\rho^2}} \exp(-\frac{1}{2(1-\rho)}U_i^2 - \frac{1}{2(1+\rho)}V_i^2)}{\frac{1}{2\pi} \exp(-\frac{1}{2}U_i^2 - \frac{1}{2}V_i^2)} \\ &= 1 - \varepsilon + \varepsilon(1 - \rho^2)^{-1/2} \exp(-\frac{\rho}{2(1-\rho)}U_i^2 + \frac{\rho}{2(1+\rho)}V_i^2). \end{aligned}$$

The risk of the likelihood ratio test is equal to [17, Problem 3.10]

$$\text{risk}(L) := 1 - \frac{1}{2} \mathbb{E}_0[|L - 1|].$$

We show that  $\text{risk}(L) = 1 + o(1)$  under each of the stated conditions. We consider each regime in turn.

*Dense regime.* It turns out that it suffices to bound the second moment. Indeed, using the Cauchy–Schwarz inequality, we have

$$\text{risk}(L) \geq 1 - \frac{1}{2} \sqrt{\mathbb{E}_0[L^2] - 1},$$

reducing the task to showing that  $\mathbb{E}_0[L^2] \leq 1 + o(1)$ . We have

$$\mathbb{E}_0[L^2] = \prod_{i=1}^n \mathbb{E}_0[L_i^2] = (\mathbb{E}_0[L_1^2])^n$$

where

$$\begin{aligned} \mathbb{E}_0[L_1^2] &= \mathbb{E}_0 \left[ \left( 1 - \varepsilon + \varepsilon(1 - \rho^2)^{-1/2} \exp\left(-\frac{\rho}{2(1-\rho)}U_1^2 + \frac{\rho}{2(1+\rho)}V_1^2\right) \right)^2 \right] \\ &= (1 - \varepsilon)^2 + 2(1 - \varepsilon)\varepsilon \\ &\quad + \varepsilon^2(1 - \rho^2)^{-1} \mathbb{E}_0 \left[ \exp\left(-\frac{\rho}{(1-\rho)}U_1^2\right) \right] \mathbb{E}_0 \left[ \exp\left(\frac{\rho}{(1+\rho)}V_1^2\right) \right] \\ &= 1 - \varepsilon^2 + \varepsilon^2(1 - \rho^2)^{-1} \mathbb{E}_0 \left[ \exp\left(-\frac{\rho}{(1-\rho)}U_1^2\right) \right] \mathbb{E}_0 \left[ \exp\left(\frac{\rho}{(1+\rho)}V_1^2\right) \right]. \end{aligned}$$

For the third term, we have

$$\mathbb{E}_0 \left[ \exp\left(-\frac{\rho}{(1-\rho)}U_1^2\right) \right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{\rho}{1-\rho}u^2 - \frac{1}{2}u^2} du = \sqrt{\frac{1-\rho}{1+\rho}},$$

and

$$\mathbb{E}_0 \left[ \exp\left(\frac{\rho}{(1+\rho)}V_1^2\right) \right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{\rho}{1+\rho}v^2 - \frac{1}{2}v^2} dv = \sqrt{\frac{1+\rho}{1-\rho}}.$$

Hence, we have

$$\mathbb{E}_0[L_1^2] = 1 + \varepsilon^2 \rho^2 / (1 - \rho^2),$$

and, therefore,

$$\mathbb{E}_0[L^2] = [1 + \varepsilon^2 \rho^2 / (1 - \rho^2)]^n \leq \exp [n \varepsilon^2 \rho^2 / (1 - \rho^2)],$$

so that  $\mathbb{E}_0[L^2] \leq 1 + o(1)$  when

$$n \varepsilon^2 \rho^2 = o(1),$$

since  $\rho$  is assumed to be bounded away from 1. Under the specified parameterization, this happens exactly when  $\gamma > 1/2 - \beta$ .

*Sparse regime.* It turns out that simply bounding the second moment, as we did above, does not suffice. Instead, we truncate the likelihood and study the behavior of its first two moments. Define the indicator variable  $D_i = \mathbb{I}\{|V_i| \leq \sqrt{2 \log n}\}$  and the corresponding truncated likelihood ratio

$$\bar{L} = \prod_{i=1}^n \bar{L}_i, \quad \bar{L}_i := L_i D_i.$$

Using the triangle inequality, the fact that  $\bar{L} \leq L$ , and the Cauchy–Schwarz inequality, we have the following upper bound:

$$\begin{aligned} \mathbb{E}_0 |L - 1| &\leq \mathbb{E}_0 |\bar{L} - 1| + \mathbb{E}_0(L - \bar{L}) \\ &\leq [\mathbb{E}_0[\bar{L}^2] - 1 + 2(1 - \mathbb{E}_0[\bar{L}])]^{1/2} + (1 - \mathbb{E}_0[\bar{L}]), \end{aligned}$$

so that  $\text{risk}(L) = 1 + o(1)$  when  $\mathbb{E}_0[\bar{L}^2] \leq 1 + o(1)$  and  $\mathbb{E}_0[\bar{L}] \geq 1 - o(1)$ .

For the first moment, we have

$$\mathbb{E}_0[\bar{L}] = \prod_{i=1}^n \mathbb{E}_0[\bar{L}_i] = (\mathbb{E}_0[\bar{L}_1])^n$$

where, using the independence of  $U_1$  and  $V_1$ , and taking the expectation with respect to  $U_1$  first,

$$\begin{aligned} \mathbb{E}_0[\bar{L}_1] &= \mathbb{E}_0 \left[ \left( 1 - \varepsilon + \varepsilon(1 + \rho)^{-1/2} \exp\left(\frac{\rho}{2(1+\rho)}V_1^2\right) \right) D_1 \right] \\ &= (1 - \varepsilon)\Psi(\sqrt{2 \log n}) + \varepsilon\Psi(\sqrt{2 \log n}/\sqrt{1 + \rho}) \\ &= (1 - \varepsilon)(1 - O(n^{-1}/\sqrt{\log n})) + \varepsilon(1 - O(n^{-1/(1+\rho)}/\sqrt{\log n})) \\ &= 1 - o(1/n) - o(\varepsilon n^{-1/(1+\rho)}), \end{aligned}$$

where, for  $t \geq 0$ ,

$$\Psi(t) = \mathbb{P}(|\mathcal{N}(0, 1)| \leq t) = 2\Phi(t) - 1 = \int_{-t}^t \frac{e^{-s^2/2}}{\sqrt{2\pi}} ds,$$

and we used the fact that  $1 - \Psi(t) \asymp e^{-t^2/2}/t$  when  $t \rightarrow \infty$ . Since  $\varepsilon = n^{-\beta}$  with  $\beta > 1/2$  in the sparse regime, for  $\rho$  sufficiently close to 1,  $\varepsilon n^{-1/(1+\rho)} \leq 1/n$ , in which case  $\mathbb{E}_0[\bar{L}_1] \geq 1 - o(1/n)$ . This yields

$$\mathbb{E}_0[\bar{L}] \geq (1 - o(1/n))^n = 1 - o(1).$$

For the second moment, we have

$$\mathbb{E}_0[\bar{L}^2] = \prod_{i=1}^n \mathbb{E}_0[\bar{L}_i^2] = \mathbb{E}_0[\bar{L}_1^2]^n,$$

where

$$\begin{aligned} \mathbb{E}_0[\bar{L}_1^2] &= \mathbb{E}_0 \left[ \left( 1 - \varepsilon + \varepsilon(1 - \rho^2)^{-1/2} \exp\left(-\frac{\rho}{2(1-\rho)}U_1^2 + \frac{\rho}{2(1+\rho)}V_1^2\right) \right)^2 D_1 \right] \\ &= (1 - \varepsilon)^2\Psi(\sqrt{2 \log n}) + 2(1 - \varepsilon)\varepsilon\Psi(\sqrt{2 \log n}/\sqrt{1 + \rho}) \\ &\quad + \varepsilon^2(1 - \rho^2)^{-1} \mathbb{E}_0[\exp(-\frac{\rho}{(1-\rho)}U_1^2)] \mathbb{E}_0[\exp(\frac{\rho}{(1+\rho)}V_1^2)D_1]. \end{aligned}$$

The sum of first two terms is bounded from above by  $(1 - \varepsilon)^2 + 2(1 - \varepsilon)\varepsilon = 1 - \varepsilon^2$ .

For the third term, we have

$$\mathbb{E}_0[\exp(-\frac{\rho}{(1-\rho)}U_1^2)] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{\rho}{1-\rho}u^2 - \frac{1}{2}u^2} du = \sqrt{\frac{1-\rho}{1+\rho}},$$



and

$$\mathbb{E}_0[\exp(\frac{\rho}{1+\rho} V_1^2) D_1] = \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{2\log n}}^{\sqrt{2\log n}} e^{\frac{\rho}{1+\rho} v^2 - \frac{1}{2} v^2} dv \leq \frac{1}{\sqrt{2\pi}} 2\sqrt{2\log n},$$

using the fact that  $\rho \leq 1$ . Hence,

$$\begin{aligned} \mathbb{E}_0[\bar{L}_1^2] &\leq 1 - \varepsilon^2 + \varepsilon^2(1 - \rho^2)^{-1} \sqrt{\frac{1 - \rho}{1 + \rho}} \frac{1}{\sqrt{2\pi}} 2\sqrt{2\log n} \\ &\leq 1 + \varepsilon^2(1 - \rho)^{-1/2}(\log n)^{1/2}, \end{aligned}$$

when  $\rho$  is sufficiently close to 1. This in turn yields the following bound

$$\mathbb{E}_0[\bar{L}^2] \leq [1 + \varepsilon^2(1 - \rho)^{-1/2}(\log n)^{1/2}]^n \leq \exp [n\varepsilon^2(1 - \rho)^{-1/2}(\log n)^{1/2}],$$

so that  $\mathbb{E}_0[\bar{L}^2] \leq 1 + o(1)$  when

$$n\varepsilon^2(1 - \rho)^{-1/2}(\log n)^{1/2} = o(1).$$

Under the specified parameterization, this happens exactly when  $\gamma < 4\beta - 2$ .  $\square$

In the dense regime, with  $\rho$  parameterized as in (6), we say that a test achieves the detection boundary if it is asymptotically powerful when  $\gamma < 1/2 - \beta$ , and in the sparse regime, with  $\rho$  parameterized as in (7), we say that a test achieves the detection boundary if it is asymptotically powerful when  $\gamma > 4(\beta - 1/2)$ .

### 2.2. The covariance test

Recall that the covariance test rejects for large values of  $T_n := \sum_{i=1}^n X_i Y_i$ , calibrated under the null where  $X_1, \dots, X_n, Y_1, \dots, Y_n$  are iid standard normal.

**Proposition 1.** *For the testing problem (2), the covariance test achieves the detection boundary in the dense regime, while it is asymptotically powerless in the sparse regime.*

*Proof.* We divide the proof into the two regimes.

*Dense regime.* Under  $\mathcal{H}_0$ , we have

$$\begin{aligned} \mathbb{E}_0(T_n) &= n \mathbb{E}_0(X_1 Y_1) = n \mathbb{E}_0(X_1) \mathbb{E}_0(Y_1) = 0, \\ \text{Var}_0(T_n) &= n \text{Var}_0(X_1 Y_1) = n \mathbb{E}_0(X_1^2) \mathbb{E}_0(Y_1^2) = n, \end{aligned}$$

so that, by Chebyshev's inequality,

$$\mathbb{P}_0(|T_n| \geq a_n \sqrt{n}) \rightarrow 0,$$

for any sequence  $(a_n)$  diverging to infinity.

Under  $\mathcal{H}_1$ , we have

$$\mathbb{E}_1(T_n) = n \mathbb{E}_1(X_1 Y_1) = n\varepsilon\rho,$$

$$\text{Var}_1(T_n) = n \text{Var}_1(X_1 Y_1) = n(1 + 2\varepsilon\rho^2 - \varepsilon^2\rho^2) \leq 3n,$$

so that, by Chebyshev's inequality,

$$\mathbb{P}_1(|T_n - n\varepsilon\rho| \geq a_n\sqrt{n}) \rightarrow 0.$$

Thus the test with rejection region  $\{T_n \geq a_n\sqrt{n}\}$  is asymptotically powerful when

$$\sqrt{n}\varepsilon\rho \geq 2a_n.$$

If we choose  $a_n = \log n$ , for example, and  $\rho$  is parameterized as in (6), this happens for  $n$  large enough when  $\gamma < 1/2 - \beta$ .

*Sparse regime.* To prove that the covariance test is asymptotically powerless when  $\beta > 1/2$ , we show that, under  $\mathcal{H}_1$ ,  $T_n$  converges to the same limiting distribution as under  $\mathcal{H}_0$ .

Under  $\mathcal{H}_0$ , by the central limit theorem,

$$\frac{T_n}{\sqrt{n}} \rightarrow \mathcal{N}(0, 1).$$

Under  $\mathcal{H}_1$  the distribution of the  $(X_i, Y_i)$ 's (which remain iid) depends on  $n$ , but the condition for applying Lyapunov's central limit theorem are satisfied since

$$\mathbb{E}_1[(X_i Y_i - \varepsilon\rho)^4] \leq 8(\mathbb{E}_1[(X_i Y_i)^4] + (\varepsilon\rho)^4),$$

with  $(\varepsilon\rho)^4 \leq 1$  and

$$\mathbb{E}_1[(X_i Y_i)^4] \leq [\mathbb{E}_1(X_i^8) \mathbb{E}_1(Y_i^8)]^{1/2} = \mathbb{E}(Z^8) = \text{const},$$

where  $Z \sim \mathcal{N}(0, 1)$  and the inequality is Cauchy-Schwarz's, while

$$\text{Var}_1(X_i Y_i) = 1 + 2\varepsilon\rho^2 - \varepsilon^2\rho^2 \geq 1,$$

so that the test statistic still converges weakly to a normal distribution,

$$\frac{T_n - \mathbb{E}_1(T_n)}{\sqrt{\text{Var}_1(T_n)}} \rightarrow \mathcal{N}(0, 1).$$

In the present regime, we have

$$\mathbb{E}_1(T_n) = n\varepsilon\rho, \quad \text{Var}_1(T_n) = n(1 + 2\varepsilon\rho^2 - \varepsilon^2\rho^2),$$

so that  $\mathbb{E}_1(T_n)/\sqrt{\text{Var}_1(T_n)} \rightarrow 0$  and  $\text{Var}_1(T_n) \sim n$ , and thus we conclude by Slutsky's theorem that  $T_n/\sqrt{n} \rightarrow \mathcal{N}(0, 1)$ .  $\square$

*Remark 1.* There are good reasons to consider the covariance test in this specific form since the means and variances are known. It is worth pointing out that the Pearson correlation test, which is more standard in practice since it does not require knowledge of the means or variances, has the same asymptotic power properties.

### 2.3. The higher criticism test and the extremes test

Define  $U_i = (X_i - Y_i)/\sqrt{2}$ , and note that

$$U_1, \dots, U_n \stackrel{\text{iid}}{\sim} (1 - \varepsilon)\mathcal{N}(0, 1) + \varepsilon\mathcal{N}(0, 1 - \rho).$$

Seen through the  $U_i$ 's, the problem becomes that of detecting a sparse contamination where the effect is in the variance. We recently studied this problem in detail [1], extending previous work by Cai et al [6], who considered a setting where the effect is both in the mean and variance. Borrowing from our prior work, we consider a higher criticism test, already defined in (3), and an extremes test, which rejects for small values of  $\min_i |U_i|$ .

**Proposition 2.** *For the testing problem (2), the higher criticism test achieves the detection boundary in the dense and sparse regimes.*

*Proof.* Set  $\sigma^2 = 1 - \rho$ , which is the variance of the contaminated component. In our prior work [1, Prop 3], we showed that the higher criticism test as defined in (3) is asymptotically powerful when

- (a)  $\sigma^2 = n^{-\gamma}$  with  $\gamma > 0$  fixed such that  $\gamma > 4(\beta - 1/2)$ ;
- (b)  $|\sigma^2 - 1| = n^{-\gamma}$  with  $\gamma > 0$  fixed such that  $\gamma < 1/2 - \beta$ .

This can be directly translated into the present setting, yielding the stated result.  $\square$

**Proposition 3.** *For the testing problem (2), the extremes test is asymptotically powerless when  $\rho$  is bounded away from 1, while when  $\varepsilon$  parameterized as in (4) and  $\rho$  parameterized as in (7), it is asymptotically powerful when  $\gamma > 2\beta$ , and asymptotically powerless when  $\gamma < 2\beta$ .*

*Proof.* This is also a direct corollary from our prior work our prior work [1, Prop. 2].  $\square$

Thus the extremes test is grossly suboptimal in the dense regime, while it is suboptimal in the sparse regime due to the fact that  $2\beta - 4(\beta - 1/2) = 2 - 2\beta > 0$ .

*Remark 2.* The higher criticism and extremes tests are both based on the  $U_i$ 's. This was convenient as it reduced the problem of testing for independence to the problem of testing for a change in variance (both in a contamination model). However, reducing the original data, meaning the  $(X_i, Y_i)$ 's, to the  $U_i$ 's implies a loss of information. Indeed, a lossless reduction would be from the  $(X_i, Y_i)$ 's to the  $(U_i, V_i)$ 's, where  $V_i := (X_i + Y_i)/\sqrt{2}$ , with joint distribution given in (8). It just turns out that ignoring the  $V_i$ 's does not lead to any loss in first-order asymptotic power.

### 2.4. Numerical experiments

We performed some numerical experiments to investigate the finite sample performance of the tests considered here: the likelihood ratio test, the Pearson correlation test (instead of the covariance test from a practical point of view),

the extremes test, the higher criticism test, and also a plug-in version of the higher criticism test where the parameters of the bivariate normal distribution (the two means and two variances) are estimated under the null. The sample size  $n$  is set large to  $n = 10^6$  in order to capture the large-sample behavior of these tests. We tried four sparsity levels, setting  $\beta \in \{0.2, 0.4, 0.6, 0.8\}$ . The p-values for each test are computed as follows:

- (a) For the *likelihood ratio test*, the p-values are estimated based on  $10^3$  permutations.
- (b) For the *higher criticism test* and the *plug-in higher criticism test*, the p-values are estimated based on 200 permutations.
- (c) For the *extremes test*, we used the exact null distribution, which is available in a closed form.
- (d) For the *Pearson correlation test*, the p-values are from the limiting distribution.

For each scenario, we repeated the process 200 times and calculated the fraction of p-values smaller than 0.05, representing the empirical power at the 0.05 level.

The results of this experiment are reported in Figure 1 and are broadly consistent with the theory developed earlier in this section. Though we show that the higher criticism test is first-order comparable to the likelihood ratio test in the dense regime, even with a large sample, its power is much lower. The Pearson correlation test does better in that regime. The plug-in higher criticism test has a similar performance as the higher criticism test in the dense regime, while it loses some power in the moderately sparse regime, and is powerless in the very sparse regime.

### 3. Gaussian mixture copula model

In this section we turn to the Gaussian mixture copula model introduced in (5). The setting is thus nonparametric, since the marginal distributions are completely unknown, and standard invariance considerations [17, Ch 6] lead us to consider test procedures that are based on the ranks. For this, we let  $R_i$  denote the rank of  $X_i$  among  $\{X_1, \dots, X_n\}$ , and similarly, we let  $S_i$  denote the rank of  $Y_i$  among  $\{Y_1, \dots, Y_n\}$ . (The ranks are in increasing order, say.)

Although not strictly necessary, we will assume that  $F$  and  $G$  in (5) are strictly increasing and continuous. In that case, the ranks are invariant with respect to transformations of the form  $(x, y) \mapsto (p(x), q(y))$  with  $p$  and  $q$  strictly increasing on the real line. In particular, for the rank tests that follow, this allows us to reduce their analysis under (5) to their analysis under (1).

#### 3.1. The covariance rank test

The covariance rank test is the analog of the covariance test of Section 2.2. It rejects for large values of  $T_n := \sum_i R_i S_i$  (redefined). As is well-known, this is equivalent to rejecting for large values of the Spearman rank correlation.

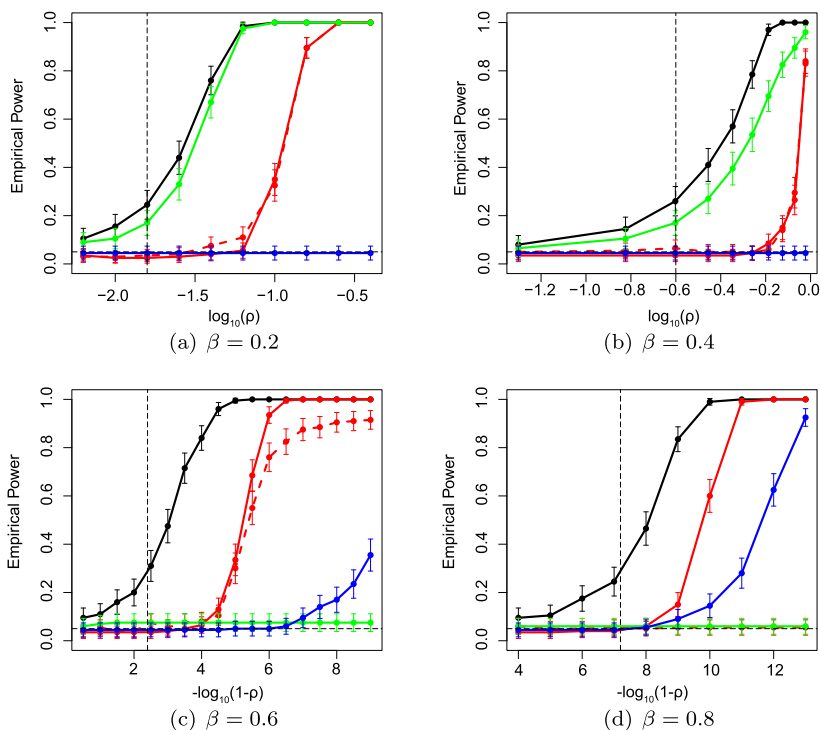


FIG 1. Empirical power comparison with 95% error bars for the likelihood ratio test (black), the Pearson correlation test (green), the extremes test (blue), the higher criticism test (red, solid) and the plug-in higher criticism test (red, dashed). (a) Dense regime where  $\beta = 0.2$ . (b) Dense regime where  $\beta = 0.4$ . (c) Sparse regime where  $\beta = 0.6$  and  $\rho \rightarrow 1$ . (d) Sparse regime where  $\beta = 0.8$  and  $\rho \rightarrow 1$ . The horizontal line marks the level (set at 0.05) and the vertical line marks the asymptotic detection boundary derived earlier. The sample size is  $n = 10^6$  and the power curves and error bars are based on 200 replications.

**Proposition 4.** For the testing problem (2) under the model (5), the covariance rank test achieves the detection boundary in the dense regime, while it is asymptotically powerless in the sparse regime.

*Proof.* We again divide the proof into the two regimes.

*Dense regime.* We start by considering the null hypothesis  $\mathcal{H}_0$ . From [13, Eq. 3.11–3.12, Ch. 11], we have

$$\begin{aligned} \mathbb{E}_0(T_n) &= n(n+1)^2/4 = n^3/4 + O(n^2), \\ \text{Var}_0(T_n) &= n^2(n-1)(n+1)^2/144 \asymp n^5, \end{aligned} \tag{9}$$

so that, using Chebyshev’s inequality,

$$\mathbb{P}_0(T_n \geq n^3/4 + a_n n^{5/2}) \rightarrow 0,$$

for any sequence  $(a_n)$  diverging to infinity.

We now turn to the alternative hypothesis  $\mathcal{H}_1$ . For convenience, we assume that the ranks run from 0 to  $n - 1$ . This does not change the test procedure since  $T_n = -\frac{1}{2} \sum_i (R_i - S_i)^2 + \text{const}$ , but makes the derivations somewhat less cumbersome. In particular, we have

$$R_i = \sum_{j=1}^n A_{ij}, \quad A_{ij} := \mathbb{I}\{X_i > X_j\},$$

$$S_i = \sum_{j=1}^n B_{ij}, \quad B_{ij} := \mathbb{I}\{Y_i > Y_j\},$$

so that

$$T_n = \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n A_{ij} B_{ik}.$$

For the expectation, we have

$$\begin{aligned} \mathbb{E}_1(T_n) &= n(n-1)(n-2) \mathbb{E}_1[A_{12}B_{13}] + O(n^2) \\ &= n^3 \mathbb{E}_1[A_{12}B_{13}] + O(n^2). \end{aligned}$$

The expectation is with respect to  $(X_1, Y_1), X_2, Y_3$  independent, with  $(X_1, Y_1)$  drawn from the mixture (1), and  $X_2$  and  $Y_3$  standard normal. Let  $U = (X_1 - X_2)/\sqrt{2}$  and  $V = (Y_1 - Y_3)/\sqrt{2}$ , so that  $\mathbb{E}_1[A_{12}B_{13}] = \mathbb{P}_1(U > 0, V > 0)$ . We note that  $(U, V)$  is bivariate normal with standard marginals. Moreover, when  $(X_1, Y_1)$  comes from the main component,  $U$  and  $V$  are uncorrelated, and therefore independent; while when  $(X_1, Y_1)$  comes from the contaminated component,  $U$  and  $V$  have correlation  $\rho/2$ . Therefore,

$$\mathbb{E}_1[A_{12}B_{13}] = (1 - \varepsilon)\Lambda(0) + \varepsilon\Lambda(\rho/2),$$

where  $\Lambda(\rho) = \mathbb{P}(U > 0, V > 0)$  under  $(U, V) \sim \mathcal{N}(0, \Sigma_\rho)$ . We immediately have  $\Lambda(0) = 1/4$ , and in general,<sup>2</sup>

$$\Lambda(\rho) = \frac{1}{4} + \frac{1}{2\pi} \sin^{-1}(\rho).$$

We conclude that

$$\begin{aligned} \mathbb{E}_1(T_n) &= n^3 \left[ \frac{1}{4} + \frac{1}{2\pi} \varepsilon \sin^{-1}(\rho/2) \right] + O(n^2) \\ &\geq \frac{1}{4} n^3 + \frac{1}{4\pi} n^3 \varepsilon \rho + O(n^2), \end{aligned}$$

using the fact that  $\sin^{-1}(a) \geq a$  for all  $a \geq 0$ . For the variance, we start with the second moment

$$\mathbb{E}_1(T_n^2) = n(n-1) \cdots (n-5) \mathbb{E}_1[A_{12}B_{13}A_{45}B_{46}] + O(n^5)$$

---

<sup>2</sup>This identity is well-known, and not hard to prove (<https://math.stackexchange.com/questions/255368/getting-px0-y0-for-a-bivariate-distribution>). It also appears, for example, in [21, Lem. 1].

$$= n^6 \mathbb{E}_1[A_{12}B_{13}A_{45}B_{46}] + O(n^5),$$

which then implies that

$$\begin{aligned} \text{Var}_1(T_n) &= n^6 \mathbb{E}_1[A_{12}B_{13}A_{45}B_{46}] + O(n^5) - \left[ n^3 \mathbb{E}_1[A_{12}B_{13}] + O(n^2) \right]^2 \\ &= O(n^5), \end{aligned}$$

the same bound we had for  $\text{Var}_0(T_n)$ . Thus, by Chebyshev's inequality, we have

$$\mathbb{P}_1 \left( T_n \leq \frac{1}{4}n^3 + \frac{1}{4\pi}n^3\varepsilon\rho - a_n n^{5/2} \right) \rightarrow 0,$$

for any sequence  $(a_n)$  diverging to infinity.

We consider the test with rejection region  $\{T_n \geq n^3/4 + a_n n^{5/2}\}$ . Our analysis implies that this test is asymptotically powerful when

$$n^3\varepsilon\rho/4\pi \geq 2a_n n^{5/2}.$$

If we choose  $a_n = \log n$ , for example, and  $\rho$  is parameterized as in (6), this happens for  $n$  large enough when  $\gamma < 1/2 - \beta$ .

*Sparse regime.* To prove that the covariance rank test is asymptotically powerless when  $\beta > 1/2$ , similarly as the covariance test, we show that, under  $\mathcal{H}_1$ ,  $T_n$  converges to the same limiting distribution as under  $\mathcal{H}_0$ . Under  $\mathcal{H}_0$ , we have [13, Ch. 11],

$$\frac{T_n - \zeta_n}{\tau_n} \rightharpoonup \mathcal{N}(0, 1), \quad n \rightarrow \infty, \tag{10}$$

where  $\zeta_n := \mathbb{E}_0(T_n)$  and  $\tau_n^2 := \text{Var}_0(T_n)$ . We place ourselves under  $\mathcal{H}_1$ , and show that (10) continues to hold. For this we use a simple coupling. We couple  $T_n$  with a new statistic  $T'_n$ , defined just like  $T_n$ , except that, for each pair  $(X_i, Y_i)$  drawn from the contaminated component, we replace  $Y_i$  by  $Y'_i \sim \mathcal{N}(0, 1)$  independent of  $X_i$  and any other variable. Let  $M$  denote the number of pairs drawn from the contaminated component, and note that  $M$  is random, having the binomial distribution with parameters  $(n, \varepsilon)$ . It's not hard to show that  $|T_n - T'_n| \leq Mn^2$ , so that  $|T_n - T'_n| = O_P(n^3\varepsilon)$ . And by construction,  $T'_n$  has the same distribution as  $T_n$  under  $\mathcal{H}_0$ . We use this in what follows

$$\frac{T_n - \zeta_n}{\tau_n} = \frac{T'_n - \zeta_n}{\tau_n} + \frac{T_n - T'_n}{\tau_n},$$

where, on the RHS, the first term converges weakly to the standard normal distribution, while the second term is  $= O_P(n^3\varepsilon/\tau_n) = o_P(1)$ , since  $\varepsilon = n^{1-\beta}$  with  $\beta > 1/2$  and  $\tau_n \asymp n^{5/2}$  by (9). We thus conclude that (10) with an application of Slutsky's theorem.  $\square$

### 3.2. The higher criticism rank test

The analog of the higher criticism test of (3) is a higher criticism based on the pairwise differences in ranks,  $D_i := |R_i - S_i|$ . To be specific, we define

$$\text{HC}_{\text{rank}} = \max_{0 \leq t \leq n/2} \frac{\sum_{i=1}^n \mathbb{I}\{D_i \leq t\} - nu(t)}{\sqrt{nu(t)(1-u(t))}},$$

where  $u(t)$  is the probability  $\mathbb{P}_0(D_i \leq t)$ , which can be expressed in closed form as

$$u(t) = \frac{n^2 - (n-t)(n-t-1)}{n^2} = \frac{n(2t+1) - t(t+1)}{n^2}.$$

Note that in this definition the denominator is only an approximation to the standard deviation of the numerator. The standard deviation has a closed-form expression which can be derived from a more general result of Hoeffding [15, Th. 2], but it is cumbersome and relatively costly to compute (although its computation is only done once for each  $n$ ). Also, there is a fair amount of flexibility in the choice of range of thresholds  $t$  considered. This particular choice seems to work well enough. As any other rank test, it is calibrated by permutation (or Monte Carlo if there are no ties in the data).

**Theorem 2.** *For the testing problem (2) under the model (5), the higher criticism rank test achieves the detection boundary in the dense and in the moderately sparse regimes.*

*Proof.* As usual, we first control the test statistic under the null, and then analyze its behavior under the alternative.

*Under the null hypothesis*

We start with the situation under the null hypothesis  $\mathcal{H}_0$ , where we show that  $\text{HC}_{\text{rank}}$  is of order at most  $O(\log n)$  based on the concentration inequality for randomly permuted sums. Fixing critical value  $t$ , define

$$a_{i,j} = \mathbb{I}\{|i-j| \leq t\}, \quad \text{for } 1 \leq i, j \leq n.$$

Since  $X$  is independent of  $Y$ , as we are under the null, we have that

$$\Delta(t) := \sum_{i=1}^n \mathbb{I}\{D_i \leq t\} \tag{11}$$

has the same distribution as  $A_n := \sum_{i=1}^n a_{i, \pi_n(i)}$  when  $\pi_n$  is a uniformly distributed random permutation of  $[n] := \{1, \dots, n\}$ . Note that

$$\mathbb{E}(A_n) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n a_{i,j} = \frac{n(2t+1) - t(t+1)}{n} = nu(t). \tag{12}$$

By [8, Prop. 1.1],

$$\mathbb{P}(|A_n - \mathbb{E}(A_n)| \geq b) \leq 2 \exp\left(-\frac{b^2}{4\mathbb{E}(A_n) + 2b}\right). \tag{13}$$



This implies that, for  $q \geq 1$ ,

$$\begin{aligned} & \mathbb{P}_0 (\Delta(t) \geq nu(t) + q\sqrt{nu(t)(1-u(t))}) \\ & \leq 2 \exp \left( -\frac{q^2 nu(t)(1-u(t))}{4nu(t) + 2q\sqrt{nu(t)(1-u(t))}} \right) \\ & \leq 2 \exp(-q/c_1), \end{aligned}$$

for some other constant  $c_1 > 0$ , using the fact that  $1/n \leq u(t) \leq 3/4 + 1/2n$  when  $0 \leq t \leq n/2$ , which is the range of  $t$ 's we are considering. Hence, choosing  $q = 2c_1 \log n$  and using the union bound, we have

$$\begin{aligned} \mathbb{P}_0(\text{HC}_{\text{rank}} \geq q) & \leq \sum_{t \leq n/2} \mathbb{P}_0 (\Delta(t) \geq nu(t) + q\sqrt{nu(t)(1-u(t))}) \\ & \leq 2(n/2 + 1) \exp(-q/c_1) \asymp 1/n \rightarrow 0. \end{aligned}$$

*Under the alternative hypothesis*

We now consider the alternative  $\mathcal{H}_1$ , and show that  $\text{HC}_{\text{rank}} \gg \log n$  in probability under the stated condition. For this, it suffices to find some  $t = t_n \leq n/2$  such that, for some  $q = q_n \gg \log n$ ,

$$\Delta(t) \geq nu(t) + q\sqrt{nu(t)(1-u(t))}, \tag{14}$$

with probability tending to 1 (under  $\mathcal{H}_1$ ).

Since rank-based methods are invariant with respect to increasing transformations, in the following analysis we simply assume that  $F = G = \Phi$ .

*Dense regime.* Define  $\hat{F}(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}\{X_i \leq x\}$  and  $\hat{G}(y) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}\{Y_i \leq y\}$ . These empirical distribution functions are useful because, by definition,  $R_i = n\hat{F}(X_i)$  and  $S_i = n\hat{G}(Y_i)$ , so that

$$\begin{aligned} D_i/n & = |R_i - S_i|/n \\ & = |\hat{F}(X_i) - \hat{G}(Y_i)| \\ & \leq |\hat{F}(X_i) - \Phi(X_i)| + |\Phi(X_i) - \Phi(Y_i)| + |\Phi(Y_i) - \hat{G}(Y_i)| \\ & \leq \underbrace{|\Phi(X_i) - \Phi(Y_i)|}_{M_i} + \underbrace{\|\hat{F} - \Phi\|_\infty + \|\hat{G} - \Phi\|_\infty}_K. \end{aligned}$$

This gives

$$\Delta(t) \geq \mathbb{I}\{K \leq k/n\} \Lambda(t), \quad \Lambda(t) := \sum_{i=1}^n \mathbb{I}\{M_i \leq (t-k)/n\}. \tag{15}$$

By the Dvoretzky–Kiefer–Wolfowitz (DKW) concentration inequality, there is a universal constant  $c_0$  such that, for any  $b \geq 0$ ,

$$\mathbb{P}(K \geq b) \leq c_0 \exp(-nb^2/c_0).$$

We choose  $k = (\log n)\sqrt{n}$ , and with that choice we have that  $\mathbb{I}\{K \leq k/n\} = 1 - Q_n$ , where  $Q_n$  is Bernoulli with parameter bounded by  $\eta := c_0 \exp(-(\log n)^2/c_0)$  (so that  $Q_n = O_P(\eta)$ ).

As for the sum, the  $M_i$  are iid, and for an observation  $(X_i, Y_i)$  that comes from the null component,  $X_i, Y_i$  are iid standard normal, while when it comes from the contaminated component,  $X_i, Y_i$  are still marginally standard normal but no longer independent:  $Y_i = \sqrt{1 - \rho^2} \tilde{Y}_i + \rho X_i$ , where  $\tilde{Y}_i$  is independent of  $X_i$  and also standard normal. We thus have

$$\mathbb{P}_1(M_i \leq s) = (1 - \varepsilon)v_s(0) + \varepsilon v_s(\rho),$$

where

$$v_s(\rho) := \mathbb{E}[f_s(Z, \sqrt{1 - \rho^2}Z' + \rho Z)],$$

where in the expectation  $Z, Z'$  are iid standard normal, and  $f_s(z, z') := \mathbb{I}\{|\Phi(z) - \Phi(z')| \leq s\}$  is bounded and measurable. Elementary calculations show that  $v_s(0) = 1 - (1 - s)^2$ , and an application of Lemma 1 shows that  $v_s$  is infinitely differentiable, with derivative at 0 equal to  $\mathbb{E}[f_s(Z, Z')ZZ']$ , and second derivative uniformly bounded over  $[-1/2, 1/2]$  by some numerical constant, say  $c_2$ , independently of  $s$ . Recalling that  $\rho$  is small in the present regime, a Taylor development based on the above gives

$$v_s(\rho) \geq 1 - (1 - s)^2 + v'_s(0)\rho - c_2\rho^2/2, \quad \rho \in [-1/2, 1/2].$$

In the dense regime, remember that  $0 < \beta < 1/2$  and  $\rho = n^{-\gamma}$ . We place ourselves above the detection boundary, meaning that we fix  $\gamma < 1/2 - \beta$ . Here we choose  $t = n/2$  (assumed to be an integer for convenience), let  $s = (t - k)/n = 1/2 - k/n$ . We note that  $v'_s(0)$  is continuous in  $s$  (by dominated convergence), and because  $s \rightarrow 1/2$  in our setting, we have

$$v'_s(0) \rightarrow v'_{1/2}(0) = \mathbb{E}[\mathbb{I}\{|\Phi(Z) - \Phi(Z')| \leq 1/2\}ZZ'] =: c_1 > 0.$$

Indeed, using the fact that

$$|\Phi(z) - \Phi(z')| \leq 1/2 \Leftrightarrow (\Phi(z) - 1/2) \vee 0 \leq \Phi(z') \leq (\Phi(z) + 1/2) \wedge 1,$$

with  $\Phi(z) \leq 1/2$  if and only if  $z \leq 0$ , we have

$$\begin{aligned} c_1 &= \int_0^\infty \underbrace{\int_{\Phi^{-1}(\Phi(z)-1/2)}^\infty \phi(z')dz'}_{>0} \underbrace{\phi(z)z dz}_{>0} \\ &+ \int_{-\infty}^0 \underbrace{\int_{-\infty}^{\Phi^{-1}(\Phi(z)+1/2)} \phi(z')dz'}_{<0} \underbrace{\phi(z)z dz}_{<0}, \end{aligned}$$

where the inner integrals are positive by the fact that  $\phi$  is symmetric, and the inequalities are indeed strict except when  $z = 0$ .

Thus, eventually (as  $n \rightarrow \infty$ ),

$$v_s(\rho) \geq 1 - (1 - s)^2 + (c_1/2)\rho.$$

Thus, an application of Chebyshev's inequality gives

$$\Lambda(n/2) \geq n[(1 - \varepsilon)v_s(0) + \varepsilon v_s(\rho)] + O_P(\sqrt{n}).$$

Putting everything together, we have

$$\begin{aligned} \Delta(t) - nu(t) &= (1 + O_P(\eta))n[(1 - \varepsilon)v_s(0) + \varepsilon v_s(\rho)] + O_P(\sqrt{n}) - nu(t) \\ &\geq n[1 - (1 - (t - k)/n)^2 - u(t)] + n\varepsilon(c_1/2)\rho + O_P(n\eta) + O_P(\sqrt{n}) \\ &= n\varepsilon(c_1/2)\rho + O_P((\log n)\sqrt{n}), \end{aligned}$$

using the fact that  $\eta = o(1/n^2)$ . For (14) to hold it thus suffices that  $n\varepsilon\rho \gg (\log n)\sqrt{n}$ , which is the case since  $n\varepsilon\rho = n^{1-\beta-\gamma}$  with  $1 - \beta - \gamma > 1/2$ .

*Moderately sparse regime.* Let  $I_0$  and  $I_1$  index the observations coming from the null and contaminated components, respectively. We have

$$\Delta(t) = \sum_{i \in I_0} \mathbb{I}\{D_i \leq t\} + \sum_{i \in I_1} \mathbb{I}\{D_i \leq t\} =: \Delta_0(t) + \Delta_1(t). \tag{16}$$

We lower bound both terms on the right-hand side, starting with  $\Delta_0(t)$ . To do this, we consider a slightly smaller threshold, specifically  $t_0 = (1 - \omega)t$  with  $\omega = o(1)$  specified below, and compare  $\Delta_0(t)$  with  $\Delta^0(t_0) := \sum_{i \in I_0} \mathbb{I}\{D_i^0 \leq t_0\}$ , where  $D_i^0 := |R_i^0 - S_i^0|$  with  $R_i^0$  denoting the rank of  $X_i$  among  $\{X_j : j \in I_0\}$  and  $S_i^0$  denoting the rank of  $Y_i$  among  $\{Y_j : j \in I_0\}$ . Conditional on  $|I_0| = n_0$ ,  $\Delta^0(t_0)$  has the same distribution as  $\Delta(t_0)$  in (11) under the null hypothesis but with  $n$  replaced by  $n_0$ , so that from (12) we deduce that it has expectation

$$\mu := (n_0(2t_0 + 1) - t_0(t_0 + 1))/n_0,$$

and from (13) that

$$\Delta^0(t_0) \geq \mu - 8(\log n)\sqrt{\mu \vee \log n}$$

with probability at least  $1 - 2/n$  when  $n$  is large enough. (Again, this is conditional on  $|I_0| = n_0$ .) Because  $\varepsilon \ll n^{-1/2}$  in the present regime, we have  $|I_0| \geq n - (\log n)\sqrt{n}$  with probability at least  $1 - 1/n$  when  $n$  is large enough. Also, we will choose  $t$  below such that  $\sqrt{n} \ll t \ll n$ , and  $\omega$  such that  $\omega \ll 1$ , so that  $t_0 \sim t$ . Together, this implies that

$$\Delta^0(t_0) \geq 2t_0 + 1 - \frac{t_0(t_0 + 1)}{n - (\log n)\sqrt{n}} - 8(\log n)\sqrt{2t_0 + 1} = 2t_0 - \frac{t_0^2}{n} - O((\log n)\sqrt{t}),$$

eventually, with probability at least  $1 - 3/n$ .

We now claim that, with probability tending to 1,  $\Delta_0(t) \geq \Delta^0(t_0)$ . Indeed, by definition of the ranks  $R_i$  and modified ranks  $R_i^0$ , we have

$$R_i - R_i^0 = \sum_{j \in I_1} \mathbb{I}\{X_j \leq X_i\} = |I_1| \hat{F}_1(X_i),$$

where  $\hat{F}_1(x) := \frac{1}{|I_1|} \sum_{j \in I_1} \mathbb{I}\{X_j \leq x\}$  is the empirical distribution function associated with the contaminated  $X$  observations. In particular, when  $|I_0| = n_0$ , so that  $|I_1| = n - n_0 =: n_1$ , we have

$$|R_i - R_i^0 - n_1 \Phi(X_i)| \leq n_1 \|\hat{F}_1 - \Phi\|_\infty,$$

valid for all  $i \in I_0$ . At the same time, and with analogous notation, we also have

$$|S_i - S_i^0 - n_1 \Phi(Y_i)| \leq n_1 \|\hat{G}_1 - \Phi\|_\infty,$$

valid for all  $i \in I_0$ . Combining these, we obtain

$$\underbrace{|R_i - S_i|}_{D_i} \leq \underbrace{|R_i^0 - S_i^0|}_{D_i^0} + n_1 |\Phi(X_i) - \Phi(Y_i)| + n_1 \underbrace{(\|\hat{F}_1 - \Phi\|_\infty + \|\hat{G}_1 - \Phi\|_\infty)}_{=: K_1},$$

valid for all  $i \in I_0$ . Letting  $\hat{F}_0$  denote the empirical distribution function of  $\{X_i : i \in I_0\}$  and  $\hat{G}_0$  denote that of  $\{Y_i : i \in I_0\}$ , we have

$$|\Phi(X_i) - \Phi(Y_i)| \leq \underbrace{|\hat{F}_0(X_i) - \hat{G}_0(Y_i)|}_{D_i^0/n_0} + \underbrace{\|\hat{F}_0 - \Phi\|_\infty + \|\hat{G}_0 - \Phi\|_\infty}_{=: K_0}, \quad (17)$$

valid for all  $i \in I_0$ . Note that this is conditional on  $|I_0| = n_0$  and that the distributions of  $K_0$  and  $K_1$  depend (implicitly) on  $n_0$  (and  $n_1$ ). We conclude that, conditional on  $|I_0| = n_0$ , for any  $i \in I_0$ ,

$$D_i \leq (n/n_0)D_i^0 + n_1(K_0 + K_1). \quad (18)$$

Applying the DKW inequality with the tight constant, we have that  $K_0 \leq (\log n)/\sqrt{n_0}$  and  $K_1 \leq (\log n)/\sqrt{n_1}$  with probability at least  $1 - 2/n$  when  $n$  is large enough, and when this is the case,  $D_i \leq (n/n_0)D_i^0 + 2(\log n)\sqrt{n_1}$ , assuming that  $n_0 \geq n_1$ . This is given  $|I_0| = n_0$  and (therefore)  $|I_1| = n_1$ , and we also know that  $|I_0| \geq n - (\log n)\sqrt{n}$  and  $|I_1| \leq 2n\varepsilon$  with probability at least  $1 - 1/n$  when  $n$  is large enough. (We are using that  $|I_1| \sim \text{Bin}(n, \varepsilon)$  with  $n\varepsilon = n^{1-\beta}$  with  $\beta < 1$ .) Hence, with probability at least  $1 - 3/n$ ,

$$D_i \leq \frac{nD_i^0}{n - (\log n)\sqrt{n}} + 2(\log n)\sqrt{2n\varepsilon},$$

for any  $i \in I_0$ . In particular, if we choose  $\omega = (\log n)^2 \max(1/\sqrt{n}, \sqrt{n\varepsilon}/t)$ , then, with probability at least  $1 - 2/n$  when  $n$  is large enough,  $D_i^0 \leq t_0$  implies that  $D_i \leq t$  for any  $i \in I_0$ , implying that  $\Delta_0(t) \geq \Delta^0(t_0)$ .

We thus conclude that

$$\Delta_0(t) \geq 2t_0 - t_0^2/n - O_P((\log n)\sqrt{t}).$$

As for  $\Delta_1(t)$ , as in (15), we have

$$\Delta_1(t) \geq \mathbb{I}\{K \leq k/n\}\Lambda_1(t), \quad \Lambda_1(t) := \sum_{i \in I_1} \mathbb{I}\{M_i \leq (t - k)/n\}.$$

We choose  $k = (\log n)\sqrt{n}$  as we did before, so that  $\mathbb{I}\{K \leq k/n\} = 1 + O_P(\eta)$ , with the same  $\eta$  defined previously. As for the sum,  $\Lambda_1(t)$  has the same distribution as  $\sum_{i=1}^B \mathbb{I}\{\tilde{M}_i \leq (t - k)/n\}$ , with  $B$  binomial with parameters  $(n, \varepsilon)$  and  $\tilde{M}_i = |\Phi(\tilde{X}_i) - \Phi(\tilde{Y}_i)|$  with  $(\tilde{X}_i, \tilde{Y}_i)$  iid normal with standard normal marginals and correlation  $\rho$ . In particular,

$$\tilde{M}_i \leq \frac{1}{\sqrt{2\pi}}|\tilde{X}_i - \tilde{Y}_i| =: \frac{1}{\sqrt{\pi}}|\tilde{U}_i|,$$

by the fact that  $\Phi$  has derivative bounded by  $1/\sqrt{2\pi}$  everywhere, and where  $\tilde{U}_i \sim \mathcal{N}(0, 1 - \rho)$ , and simple calculations give

$$v(s) := \mathbb{P}(\tilde{M}_i \leq s) \geq \Psi\left(\frac{\sqrt{\pi}s}{\sqrt{1-\rho}}\right) =: \lambda(s), \quad s \in [0, 1].$$

We thus have

$$\mathbb{E}_1(\Lambda_1(t)) = n\varepsilon v((t - k)/n),$$

and

$$\text{Var}_1(\Lambda_1(t)) = \text{Var}(B)v((t - k)/n)^2 + \mathbb{E}(B)v((t - k)/n) \leq 2n\varepsilon v((t - k)/n),$$

and applying Chebyshev's inequality, we thus have

$$\begin{aligned} \Lambda_1(t) &= n\varepsilon v((t - k)/n) + O(\sqrt{n\varepsilon v((t - k)/n)}) \\ &\geq (1 + o_P(1))n\varepsilon \lambda((t - k)/n), \end{aligned}$$

as long as the right-hand side diverges.

In the moderately sparse regime, remember that  $1/2 < \beta < 3/4$  and  $\rho = 1 - n^{-\gamma}$ . We place ourselves just above the detection boundary, meaning that we fix  $\gamma > 4(\beta - 1/2)$ . We focus on the harder sub-case where, in addition,  $\gamma < 2\beta$ . In that case, we can fix  $a$  such that  $1/2 > a > \gamma/2$  and  $1/2 - \beta + \gamma/2 - a/2 > 0$ , and set  $t = \lfloor n^{1-a} \rfloor$ . Note that such a real number  $a$  exists, and that  $t \leq n/2$  with  $t \gg k$ . We also have  $n\varepsilon = n^{1-\beta}$  and  $u(t) \asymp t/n \asymp n^{-a}$ , as well as

$$\lambda((t - k)/n) = \Psi\left(\frac{\sqrt{\pi}(t - k)}{n\sqrt{1-\rho}}\right) \asymp n^{\gamma/2-a}, \quad \text{since } \frac{\sqrt{\pi}(t - k)}{n\sqrt{1-\rho}} \asymp n^{\gamma/2-a} \rightarrow 0,$$

and  $\Psi$  is differentiable at 0 with positive derivative. In particular,  $n\varepsilon \lambda((t - k)/n) \asymp n^{1-\beta+\gamma/2-a} \rightarrow \infty$ . Putting everything together, we have

$$\Delta(t) - nu(t) \geq 2t_0 - t_0^2/n - O_P((\log n)\sqrt{t}) + (1 + o_P(1))n\varepsilon \lambda((t - k)/n)$$

$$\begin{aligned}
 & - (2t + 1 - t(t + 1)/n) \\
 & = -O_P((\log n)\sqrt{t}) + n\varepsilon(1 + o_P(1))\lambda((t - k)/n),
 \end{aligned}$$

after some simplifications, using the definition of  $\omega$  above and the fact that  $\sqrt{n} \ll t \ll n$ . For (14) to hold, it is thus enough to have  $n\varepsilon\lambda((t - k)/n) \gg (\log n)\sqrt{t}$ , which is the case since

$$\frac{n\varepsilon\lambda((t - k)/n)}{\sqrt{t}} \asymp \frac{n^{1-\beta+\gamma/2-a}}{n^{1/2-a/2}} = n^{1/2-\beta+\gamma/2-a/2},$$

with  $1/2 - \beta + \gamma/2 - a/2 > 0$  by our choice of  $a$ . □

**Lemma 1.** *Let  $A, B$  be iid standard normal, and for  $f: \mathbb{R}^2 \rightarrow [0, 1]$  measurable and  $r \in [-1, 1]$ , define  $\Gamma_f(r) = \mathbb{E}[f(A, \sqrt{1 - r^2}B + rA)]$ . Then  $\Gamma_f$  is infinitely differentiable, with  $\Gamma'_f(0) = \mathbb{E}[f(A, B)AB]$ , and with  $\sup_{|r| \leq 1/2} |\Gamma''_f(r)|$  bounded by some numerical constant (independent of  $f$ ).*

*Proof.* We have

$$\Gamma_f(r) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(a, b)\phi(a, b; r)dadb,$$

where

$$\phi(a, b; r) := \frac{\exp[-(a^2 - 2rab + b^2)/(2 - 2r^2)]}{2\pi\sqrt{1 - r^2}}.$$

An application of the dominated convergence theorem allows us to differentiate under the integral at will. In particular,

$$\Gamma_f^{(k)}(r) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(a, b)\partial_r^k \phi(a, b; r)dadb.$$

Elementary calculations show that  $\partial_r \phi(a, b; 0) = (2\pi)^{-1}ab \exp[-(a^2 + b^2)/2]$ . We also obtain

$$|\Gamma''_f(r)| \leq \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |\partial_r^2 \phi(a, b; r)|dadb,$$

which is easily seen to uniformly bounded for  $|r| \leq 1/2$ . □

It is natural to wonder whether the higher criticism rank test has some power in the very sparse regime. The following indicates that it is powerless in that regime.

**Proposition 5.** *Consider the very sparse regime in the most extreme case where  $\rho = 1$ . In that setting, any test that rejects for large values of  $\Delta(t) := \sum_{i=1}^n \mathbb{I}\{D_i \leq t\}$  (where the threshold  $t$  is allowed to vary with  $n$ ) is asymptotically powerless.*

*Proof.* By a compactness argument, we may assume that either  $t \rightarrow \infty$  or  $t$  is constant (as  $n$  varies). We start with the former and address the latter at the

end. We focus on the case where  $t \ll n$ , as the case where  $t \asymp n$  can be dealt with in a very similar fashion.

*Under the null hypothesis*

We first consider the behavior of  $\Delta(t)$  under the null hypothesis, and argue that  $\Delta(t)$  is asymptotically normally distributed. This is based on an application of a combinatorial central limit theorem due to Hoeffding [15]. Remember that under  $\mathcal{H}_0$ ,  $\Delta(t)$  has the distribution of  $A_n = \sum_{i=1}^n a_{i, \pi_n(i)}$  when  $\pi_n$  is a uniformly distributed random permutation of  $[n]$  and  $a_{i,j} = \mathbb{I}\{|i - j| \leq t\}$ . We saw that

$$\mathbb{E}(A_n) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n a_{i,j} = \frac{n(2t + 1) - t(t + 1)}{n} = nu(t),$$

and, as derived in [15], we also have

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

where

$$d_{i,j} = a_{i,j} - \frac{1}{n} \sum_{g=1}^n a_{g,j} - \frac{1}{n} \sum_{h=1}^n a_{i,h} + \frac{1}{n^2} \sum_{g=1}^n \sum_{h=1}^n a_{g,h}.$$

[15, Th. 3] implies that  $A_n$  is asymptotically normal when

$$\frac{\max_{i,j \in [n]} d_{i,j}^2}{\frac{1}{n^2} \sum_{i \in [n]} \sum_{j \in [n]} d_{i,j}^2} \rightarrow 0.$$

Elementary but somewhat tedious calculations yield that this is the case if and only if  $t \rightarrow \infty$ , which we assume. Further elementary calculations, in part similar to some appearing in the proof of Theorem 2, yield that

$$\frac{\text{Var}(A_n)}{nu(t)(1 - u(t))} \rightarrow 1.$$

We thus have, under the null hypothesis,

$$\frac{\Delta(t) - nu(t)}{\sqrt{nu(t)(1 - u(t))}} \rightarrow \mathcal{N}(0, 1),$$

and therefore, together with the fact that  $1 \ll t \ll n$ , we conclude that

$$\frac{\Delta(t) - 2t + t^2/n}{\sqrt{2t}} \rightarrow \mathcal{N}(0, 1), \tag{19}$$

again under the null hypothesis.

*Under the alternative hypothesis*

We now consider the alternative, again in the very sparse regime and in the most advantageous case where  $\rho = 1$ , and show that the same weak limit holds. For this, we follow the arguments of the proof of Theorem 2 in the moderately sparse regime, although in the reverse direction so-to-speak. We use the same notation.

Starting from the decomposition (16), we have

$$\frac{\Delta(t) - 2t + t^2/n}{\sqrt{2t}} = \frac{\Delta_0(t) - 2t + t^2/n}{\sqrt{2t}} + \frac{\Delta_1(t)}{\sqrt{2t}}. \quad (20)$$

In what follows, we first show that the first term on the RHS is asymptotically standard normal, and then we show that the second term converges to 0 in probability.

*First term in (20).* For  $i \in I_0$ , as in (18) but in reverse, we have

$$\begin{aligned} D_i^0 &\leq (1 + |I_1|/n)D_i + |I_1|(K_0 + K_1) \\ &\leq (1 + \varepsilon + (\log n)\sqrt{\varepsilon/n})D_i + (\log n)\sqrt{n\varepsilon}, \end{aligned}$$

with probability tending to 1 uniformly over  $i \in I_0$ . Assuming this is true, then  $D_i \leq t$  implies that

$$\begin{aligned} D_i^0 &\leq (1 + \varepsilon + (\log n)\sqrt{\varepsilon/n})t + (\log n)\sqrt{n\varepsilon} \\ &\leq t_0 := (1 + \varepsilon)t + 2(\log n)\sqrt{n\varepsilon}. \end{aligned}$$

Hence, with probability tending to 1,

$$\Delta_0(t) \leq \Delta^0(t_0).$$

As before, conditional on  $|I_0| = n_0$ ,  $\Delta^0(t_0)$  has the same distribution as  $\Delta(t_0)$  in (11) under the null hypothesis but with  $n$  replaced by  $n_0$ . This, the fact that  $|I_0| \geq n - O_P(\sqrt{n})$ , and (19), implies that

$$\frac{\Delta^0(t_0) - 2t_0 + t_0^2/n}{\sqrt{2t_0}} \rightarrow \mathcal{N}(0, 1).$$

We used the fact that  $t_0^2/n \leq t_0^2/|I_0| \leq t_0^2/(n - O(\sqrt{n}))$ , which implies that

$$\frac{t_0^2/|I_0|}{\sqrt{t_0}} = \frac{t_0^2/n}{\sqrt{t_0}} + \underbrace{O(t_0\sqrt{t_0}/n\sqrt{n})}_{o(1)},$$

where the  $O$  term is  $o(1)$  by the fact that  $t_0/n = o(1)$ . Continuing, with probability tending to 1, we have

$$\frac{\Delta_0(t) - 2t + t^2/n}{\sqrt{2t}} \leq \frac{\Delta^0(t_0) - 2t + t^2/n}{\sqrt{2t}}$$



$$\begin{aligned}
 &= \sqrt{t_0/t} \frac{\Delta^0(t_0) - 2t_0 + t_0^2/n}{\sqrt{2t_0}} + \frac{2t_0 - t_0^2/n - 2t + t^2/n}{\sqrt{2t}} \\
 &\rightarrow \mathcal{N}(0, 1),
 \end{aligned} \tag{21}$$

whenever  $t_0/t \rightarrow 1$  and  $(t_0 - t)/\sqrt{t} \rightarrow 0$  (using the fact that  $t \leq t_0 \ll n$ ). This is the case exactly when  $t \gg (\log n)^2 n \varepsilon$ .

We now consider the complementary case. In fact, what follows applies when  $t \leq \sqrt{n}$ . We use a slightly different strategy. Recall that, for  $i \in I_0$ ,

$$\begin{aligned}
 R_i - R_i^0 &= \sum_{j \in I_1} \mathbb{I}\{X_j \leq X_i\}, \\
 S_i - S_i^0 &= \sum_{j \in I_1} \mathbb{I}\{Y_j \leq Y_i\},
 \end{aligned}$$

and combined with the triangle inequality, and recalling that  $X_j = Y_j$  when  $j \in I_1$ , we have

$$|D_i - D_i^0| \leq W_i := \sum_{j \in I_1} \mathbb{I}\{X_i \wedge Y_i \leq X_j \leq X_i \vee Y_i\}.$$

Consider the event

$$\Omega = \{|I_1| \leq 2n\varepsilon, K_0 \leq (\log n)/\sqrt{n}\},$$

which happens with probability tending to one. Given  $\Omega$ , we have

$$\begin{aligned}
 \{D_i \leq t\} &= \{D_i \leq t, D_i^0 \leq t\} \cup \{D_i \leq t, D_i^0 > t\} \\
 &\subset \{D_i^0 \leq t\} \cup \{W_i \geq D_i^0 - t, 2n\varepsilon + t \geq D_i^0 > t\},
 \end{aligned}$$

using the fact that  $D_i^0 \leq D_i + |I_1|$ , so that

$$\Delta_0(t) \leq \Delta^0(t) + \sum_{i \in I_0} \mathbb{I}\{W_i \geq D_i^0 - t, 2n\varepsilon + t \geq D_i^0 > t\}. \tag{22}$$

Given  $\{(X_k, Y_k) : k \in I_0\}$ , and conditional on  $(|I_0|, |I_1|) = (n_0, n_1)$ ,  $W_i$  is binomial with parameters  $n_1$  and  $P_i := |\Phi(X_i) - \Phi(Y_i)|$ . As in (17), the latter is bounded by  $D_i^0/n + K_0$ , which itself is bounded (eventually) by  $2(\log n)/\sqrt{n}$  under  $\Omega$  when  $D_i^0 = d$  with  $d \leq t + 2n\varepsilon$  (since we work under the assumption that  $t \leq \sqrt{n}$ ). Thus, for such a  $d$ , eventually,

$$\begin{aligned}
 \mathbb{P}(W_i \geq w \mid \Omega, D_i^0 = d) &\leq \mathbb{E} \left[ \mathbb{P}(W_i \geq w \mid \Omega, D_i^0 = d, (X_k, Y_k)_{k \in I_0}) \right] \\
 &\leq 2 \mathbb{P}(W_i \geq w \mid P_i \leq 2(\log n)/\sqrt{n}) \\
 &\leq 2 \text{Prob}(\text{Bin}(2n\varepsilon, 2(\log n)/\sqrt{n}) \geq w) \\
 &\leq c_0(n\varepsilon \times (\log n)/\sqrt{n})^w,
 \end{aligned}$$

where  $c_0$  is a universal constant. The factor of 2 in the second inequality comes from de-conditioning from  $\{K_0 \leq (\log n)/\sqrt{n}\}$ . In the last line we used the fact

that  $\text{Prob}(\text{Bin}(m, q) \geq k) \leq \binom{m}{k} q^k$ , referred to as the Giné–Zinn inequality in [9]. We also have

$$\mathbb{P}(D_i^0 = d \mid \Omega) \leq 2\mathbb{P}(D_i^0 = d \mid |I_1| \leq 2n\varepsilon) \leq 2\frac{2}{n - 2n\varepsilon} \leq \frac{5}{n},$$

eventually, using the fact that  $\mathbb{P}(D_i^0 = d \mid |I_0| = n_0) \leq 2/n_0$ . Together, this yields

$$\begin{aligned} & \mathbb{P}(W_i \geq D_i^0 - t, 2n\varepsilon + t \geq D_i^0 > t \mid \Omega) \\ & \leq \sum_{d \geq t+1}^{t+2\lfloor n\varepsilon \rfloor} \mathbb{P}(W_i \geq d - t \mid \Omega, D_i^0 = d) \times \mathbb{P}(D_i^0 = d \mid \Omega) \\ & \leq c_1 \sum_{d \geq t+1}^{t+2\lfloor n\varepsilon \rfloor} ((\log n)\sqrt{n\varepsilon})^{d-t} \times \frac{1}{n}, \\ & \leq c_1 \times \frac{2}{n} \times ((\log n)\sqrt{n\varepsilon}). \end{aligned}$$

Hence, the second term on the RHS of (22) has expectation of order at most  $n$  times the last term in our last derivations, which is of order at most  $(\log n)\sqrt{n\varepsilon} = o(1)$ . Since that term is integer-valued, this implies that  $\Delta_0(t) \leq \Delta^0(t)$  with probability tending to one. In particular, (21) applies.

*Second term in (20).* Consider  $i \in I_1$ . Because  $\rho = 1$ , we have  $X_i = Y_i$ , and conditional on  $X_i = z$ ,  $R_i - 1$  and  $S_i - 1$  are iid with distribution  $\text{Bin}(n - 1, p)$  where  $p := \Phi(z)$ . In particular,  $D_i$  has the distribution of  $|U - V|$  where  $U$  and  $V$  are iid with distribution  $\text{Bin}(n - 1, P)$  and  $P \sim \text{Unif}(0, 1)$ . Let  $u_2(t)$  denote the probability that  $D_i \leq t$ . We want to bound  $u_2(t)$  from above.

For  $p \in [0, 1]$ , define  $g(p)$  as the probability that  $|U - V| \leq t$  when  $U$  and  $V$  are iid  $\text{Bin}(n - 1, p)$ , and note that  $u_2(t) = \int_0^1 g(p) dp$ . Define  $\sigma^2 = 2(n - 1)p(1 - p)$ , which is the variance of  $U - V$ , and also  $h(a) = \mathbb{P}((U - V)/\sigma \leq a)$ . Using the fact that  $U - V$  is integer valued, we have

$$g(p) = h(t/\sigma) - h(-(t + 1)/\sigma) \leq \Phi(t/\sigma) - \Phi(-(t + 1)/\sigma) + 2\|h - \Phi\|_\infty.$$

Where  $\Phi$  is the standard normal distribution function. Because  $\Phi$  has derivative bounded by  $1/\sqrt{2\pi}$  everywhere, the first term on the RHS is  $= O(t/\sigma)$ . For the second term, we use the Berry–Esseen inequality (seeing  $U$  and  $V$ , each, as the sum of  $n - 1$  iid  $\text{Ber}(p)$  random variables), to get that it is  $= O(1/\sigma)$ . Therefore, since  $t \geq 1$ , there is a universal constant  $c_0$  such that  $g(p) \leq c_0 t/\sigma$ . Of course, being a probability, we also have  $g(p) \leq 1$ . Hence,

$$u_2(t) = \int_0^1 g(p) dp \leq \int_0^1 \left( 1 \wedge \frac{c_0 t}{2(n - 1)p(1 - p)} \right) dp \asymp 1 \wedge t/\sqrt{n}.$$

Now, by Markov's inequality, and the fact that  $|I_1|$  is binomial with parameters  $(n, \varepsilon)$ , the second term in (20) is

$$= \frac{O_P(n\varepsilon)O_P(u_2(t))}{\sqrt{nu(t)(1 - u(t))}} \asymp \frac{n^{1-\beta}(1 \wedge t/\sqrt{n})}{\sqrt{t}} \asymp n^{1-\beta}t^{-1/2} \wedge n^{1/2-\beta}t^{1/2} \rightarrow 0,$$

for any choice of  $t$  when  $\beta > 3/4$  (very sparse regime).

*Special case:  $t$  constant.* When  $t$  is constant, the null distribution of  $\Delta(t)$  is known to converge to the Poisson distribution of mean  $2t + 1$ . (See [3, Exa 1.3], which is only cosmetically different.) The control under the alternative can be secured in exactly the same way. In particular, it holds that  $\Delta_0(t) \leq \Delta^0(t)$  with probability tending to one, with  $\Delta^0(t)$  having the same asymptotic distribution (Poisson with mean  $2t + 1$ ).  $\square$

### 3.3. Numerical experiments

We consider the same setting as in Section 2 and compare the two nonparametric tests, the covariance rank test and the higher criticism rank test, to the parametric tests. The p-values for the higher criticism rank test are obtained based on  $10^5$  permutations, while the p-values for the covariance rank test are taken from the limiting distribution based on its correspondence with the Spearman rank correlation.

The results are presented in Figure 2. In finite samples, the higher criticism rank test exhibits substantially more power than the higher criticism in the dense and moderately sparse regime. We have no good explanation for this rather surprising phenomenon. However, the higher criticism rank test has no power in the very sparse regime, and neither does the covariance rank test.

## 4. Discussion

**The power residing in the  $V_i$**  In Proposition 2 we established that the higher criticism test based on  $U_1, \dots, U_n$  achieves the detection boundary in the Gaussian mixture model. It is natural, however, to ask whether one could do better in finite samples by also utilizing  $V_1, \dots, V_n$ . We performed some side experiments to quantify this by comparing the full LRT, meaning the LRT based on  $(U_1, V_1), \dots, (U_n, V_n)$ , the LRT based on  $U_1, \dots, U_n$  only, and the LRT based on  $V_1, \dots, V_n$  only. We did so in the same parametric setting of Section 2.4. The results are reported in Figure 3, and can be to some extent anticipated from our previous work [1]. In a nutshell, in the dense regime, what matters is the deviation of the variance from 1, and this is felt by all tests, so that the  $U$ -LRT and the  $V$ -LRT are seen to be also as powerful as the full LRT. In the sparse regime, however, we can see that the  $V$ -LRT has essentially no power. This is due to the fact that the  $V_i$ 's in that case have variance  $1 + \rho$ , which is bounded from above by 2, so that no test depending on the  $V_i$ 's can have any power as we show in [1]. The  $U$ -LRT, which we know to be asymptotically optimal to first order, remains competitive, although now clearly less powerful than the full LRT.

**The power of rank tests in the very sparse regime** In Proposition 5 we argued, we hope convincingly, that no test that resembles the higher criticism rank test has any power in the very sparse regime ( $\beta > 3/4$ ). This seems clear from the experiments reported in Figure 2. This begs the question of whether

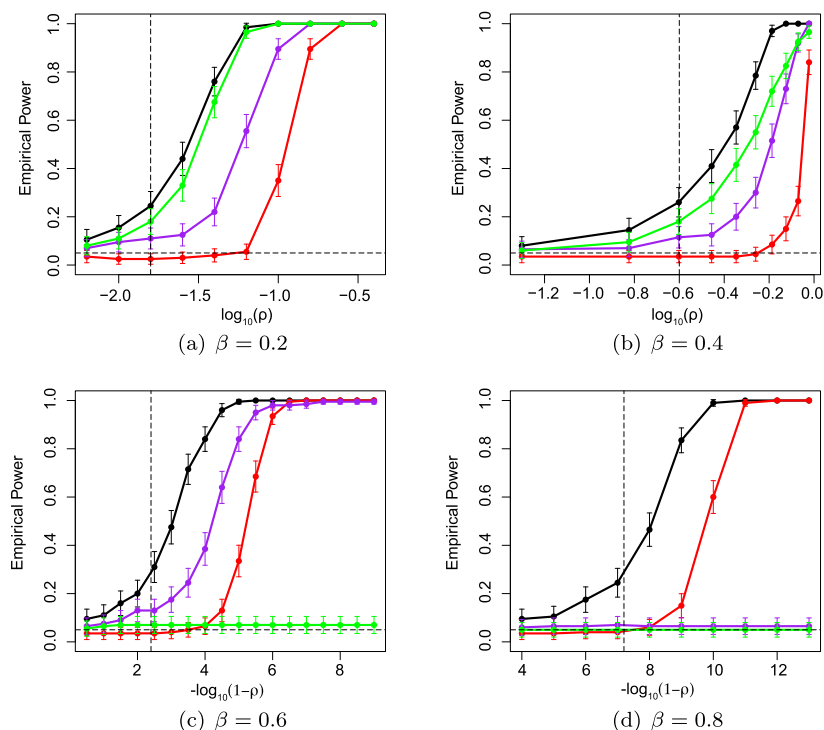


FIG 2. Empirical power comparison with 95% error bars for the likelihood ratio test (black), the covariance rank test (green), the higher criticism test (red) and the higher criticism rank test (purple). (a) Dense regime where  $\beta = 0.2$ . (b) Dense regime where  $\beta = 0.4$ . (c) Sparse regime where  $\beta = 0.6$  and  $\rho \rightarrow 1$ . (d) Sparse regime where  $\beta = 0.8$  and  $\rho \rightarrow 1$ . The horizontal line marks the level (set at 0.05) and the vertical line marks the asymptotic detection boundary derived earlier. The sample size is  $n = 10^6$  and the power curves and error bars are based on 200 replications.

there are any rank tests that have any (asymptotic) power in the very sparse regime. We do not know the answer to that question, but are willing to conjecture that there are no such tests.

**The two-sided problem** We focused on the one-sided setting (1), effectively testing  $\rho = 0$  versus  $\rho > 0$ . Knowing the sign of  $\rho$  is not crucial, as one can apply a one-sided test for  $\rho > 0$  to the transformed data  $(X_1, -Y_1), \dots, (X_n, -Y_n)$ . Less trivial is the case where there are three components

$$(X, Y) \sim (1 - \varepsilon)\mathcal{N}(0, \mathbf{I}) + \frac{\varepsilon}{2}\mathcal{N}(0, \Sigma_\rho) + \frac{\varepsilon}{2}\mathcal{N}(0, \Sigma_{-\rho}).$$

We did not look at this model, in part because we wanted to test against a monotonic association (in the contamination component), which is perhaps the most popular alternative in a nonparametric context.

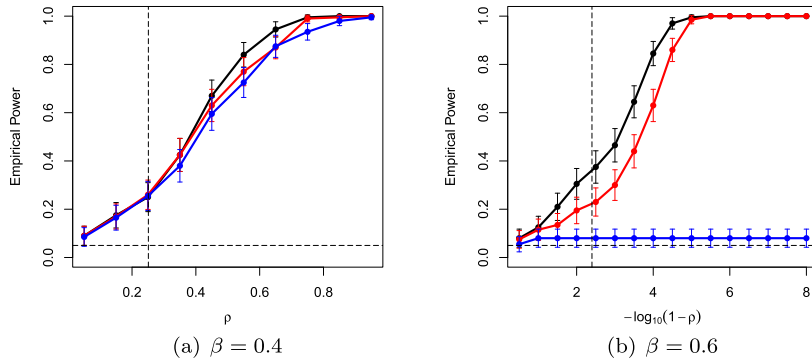


FIG 3. Empirical power comparison with 95% error bars for the full LRT (black), the U-LRT (red) and the V-LRT (blue). (a) Dense regime where  $\beta = 0.4$ . (b) Sparse regime where  $\beta = 0.6$  and  $\rho \rightarrow 1$ . The horizontal line marks the level (set at 0.05) and the vertical line marks the asymptotic detection boundary derived earlier. The sample size is  $n = 10^6$  and the power curves and error bars are based on 200 replications.

## Acknowledgements

We are grateful to two anonymous referees for asking pertinent questions that helped improved the paper, and for asking about a technical point in the proof of Theorem 2 lacking rigor.

## References

- [1] ARIAS-CASTRO, E. and HUANG, R. (2018). The Sparse Variance Contamination Model. *arXiv preprint arXiv:1807.10785*.
- [2] ARIAS-CASTRO, E. and WANG, M. (2016). Distribution-free tests for sparse heterogeneous mixtures. *TEST* **26** 71–94. [MR3613606](#)
- [3] ARRATIA, R., GOLDSTEIN, L. and GORDON, L. (1990). Poisson approximation and the Chen–Stein method. *Statistical Science* 403–424. [MR1092983](#)
- [4] BILGRAU, A. E., ERIKSEN, P. S., RASMUSSEN, J. G., JOHNSEN, H. E., DYBKÆR, K. and BØGSTED, M. (2016). GMCM: Unsupervised clustering and meta-analysis using Gaussian mixture copula models. *Journal of Statistical Software* **70** 1–23.
- [5] BOGOMOLOV, M. and HELLER, R. (2018). Assessing replicability of findings across two studies of multiple features. *Biometrika* **105** 505–516. [MR3842881](#)
- [6] CAI, T. T., JENG, X. J. and JIN, J. (2011). Optimal detection of heterogeneous and heteroscedastic mixtures. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **73** 629–662. [MR2867452](#)
- [7] CAI, T. T. and WU, Y. (2014). Optimal detection of sparse mixtures against a given null distribution. *IEEE Transactions on Information Theory* **60** 2217–2232. [MR3181520](#)

- [8] CHATTERJEE, S. (2007). Stein's method for concentration inequalities. *Probability Theory and Related Fields* **138** 305–321. [MR2288072](#)
- [9] DASGUPTA, A. (2008). *Asymptotic Theory of Statistics and Probability*. Springer. [MR2664452](#)
- [10] DONOHO, D. and JIN, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *The Annals of Statistics* **32** 962–994. [MR2065195](#)
- [11] DONOHO, D. and JIN, J. (2008). Higher criticism thresholding: Optimal feature selection when useful features are rare and weak. *Proceedings of the National Academy of Sciences* **105** 14790–14795. [MR2520682](#)
- [12] DONOHO, D. and JIN, J. (2015). Higher criticism for large-scale inference, especially for rare and weak effects. *Statistical Science* **30** 1–25. [MR3317751](#)
- [13] GIBBONS, J. D. and CHAKRABORTI, S. (2003). *Nonparametric Statistical Inference* **168**, 4th ed. Marcel Dekker, Inc. [MR2064386](#)
- [14] HALL, P. and JIN, J. (2010). Innovated higher criticism for detecting sparse signals in correlated noise. *The Annals of Statistics* **38** 1686–1732. [MR2662357](#)
- [15] Hoeffding, W. (1951). A combinatorial central limit theorem. *The Annals of Mathematical Statistics* **22** 558–566. [MR0044058](#)
- [16] INGSTER, Y. I. (1997). Some problems of hypothesis testing leading to infinitely divisible distributions. *Mathematical Methods of Statistics* **6** 47–69. [MR1456646](#)
- [17] LEHMANN, E. L. and ROMANO, J. P. (2005). *Testing Statistical Hypotheses*, 3rd ed. Springer. [MR2135927](#)
- [18] LI, Q., BROWN, J. B., HUANG, H. and BICKEL, P. J. (2011). Measuring reproducibility of high-throughput experiments. *The Annals of Applied Statistics* **5** 1752–1779. [MR2884921](#)
- [19] MOSCOVICH, A., NADLER, B. and SPIEGELMAN, C. (2016). On the exact Berk–Jones statistics and their p-value calculation. *Electronic Journal of Statistics* **10** 2329–2354. [MR3544289](#)
- [20] REDEKOP, W. K. and MLADSI, D. (2013). The faces of personalized medicine: A framework for understanding its meaning and scope. *Value in Health* **16** S4–S9.
- [21] XU, W., HOU, Y., HUNG, Y. S. and ZOU, Y. (2013). A comparative analysis of Spearman's rho and Kendall's tau in normal and contaminated normal models. *Signal Processing* **93** 261–276.
- [22] ZHAO, S. D. (2015). False discovery rate control for identifying simultaneous signals. *arXiv preprint arXiv:1512.04499*.
- [23] ZHAO, S. D., CAI, T. T. and LI, H. (2017). Optimal detection of weak positive latent dependence between two sequences of multiple tests. *Journal of Multivariate Analysis* **160** 169–184. [MR3688697](#)