

Optimal detection of weak positive dependence between two mixture distributions

Sihai Dave Zhao*

Department of Statistics,
University of Illinois at Urbana-Champaign

and

T. Tony Cai

Department of Statistics, The Wharton School,
University of Pennsylvania

and

Hongzhe Li

Department of Biostatistics and Epidemiology,
University of Pennsylvania Perelman School of Medicine

December 4, 2014

Abstract

This paper studies the problem of detecting dependence between two mixture distributions, motivated by questions arising from statistical genomics. The fundamental limits of detecting weak positive dependence are derived and an oracle test statistic is proposed. It is shown that for mixture distributions whose components are stochastically ordered, the oracle test statistic is asymptotically optimal. Connections are drawn between dependency detection and signal detection, where the goal of the latter is to detect the presence of non-null components in a single mixture distribution. It is shown that the oracle test for dependency can also be used as a signal detection procedure in the two-sample setting, and there can achieve detection even when detection using each sample separately is provably impossible. A nonparametric data-adaptive test statistic is then proposed, and its closed-form asymptotic distribution under the null hypothesis of independence is established. Simulations

*The authors gratefully acknowledge Yun Li and Dr. Yihong Wu for their helpful discussions. This work was supported by NIH grants R01 GM097505 and R01 CA127334. The research of Tony Cai was supported in part by NSF Grant DMS-1208982 and DMS-1403708, and NIH Grant R01 CA127334.

show that the adaptive procedure performs as well as the oracle test statistic, and that both can be more powerful than existing methods. In an application to the analysis of the shared genetic basis of psychiatric disorders, the adaptive test is able to detect genetic relationships not detected by other procedures.

Keywords: Dependency detection; Multiple comparisons; Signal detection

1 Introduction

The problem of detecting dependence between two sequences of test statistics arises frequently in many settings. Three canonical examples from statistical genomics illustrate how dependency detection is applied. First, in gene expression profiling it is often of interest to determine whether a given gene set is enriched with differentially expressed genes (Subramanian et al. 2005). When the gene set is created by selecting differentially expressed genes from another experiment, testing for enrichment is equivalent to testing whether the differential expression test statistics from the two experiments are correlated; see Rhinn et al. (2013) for an example from Alzheimer’s research. Second, in genetics it is often of interest to jointly analyze genome-wide association studies of two different traits to determine whether the traits are genetically related; see Bhattacharjee et al. (2012) and Andreassen et al. (2013) for examples from cancer and psychiatric disorder research. A positive correlation between the SNP-trait test statistics indicates some degree of genetic similarity. Finally, in the integrative analysis of genomic data it is often of interest to determine whether there are any SNPs that are associated with both a gene and a phenotype, in order to better understand the underlying molecular mechanisms (Nicolae et al. 2010, He et al. 2013, Ware et al. 2013). Testing for simultaneous associations can be framed as testing for correlation between the SNP-gene and SNP-phenotype test statistics.

This paper focuses on detecting a particular type of dependence between two random variables U_i and V_i that marginally follow mixture distributions, given observations $(U_i, V_i), i = 1, \dots, n$. Specifically, introduce unobserved indicator variables $X_i \sim Ber(\pi_U)$ and $Y_i \sim Ber(\pi_V)$ such that

$$\begin{aligned} U_i | X_i = 0 &\sim F_0^U, & U_i | X_i = 1 &\sim F_n^U, \\ V_i | Y_i = 0 &\sim F_0^V, & V_i | Y_i = 1 &\sim F_n^V, \end{aligned} \tag{1}$$

where U_i and V_i are independent conditional on X_i and Y_i . The F_0^U, F_n^U, F_0^V , and F_n^V are assumed to be distribution functions of continuous random variables. In other words, the marginal distributions of U_i and V_i are $U_i \sim F_U = (1 - \pi_U)F_0^U + \pi_U F_n^U$ and $V_i \sim F_V = (1 - \pi_V)F_0^V + \pi_V F_n^V$. In the asymptotic analysis in Section 4, the F_n^U and F_n^V will be allowed to depend on n .

Table 1: Bivariate distribution of (X_i, Y_i) under null and alternative hypotheses

	H_0		$H_A : \epsilon \neq \pi_U \pi_V$	
	$Y_i = 0$	$Y_i = 1$	$Y_i = 0$	$Y_i = 1$
$X_i = 0$	$(1 - \pi_U)(1 - \pi_V)$	$(1 - \pi_U)\pi_V$	$1 - \pi_U - \pi_V + \epsilon$	$\pi_V - \epsilon$
$X_i = 1$	$\pi_U(1 - \pi_V)$	$\pi_U \pi_V$	$\pi_U - \epsilon$	ϵ

These mixture models are especially appropriate when U_i and V_i are test statistics, as in the statistical genomics applications. The F_0^U and F_0^V can be thought of as the null distributions of the test statistics, and the F_n^U and F_n^V can be thought of as the alternative distributions. The X_i and Y_i are latent indicators for whether the i th test statistics are null or not. Marginally, proportions $1 - \pi_U$ and $1 - \pi_V$ of the U_i and V_i , respectively, are null and proportions π_U and π_V are non-null. Modeling multiple test statistics using mixture distributions is common in the large-scale hypothesis testing literature (Storey & Tibshirani 2003, Donoho & Jin 2004, Cai et al. 2007, Efron 2010).

This paper studies the problem of detecting dependence between the unobserved X_i and Y_i . Under (1), this is equivalent to detecting dependence between U_i and V_i . Testing for dependence is not a new problem; see Section 2 for a brief summary of relevant previous work. However, it is difficult to find an omnibus test that is optimally powerful against all possible alternatives. On the other hand, the mixture model structure (1) gives rise to a particular dependence alternative. Table 1 shows that in general, the joint distribution of the tuple (U_i, V_i) is $(U_i, V_i) \sim (1 - \pi_U - \pi_V + \epsilon)F_0^U F_0^V + (\pi_V - \epsilon)F_0^U F_n^V + (\pi_U - \epsilon)F_n^U F_0^V + \epsilon F_n^U F_n^V$, where $0 < \epsilon \leq \pi_U \wedge \pi_V$. Detecting dependence amounts to testing whether ϵ equals $\pi_U \pi_V$. This paper is concerned with the alternative of weak positive dependence:

$$H_0 : \epsilon = \pi_U \pi_V \quad \text{vs.} \quad H_A : \epsilon > \pi_U \pi_V, \quad (2)$$

where ϵ is only very slightly larger than $\pi_U \pi_V$. Dependency detection is very challenging in this setting.

Detecting weak positive dependence is important for two reasons. First, it is exactly the alternative of interest in the statistical genomics problems mentioned above. For example, if U_i and V_i are the test statistics from two different experiments, when a gene or a SNP which has a non-null effect in one experiment is more likely to also have a non-null effect in

the other experiment than by random chance alone, the U_i and V_i are positively dependent. Generally it is expected that only a very small proportion of genes or SNPs exhibit non-null effects in both experiments, which makes the dependence weak. The notion of weak dependence corresponds to the sparsity assumption of large-scale hypothesis testing, which assumes that only a small proportion of the tests are non-null. Second, a dependency detection procedure can also be used for signal detection, an important problem which has been the subject of a great deal of research; see Section 2. The signal detection problem is to determine whether $\pi_U > 0$ given only the U_i . In the dependency detection setting, if it is found that U_i and V_i are weakly positively dependent, then π_U must be non-zero.

This paper makes three contributions. First, Section 3 proposes an oracle test for detecting dependence. Simply implementing a likelihood ratio test is impossible because the distribution functions F_0^U , F_n^V , F_0^U , and F_n^V , and the parameters π_U and π_V , are all unknown. Section 4 shows that when the non-null and null distributions are stochastically ordered, the oracle test is asymptotically optimal for distinguishing between the independence null and the weak positive dependence alternative. In addition, Section 5 shows that the oracle test, interpreted as a signal detection procedure, can asymptotically detect non-null components in the U_i and V_i even when signal detection is provably impossible given only U_i or V_i alone. Finally, Section 6 proposes an adaptive version of the oracle test that is entirely data-driven, and is distribution-free in that its asymptotic null distribution that does not depend on any unknown parameters. Furthermore, its asymptotic distribution can be expressed in closed form, so p -values can be computed without resorting to permutation. In the remainder of the paper, Sections 7 and 8 show that in simulated data and in a study of the genetic relatedness of psychiatric disorders, the adaptive test performed as well as the oracle test in practice, and both tests could outperform competing procedures. Section 9 concludes with a discussion of future work. Proofs are found in the Supplementary Material.

2 Previous work

In statistical genomics, a popular method for testing for weak positive dependence is to define thresholds τ_U and τ_V and then to use the hypergeometric distribution to test for

dependence between $I(|U_i| \geq \tau_U)$ and $I(|V_i| \geq \tau_V)$ (Goeman & Bühlmann 2007, Rivals et al. 2007). The motivation is that these indicator functions estimate the latent indicators X_i and Y_i defined in (1). However, it is unclear how best to choose the thresholds. He et al. (2013) developed a Bayesian version of this enrichment approach and avoided choosing thresholds by specifying prior distributions on all unknown parameters. However, their formulation requires knowing the distributions of U_i and V_i , as well as having reasonable priors for π_U and π_V . Another test for dependence is to test the Spearman correlation between U_i and V_i , but this method may not be sensitive enough to detect weak dependence. Recent interest in the statistical literature has centered on detecting arbitrary types of dependence (Székely et al. 2009, Reshef et al. 2011); see in particular Heller et al. (2014). In the actuarial sciences, a great deal of work has gone into methods for detecting positive dependence; see Ledwina & Wylupek (2014) and references therein. In contrast, the focus of the present paper is to detect a specific type of dependence (2). Furthermore, there has been little prior work on characterizing the fundamental limits of dependency detection, or on the asymptotic optimality of existing procedures.

The statistical framework used in this paper to analyze the dependency detection problem is also found in the signal detection literature, where the problem is to test

$$U_i \sim F_0^U \quad \text{vs.} \quad U_i \sim (1 - \pi_U)F_0^U + \pi_U F_n^U$$

given only one sequence $U_i, i = 1, \dots, n$, where F_0^U is known but π_U and F_n^U are unknown. The fundamental limits of detection for this problem, namely, conditions on π_U , F_0^U , and F_n^U that render detection by any test impossible regardless of sample size, have been derived (Ingster 1997, 2002*a,b*, Donoho & Jin 2004, Cai et al. 2011, Cai & Wu 2014). Asymptotically optimal tests, which can detect $\pi_U > 0$ as $n \rightarrow \infty$ for any π_U , F_0^U , and F_n^U above the limits of detection, have also been developed (Ingster 1997, 2002*a,b*, Donoho & Jin 2004, Jager & Wellner 2007). In particular, special attention has been paid to sparse mixtures, where π_U is very close to zero, because standard tests fail in this regime. Notably, the higher criticism test developed by Donoho & Jin (2004) can still detect the presence of non-null components in U_i for both sparse and non-sparse mixtures anywhere above the detection boundary (Cai et al. 2011). Section 4 presents a similar analysis of the dependency test proposed here.

The obvious difference between signal detection and dependency detection is that the latter involves two sequences U_i and V_i , while the former involves only one. These problems are nevertheless closely related. When π_U and π_V are small, as in the sparse mixture setting, $\pi_U\pi_V \approx 0$, so dependency detection (2) amounts to testing $\epsilon = 0$ versus $\epsilon > 0$. Assuming that the magnitudes of U_i and V_i are large when they are non-null, Zhao et al. (2014) defined $T_i = \max(|U_i|, |V_i|)$, which should be large only if both U_i and V_i are non-null. They then modeled the T_i using a mixture distribution and performed signal detection on the T_i . Taking the pairwise maximum is not the only scalar summary of (U_i, V_i) , and Phillips & Ghosh (2013) evaluate three other summary statistics. However, it is unclear how to choose the best summary statistic. The method proposed in this paper deals directly with the bivariate (U_i, V_i) and will be shown to be asymptotically optimal for weak positive dependence regardless of the magnitudes of π_U and π_V .

The work in this paper is most closely related to recent research on correlation detection (Arias-Castro, Bubeck & Lugosi 2012, Arias-Castro, Bubeck, Lugosi et al. 2012). Those authors also studied fundamental detection limits as well as asymptotically optimal tests. However, they were concerned with multivariate Gaussian distributions, while here the focus is on mixture distributions with a particular type of dependence structure.

3 Oracle test statistic

3.1 Test statistic

Testing (2) is equivalent to testing for dependence between U_i and V_i . Therefore define the empirical bivariate survival function

$$\hat{S}_{UV}(u, v) = \frac{1}{n} \sum_{i=1}^n I(U_i \geq u, V_i \geq v),$$

and true marginal survival functions $S_U = 1 - F_U$ and $S_V = 1 - F_V$. The proposed oracle test statistic is

$$\mathcal{D}_n = \sup_{-\infty < u, v < \infty} \frac{n^{1/2}}{(\log n)^{1/2}} \frac{|\hat{S}_{UV}(u, v) - S_U(u)S_V(v)|}{\{S_U(u)S_V(v) - S_U^2(u)S_V^2(v)\}^{1/2}}. \quad (3)$$

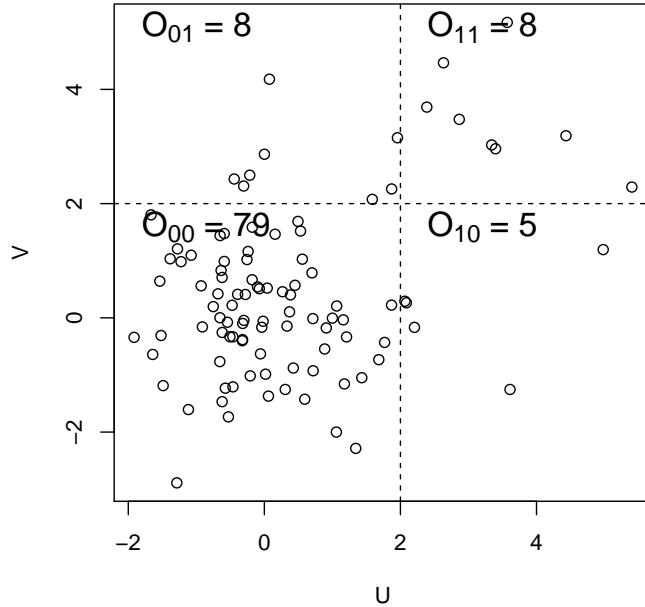


Figure 1: The 2×2 table induced in the observed data by the tuple $(u, v) = (2, 2)$. The number of points in each cell of the table are denoted by O_{jk} , $j = 0, 1$ and $k = 0, 1$.

This is an oracle in the sense that S_U and S_V are usually unknown in practice. Section 6 proposes an adaptive test statistic which estimates these marginal survival functions from the data.

The quantity $\hat{S}_{UV}(u, v) - S_U(u)S_V(v)$ in the numerator of \mathcal{D}_n is natural given the definition of statistical independence, but there is a useful alternative interpretation. Figure 1 is a scatterplot of 100 realizations from the following data-generating mechanism, which follows the form of (1):

$$\begin{aligned}
 U_i \mid X_i = 0 &\sim N(0, 1), & U_i \mid X_i = 1 &\sim N(3, 1), \\
 V_i \mid Y_i = 0 &\sim N(0, 1), & V_i \mid Y_i = 1 &\sim N(3, 1), \\
 (X_i, Y_i) &= \begin{cases} (1, 1) & \text{with probability } 0.1, \\ (1, 0) & \text{with probability } 0.05, \\ (0, 1) & \text{with probability } 0.05, \\ (0, 0) & \text{with probability } 0.8. \end{cases}
 \end{aligned}$$

The figure illustrates that any tuple (u, v) divides the observed data into a 2×2 table. With $O_{jk}(u, v)$, $j = 0, 1$ and $k = 0, 1$ denoting the number of observations in each cell of the table

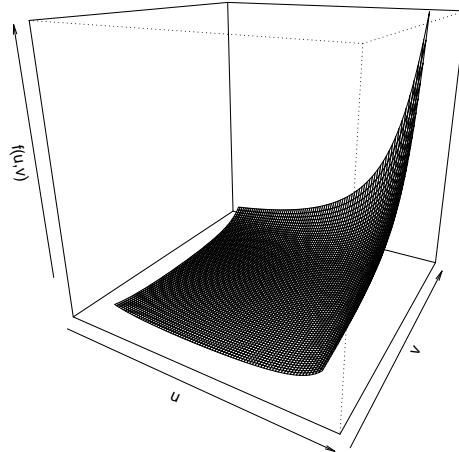


Figure 2: Plot of the weight function $f(u, v) = \{S_U(u)S_V(v) - S_U^2(u)S_V^2(v)\}^{-1/2}$ used in \mathcal{D}_n .

induced by (u, v) , Blum et al. (1961) recognized that the quantity $\hat{S}_{UV}(u, v) - S_U(u)S_V(v)$ is closely related to the quantity $O_{11}(u, v)O_{00}(u, v) - O_{10}(u, v)O_{01}(u, v)$. Thus \mathcal{D}_n is intimately connected to testing for independence using the 2×2 table induced by (u, v) .

It is difficult to know *a priori* which tuple (u, v) will give the most powerful test. For example, the table in Figure 1 induced by $(u, v) = (2, 2)$ suggests that the latent indicators X_i and Y_i are dependent, but the table generated by $(u, v) = (0, 0)$ would look far less convincing. This problem is solved in \mathcal{D}_n by taking the supremum of the test statistic over all possible (u, v) . The best choice of (u, v) depends on the distributions F_0^U , F_0^V , F_n^U , and F_n^V , the proportions π_U , and π_V , and the degree of dependency ϵ . Taking the supremum over (u, v) allows \mathcal{D}_n to adapt to any combination of these unknown parameters.

The denominator of \mathcal{D}_n , which is natural in that it is the variance of \hat{S}_{UV} under the null hypothesis of independence, acts as a weight function that controls the power of the test statistic. Other weight functions are also possible, as discussed in Section 3.2. Figure 2 plots the inverse of the denominator. It implies that when U_i and V_i are dependent, \mathcal{D}_n will be largest when the optimal 2×2 table is induced by large u and v , as this is where the weight function is the largest. This occurs when observations U_i and V_i from the non-null components tend to be larger than the observations from the null components. Thus \mathcal{D}_n therefore has a preference for large U_i and V_i . In Section 4.3 another test statistic is

proposed that eliminates this directional preference.

3.2 Relationship to other work

The proposed oracle \mathcal{D}_n is a goodness-of-fit-type statistic. In this sense it is related to the higher criticism statistic of Donoho & Jin (2004), used in signal detection. The higher criticism statistic for detecting non-null components in the U_i is defined as

$$\sup_{U_{(1)} \leq u < U_{(n)}} n^{1/2} \frac{|\hat{F}_U(u) - F_0^U(u)|}{[F_0^U(u)\{1 - F_0^U(u)\}]^{1/2}}, \quad (4)$$

where F_0^U is the distribution of the non-null component of U_i and $\hat{F}_U(u)$ is the empirical distribution function. Jager & Wellner (2007) developed alternative goodness-of-fit-type tests for the signal detection problem, which could perhaps also lead to alternative tests for dependence.

Many existing tests for dependence are based on the numerator of \mathcal{D}_n , starting with Hoeffding (1948). Blum et al. (1961) recognized the connection to 2×2 tables. The enrichment procedure mentioned in Section 2 is equivalent to choosing $u = \tau_U$ and $v = \tau_V$ for some pre-defined thresholds τ_U and τ_V and then conducting a Fisher exact test using the induced table. Instead of pre-selecting (u, v) , Thas & Ottoy (2004) proposed a statistic that integrates over all (u, v) ; their statistic turns out to be equivalent to summing the Pearson chi-square test statistics calculated from each 2×2 table induced by the observed tuples (U_i, V_i) . Heller et al. (2014) proposed using the maximum of all Pearson test statistics, instead of the sum. Thas & Ottoy (2004) and Heller et al. (2014) also considered using $k \times k$ tables for $k > 2$, and Heller et al. (2014) showed that these tests are consistent and can have greater power. The \mathcal{D}_n proposed in (3) is closely related to but different from all of these statistics.

The denominator of \mathcal{D}_n is only one possible weight function. Without a weight \mathcal{D}_n is similar to a Kolmogorov-Smirnov-type test statistic, which in fact was studied by Scaillet (2005). The test statistic of Thas & Ottoy (2004) uses the weight function

$$[S_U(u)\{1 - S_U(u)\}S_V(v)\{1 - S_V(v)\}]^{-1/2},$$

which gives equal preference to large and small u and v . It would be possible to replace the weight function in \mathcal{D}_n with this weight. However, characterizing the asymptotic properties

of the resulting statistic requires a law of the iterated logarithm-type result, and calculating p -values in finite samples requires the asymptotic distribution of a data-adaptive version. These are difficult to derive for the Thas & Ottoy (2004) weight function. In contrast, both of these results are available for \mathcal{D}_n ; see Lemma 1 and Theorem 6.

4 Asymptotic theory

4.1 Asymptotic testing framework

To evaluate the asymptotic properties of the weak positive dependency detection problem (2) and the test statistic \mathcal{D}_n (3), calibrate

$$\begin{aligned} \pi_U &= n^{-\beta_U}, & 0 \leq \beta_U \leq 1, \\ \pi_V &= n^{-\beta_V}, & 0 \leq \beta_V \leq 1, \\ \epsilon &= \pi_U \pi_V + n^{-\beta}, & 1/2 < \beta < 1, (\beta_U \vee \beta_V) \leq \beta. \end{aligned} \tag{5}$$

The calibrations of π_U and π_V imply that there are no restrictions on the proportions of non-null signals in U_i and V_i , though of course when either proportion equals zero or one the U_i and V_i are independent. In the signal detection setting, which deals only with a single sequence U_i , the settings $0 < \beta_U < 1/2$ and $1/2 < \beta_U$ are termed the “moderately sparse” and “very sparse” regimes, respectively, and are frequently treated separately in theoretical analyses (Cai et al. 2011). Zhao et al. (2014) considered simultaneous signal detection with two sequences U_i and V_i , but their theoretical results depend heavily on π_U and π_V being in the very sparse regime.

The calibration of ϵ makes specific the notion of weak positive dependence: ϵ is always larger than $\pi_U \pi_V$, so under the alternative there are more pairs of non-null signals than would be expected by chance. However, ϵ only differs from $\pi_U \pi_V$ by a small amount $n^{-\beta}$, where $\beta > 1/2$. This weak dependence regime is the analog of the very sparse regime of signal detection. In addition, since Table 1 makes clear that $\epsilon \leq (\pi_U \wedge \pi_V)$, it must be that β is at least $\beta_U \vee \beta_V$. The behavior of the proposed test under negative dependence, for example where $\epsilon = \pi_U \pi_V - n^{-\beta}$, and under stronger dependence, for example where $0 < \beta < 1/2$, is left for future work.

The performance of the proposed oracle \mathcal{D}_n will be studied using the asymptotic testing framework. For any test $\phi(U_1, \dots, U_n, V_1, \dots, V_n) : \mathbb{R}^n \times \mathbb{R}^n \rightarrow [0, 1]$ of H_0 vs. H_A , define $S_{H_0, H_A}(\phi) = \mathbb{E}_{H_0} \phi + \mathbb{E}_{H_A}(1 - \phi)$, the sum of the type I and II errors. For certain values of the parameters β_U , β_V , and β and certain distribution functions F_0^U , F_0^V , F_n^U , and F_n^V , $S_{H_0, H_A} \rightarrow 1$ as $n \rightarrow \infty$ for all tests ϕ ; this is called the undetectable region and characterizes the fundamental limit of dependency detection. For a particular test, there may be some parameter values and distribution functions for which $S_{H_0, H_A} \rightarrow 0$; this is referred to as the detectable region of that test. Outside of the undetectable region, the likelihood ratio test minimizes S_{H_0, H_A} over all tests, but cannot be implemented in practice because the null and alternative hypotheses contain many unknown parameters. It will be shown below that when the components of the mixture distributions of U_i and V_i are stochastically ordered, the oracle \mathcal{D}_n , with a suitable critical value, has $S_{H_0, H_A} \rightarrow 0$ everywhere outside of the undetectable region, making it asymptotically optimal.

4.2 Assumptions

The results derived below in Section 4.3 hold for distribution functions satisfying two assumptions.

Assumption 1 *The distributions F_0^U , F_0^V , F_n^U , and F_n^V are absolutely continuous with respect to the Lebesgue measure and have continuous densities f_0^U , f_0^V , f_n^U , and f_n^V .*

Assumption 2 *The log-likelihood ratios $\ell_n^U = \log(f_n^U / f_0^U)$ and $\ell_n^V = \log(f_n^V / f_0^V)$ satisfy*

$$\begin{aligned} \lim_{n \rightarrow \infty} \sup_{a \geq \log_n 2} \left| \frac{\ell_n^U \{(F_0^U)^{-1}(n^{-a})\}}{\log n} - \alpha_U^-(a) \right| &= 0, \\ \lim_{n \rightarrow \infty} \sup_{a \geq \log_n 2} \left| \frac{\ell_n^U \{(F_0^U)^{-1}(1 - n^{-a})\}}{\log n} - \alpha_U^+(a) \right| &= 0, \\ \lim_{n \rightarrow \infty} \sup_{b \geq \log_n 2} \left| \frac{\ell_n^V \{(F_0^V)^{-1}(n^{-b})\}}{\log n} - \alpha_V^-(b) \right| &= 0, \\ \lim_{n \rightarrow \infty} \sup_{b \geq \log_n 2} \left| \frac{\ell_n^V \{(F_0^V)^{-1}(1 - n^{-b})\}}{\log n} - \alpha_V^+(b) \right| &= 0, \end{aligned}$$

for measurable functions $\alpha_U^-, \alpha_U^+, \alpha_V^-, \alpha_V^+ : \mathbb{R} \rightarrow \mathbb{R}$.

Assumption 1 implies that the null and alternative distribution functions are continuously differentiable almost everywhere, and is satisfied by most commonly-used test statistics. Assumption 2 states that the asymptotic behaviors of the likelihood ratios can be uniformly approximated by polynomial functions of n . In particular, $n^{\alpha_{\bar{U}}}$ and $n^{\alpha_{\bar{V}}}$ describe their behaviors at values smaller than the median of the null distribution and $n^{\alpha_{\bar{U}}^+}$ and $n^{\alpha_{\bar{V}}^+}$ describe them at values larger than the median. The likelihood ratios are characterized separately on the left and right sides of the null in order to make the asymptotic properties of the oracle test more clear, because as described in Section 3, \mathcal{D}_n prefers the right side over the left; see Theorems 2 and 3. For this reason the α functions only approximate the log-likelihood ratios for a and b above $\log_n 2$, because $n^{-\log_n 2} = 1 - n^{-\log_n 2} = 0.5$; otherwise the α^- and α^+ functions would simply be reparametrizations of each other. Intuitively the α functions describe where observations from F_n^U and F_n^V are most likely to lie, and therefore where to look for dependent pairs, when both U_i and V_i come from the non-null components.

To illustrate Assumption 2, suppose the null component F_0^U of the U_i is $N(0, 1)$. Since $\Phi(-x) \approx -\phi(x)/x$ for large x , $\Phi\{-(2a \log n)^{1/2}\} \approx n^{-a}$ as long as $2a \log n$ is sufficiently large, which is guaranteed by the condition $a \geq \log_n 2$. Therefore the (n^{-a}) th quantile of F_0^U is $-(2a \log n)^{-1/2}$, and by similar reasoning the $(1 - n^{-a})$ th quantile is $(2a \log n)^{1/2}$. Now suppose $F_n^U \sim N\{(2r_\mu \log n)^{1/2}, \sigma^2\}$. Then the likelihood ratios obey

$$\frac{f_n^U\{(F_0^U)^{-1}(n^{-a})\}}{f_0^U\{(F_0^U)^{-1}(n^{-a})\}} = n^{a-\sigma^{-2}(-a^{1/2}-r_\mu^{1/2})^2}, \quad \frac{f_n^U\{(F_0^U)^{-1}(1-n^{-a})\}}{f_0^U\{(F_0^U)^{-1}(1-n^{-a})\}} = n^{a-\sigma^{-2}(a^{1/2}-r_\mu^{1/2})^2},$$

so $\alpha_{\bar{U}}^-(a) = a - \sigma^{-2}(a^{1/2} + r_\mu^{1/2})^2$ and $\alpha_{\bar{U}}^+(a) = a - \sigma^{-2}(a^{1/2} - r_\mu^{1/2})^2$. In this case $\alpha_{\bar{U}}^-(a) \leq 0$ while $\alpha_{\bar{U}}^+(a) \geq 0$ for all a , which implies that looking at values in the right-hand tail of F_0^U will be most useful for detecting dependency.

4.3 Optimality

The undetectable region of the dependency detection problem, where no test is capable of differentiating H_0 and H_A , is characterized by the following theorem. The characterization involves the essential supremum, which for a measurable function f and a measure μ is

defined as

$$\operatorname{ess\,sup}_x f(x) = \inf[a \in \mathbb{R} : \mu\{f(x) > a\} = 0].$$

Intuitively, the essential supremum is the supremum with probability one. In this paper the essential suprema are taken with respect to the Lebesgue measure.

Theorem 1 (Undetectable region) *Assume $F_0^U \neq F_n^U$, $F_0^V \neq F_n^V$, and Assumptions 1 and 2. Under weak positive dependence, $S_{H_0, H_A}(\phi) \rightarrow 1$ for all tests ϕ if each of the following holds:*

$$1 - 2\beta + 0 \vee \operatorname{ess\,sup}_{a>0} \{\alpha_U + (\alpha_U \wedge \beta_U) - a\} < 0, \quad (6)$$

$$1 - 2\beta + 0 \vee \operatorname{ess\,sup}_{b>0} \{\alpha_V + (\alpha_V \wedge \beta_V) - b\} < 0, \quad \text{and} \quad (7)$$

$$1 + 0 \vee \operatorname{ess\,sup}_{a,b>0} [(-\beta + \alpha_U + \alpha_V) \wedge \{-2\beta + \alpha_U + \alpha_V + (\alpha_U \wedge \beta_U) + (\alpha_V \wedge \beta_V)\} - a - b] < 0, \quad (8)$$

where $\alpha_U(a) = \alpha_U^-(a) \vee \alpha_U^+(a)$ and $\alpha_V(a) = \alpha_V^-(a) \vee \alpha_V^+(a)$.

Outside of the undetectable region, an asymptotic test based on \mathcal{D}_n (3) can be applied. A suitable critical value for \mathcal{D}_n is provided by Lemma 1, due to Einmahl & Mason (1985), who showed that $(\log \log n)^{-1} \log \mathcal{D}_n \rightarrow 1$ almost surely. This motivates the asymptotic test

$$\text{reject } H_0 \text{ when } (\log n)^{1/2} < \mathcal{D}_n. \quad (9)$$

The detectable region of (9) is characterized by the following theorem.

Theorem 2 (Detectable region) *Assume $F_0^U \neq F_n^U$, $F_0^V \neq F_n^V$, and Assumptions 1 and 2. Define*

$$p_U^-(x) = \sup_{a \geq x} \{\alpha_U^-(a) - a\}, \quad p_U^+(x) = \sup_{a \geq x} \{\alpha_U^+(a) - a\},$$

$$p_V^-(y) = \sup_{b \geq y} \{\alpha_V^-(b) - b\}, \quad p_V^+(y) = \sup_{b \geq y} \{\alpha_V^+(b) - b\}.$$

For the proposed test (9), $S_{H_0, H_A} \rightarrow 0$ if one of the following is true:

$$\sup_{\substack{x, y > 0, \\ x + y < 1}} \left\{ \frac{1}{2} - \beta + (-x) \vee p_U^+ + (-y) \vee p_V^+ + \frac{x \wedge (\beta_U - p_U^+)}{2} + \frac{y \wedge (\beta_V - p_V^+)}{2} \right\} > 0, \text{ or} \quad (10)$$

$$\sup_{\substack{x, y > 0, \\ x + y < 1}} \left\{ \frac{1}{2} - \beta + (-x) \vee p_U^- + (-y) \vee p_V^+ + \frac{y \wedge (\beta_V - p_V^+)}{2} \right\} > 0, \text{ or} \quad (11)$$

$$\sup_{\substack{x, y > 0, \\ x + y < 1}} \left\{ \frac{1}{2} - \beta + (-x) \vee p_U^+ + (-y) \vee p_V^- + \frac{x \wedge (\beta_U - p_U^+)}{2} \right\} > 0, \text{ or} \quad (12)$$

$$\sup_{\substack{x, y > 0, \\ x + y < 1}} \left\{ \frac{1}{2} - \beta + (-x) \vee p_U^- + (-y) \vee p_V^- + \frac{x \wedge \beta_U \wedge y \wedge \beta_V}{2} \right\} > 0. \quad (13)$$

The interpretations of the functions $p_U^-(x), p_U^+(x)$ and $p_V^-(y), p_V^+(y)$ come from Lemma 5, where it is shown that they characterize the asymptotic behaviors of the non-null distributions F_n^U and F_n^V evaluated at quantiles of the null distributions F_0^U and F_0^V . For example, $F_n^U \{(F_0^U)^{-1}(n^{-x})\}$ for $x \geq \log_n 2$ behaves like $n^{p_U^-(x)}$.

Theorem 3 below shows that if the non-null components of U_i and V_i are stochastically larger than the null components, the detectable region from Theorem 2 is the interior of the complement of the undetectable region from Theorem 1. The boundary that separates these two regions is called the detection boundary, and Theorem 3 shows that the proposed oracle test (9) attains the detection boundary, making it asymptotically optimal. The preference of the test for stochastically larger non-null components is a result of the asymmetry of the weight function of \mathcal{D}_n , as discussed in Section 3.1 and illustrated in Figure 2.

Theorem 3 (Optimality) *Under weak positive dependence, if $F_n^U(u) \leq F_0^U(u)$ for all u and $F_n^V(v) \leq F_0^V(v)$ for all v , test (9) is asymptotically optimal.*

The stochastic ordering condition on U_i and V_i is stronger than what is actually needed for optimality, but is sensible when U_i and V_i are test statistics because the directions of the stochastic order of their non-null and null components are often known *a priori*. An alternative asymptotic test can be used when it is known only that the non-null and null components are stochastically ordered, but the directions are unknown. Let $\mathcal{D}_n^i, i = 1, \dots, 4$ denote \mathcal{D}_n calculated for $(U_i, V_i), (U_i, -V_i), (-U_i, V_i),$ and $(-U_i, -V_i),$ respectively, and

define the asymptotic test

$$\text{reject } H_0 \text{ when } (\log n)^{1/2} < \mathcal{D}_n^* = \max_{i=1,\dots,4} \mathcal{D}_n^i. \quad (14)$$

One of these four transformations, say i^* , will be such that the non-null components are stochastically larger than the null components, and \mathcal{D}_n^* will be no smaller than $\mathcal{D}_n^{i^*}$. It is also easy to show that \mathcal{D}_n^* is asymptotically almost always less than $(\log n)^{1/2}$, leading to the following theorem.

Theorem 4 *Under weak positive dependence, if F_n^U and F_0^U are stochastically ordered, and F_n^V and F_0^V are also stochastically ordered, test (14) is asymptotically optimal.*

To be concrete, some quantities used in Theorems 1 and 2 are calculated here for two sequences of normal mixtures when $\beta_U, \beta_V > 1/2$:

$$\begin{aligned} U_i &\sim (1 - n^{-\beta_U})N(0, 1) + n^{-\beta_U} N\{(2r_\mu \log n)^{1/2}, 1\}, \\ V_i &\sim (1 - n^{-\beta_V})N(0, 1) + n^{-\beta_V} N\{(2r_\nu \log n)^{1/2}, 1\}, \end{aligned} \quad (15)$$

for $r_\mu, r_\nu > 0$. Then the functions α_U and α_V in Theorem 1 are

$$\alpha_U(a) = a - (a^{1/2} - r_\mu^{1/2})^2, \quad \alpha_V(b) = b - (b^{1/2} - r_\nu^{1/2})^2,$$

as derived in Section 4.2, and the functions p_U and p_V from Theorem 2 are therefore

$$p_U(x) = -\{(x^{1/2} - r_\mu^{1/2})_+\}^2, \quad p_V(y) = -\{(y^{1/2} - r_\nu^{1/2})_+\}^2.$$

In practice things can be much more complicated than example (15). For instance, when U_i and V_i are test statistics, it is usually untenable to assume that the non-null effects in one sequence are all identically distributed. It is more likely that each non-null effect has a different mean μ_i or ν_i . On the other hand, taking an empirical Bayes interpretation, these means can be considered as random draws from some hyper-distributions, making F_n^U and F_n^V convolutions of a normal density and the hyper-distributions. Another complication is that when the U_i and V_i are one-tailed test statistics, the μ_i and/or the ν_i can have different signs for different i . The hyper-distributions will then not satisfy the stochastic ordering conditions of Theorems 3 and 4. On the other hand, the U_i and V_i can be converted to two-tailed test statistics by taking the absolute value or the square, which will satisfy

the stochastic ordering. Finally, the null and non-null distributions of U_i and V_i can be arbitrarily complicated in practice. One such situation will be simulated in Section 7.3. This will pose no problem for the adaptive test discussed in Section 6 because there the null and non-null distributions are empirically estimated from the data.

4.4 Implications

There are a number of interesting implications of Theorems 1 and 2 that will be introduced here and then studied in simulations in Section 7. Most obviously, dependency detection is easier for smaller β , corresponding to stronger dependence. It is also easier for larger α_U , α_V , p_U , and p_V , corresponding to larger differences between the null and alternative distributions. For example, under the normal mixture model (15) with r_μ and r_ν big enough, the inequality (10) that characterizes part of the detectable region becomes

$$\sup_{\substack{x,y \in (0,1), \\ x+y < 1}} \left[\frac{1}{2} - \beta - \{(x^{1/2} - r_\mu^{1/2})_+\}^2 - \{(y^{1/2} - r_\nu^{1/2})_+\}^2 + \frac{x+y}{2} \right] = 1 - \beta > 0.$$

In this case dependency is always asymptotically detectable, since $\beta < 1$ from (5).

Next, for fixed β and α_U and α_V , dependency detection is easier for large β_U and β_V , corresponding to sparser individual sequences U_i and V_i , with the caveat that $(\beta_U \vee \beta_V) \leq \beta$ from (5). For example, for large enough β_U and β_V , inequality (10) takes the same form as above even if r_μ and r_ν are not that large. The reason is that in dependency detection, when there are very few non-null signals in the individual sequences, the presence of even a single i for which U_i and V_i are both non-null provides significant evidence for dependence. This is true even in the noiseless case when the latent indicators X_i and Y_i from (1) are directly observed.

Finally, the oracle test statistic can detect weak positive dependency even if in one of the sequences, say in the U_i , the non-null and null components are very similar. Detection is possible as long as the non-null and null distributions in the other sequence are stochastically ordered and sufficiently separated. This is reflected in the inequalities (6) and (7) that characterize part of the undetectable region. For example, suppose in the normal mixture model (15) that r_μ is very close to zero. This makes F_0^U and F_n^U extremely similar, and it would seem that detecting dependency would be difficult. However, now suppose that

$r_\nu = 1$. Then when (7) is false,

$$0 < 1 - 2\beta + 0 \vee \sup_{b>0} \{ -(b^{1/2} - 1)^2 + \beta_V \} \wedge (-b + 4b^{1/2} - 2) = 1 - 2\beta + \beta_V.$$

The equality follows because the function $-(b^{1/2} - 1)^2 + \beta_V$ has a maximum value of β_V at $b = 1$, where the function $-b + 4b^{1/2} - 2$ equals one. Since $\beta_V < 1$ from (5), the quantity inside the supremum above can never be larger than β_V . Therefore dependency is detectable when $(1 + \beta_V)/2 > \beta$, even if r_μ is close to zero.

This last phenomenon may be unexpected. Intuitively, it occurs because if the non-null V_i are strong enough to be easily identified, dependency can be detected simply by checking whether the U_i paired with those V_i are also non-null. Consider the case under the normal mixture model where only V_1 is non-null, in which case if U_1 is also non-null it is highly likely that the sequences U_i and V_i are dependent. If ν_n is extremely large, it is easy to identify $i = 1$ as the non-null signal in the V_i . Then testing for dependency amounts to testing whether U_1 has mean zero. If U_1 is truly non-null then it has distribution $N\{(2r_\mu \log n)^{1/2}, 1\}$, and the test will easily reject for large n .

5 Signal detection via dependency detection

The properties of the proposed oracle test (9) are especially interesting when dependency detection is used as signal detection, as briefly discussed in Sections 1 and 2. Specifically, the standard signal detection problem tests whether one sequence, say U_i , contains observations from two different distributions. If the U_i are found to be positively dependent with another sequence V_i , both U_i and V_i must contain null and non-null components. Therefore (9) can be used as a signal detection procedure for U_i and V_i simultaneously.

This approach to signal detection has some advantages. First, as discussed in Section 4.4, for a fixed level of positive dependence, in other words for a fixed β , dependency detection gets easier as the mixtures U_i and V_i get more sparse. This is in direct contrast to the signal detection problem, which is more difficult when the mixture is sparse. When the mixture is too sparse and the signal is weak, it can be proved that no test can detect the presence of the non-null component given only one sequence. Theorems 1 and 2 show, however, that when another sequence is available, the non-null component of even a

weak signal can be detected by dependency detection if the two sequences are sufficiently positively dependent.

This is related to the second advantage of dependency detection: it can detect non-null components that can never be detected given only a single sequence, even in nearly dense settings. A concrete example is

$$\begin{aligned} U_i &\sim (1 - n^{-\beta_U})N(0, 1) + n^{-\beta_U} N[\{(2\beta_U - 1) \log n\}^{1/2}, 1], \\ V_i &\sim (1 - n^{-\beta_V})N(0, 1) + n^{-\beta_V} N\{(2 \log n)^{1/2}, 1\}, \\ 1/2 &< \beta_U \wedge \beta_V, (1 + \beta_V)/2 > \beta. \end{aligned} \tag{16}$$

The results of Ingster (1997) and Donoho & Jin (2004) imply that the non-null component of U_i is not detectable using the sequence U_i alone. However, as discussed in Section 4.4, the proposed test can detect dependency between U_i and V_i , and therefore can detect the presence of non-null signals in U_i . This is demonstrated in simulated data in Section 7.4.

The advantages of dependency detection arise from the pairing information available from the bivariate observations (U_i, V_i) . In order to improve upon standard signal detection methods for U_i , a sequence V_i must be found such that a particular V_i is likely to be non-null when the U_i that it is paired with is non-null. The locations of the non-null i do not need to be known. On the other hand, if the non-null V_i are easily identified, the problem of checking whether the correspondingly U_i are also non-null is made significantly easier.

6 Data-adaptive test statistic

6.1 Test statistic

The oracle test statistic \mathcal{D}_n defined in (3) cannot be calculated when the true marginal survival functions S_U and S_V are unknown. On the other hand, they can be estimated from the data using the empirical survival functions

$$\hat{S}_U(u) = \frac{1}{n} \sum_{i=1}^n I(U_i \geq u), \quad \hat{S}_V(v) = \frac{1}{n} \sum_{i=1}^n I(V_i \geq v).$$

The proposed data-adaptive test statistic is

$$\hat{\mathcal{D}}_n = \sup_{\substack{U_{(1)} < u \leq U_{(n)}, \\ V_{(1)} < v \leq V_{(n)}}} \frac{n^{1/2}}{(\log n)^{1/2}} \frac{|\hat{S}_{UV}(u, v) - \hat{S}_U(u)\hat{S}_V(v)|}{\{\hat{S}_U(u)\hat{S}_V(v) - \hat{S}_U^2(u)\hat{S}_V^2(v)\}^{1/2}}. \tag{17}$$

This adaptive test statistic can be used for dependency detection in practice. When used for signal detection, as in Section 5, has an additional advantage over other detection procedures: it does not require knowledge of the null components F_0^U and F_0^V or the non-null components F_n^U and F_n^V . In contrast, the higher criticism statistic (4) cannot be calculated if F_0^U is unknown. Arias-Castro & Wang (2013) proposed nonparametric signal detection procedures but needed to assume that the null component is symmetric about zero while the non-null component has positive median.

6.2 Implementation

A simple algorithm for calculating $\hat{\mathcal{D}}_n$ requires only $O(n^2)$ operations: the U_i and V_i are first sorted using quicksort, which on average requires $O(n \log n)$ operations and at most requires $O(n^2)$. Let $U_{(i)}$ and $V_{(i)}$ denote the order statistics of U_i and V_i , respectively. Next, the algorithm iterates from the largest to the smallest $U_{(i)}$, where for each i it iterates from the largest to the smallest $V_{(j)}$ in order to calculate

$$D_{ij} = \frac{n^{1/2}}{(\log n)^{1/2}} \frac{|\hat{S}_{UV}(U_{(i)}, V_{(j)}) - \hat{S}_U(U_{(i)})\hat{S}_V(V_{(j)})|}{\{\hat{S}_U(U_{(i)})\hat{S}_V(V_{(j)}) - \hat{S}_U(U_{(i)})^2\hat{S}_V(V_{(j)})^2\}^{1/2}}$$

for all $i, j = 1, \dots, n$. Finally, $\hat{\mathcal{D}}_n = \max_{ij} D_{ij}$. Heller et al. (2014) proposed a similar algorithm with the same computational complexity.

When the non-null components are known to be stochastically larger than the null components, the algorithm only needs to iterate over $U_{(n-n_1)}, \dots, U_{(n)}$ and $V_{(n-n_2)}, \dots, V_{(n)}$, where n_1 and n_2 can be close to n , because D_{ij} is likely to achieve its supremum at large i and j . Even if the true maximum D_{ij} is not attained in $i \geq n - n_1, j \geq n - n_2$, the largest D_{ij} over this restricted area may still be large enough to reject the null hypothesis. In other words, this truncated iteration will provide a conservative p -value, which can still be powerful. This algorithm has been implemented as a C function called by R. As an example, calculating $\hat{\mathcal{D}}_n$ with $n = 10^7$ and $n_1 = n_2 = 10^4$ took 29 seconds on a laptop with a 2.5 GHz Intel Core i5 processor with 8 GB RAM.

6.3 Finite-sample p -values

Einmahl (1996) derived a closed-form expression for the asymptotic null distribution of the

oracle test statistic \mathcal{D}_n (3), which is reproduced in Lemma 2. The following theorem shows that under certain conditions on the true bivariate distribution of (U_i, V_i) , \mathcal{D}_n is very close to $\widehat{\mathcal{D}}_n$.

Theorem 5 *Define $U'_i = -U_i$ and $V'_i = -V_i$, let $\widehat{F}_{U'V'}$ be the corresponding bivariate empirical distribution function, and let $F_{U'}$ and $F_{V'}$ be the true univariate distribution functions. Under Assumption 1, for $\widehat{\mathcal{D}}_n$ defined in (17),*

$$\widehat{\mathcal{D}}_n = \mathcal{D}_n + o_P(1),$$

Combining Theorem 5 and Lemma 2 leads to the following result, which is used to obtain p -values in the simulations and data analysis in Sections 7 and 8.

Theorem 6 *Under Assumption 1, under the null hypothesis where U_i and V_i are independent,*

$$P_{H_0}(\widehat{\mathcal{D}}_n > x) \longrightarrow 1 - \exp(-x^2),$$

with $\widehat{\mathcal{D}}_n$ defined in (17).

Theorem 6 can be used when the stochastic order of the non-null and null components of U_i and V_i are known *a priori*. For example, if U_i and V_i are two-tailed test statistics, their non-null components will be stochastically larger than their nulls. When U_i and V_i are p -values, their non-null components will be stochastically smaller. By multiplying U_i or V_i by negative one if necessary, the test statistics can be transformed so that the conditions of Theorem 3 hold. This is the approach taken in the simulations and data analysis in Sections 7 and 8.

When the directions of the stochastic ordering are unknown, the appropriate oracle test statistic is \mathcal{D}_n^* , defined in (14). A data-adaptive version of \mathcal{D}_n^* can also be defined. However, \mathcal{D}_n^* is the maximum of four highly correlated random variables, so its adaptive version has a complicated null distribution. In theory its null could be simulated by fixing the indices of the U_i and permuting those of the V_i , but this is computationally inconvenient when n is large. The most convenient option is to use a Bonferroni adjustment, which would give a conservative but valid p -value.

7 Simulations

7.1 Simulation settings and competing methods

Simulations were conducted to numerically explore the performance of dependency detection based on the oracle and adaptive test statistics proposed in (3) and (17). These simulation settings were designed to illustrate the theoretical properties discussed in Sections 4.4 and 5, and to mimic situations that might arise in statistical genomics applications. Three sets of simulations were conducted, described in Sections 7.2–7.4. Each had $n = 100,000$ and was repeated 200 times.

Under the null hypothesis, the latent indicators X_i and Y_i were independently sampled from Bernoulli random variables with parameters $n^{-\beta_U}$ and $n^{-\beta_V}$, respectively. Under the alternative hypothesis, bivariate (X_i, Y_i) were sampled according to Table 1. The values of β_U and β_V varied across the three sets of simulations. The positions of the null and non-null U_i and V_i were not changed throughout the replications, but in each simulation new observations were drawn from the appropriate distributions. This was done because in real applications, for example in statistical genomics, whether or a not a gene or a SNP exhibits a non-null effect remains the same across repeated samples. It is only the value of the test statistic that is random, which is the case in these simulations. The U_i and V_i were simulated to mimic two-tailed test statistics.

To implement the proposed tests, it was assumed that the stochastic ordering of the non-null and null components of the U_i and V_i was known *a priori*, as discussed in Section 6.3. For computational convenience the supremum in the adaptive $\hat{\mathcal{D}}_n$ was taken only over the 5,000 largest U_i and V_i , as discussed in Section 6.2, and the same was done for the oracle \mathcal{D}_n . Lemma 2 and Theorem 6 were used to calculate p -values.

The proposed tests were compared to two competing methods. The first is Spearman’s correlation, which is robust to the distribution of the U_i and V_i . The second is the method of Zhao et al. (2014) mentioned in Section 2. They first calculated $T_i = |U_i| \wedge |V_i|$, and their test statistic is $M_n = \max_i T_i$, so that M_n will be large only if there exist i such that both U_i and V_i are simultaneously non-null. This will be referred to as the “max test”. The null distribution of M_n can be obtained by permuting the indices of only the U_i or the

Table 2: Type I errors at nominal $\alpha = 0.05$ for simulation setting I; see Section 7.2

Test	β_U, β_V					
	0.3,0.3	0.3,0.5	0.3,0.7	0.5,0.5	0.5,0.7	0.7,0.7
Oracle \mathcal{D}_n	0.05	0.06	0.09	0.02	0.02	0.04
Adaptive $\hat{\mathcal{D}}_n$	0.03	0.04	0.04	0.01	0.02	0.01
Spearman	0.04	0.04	0.04	0.06	0.04	0.03
Max	0.04	0.07	0.09	0.01	0.04	0.04

V_i , which induces independence. Zhao et al. (2014) gave a simple analytic expression based on the hypergeometric distribution to obtain this p -value without actually implementing the permutation: the p -value is defined to be the proportion of permutations in which the calculated M_n exceeds the observed M_n . This is equal to the probability that at least one of the U_i with magnitude at least M_n is permuted such that it is paired with one of the V_i with magnitude at least M_n . Thus if there are k indices such that $|U_i| \geq M_n$ and m indices such that $|V_i| \geq M_n$, the permutation p -value of the max test is

$$1 - \binom{m}{0} \binom{n-m}{k} \binom{n}{k}^{-1}.$$

Others possible competing methods were not implemented. The enrichment procedure described in Section 2 was not implemented because requires pre-selected thresholds τ_U and τ_V . The Bayesian procedure of He et al. (2013) is too computationally inconvenient for large-scale simulations. The tests of Heller et al. (2014) are perhaps the most natural competitors to the proposed procedure, but their implementation cannot handle an n as large as the one considered here.

7.2 Setting I: sequence sparsity and dependence strength

For i such that $X_i = 0$, U_i was drawn from $|N(0,1)|$, and otherwise was drawn from $|N(3,1)|$. Similarly, $V_i \sim |N(0,1)|$ when $Y_i = 0$ and $V_i \sim |N(3,1)|$ when $Y_i = 1$. For comparison, a Z -statistic equal to three corresponds to a two-tailed p -value of 0.0027. The sparsity parameters β_u and β_V equaled either 0.3, 0.5, or 0.7. The dependency parameter β equaled either 0.305, 0.505, or 0.705.

Table 3: Powers at nominal $\alpha = 0.05$ for simulation setting I; see Section 7.2

β_U, β_V	Test	β		
		0.305	0.505	0.705
0.3,0.3	Oracle \mathcal{D}_n	1.00	0.89	0.03
	Adaptive $\widehat{\mathcal{D}}_n$	1.00	0.88	0.01
	Spearman	1.00	0.73	0.06
	Max	0.91	0.21	0.06
0.3,0.5	Oracle \mathcal{D}_n		1.00	0.12
	Adaptive $\widehat{\mathcal{D}}_n$		1.00	0.09
	Spearman		0.74	0.06
	Max		0.88	0.19
0.3,0.7	Oracle \mathcal{D}_n			0.47
	Adaptive $\widehat{\mathcal{D}}_n$			0.36
	Spearman			0.06
	Max			0.49
0.5,0.5	Oracle \mathcal{D}_n		1.00	0.64
	Adaptive $\widehat{\mathcal{D}}_n$		1.00	0.61
	Spearman		0.67	0.07
	Max		1.00	0.78
0.5,0.7	Oracle \mathcal{D}_n			0.94
	Adaptive $\widehat{\mathcal{D}}_n$			0.94
	Spearman			0.07
	Max			0.93
0.7,0.7	Oracle \mathcal{D}_n			0.99
	Adaptive $\widehat{\mathcal{D}}_n$			0.99
	Spearman			0.06
	Max			0.97

Table 2 reports the type I errors at a nominal $\alpha = 0.05$ achieved by the competing methods for different combinations of β_U and β_V . The method proposed in this paper and Spearman’s correlation controlled type I error at the nominal rate for all combinations, while the max test of Zhao et al. (2014) had trouble when U_i was moderately sparse but V_i was very sparse. Table 3 reports the powers of the competing methods to detect dependency for different combinations of β_U , β_V , and β . Some of the cells of the table are empty because β must be greater than the larger of β_U and β_V . As Theorem 1 predicted, detection was most difficult when the U_i and V_i were moderately sparse and dependency was weak.

When the U_i and V_i were moderately sparse, with $\beta_U = \beta_V = 0.3$, Spearman’s correlation was usually the most powerful. However, the proposed dependency detection procedures had comparable power. All three tests dramatically outperformed the max test of Zhao et al. (2014), which was designed for very sparse U_i and V_i . When at least one of the U_i or V_i was very sparse, with either $\beta_U \geq 0.5$ or $\beta_V \geq 0.5$, the max test was usually the most powerful. However, again the proposed dependency detection procedures had comparable power, and all three dramatically outperformed Spearman’s correlation. Table 3 shows that the oracle and adaptive tests had similar performances. They were powerful for both large and small β_U , β_V , and β , and consistently performed as well as or better than competing methods.

7.3 Setting II: complicated distributions

The U_i and V_i were very sparse, with $\beta_U = \beta_V = 0.5$. Dependency was relative weak, with $\beta = 0.705$, as is often the case in genomics studies. For i such that $X_i = 0$, U_i was the absolute value of a central t distribution with 19 degrees of freedom, and otherwise was the absolute value of a non-central t with 19 degrees of freedom and noncentrality parameter equal to 3.5. For comparison, a t -statistic with 19 degrees of freedom and equal to 3.5 corresponds to a two-tailed p -value of 0.0024.

The V_i were generated as follows. An observation from a p -variate mean-zero normal distribution with a compound symmetry covariance matrix with parameter 0.5 was first generated, where p equaled one, 10, or 50. Next, each component of this observation was transformed to have a marginal t distribution with 19 degrees of freedom. For i such that

Table 4: Type I errors and powers at nominal $\alpha = 0.05$ for simulation setting II; see Section 7.3

Test	Type I errors			Powers		
	p			p		
	1	10	50	1	10	50
$\widehat{\mathcal{D}}_n$	0.01	0.00	0.01	0.51	0.71	0.80
Spearman	0.07	0.04	0.07	0.07	0.07	0.06
Max	0.06	0.00	0.00	0.21	0.00	0.00

$Y_i = 0$, this t distribution had a noncentrality parameter of zero, and otherwise had a noncentrality parameter equal to 3.5. Finally, V_i was calculated by taking the sum of the squared t -transformed components.

The construction of V_i was motivated by a problem in statistical genomics. Sometimes it is of interest to test whether there are SNPs that are simultaneously associated with a trait and with the expressions of any, some, or all of an entire set of genes, in order to test whether the gene set as a whole is associated with the trait. The p -dimensional vector generated in the calculation of V_i models the test statistics associated with each of p genes. The compound symmetry covariance models the covariance between the genes. Taking the sum of the squares of the transformed components models one way to combine the individual SNP-gene test statistics. The final distribution of the V_i is very complicated and unknown in practice, especially because the correlation between genes in the gene set is usually unknown.

The oracle test was not implemented because of the complicated distribution of V_i , but the other competing methods were implemented as in Section 7.2. Table 4 reports the type I errors and powers at a nominal $\alpha = 0.05$ for different values of p , i.e. different gene set sizes. All methods were successful at controlling the type I error, and the proposed dependency detection test was far more powerful than the other methods. The poor performance of Spearman’s correlation with very sparse U_i and V_i should be expected given the results of Section 7.2. The poor performance of the max test of Zhao et al. (2014) when $p \neq 1$ is due to the fact that the U_i and V_i had different variances even when both were null signals. Thus the pairwise maximum used in the max test, which implicitly assumes that the U_i

Table 5: Type I errors and powers at nominal $\alpha = 0.05$ for simulation setting III; see Section 7.4

Test	Type I errors			Powers		
	β_U			β_U		
	0.51	0.6	0.7	0.51	0.6	0.7
Oracle \mathcal{D}_n	0.06	0.04	0.08	0.12	0.48	0.62
Adaptive $\widehat{\mathcal{D}}_n$	0.03	0.04	0.05	0.10	0.41	0.56
Spearman	0.03	0.01	0.03	0.03	0.10	0.05
Max	0.07	0.06	0.08	0.14	0.64	0.74

and V_i are of comparable scales, is no longer sensible. Good performance of their test can be recovered by standardizing U_i and V_i to have unit variance. Nevertheless, even when $p = 1$ the proposed adaptive $\widehat{\mathcal{D}}_n$ outperformed the max test.

Table 4 also shows that the power increased for increasing p . This is because for i such that $Y_i = 1$, each of the p components was simulated to have signal, which was aggregated into V_i by taking the sum of their squares. This shows the potential gain in power than can be achieved by using gene sets in integrative genomics.

7.4 Detecting the undetectable

The U_i and V_i were generated according to the mixture model (16) discussed in Section 5. Specifically, for i such that $X_i = 0$, U_i was drawn from $|N(0, 1)|$ and otherwise was drawn from $|N[\{(2\beta_U - 1) \log n\}^{1/2}, 1]|$, with β_U equaling 0.51, 0.6, or 0.7. These choices give non-null Z -statistic means of 0.48, 1.51, and 2.15, which correspond to two-tailed p -values of 0.63, 0.13, and 0.03. Similarly, $V_i \sim |N(0, 1)|$ when $Y_i = 0$ and $V_i \sim |N\{(2 \log n)^{1/2}, 1\}|$ when $Y_i = 1$. Throughout these simulations V_i was very sparse, with $\beta_V = 0.5$. The dependency parameter β was set to $\beta_U \vee \beta_V + 0.01$. As discussed in Section 5, the work of Ingster (1997) and Donoho & Jin (2004) imply that the components with $X_i = 1$ are undetectable by an test given only the U_i .

Table 5 reports the type I errors and powers at a nominal $\alpha = 0.05$ for different sparsity levels of U_i . Both proposed tests, and the max test of Zhao et al. (2014), had power

to detect the presence of the non-null component of U_i , though the max had type I errors slightly larger than the nominal α . Again the oracle and adaptive tests performed similarly.

8 Data analysis

The adaptive dependency detection test statistic $\hat{\mathcal{D}}_n$ (17) was applied to study the genetic relationships between five psychiatric disorders: attention-deficit disorder, autism-spectrum disorders, bipolar disorder, major depressive disorder, and schizophrenia. Pairs of genetically related diseases can be jointly studied to discover pleiotropic SNPs as in Bhattacharjee et al. (2012), which may lead to a better understanding of the common underlying biological processes. The Cross-Disorder Group of the Psychiatric Genomics Consortium also studied this issue (Cross-Disorder Group of the Psychiatric Genomics Consortium et al. 2013*a,b*), and has made available meta-analysis p -values from genome-wide association studies of all five disorders; see <http://www.med.unc.edu/pgc/downloads>.

SNP association p -values were available for each of the five disorders for 1,219,805 genotyped and imputed SNPs. Many of these SNPs are highly correlated due to linkage disequilibrium. To account for this, the SNPs were pruned by physical distance. This was done for each chromosome by starting with the 5'-most SNP, proceeding in the 3' direction, and keeping only SNPs separated by at least 250 kilobases. The results of Dawson et al. (2002) suggest that the average r^2 between the remaining SNPs is only around 0.05. This pruning left $n = 10,488$ roughly independent SNPs.

The proposed $\hat{\mathcal{D}}_n$ was applied to each pair of disorders after converting the p -values to Z -scores. Non-null Z -scores are stochastically larger than null Z -scores, making $\hat{\mathcal{D}}_n$ asymptotically optimal. For comparison, Spearman's correlation and the max test of Zhao et al. (2014) were also performed, and the results are reported in Table 6. The proposed test suggested that bipolar disorder might be genetically related to major depressive disorder and schizophrenia, while the max test suggested only the latter relationship and Spearman's correlation identified no marginally significant disorder pairs. However, it is important to correct for having performed 10 pairwise tests. It is difficult to accurately adjust for multiple testing because of the complicated correlation of the tests, so a conservative approach is to use a Bonferroni threshold of 0.005. With this threshold none of the disorder pairs were

Table 6: Dependency detection p -values for all pairs of disorders

Disorder pair		Adaptive $\hat{\mathcal{D}}_n$	Spearman	Max
ADD	Autism	0.8142	0.1461	0.5745
ADD	Bipolar	0.2774	0.5514	0.1533
ADD	MDD	0.6701	0.2349	0.9637
ADD	Schizophrenia	0.6815	0.0771	0.8812
Autism	Bipolar	0.2902	0.4288	0.6705
Autism	MDD	0.4189	0.3418	0.1689
Autism	Schizophrenia	0.4603	0.0631	0.2391
Bipolar	MDD	0.0306*	0.7427	0.0825
Bipolar	Schizophrenia	0.0289*	0.1653	0.0310*
MDD	Schizophrenia	0.2447	0.2814	0.1904

ADD: attention-deficit disorder; MDD: major depressive disorder; *marginally significant.

significant.

On the other hand, there is an alternative method of analysis: it is also of interest to test whether each disorder is related to *any* of the remaining disorders. Specifically, let Z_i^j be the Z -score of the i th SNP for the j th disorder. Now let $U_i = Z_i^j$ and $V_i = \sum_{k \neq j} (Z_i^k)^2$; this is similar to the test statistics used in simulation setting II in Section 7.3. Now dependency detection can be applied to U_i and V_i . If dependency is detected, it can be concluded that disorder j is genetically related to at least one of the remaining disorders. One advantage of this analysis is that only 5 tests are required, reducing the multiple testing burden.

Table 7: Dependency detection p -values for each disorder against all remaining disorders

Disorder	Adaptive $\hat{\mathcal{D}}_n$	Spearman	Max
ADD	0.6577	0.4433	0.8444
Autism	0.1897	0.0034	0.8636
Bipolar	0.0035	0.3395	0.3954
MDD	0.0306*	0.5090	0.9938
Schizophrenia	0.0289*	0.0260*	0.3212

ADD: attention-deficit disorder; MDD: major depressive disorder; **bold** p -values pass Bonferroni correction; *marginally significant.

Table 7 reports the results of testing each disorder against the entire group of remaining disorders. The proposed $\hat{\mathcal{D}}_n$ found that bipolar disorder was highly genetically related to at least one of the remaining disorders. Furthermore, the corresponding p -value is much smaller than any of the pairwise p -values associated with bipolar disorder given in Table 6. This suggests that there is more than one disorder that is genetically related to bipolar disorder, but that the individual genetic relationships are too weak to be detected separately. In light of the pairwise p -values, it seems that bipolar disorder, major depressive disorder, and schizophrenia constitute a trio of highly related disorders. This is in keeping with previous findings (Cross-Disorder Group of the Psychiatric Genomics Consortium et al. 2013a). Spearman’s correlation found a similar relationship between autism and the remaining diseases, though this is not reflected in any of the pairwise analyses. These results show that grouping disorders can be a valuable analysis strategy.

9 Discussion

In some applications, for example in the psychiatric disorder example in Section 8, the n observations in one or both sequences may not be independent. In other words, the U_i may not be independent across i , and the V_i may not be independent either. For example, if U_i is a test statistic for association between the i th gene and a disease, the different U_i will be dependent because different genes can be highly correlated with each other. Characterizing the behavior of $\hat{\mathcal{D}}_n$, especially its asymptotic null distribution, for dependent components of U_i and V_i is of great practical interest. Hall & Jin (2010) studied this problem in the signal detection setting, and a related approach may work well for the dependency detection setting.

As discussed in Section 5, when two paired sequences U_i and V_i are available, using dependency detection as a method of signal detection can be more powerful than signal detection methods that use only one sequence at a time. A natural question is whether this advantage extends to signal identification, specifically to identifying signals that are simultaneously present in both sequences. One approach would be to identify signals in the U_i and V_i separately, and then to record the signals common to both sequences. Signal identification in this one-sample setting has been extensively studied. Xie et al. (2011)

established the fundamental limits of signal recovery, and Cai & Sun (2014) developed data-driven procedures for signal screening and discovery and also established phase diagrams to characterize the fundamental limits of signal identification. One of the implications of the findings in Section 5 is that having two paired sequences may allow the identification of simultaneous signals that are below these limits.

It may sometimes be useful to detect dependency or identify signals using more than two sequences. For example, in the psychiatric disorder example in Section 8, it may be desirable to test whether any of the five disorders are genetically related without testing all pairs. In principle the proposed test statistic could be extended to higher dimensions, and in fact Einmahl (1996) derived the asymptotic null distribution of the oracle \mathcal{D}_n for any dimension. However, it becomes computationally cumbersome to calculate a multivariate empirical distribution function even in relatively low dimensions, so a different test statistic is needed.

A related problem is that nonparametric estimation of multivariate distribution functions is well-known to be difficult. Even in the bivariate setting, a large number of observations are needed before \hat{S}_{UV} can be a useful estimate for the true bivariate survival function. In cases where something is known about the null and non-null components of U_i and V_i , and when those sequences are dense enough that estimation of their mixture proportions is feasible, a parametric dependency detection procedure may have greater power than the method proposed here.

Finally, having two or more paired sequences opens up a number of interesting new statistical questions, some of which are especially relevant for genomics research. For example, it is sometimes desirable to identify genes that are expressed only in one cell type and not another. Letting U_i and V_i denote the expression levels of gene i in the two cell types, the task becomes to identify signals that are present only in the U_i but not the V_i . As another example, it is sometimes also desirable to determine whether genes that are correlated in one tissue type are also correlated in another. In this case, the task becomes to detect differences in the supports of two correlations matrices, instead of two sequences of test statistics. Ideas similar to those described in this paper may be applicable to these new problems.

SUPPLEMENTARY MATERIAL

Proofs: Lemmas and proofs of theorems. (pdf file)

10 Lemmas

The following lemmas are used to prove the results in this paper. Some useful results from (Einmahl & Mason 1985), (Einmahl 1996), and Cai & Wu (2014) are reproduced here for completeness.

Lemma 1 (Einmahl & Mason 1985) *Let \mathcal{D}_n be defined as in (3). If U_i and V_i are independent, then*

$$\limsup_{n \rightarrow \infty} \frac{\log \mathcal{D}_n}{\log \log n} \stackrel{a.s.}{=} 1.$$

Lemma 2 (Einmahl 1996) *Let \mathcal{D}_n be defined as in (3). If U_i and V_i are independent, then*

$$(\log n)^{-1/2} \mathcal{D}_n \xrightarrow{d} E^{-1/2},$$

where E is a standard exponential random variable.

Lemma 3 (Lemma 3 of Cai & Wu (2014)) *Let (X, \mathcal{F}, ν) be a measure space. Let $F : X \times \mathbb{R}_+ \rightarrow \mathbb{R}_+$ be measurable. Assume that*

$$\lim_{M \rightarrow \infty} \frac{\log F(x, M)}{M} = f(x)$$

holds uniformly in $x \in X$ for some measurable $f : X \rightarrow \mathbb{R}$. If

$$\int_X \exp(M_0 f) d\nu < \infty$$

for some $M_0 > 0$, then

$$\lim_{M \rightarrow \infty} \frac{1}{M} \log \int_X F(x, M) d\nu = \operatorname{ess\,sup}_{x \in X} f(x).$$

Lemma 4 (Lemma 4.2 of Cai & Wu (2014)) *For any $0 \leq t$, $(2^{1/2} - 1)^2 t \wedge t^2 \leq \{1 - (1 + t)^{1/2}\}^2 \leq t \wedge t^2$.*

Lemma 5 Under Assumptions 1 and 2, for $x, y \geq \log_n 2$,

$$\begin{aligned} F_n^U \{F_0^{-1}(n^{-x})\} &= n^{\text{ess sup}_{x \leq a} \{\alpha_{\bar{U}}^-(a) - a\} + o(1)}, & F_n^U \{F_0^{-1}(1 - n^{-x})\} &= 1 - n^{\text{ess sup}_{x \leq a} \{\alpha_{\bar{U}}^+(a) - a\} + o(1)}, \\ F_n^V \{F_0^{-1}(n^{-y})\} &= n^{\text{ess sup}_{y \leq b} \{\alpha_{\bar{V}}^-(b) - b\} + o(1)}, & F_n^V \{F_0^{-1}(1 - n^{-y})\} &= 1 - n^{\text{ess sup}_{y \leq b} \{\alpha_{\bar{V}}^+(b) - b\} + o(1)}. \end{aligned}$$

Proof. This lemma is similar to Lemma 6 of Cai & Wu (2014). When $x \geq \log_n 2$, using Assumption 1 and making the change of variables $u \mapsto (F_0^U)^{-1}(n^{-a})$ implies

$$du = -\frac{F_0^U \{(F_0^U)^{-1}(n^{-a})\} \log n}{f_0^U \{(F_0^U)^{-1}(n^{-a})\}} da = -\frac{n^{-a} \log n}{f_0^U \{(F_0^U)^{-1}(n^{-a})\}} da.$$

Therefore

$$\begin{aligned} F_n^U \{(F_0^U)^{-1}(n^{-x})\} &= \int_{-\infty}^{(F_0^U)^{-1}(n^{-x})} f_n^U(u) du = -\log n \int_{\infty}^x \exp[\ell_n \{(F_0^U)^{-1}(n^{-a})\}] n^{-a} da \\ &= \log n \int_x^{\infty} n^{\alpha_{\bar{U}}^-(a) - a + o(1)} da = n^{\text{ess sup}_{x \leq a} \{\alpha_{\bar{U}}^-(a) - a\} + o(1)}. \end{aligned}$$

where the third equality follows from Assumption 2 and the last equality follows from Lemma 3.

Similarly, making the change of variables $u \mapsto (F_0^U)^{-1}(1 - n^{-a})$ implies

$$du = \frac{n^{-a} \log n}{f_0^U \{(F_0^U)^{-1}(n^{-a})\}} da.$$

Therefore

$$\begin{aligned} F_n^U \{(F_0^U)^{-1}(1 - n^{-x})\} &= \int_{-\infty}^{(F_0^U)^{-1}(1 - n^{-x})} f_n^U(u) du = 1 - \int_{(F_0^U)^{-1}(1 - n^{-x})}^{\infty} f_n^U(u) du \\ &= 1 - \log n \int_x^{\infty} \exp[\ell_n \{(F_0^U)^{-1}(1 - n^{-a})\}] n^{-a} da \\ &= 1 - \log n \int_x^{\infty} n^{\alpha_{\bar{U}}^+(a) - a + o(1)} da = 1 - n^{\text{ess sup}_{x \leq a} \{\alpha_{\bar{U}}^+(a) - a\} + o(1)}. \end{aligned}$$

The proofs for F_n^V when $y \geq \log_n 2$ are exactly analogous. \square

Lemma 6 Under Assumptions 1 and 2,

$$\begin{aligned} \text{ess sup}_{a \geq \log_n 2} \{\alpha_{\bar{U}}^-(a) - a\} &= 0, & \text{ess sup}_{a \geq \log_n 2} \{\alpha_{\bar{U}}^+(a) - a\} &= 0, \\ \text{ess sup}_{b \geq \log_n 2} \{\alpha_{\bar{V}}^-(b) - b\} &= 0, & \text{ess sup}_{b \geq \log_n 2} \{\alpha_{\bar{V}}^+(b) - b\} &= 0, \end{aligned}$$

Proof. This lemma is similar to Lemma 2 of Cai & Wu (2014). Following Lemma 5,

$$(\log n)^{-1} = (\log n)^{-1} \left\{ \int_{-\infty}^{(F_0^U)^{-1}(0.5)} f_n^U(u) du + \int_{(F_0^U)^{-1}(0.5)}^{\infty} f_n^U(u) du \right\}.$$

For the first integral, make the change of variables $u \mapsto (F_0^U)^{-1}(n^{-a})$, and for the second integral use $u \mapsto (F_0^U)^{-1}(1 - n^{-a})$. Similar to Lemma 5,

$$\begin{aligned} (\log n)^{-1} &= \int_{\log_n 2}^{\infty} n^{\alpha_U^-(a)-a+o(1)} da + \int_{\log_n 2}^{\infty} n^{\alpha_U^+(a)-a+o(1)} da \\ &= n^{\text{ess sup}_{a \geq \log_n 2} \{\alpha_U^-(a)-a+o(1)\}} + n^{\text{ess sup}_{a \geq \log_n 2} \{\alpha_U^+(a)-a+o(1)\}}. \end{aligned}$$

Both terms in the sum are positive, so both essential suprema must equal zero. The proof for α_V^- and α_V^+ is analogous. \square

Lemma 7 For any function $f(x)$ and constants c_1 and c_2 ,

$$\{\sup_x f(x)\} \wedge [\sup_x \{c_1 f(x) + c_2\}] = \sup_x [f(x) \wedge \{c_1 f(x) + c_2\}].$$

Proof. First it is clear that

$$\{\sup_x f(x)\} \wedge [\sup_x \{c_1 f(x) + c_2\}] \geq \sup_x [f(x) \wedge \{c_1 f(x) + c_2\}].$$

Now fix $\epsilon > 0$. By the definition of the supremum, there exist x_1 and x_2 such that

$$f(x_1) > \sup_x f(x) - \epsilon, \quad c_1 f(x_2) + c_2 > c_1 \sup_x f(x) + c_2 - \epsilon.$$

Complete the proof by defining x^* equal either x_1 or x_2 such that $f(x^*) \geq f(x_1) \vee f(x_2)$.

Then

$$f(x^*) \wedge \{c_1 f(x^*) + c_2\} \geq f(x_1) \wedge \{c_1 f(x_2) + c_2\} \geq \{\sup_x f(x) - \epsilon\} \wedge \{c_1 \sup_x f(x) + c_2 - \epsilon\}. \square$$

11 Proof of Theorem 1

The squared Hellinger distance between two distributions P_0 and P_1 , with densities p_0 and p_1 with respect to the Lebesgue measure μ , is defined as

$$H^2(P_0, P_1) = \frac{1}{2} \int (p_0^{1/2} - p_1^{1/2})^2 d\mu.$$

If P_0 and P_1 are the distributions of (U_i, V_i) under H_0 and H_A , respectively, then by Theorem 13.1.3 of Lehmann & Romano (2005), $S_{H_0, H_A}(\phi) \rightarrow 1$ for all tests ϕ if $nH^2(P_0, P_1) \rightarrow 0$. It remains to show that conditions (6)–(8) imply $H^2(P_0, P_1) = o(n^{-1})$.

For compactness of notation define the function

$$q(u, v) = \left(1 - \left[1 + \frac{n^{-\beta}(L_n^U - 1)(L_n^V - 1)}{\{1 + n^{-\beta u}(L_n^U - 1)\}\{1 + n^{-\beta v}(L_n^V - 1)\}} \right]^{1/2} \right)^2,$$

where $L_n^U = f_n^U/f_0^U$ and $L_n^V = f_n^V/f_0^V$ are likelihood ratios. Next define the sets

$$\mathcal{I}_1 = \{u, v : L_n^U(u) < 1, L_n^V(v) < 1\},$$

$$\mathcal{I}_2 = \{u, v : 1 \leq L_n^U(u), L_n^V(v) < 1\},$$

$$\mathcal{I}_3 = \{u, v : L_n^U(u) < 1, 1 \leq L_n^V(v)\},$$

$$\mathcal{I}_4 = \{u, v : 1 \leq L_n^U(u), 1 \leq L_n^V(v)\}.$$

By definition L_n^U and L_n^V are always positive. Then the squared Hellinger distance satisfies

$$2H^2(P_0, P_1) = \int_{\mathcal{I}_1} q f_U f_V dudv + \int_{\mathcal{I}_2} q f_U f_V dudv + \int_{\mathcal{I}_3} q f_U f_V dudv + \int_{\mathcal{I}_4} q f_U f_V dudv,$$

where $f_U = (1 - \pi_U)f_0^U + \pi_U f_n^U$ and $f_V = (1 - \pi_V)f_0^V + \pi_V f_n^V$ are the marginal densities of U_i and V_i . Each one of these four integrals can be bounded separately.

First, on \mathcal{I}_1 the term inside the square root in $q(u, v)$ is always larger than one for $n > 1$, and is maximized when $L_n^U = L_n^V = 0$. Therefore

$$\begin{aligned} \int_{\mathcal{I}_1} q f_U f_V dudv &\leq \int_{\mathcal{I}_1} \left(1 - \left[1 + \frac{n^{-\beta}}{(1 - n^{-\beta u})(1 - n^{-\beta v})} \right]^{1/2} \right)^2 f_U f_V dudv \\ &\leq \frac{n^{-\beta}}{(1 - n^{-\beta u})(1 - n^{-\beta v})} \wedge \frac{n^{-2\beta}}{(1 - n^{-\beta u})^2(1 - n^{-\beta v})^2} \\ &= o(n^{-1}), \end{aligned}$$

where the second inequality comes from Lemma 4 and the last equality follows because $\beta > 1/2$ under weak dependence.

Next, to upper-bound $q(u, v)$ on \mathcal{I}_2 , it is easy to show that $\partial q/\partial L_n^V \leq 0$, which implies

that q is maximized when $L_n^V = 0$. Therefore

$$\begin{aligned} \int_{\mathcal{I}_2} q f_U f_V dudv &\leq \int_{\mathcal{I}_2} \left[1 - \left\{ 1 - \frac{n^{-\beta}}{1 - n^{-\beta v}} \frac{L_n^U - 1}{1 + n^{-\beta v} (L_n^U - 1)} \right\}^{1/2} \right]^2 f_U f_V dudv \\ &\leq \frac{n^{-2\beta}}{(1 - n^{-\beta v})^2} \int_{\mathcal{I}_2} \frac{(L_n^U - 1)^2}{\{1 + n^{-\beta v} (L_n^U - 1)\}^2} f_U f_V dudv \\ &= \frac{n^{-2\beta}}{(1 - n^{-\beta v})^2} \int_{\mathcal{I}_2} \frac{(L_n^U - 1)^2}{1 + n^{-\beta v} (L_n^U - 1)} f_0^U f_V dudv. \end{aligned}$$

where the second inequality follows the facts that $\{1 - (1 - x)^{1/2}\}^2 < x^2$ for $x \in [0, 1]$.

The set \mathcal{I}_2 can be divided into disjoint subsets

$$\begin{aligned} \mathcal{I}_{21} &= \{u, v : 1 \leq L_n^U(u), u \leq (F_0^U)^{-1}(0.5), L_n^V(v) < 1\}, \\ \mathcal{I}_{22} &= \{u, v : 1 \leq L_n^U(u), u > (F_0^U)^{-1}(0.5), L_n^V(v) < 1\}. \end{aligned}$$

On \mathcal{I}_{21} make the change of variables $u \mapsto (F_0^U)^{-1}(n^{-a})$, $a \geq \log_n 2$, such that by Assumption 1,

$$du = -\log n \frac{F_0^U(u)}{f_0^U(u)} da = -\frac{n^{-a} \log n}{f_0^U\{(F_0^U)^{-1}(n^{-a})\}} da.$$

On \mathcal{I}_{22} use $u \mapsto (F_0^U)^{-1}(n^{1-a})$, $a > \log_n 2$, which implies

$$du = \frac{n^{-a} \log n}{f_0^U\{(F_0^U)^{-1}(n^{-a})\}} da.$$

Finally, Assumption 2 implies that for n sufficiently large, there is a small $\delta > 0$ such that $a \geq \log_n 2$, $L_n^U\{(F_0^U)^{-1}(n^{-a})\} \leq n^{\alpha_V^-(a)+\delta}$ and $L_n^U\{(F_0^U)^{-1}(1 - n^{-a})\} \leq n^{\alpha_V^+(a)+\delta}$.

Therefore for n large enough and a generic constant C_n that contains a $\log n$ factor,

$$\begin{aligned} \int_{\mathcal{I}_2} q f_U f_V dudv &\leq C_n n^{-2\beta} \int_{\mathcal{I}_{21}} \frac{[L_n^U\{(F_0^U)^{-1}(n^{-a})\} - 1]^2}{1 + n^{-\beta v} [L_n^U\{(F_0^U)^{-1}(n^{-a})\} - 1]} n^{-a} f_V dadv + \\ &\quad C_n n^{-2\beta} \int_{\mathcal{I}_{22}} \frac{[L_n^U\{(F_0^U)^{-1}(1 - n^{-a})\} - 1]^2}{1 + n^{-\beta v} [L_n^U\{(F_0^U)^{-1}(1 - n^{-a})\} - 1]} n^{-a} f_V dadv \\ &\leq C_n n^{-2\beta} \int_{\left\{ \begin{array}{l} 0 \leq \alpha_V^-(a) + \delta, \\ a \geq \log_n 2 \end{array} \right\}} \frac{(n^{\alpha_V^-(a)+\delta} - 1)^2}{1 + n^{-\beta v} (n^{\alpha_V^-(a)+\delta} - 1)} n^{-a} da + \\ &\quad C_n n^{-2\beta} \int_{\left\{ \begin{array}{l} 0 \leq \alpha_V^+(a) + \delta, \\ a > \log_n 2 \end{array} \right\}} \frac{(n^{\alpha_V^+(a)+\delta} - 1)^2}{1 + n^{-\beta v} (n^{\alpha_V^+(a)+\delta} - 1)} n^{-a} da. \end{aligned}$$

If both $\alpha_V^-(a) < 0$ and $\alpha_V^+(a) < 0$ for all $a \geq \log_n 2$, both integrals above are equal to zero. Otherwise, $\alpha_V(a) = \alpha_V^-(a) \vee \alpha_V^+(a) > 0$ for some subset of $\{a \geq \log_n 2\}$ with positive

Lebesgue measure. Therefore

$$\begin{aligned} \int_{\mathcal{I}_2} qf_U f_V dudv &\leq C_n n^{-2\beta} \int_{\log_n 2} \left\{ \frac{n^{2(\alpha_U^- + \delta)}}{1 + n^{-\beta_U + \alpha_U^- + \delta}} n^{-a} + \frac{n^{2(\alpha_V^+ + \delta)}}{1 + n^{-\beta_V + \alpha_V^+ + \delta}} n^{-a} \right\} da \\ &\leq C_n n^{-2\beta} \int_{\log_n 2} n^{(\alpha_U^- \vee \alpha_V^+ + \delta) + \{(\alpha_U^- \vee \alpha_V^+ + \delta) \wedge \beta_U\} - a} da, \end{aligned}$$

and by Lemma 3 and (6),

$$\int_{\mathcal{I}_2} qf_U f_V dudv \leq C_n n^{-2\beta + \text{ess sup}_{a \geq \log_n 2} [(\alpha_U + \delta) + \{(\alpha_U + \delta) \wedge \beta_U\} - a]} = o(n^{-1}).$$

Similar reasoning shows that under (7),

$$\int_{\mathcal{I}_3} qf_U f_V dudv = o(n^{-1}).$$

To complete the proof the integral of $q(u, v)$ over \mathcal{I}_4 must be upper-bounded. Similar to before, \mathcal{I}_4 can be divided into disjoint subsets

$$\begin{aligned} \mathcal{I}_{41} &= \{u, v : 1 \leq L_n^U(u), u \leq (F_0^U)^{-1}(0.5), 1 \leq L_n^V(v), v \leq (F_0^V)^{-1}(0.5)\}, \\ \mathcal{I}_{42} &= \{u, v : 1 \leq L_n^U(u), u > (F_0^U)^{-1}(0.5), 1 \leq L_n^V(v), v \leq (F_0^V)^{-1}(0.5)\}, \\ \mathcal{I}_{43} &= \{u, v : 1 \leq L_n^U(u), u \leq (F_0^U)^{-1}(0.5), 1 \leq L_n^V(v), v > (F_0^V)^{-1}(0.5)\}, \\ \mathcal{I}_{44} &= \{u, v : 1 \leq L_n^U(u), u > (F_0^U)^{-1}(0.5), 1 \leq L_n^V(v), v > (F_0^V)^{-1}(0.5)\}, \end{aligned}$$

On \mathcal{I}_{41} let $u \mapsto (F_0^U)^{-1}(n^{-a})$ and $v \mapsto (F_0^V)^{-1}(n^{-b})$. If either $\alpha_U^-(a) < 0$ or $\alpha_V^-(b) < 0$ for all $a, b \geq \log_n 2$, the integral over \mathcal{I}_{41} equals zero. Otherwise,

$$\begin{aligned} \int_{\mathcal{I}_{41}} qf_U f_V dudv &\leq C_n \int_{\{a, b \geq \log_n 2\}} \left[1 - \left\{ 1 + \frac{n^{-\beta + \alpha_U^- + \alpha_V^- + 2\delta}}{(1 + n^{-\beta_U + \alpha_U^- + \delta})(1 + n^{-\beta_V + \alpha_V^- + \delta})} \right\}^{1/2} \right]^2 \\ &\quad (1 + n^{-\beta_U + \alpha_U^- + \delta})(1 + n^{-\beta_V + \alpha_V^- + \delta}) n^{-a-b} dadb \\ &\leq C_n \int_{\{a, b \geq \log_n 2\}} \left\{ n^{-\beta + \alpha_U^- + \alpha_V^- + 2\delta} \wedge \frac{n^{-2\beta + 2\alpha_U^- + 2\alpha_V^- + 4\delta}}{(1 + n^{-\beta_U + \alpha_U^- + \delta})(1 + n^{-\beta_V + \alpha_V^- + \delta})} \right\} \\ &\quad n^{-a-b} dadb \\ &= C_n \int_{\{a, b \geq \log_n 2\}} [n^{-\beta + \alpha_U^- + \alpha_V^- + 2\delta} \wedge n^{-2\beta + \alpha_U^- + \alpha_V^- + 2\delta + \{(\alpha_U^- + \delta) \wedge \beta_U\} + \{(\alpha_V^- + \delta) \wedge \beta_V\}}] \\ &\quad n^{-a-b} dadb, \end{aligned}$$

where the second inequality is due to Lemma 4. Corresponding calculations over the other three subsets of \mathcal{I}_4 imply that when both $\alpha_U(a) = \alpha_U^-(a) \vee \alpha_U^+(a) > 0$ and $\alpha_V(b) = \alpha_V^-(b) \vee \alpha_V^+(b) > 0$ on some subset of $\{a, b \geq \log_n 2\}$ of positive Lebesgue measure,

$$\int_{\mathcal{I}_4} q f_U f_V du dv \leq C_n n^{\text{ess sup}_{a,b \geq \log_n 2} \{-\beta + \alpha_U + \alpha_V + 2\delta\} \wedge [-2\beta + \alpha_U + \alpha_V + 2\delta + \{(\alpha_U + \delta) \wedge \beta_U\} + \{(\alpha_V + \delta) \wedge \beta_V\}] - a - b},$$

which is $o(n^{-1})$ when (8) holds. \square

12 Proof of Theorem 2

Since $\mathbb{P}_{H_0}\{\mathcal{D}_n > (\log n)^{1/2}\} = o(1)$ by Lemma 1, it remains to show that $\mathbb{P}_{H_A}\{\mathcal{D}_n \leq (\log n)^{1/2}\}$ is also $o(1)$. Let $U'_i = -U_i$ and $V'_i = -V_i$, and define

$$W_n(u', v') = n^{1/2} \frac{\sum_i I\{F_{U'}(U'_i) \leq u', F_{V'}(V'_i) \leq v'\} - u'v'}{(u'v' - u'^2v'^2)^{1/2}},$$

so that $(\log n)^{-1/2} \sup_{0 < u, v < 1} |W_n| = \mathcal{D}_n$. Then for any (u', v') ,

$$\mathbb{P}_{H_A}\{\mathcal{D}_n \leq (\log n)^{1/2}\} \leq \mathbb{P}_{H_A}\{(\log n)^{-1/2} |W_n(u', v')| \leq (\log n)^{1/2}\}.$$

By the triangle inequality and Chebyshev's inequality,

$$\begin{aligned} & \mathbb{P}_{H_A}\{(\log n)^{-1/2} |W_n(u', v')| \leq (\log n)^{1/2} + r\} \\ & \leq \mathbb{P}_{H_A}\{|W_n - \mathbb{E}_{H_A} W_n| \geq |\mathbb{E}_{H_A} W_n| - \log n - r(\log n)^{1/2}\} \\ & \leq \frac{\text{var}_{H_A} W_n}{\{|\mathbb{E}_{H_A} W_n| - \log n - r(\log n)^{1/2}\}^2}. \end{aligned}$$

Therefore the desired result follows if there exists a $(u', v') \in (0, 1) \times (0, 1)$ such that

$$\log n / |\mathbb{E}_{H_A} W_n| \rightarrow 0, \tag{18}$$

$$\text{var}_{H_A}(W_n) / |\mathbb{E}_{H_A} W_n|^2 \rightarrow 0. \tag{19}$$

To check for such a (u', v') , divide $(0, 1) \times (0, 1)$ into four quadrants

$$\begin{aligned} \mathcal{Q}_1 &= [u', v' : u' \leq F_{U'}\{-(F_0^U)^{-1}(0.5)\}, v' \leq F_{V'}\{-(F_0^V)^{-1}(0.5)\}], \\ \mathcal{Q}_2 &= [u', v' : u' > F_{U'}\{-(F_0^U)^{-1}(0.5)\}, v' \leq F_{V'}\{-(F_0^V)^{-1}(0.5)\}], \\ \mathcal{Q}_3 &= [u', v' : u' \leq F_{U'}\{-(F_0^U)^{-1}(0.5)\}, v' > F_{V'}\{-(F_0^V)^{-1}(0.5)\}], \\ \mathcal{Q}_4 &= [u', v' : u' < F_{U'}\{-(F_0^U)^{-1}(0.5)\}, v' > F_{V'}\{-(F_0^V)^{-1}(0.5)\}]; \end{aligned}$$

the desired result follows if there exists a (u', v') in any of these quadrants that satisfies (18) and (19). The distribution function $F_{U'}$ satisfies

$$F_{U'}(-u) = P(U'_i \leq -u) = P(U_i \geq u) = (1 - \pi_U)\{1 - F_0^U(u)\} + \pi_U\{1 - F_n^U(u)\}$$

and a similar relation holds for $F_{V'}(-v)$. Thus letting $u' = F_{U'}(-u)$, $v' = F_{V'}(-v)$, and $F_{U'V'}$ be the true bivariate distribution of (U'_i, V'_i) ,

$$\begin{aligned} E_{H_A} W_n(u', v') &= \frac{n^{1/2}(F_{U'V'} - F_{U'}F_{V'})}{\{u'v'(1 - u'v')\}^{1/2}} \\ &= \frac{n^{1/2-\beta}\{F_n^U(u) - F_0^U(u)\}\{F_n^V(v) - F_0^V(v)\}}{\{u'v'(1 - u'v')\}^{1/2}}, \end{aligned} \quad (20)$$

$$\begin{aligned} \text{var}_{H_A} W_n(u', v') &= \frac{F_{U'V'} - F_{U'}^2 F_{V'}}{u'v'(1 - u'v')} \\ &\leq \frac{F_{U'}F_{V'} + n^{-\beta}\{F_n^U(u) - F_0^U(u)\}\{F_n^V(v) - F_0^V(v)\}}{u'v'(1 - u'v')} \\ &= \frac{1}{1 - u'v'} + \frac{(E_{H_A} W_n)^2}{n^{1-\beta}\{F_n^U(u) - F_0^U(u)\}\{F_n^V(v) - F_0^V(v)\}}. \end{aligned} \quad (21)$$

Quadrant \mathcal{Q}_1 corresponds to $u' = F_{U'}\{-(F_0^U)^{-1}(1 - n^{-x})\}$ and $v' = F_{V'}\{-(F_0^V)^{-1}(1 - n^{-y})\}$ for $x, y \geq \log_n 2$. Then from (20), Assumptions 1 and 2, and Lemma 5, for n sufficiently large and some generic constant C_n that may contain factors of $\log n$, the numerator of $|E_{H_A} W_n|$ is

$$\begin{aligned} &|n^{1/2-\beta}[F_n^U\{(F_0^U)^{-1}(1 - n^{-x})\} - 1 + n^{-x}][F_n^V\{(F_0^V)^{-1}(1 - n^{-y})\} - 1 + n^{-y}]| \\ &= |n^{1/2-\beta}\{n^{-x} - n^{p_U^+(x)+o(1)}\}\{n^{-y} - n^{p_V^+(y)+o(1)}\}| \\ &= C_n n^{1/2-\beta} n^{(-x)\vee p_U^+(x)} n^{(-y)\vee p_V^+(y)}, \end{aligned}$$

and

$$\begin{aligned} u' &= (1 - n^{-\beta_U})n^{-x} + n^{-\beta_U}[1 - F_n^U\{(F_0^U)^{-1}(1 - n^{-x})\}] \\ &= n^{-x} + n^{-\beta_U}\{n^{p_U^+(x)+o(1)} - n^{-x}\} = C_n n^{(-x)\vee(-\beta_U+p_U^+)}, \\ v' &= C_n n^{(-y)\vee(-\beta_V+p_V^+)}. \end{aligned}$$

Lemma 6 implies that $p_U^+, p_V^+ \leq 0$, so $u'v' < 1$ and $(1 - u'v') \rightarrow 1$. Therefore from (20) and

(21),

$$|E_{H_A} W_n| = C_n \frac{n^{1/2-\beta+(-x)\vee p_U^+ + (-y)\vee p_V^+}}{n^{\{(-x)\vee(-\beta_U+p_U^+)\}/2 + \{(-y)\vee(-\beta_V+p_V^+)\}/2}},$$

$$\frac{\text{var}_{H_A} W_n}{|E_{H_A} W_n|^2} = \frac{C_n}{|E_{H_A} W_n|^2} + \frac{C_n}{n^{1-\beta+(-x)\vee p_U^+ + (-y)\vee p_V^+}}.$$

When $x + y < 1$,

$$\begin{aligned} & 1 - \beta + (-x) \vee p_U^+ + (-y) \vee p_V^+ \\ & \geq \frac{1}{2} - \beta + (-x) \vee p_U^+ + (-y) \vee p_V^+ + \frac{x + y}{2} \\ & \geq \frac{1}{2} - \beta + (-x) \vee p_U^+ + (-y) \vee p_V^+ + \frac{x \wedge (\beta_U - p_U^+)}{2} + \frac{y \wedge (\beta_V - p_V^+)}{2}, \end{aligned}$$

so (10) is a sufficient condition for there to exist a $(u', v') \in \mathcal{Q}_1$ such that (18) and (19) hold.

Quadrant \mathcal{Q}_2 corresponds to $u' = F_{U'}\{-(F_0^U)^{-1}(n^{-x})\}$ and $v' = F_{V'}\{-(F_0^V)^{-1}(1 - n^{-y})\}$ for $x, y \geq \log_n 2$. Then the numerator of $|E_{H_A} W_n|$ is

$$\begin{aligned} & |n^{1/2-\beta}[F_n^U\{(F_0^U)^{-1}(n^{-x})\} - n^{-x}][F_n^V\{(F_0^V)^{-1}(1 - n^{-y})\} - 1 + n^{-y}]| \\ & = C_n n^{1/2-\beta} n^{(-x)\vee p_U^-(x)} n^{(-y)\vee p_V^+(y)}, \end{aligned}$$

and

$$\begin{aligned} u' & = (1 - n^{-\beta_U})(1 - n^{-x}) + n^{-\beta_U} \{1 - n^{p_U^-(x)+o(1)}\} \\ & = 1 - n^{-x} - n^{-\beta_U} + n^{-x-\beta_U} + n^{-\beta_U} - n^{-\beta_U+p_U^-(x)+o(1)} \\ & = 1 + C_n(n^{-x} + n^{-\beta_U}) = O(1), \\ v' & = C_n n^{(-y)\vee(-\beta_V+p_V^+)}, \end{aligned}$$

where the last equality for u' used $p_U^- \leq 0$ by Lemma 6. Again $u'v' < 1$ and $(1 - u'v') \rightarrow 1$, so from (20) and (21),

$$|E_{H_A} W_n| = C_n \frac{n^{1/2-\beta+(-x)\vee p_U^- + (-y)\vee p_V^+}}{n^{\{(-y)\vee(-\beta_V+p_V^+)\}/2}},$$

$$\frac{\text{var}_{H_A} W_n}{|E_{H_A} W_n|^2} = \frac{C_n}{|E_{H_A} W_n|^2} + \frac{C_n}{n^{1-\beta+(-x)\vee p_U^- + (-y)\vee p_V^+}}.$$

When $x + y < 1$,

$$\begin{aligned}
& 1 - \beta + (-x) \vee p_{\bar{U}} + (-y) \vee p_V^+ \\
& \geq \frac{1}{2} - \beta + (-x) \vee p_{\bar{U}} + (-y) \vee p_V^+ + \frac{y}{2} \\
& \geq \frac{1}{2} - \beta + (-x) \vee p_{\bar{U}} + (-y) \vee p_V^+ + \frac{y \wedge (\beta_V - p_V^+)}{2},
\end{aligned}$$

so (11) is a sufficient condition for there to exist a $(u', v') \in \mathcal{Q}_2$ such that (18) and (19) hold. It can be similarly be shown that (12) is a sufficient condition in \mathcal{Q}_3 .

Quadrant \mathcal{Q}_4 corresponds to $u' = F_{U'}\{-(F_0^U)^{-1}(n^{-x})\}$ and $v' = F_{V'}\{-(F_0^V)^{-1}(n^{-y})\}$ for $x, y \geq \log_n 2$. Then the numerator of $|E_{H_A} W_n|$ is

$$C_n n^{1/2-\beta} n^{(-x) \vee p_{\bar{U}}(x)} n^{(-y) \vee p_{\bar{U}}(y)},$$

and

$$u' = 1 + C_n(n^{-x} + n^{-\beta_U}), \quad v' = 1 + C_n(n^{-y} + n^{-\beta_V}).$$

Therefore $u'v' = O(1)$ and $1 - u'v'$ is of order

$$n^{-x} + n^{-\beta_U} + n^{-y} + n^{-\beta_V},$$

and

$$\begin{aligned}
|E_{H_A} W_n| &= C_n \frac{n^{1/2-\beta+(-x) \vee p_{\bar{U}}+(-y) \vee p_{\bar{V}}}}{n^{\{(-x) \vee (-\beta_U) \vee (-y) \vee (-\beta_V)\}/2}}, \\
\frac{\text{var}_{H_A} W_n}{|E_{H_A} W_n|^2} &= \frac{C_n}{|E_{H_A} W_n|^2} + \frac{C_n}{n^{1-\beta+(-x) \vee p_{\bar{U}}+(-y) \vee p_{\bar{V}}}}.
\end{aligned}$$

When $x + y < 1$,

$$\begin{aligned}
& 1 - \beta + (-x) \vee p_{\bar{U}} + (-y) \vee p_{\bar{V}} \\
& \geq \frac{1}{2} - \beta + (-x) \vee p_{\bar{U}} + (-y) \vee p_{\bar{V}} + \frac{1}{2} \\
& \geq \frac{1}{2} - \beta + (-x) \vee p_{\bar{U}} + (-y) \vee p_V^+ + \frac{x \wedge \beta_U \wedge y \wedge \beta_V}{2},
\end{aligned}$$

so (13) is a sufficient condition for there to exist a $(u', v') \in \mathcal{Q}_4$ such that (18) and (19) hold. \square

12.1 Proof of Theorem 3

It must be shown that the interior of the region where (10)–(13) are all false corresponds to the region where (6)–(8) are all true. The stochastic ordering condition implies that $F_n^U \{(F_0^U)^{-1}(n^{-x})\} \leq n^{-x}$ and $F_n^U \{(F_0^U)^{-1}(1 - n^{-x})\} \leq 1 - n^{-x}$ for $x \leq \log_n 2$. Using Lemma 5, this implies that $p_U^- \leq p_U^+$, where $p_U^-(x)$ and $p_U^+(x)$ are defined as in Theorem 2. Similarly, $p_V^- \leq p_V^+$, and therefore

$$p_U^+(x) = p_U^+(x) \vee p_U^-(x) = \operatorname{ess\,sup}_{a \geq x} \{\alpha_U^+(a) - a\} \vee \operatorname{ess\,sup}_{a \geq x} \{\alpha_U^-(a) - a\} = \operatorname{ess\,sup}_{a \geq x} \{\alpha_U(a) - a\},$$

$$p_V^+(y) = \operatorname{ess\,sup}_{b \geq y} \{\alpha_V(b) - b\},$$

where $\alpha_U(a)$ and $\alpha_V(b)$ are defined as in Theorem 1. These results also imply that the supremum term in (10) is larger than the supremum terms in (11)–(13). It remains only to show that (6)–(8) are true in the interior of the complement of (10). A useful lemma is Proposition 3.5 of Phu & Hoffmann (1996), which states that the supremum and essential supremum with respect to the Lebesgue measure are equal for lower semi-continuous functions.

First, suppose there exist positive x, y such that $x + y < 1$ and $p_U^+(x) > -x, p_V^+(y) > -y$. The interior of the complement of (10) is

$$0 > \operatorname{ess\,sup}_{\substack{x, y > 0, \\ x + y < 1}} \left[\frac{1}{2} - \beta + p_U^+(x) + p_V^+(y) + \frac{x \wedge \{\beta_U - p_U^+(x)\}}{2} + \frac{y \wedge \{\beta_V - p_V^+(y)\}}{2} \right]$$

$$= \operatorname{ess\,sup}_{\substack{x, y > 0, \\ x + y < 1}} \left[\operatorname{ess\,sup}_{a \geq x} \left\{ \alpha_U(a) - a + \frac{x}{2} \right\} \wedge \operatorname{ess\,sup}_{a \geq x} \left\{ \frac{\alpha_U(a) - a + \beta_U}{2} \right\} + \right.$$

$$\left. \operatorname{ess\,sup}_{b \geq y} \left\{ \alpha_V(b) - b + \frac{y}{2} \right\} \wedge \operatorname{ess\,sup}_{b \geq y} \left\{ \frac{\alpha_V(b) - b + \beta_V}{2} \right\} \right] + \frac{1}{2} - \beta.$$

Using Lemma 7, the essential supremum term equals

$$\operatorname{ess\,sup}_{\substack{x, y > 0, x + y < 1, \\ a \geq x, b \geq y}} \left[\left\{ \alpha_U(a) - a + \frac{x}{2} + \alpha_V(b) - b + \frac{y}{2} \right\} \wedge \left\{ \alpha_U(a) - a + \frac{x}{2} + \frac{\alpha_V(b) - b + \beta_V}{2} \right\} \wedge \right.$$

$$\left. \left\{ \frac{\alpha_U(a) - a + \beta_U}{2} + \alpha_V(b) - b + \frac{y}{2} \right\} \wedge \left\{ \frac{\alpha_U(a) - a + \beta_U}{2} + \frac{\alpha_V(b) - b + \beta_V}{2} \right\} \right],$$

and using Lemma 7 again and taking the essential suprema with respect to x and y gives

$$\operatorname{ess\,sup}_{a,b \geq 0} \left[\left\{ \alpha_U(a) - a + \alpha_V(b) - b + \frac{(a+b) \wedge 1}{2} \right\} \wedge \left\{ \alpha_U(a) - a + \frac{a \wedge 1}{2} + \frac{\alpha_V(b) - b + \beta_V}{2} \right\} \wedge \left\{ \frac{\alpha_U(a) - a + \beta_U}{2} + \alpha_V(b) - b + \frac{b \wedge 1}{2} \right\} \wedge \left\{ \frac{\alpha_U(a) - a + \beta_U}{2} + \frac{\alpha_V(b) - b + \beta_V}{2} \right\} \right].$$

Therefore one of the following holds almost everywhere on $a, b \geq 0$:

1.

$$\begin{aligned} 0 &> 1 - 2\beta + 2\alpha_U(a) - 2a + 2\alpha_V(b) - 2b + a + b \\ &\geq 1 - 2\beta + \alpha_U(a) + \{\alpha_U(a) \wedge \beta_U\} + \{\alpha_V(b) \wedge \beta_V\} - a - b; \end{aligned}$$

2.

$$\begin{aligned} 0 &> 1/2 - \beta + \alpha_U(a) - a + \alpha_V(b) - b + 1/2 \\ &= 1 - \beta + \alpha_U(a) + \alpha_V(b) - a - b; \end{aligned}$$

3.

$$\begin{aligned} 0 &> 1 - 2\beta + 2\alpha_U(a) - 2a + a + \alpha_V(b) - b + \beta_V \\ &\geq 1 - 2\beta + \alpha_U(a) + \{\alpha_U(a) \wedge \beta_U\} - a + \alpha_V(b) - b + \{\alpha_V(b) \wedge \beta_V\}; \end{aligned}$$

4.

$$\begin{aligned} 0 &> 1/2 - \beta + \alpha_U(a) - a + 1/2 + \{\alpha_V(b) - b + \beta_V\}/2 \\ &\geq 1 - \beta + \alpha_U(a) - a + \{\alpha_V(b) - b\}/2 + 0 \\ &\geq 1 - \beta + \alpha_U(a) - a + \alpha_V(b) - b, \end{aligned}$$

where the last inequality follows from Lemma 6;

5.

$$\begin{aligned} 0 &> 1 - 2\beta + \alpha_U(a) - a + \beta_U + 2\alpha_V(b) - 2b + b \\ &\geq 1 - 2\beta + \alpha_U(a) - a + \{\alpha_U(a) \wedge \beta_U\} + \alpha_V(b) + \{\alpha_V(b) \wedge \beta_V\} - b; \end{aligned}$$

6.

$$\begin{aligned} 0 &> 1/2 - \beta + \{\alpha_U(a) - a + \beta_U\}/2 + \alpha_V(b) - b + 1/2 \\ &\geq 1 - \beta + \alpha_U(a) - a + \alpha_V(b) - b, \end{aligned}$$

following the same reasoning as in possibility 4; or

7.

$$\begin{aligned} 0 &> 1 - 2\beta + \alpha_U(a) - a + \beta_U + \alpha_V(b) - b + \beta_V \\ &\geq 1 - 2\beta + \alpha_U(a) - a + \{\alpha_U(a) \wedge \beta_U\} + \alpha_V(b) + \{\alpha_V(b) \wedge \beta_V\} - b. \end{aligned}$$

Furthermore, the interior of the complement of (10) contains

$$\begin{aligned} 0 &> \operatorname{ess\,sup}_{\substack{x,y>0, \\ x+y<1}} \left[1 - 2\beta + p_U^+(x) + \frac{x \wedge \{\beta_U - p_U^+(x)\}}{2} \right], \\ 0 &> \operatorname{ess\,sup}_{\substack{x,y>0, \\ x+y<1}} \left[1 - 2\beta + p_V^+(y) + \frac{y \wedge \{\beta_V - p_V^+(y)\}}{2} \right]. \end{aligned}$$

Similar reasoning as in possibility 1 above shows that these imply

$$\begin{aligned} 0 &\geq 1 - 2\beta + \alpha_U(a) + \{\alpha_U(a) \wedge \beta_U\} - a, \\ 0 &\geq 1 - 2\beta + \alpha_V(b) + \{\alpha_V(b) \wedge \beta_V\} - b. \end{aligned}$$

Thus when there exist positive x, y such that $x + y < 1$ and $p_U^+(x) > -x, p_V^+(y) > -y$, (6)–(8) hold in the interior of the complement of (10).

Next assume there exists $x \in (0, 1)$ such that $p_U^+(x) > -x$, but $p_V^+(y) \leq -y$ for all $y \in (0, 1)$. This implies that $p_V^+(\log_n 2) < -\log_n 2$, so $\alpha_V(b) - b \leq -\log_n 2$ for all $b \geq \log_n 2$, so that (7) is always true and (8) becomes

$$1 + 0 \vee \operatorname{ess\,sup}_{a>0} [(-\beta + \alpha_U) \wedge \{-2\beta + \alpha_U + (\alpha_U \wedge \beta_U)\} - a] < 0.$$

This also implies that in the interior of the complement of (10),

$$\begin{aligned}
0 &> \operatorname{ess\,sup}_{\substack{x,y>0, \\ x+y<1}} \left[\frac{1}{2} - \beta + p_U(x) + \frac{x \wedge \{\beta_U - p_U(x)\}}{2} - \frac{y}{2} \right] \\
&= \operatorname{ess\,sup}_{\substack{x,y>0, \\ x+y<1, \\ a \geq x, b \geq y}} \left[\left\{ \alpha_U(a) - a + \frac{x-y}{2} \right\} \wedge \left\{ \frac{\alpha_U(a) - a + \beta_U - y}{2} \right\} \right] + \frac{1}{2} - \beta \\
&= \operatorname{ess\,sup}_{a \geq 0, b \geq 0} \left[\left\{ \alpha_U(a) - a + \frac{a \wedge 1}{2} \right\} \wedge \left\{ \frac{\alpha_U(a) - a + \beta_U}{2} \right\} \right] + \frac{1}{2} - \beta.
\end{aligned}$$

Therefore one of the following holds almost everywhere on $a, b \geq 0$:

1. $0 > 1 - 2\beta + 2\alpha_U(a) - 2a + a \geq 1 - 2\beta + \alpha_U(a) + \{\alpha_U(a) \wedge \beta_U\} - a$;
2. $0 > 1/2 - \beta + \alpha_U(a) - a + 1/2 \geq 1 - \beta + \alpha_U(a) - a$; or
3. $0 > 1 - 2\beta + \alpha_U(a) - a + \beta_U \geq 1 - 2\beta + \alpha_U(a) + \{\alpha_U(a) \wedge \beta_U\} - a$.

Thus when there exists $x \in (0, 1)$ such that $p_U^+(x) > -x$, but $p_V^+(y) \leq -y$ for all $y \in (0, 1)$, (6)–(8) hold in the interior of the complement of (10). Similar reasoning shows that the same is true when $p_U^+(x) \leq -x$ for all $x \in (0, 1)$ but there exists a $y \in (0, 1)$ such that $p_V^+(y) > -y$.

Finally suppose that $p_U^+(x) \leq -x$ and $p_V^+(y) \leq -y$ for all $x, y > 0, x + y < 1$. Then $\alpha_U(a) - a \leq -\log_n 2$ and $\alpha_V(b) - b \leq -\log_n 2$ for all $a, b \geq \log_n 2$, and (6)–(8) are always true, while the interior of the complement of (10) contains

$$\sup_{\substack{x,y>0, \\ x+y<1}} \left(\frac{1}{2} - \beta - \frac{x}{2} - \frac{y}{2} \right) < 0,$$

which also always holds under weak dependence. \square

13 Proof of Theorem 5

Since

$$\begin{aligned}
\hat{\mathcal{D}}_n &= \sup_{\substack{U_{(1)} < u \leq U_{(n)}, \\ V_{(1)} < v \leq V_{(n)}}} \frac{n^{1/2}}{(\log n)^{1/2}} \frac{|\hat{S}_{UV}(u, v) - \hat{S}_U(u)\hat{S}_V(v)|}{\{\hat{S}_U(u)\hat{S}_V(v) - \hat{S}_U(u)^2\hat{S}_V(v)^2\}^{1/2}} \\
&= \sup_{\substack{-U_{(1)} > u \geq -U_{(n)}, \\ -V_{(1)} > v \geq -V_{(n)}}} \frac{n^{1/2}}{(\log n)^{1/2}} \frac{|\hat{S}_{UV}(-u, -v) - \hat{S}_U(-u)\hat{S}_V(-v)|}{\{\hat{S}_U(-u)\hat{S}_V(-v) - \hat{S}_U(-u)^2\hat{S}_V(-v)^2\}^{1/2}} \\
&= \sup_{\substack{U'_{(1)} \leq u < U'_{(n)}, \\ V'_{(1)} \leq v < V'_{(n)}}} \frac{n^{1/2}}{(\log n)^{1/2}} \frac{|\hat{F}_{U'V'}(u, v) - \hat{F}_{U'}(u)\hat{F}_{V'}(v)|}{\{\hat{F}_{U'}(u)\hat{F}_{V'}(v) - \hat{F}_{U'}(u)^2\hat{F}_{V'}(v)^2\}^{1/2}} = \hat{\mathcal{D}}'_n,
\end{aligned}$$

where $\hat{F}_{U'}$ and $\hat{F}_{V'}$ are the univariate empirical distributions of U'_i and V'_i , it suffices to show that $\hat{\mathcal{D}}'_n = \mathcal{D}_n + o_P(1)$.

Define

$$\hat{C}_n(u, v) = n^{-1} \sum_i I\{\hat{F}_U(U'_i) \leq u, \hat{F}_V(V'_i) \leq v\}$$

to be the empirical copula process, following Deheuvels (1979), and \mathcal{A}_n to be the set $\{i/n\}, i = 1, \dots, (n-1)$. It is clear that

$$\hat{\mathcal{D}}'_n = \max_{u, v \in \mathcal{A}_n} \frac{1}{(\log n)^{1/2}} \frac{n^{1/2} |\hat{C}_n(u, v) - uv|}{(uv - u^2v^2)^{1/2}}.$$

First, it can be shown that the maximum can be replaced by a supremum over all $u, v \in (0, 1)$. Choose $m_n \in \mathcal{A}_n$, and consider the set $\mathcal{B}_n = \{(u, v) : \hat{C}_n(u, v) = m_n\}$. Next define the function

$$f_n(u, v) = \frac{\{\hat{C}_n(u, v) - uv\}^2}{uv - u^2v^2} = \frac{(m_n - uv)^2}{uv - (uv)^2}.$$

Since $uv \in [0, 1]$ and

$$\frac{df_n}{d(uv)} = \frac{(1 - 2m)(uv)^2 + 2m^2(uv) - m^2}{\{(uv) - (uv)^2\}^2} = \frac{(uv - m)\{(1 - 2m)uv + m\}}{\{uv - (uv)^2\}^2},$$

it can be shown that $f_n(u, v)$ is convex in uv . Therefore $f_n(uv)$ is maximized at the boundary of \mathcal{B}_n , on which either $uv = 0$, $uv = 1$, or $u, v \in \mathcal{A}_n$. When $uv < 1/n$, $m_n = 0$, and $\lim_{uv \rightarrow 0} f_n(u, v) = 0$. Similarly when $uv > (n-1)/n$, $m_n = 1$, and $\lim_{uv \rightarrow 1} f_n(u, v) = 0$ as well. Therefore

$$\hat{\mathcal{D}}'_n = \sup_{u, v \in (0, 1)} \frac{1}{(\log n)^{1/2}} \frac{n^{1/2} |\hat{C}_n(u, v) - uv|}{(uv - u^2v^2)^{1/2}}.$$

Intuitively, $\hat{\mathcal{D}}'_n$ should be close to \mathcal{D}_n because $\hat{C}_n(u, v)$ should be close to the bivariate uniform empirical process

$$C_n(u, v) = n^{-1} \sum_i I\{F_{U'}(U'_i) \leq u, F_{V'}(V'_i) \leq v\},$$

and with a change of variables \mathcal{D}_n can be written as

$$\mathcal{D}_n = \sup_{u, v \in (0, 1)} \frac{n^{1/2}}{(\log n)^{1/2}} \frac{|C_n(u, v) - uv|}{(uv - u^2v^2)^{1/2}}.$$

To make this precise, by the triangle inequality

$$|\hat{\mathcal{D}}'_n - \mathcal{D}_n| \leq \sup_{u, v \in (0, 1)} \frac{n^{1/2}}{(\log n)^{1/2}} \frac{|\hat{C}_n(u, v) - C_n(u, v)|}{(uv - u^2v^2)^{1/2}} \equiv R_n.$$

Now let \mathcal{Z}_n be the event that

$$\sup_{u, v \in (0, 1)} (\log n)^{-1/2} n^{1/2} |\hat{C}_n - C_n| = 0.$$

Therefore $R_n = 0 \times I(\mathcal{Z}_n) + R_n \times I(\mathcal{Z}_n^c)$, and it remains to show that $P(\mathcal{Z}_n) \rightarrow 1$.

Bouzebda & Zari (2013) showed that if the bivariate copula function

$$C(u, v) = P\{S_U(U_i) \leq u, S_V(V_i) \leq v\}$$

is twice continuously differentiable on $(0, 1)^2$ and its second-order partial derivatives are all continuous on $[0, 1]^2$, then

$$\begin{aligned} \sup_{u, v \in (0, 1)} n |\hat{C}_n - C_n| &\stackrel{a.s.}{=} \mathbb{K}(u, v, n) - \mathbb{K}(u, 1, n) \frac{\partial C(u, v)}{\partial u} - \mathbb{K}(1, v, n) \frac{\partial C(u, v)}{\partial v} + \\ &O\{n^{1/2-1/8}(\log n)^{3/2}\} \end{aligned}$$

for a suitably constructed Kiefer process $\mathbb{K}(u, v, t) = \mathbb{W}(u, v, t) - C(u, v)\mathbb{W}(1, 1, t)$, where \mathbb{W} is a particular Gaussian process. Under Assumption 1, the conditions on $C(u, v)$ are satisfied by the particular type of dependence (2) studied in this paper. Furthermore, the law of the iterated logarithm for Kiefer processes (Finkelstein et al. 1971) states that there is some constant C such that

$$\frac{\mathbb{K}(u, v, n)}{(2n \log \log n)^{1/2}} \stackrel{a.s.}{\rightarrow} C.$$

These facts imply that $P(\mathcal{Z}_n) \rightarrow 1$, proving the theorem. \square

References

- Andreassen, O. A., Thompson, W. K., Schork, A. J., Ripke, S., Mattingsdal, M., Kelsoe, J. R., Kendler, K. S., O'Donovan, M. C., Rujescu, D., Werge, T. et al. (2013), 'Improved detection of common variants associated with schizophrenia and bipolar disorder using pleiotropy-informed conditional false discovery rate', *PLoS genetics* **9**(4), e1003455.
- Arias-Castro, E., Bubeck, S. & Lugosi, G. (2012), 'Detecting positive correlations in a multivariate sample', *arXiv preprint arXiv:1202.5536* .
- Arias-Castro, E., Bubeck, S., Lugosi, G. et al. (2012), 'Detection of correlations', *The Annals of Statistics* **40**(1), 412–435.
- Arias-Castro, E. & Wang, M. (2013), 'Distribution-free tests for sparse heterogeneous mixtures', *arXiv preprint arXiv:1308.0346* .
- Bhattacharjee, S., Rajaraman, P., Jacobs, K. B., Wheeler, W. A., Melin, B. S., Hartge, P., Yeager, M., Chung, C. C., Chanock, S. J. & Chatterjee, N. (2012), 'A subset-based approach improves power and interpretation for the combined analysis of genetic association studies of heterogeneous traits', *The American Journal of Human Genetics* **90**(5), 821–835.
- Blum, J., Kiefer, J. & Rosenblatt, M. (1961), 'Distribution free tests of independence based on the sample distribution function', *The Annals of Mathematical Statistics* pp. 485–498.
- Bouzebda, S. & Zari, T. (2013), 'Strong approximation of empirical copula processes by gaussian processes', *Statistics* **47**(5), 1047–1063.
- Cai, T. T., Jeng, X. J. & Jin, J. (2011), 'Optimal detection of heterogeneous and heteroscedastic mixtures', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **73**(5), 629–662.
- Cai, T. T., Jin, J. & Low, M. G. (2007), 'Estimation and confidence sets for sparse normal mixtures', *The Annals of Statistics* **35**(6), 2421–2449.

- Cai, T. T. & Sun, W. (2014), ‘Optimal screening and discovery of sparse signals with applications to multistage high-throughput studies’, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* . in revision.
- Cai, T. T. & Wu, Y. (2014), ‘Optimal detection of sparse mixtures against a given null distribution’, *IEEE Trans. Inf. Theory* **60**(4), 2217–2232.
- Cross-Disorder Group of the Psychiatric Genomics Consortium et al. (2013a), ‘Genetic relationship between five psychiatric disorders estimated from genome-wide snps’, *Nature genetics* .
- Cross-Disorder Group of the Psychiatric Genomics Consortium et al. (2013b), ‘Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis’, *Lancet* **381**(9875), 1371.
- Dawson, E., Abecasis, G. R., Bumpstead, S., Chen, Y., Hunt, S., Beare, D. M., Pabial, J., Dibling, T., Tinsley, E., Kirby, S., Carter, D., Papaspyridonos, M., Livingstone, S., Ganskell, R., Löhmußaar, E., Zernant, J., Tõnisson, N., Remm, M., Mägi, R., Puurand, T., vilo, J., Kurg, A., Rice, K., Deloukas, P., Mott, R., Metspalu, A., Bentley, D. R., Cardon, L. R. & Dunham, I. (2002), ‘A first-generation linkage disequilibrium map of human chromosome 22’, *Nature* **418**(6897), 544–548.
- Deheuvels, P. (1979), ‘La fonction de dépendance empirique et ses propriétés. un test non paramétrique d’indépendance’, *Acad. Roy. Belg. Bull. Cl. Sci.(5)* **65**(6), 274–292.
- Donoho, D. & Jin, J. (2004), ‘Higher criticism for detecting sparse heterogeneous mixtures’, *The Annals of Statistics* **32**(3), 962–994.
- Efron, B. (2010), *Large-scale inference: empirical Bayes methods for estimation, testing, and prediction*, Vol. 1, Cambridge University Press.
- Einmahl, J. H. (1996), ‘Extension to higher dimensions of the jaeschke-eicker result on the standardized empirical process’, *Communications in Statistics-Theory and Methods* **25**(4), 813–822.

- Einmahl, J. H. & Mason, D. M. (1985), ‘Bounds for weighted multivariate empirical distribution functions’, *Probability Theory and Related Fields* **70**(4), 563–571.
- Finkelstein, H. et al. (1971), ‘The law of the iterated logarithm for empirical distribution’, *The Annals of Mathematical Statistics* **42**(2), 607–615.
- Goeman, J. J. & Bühlmann, P. (2007), ‘Analyzing gene expression data in terms of gene sets: methodological issues’, *Bioinformatics* **23**(8), 980–987.
- Hall, P. & Jin, J. (2010), ‘Innovated higher criticism for detecting sparse signals in correlated noise’, *The Annals of Statistics* **38**(3), 1686–1732.
- He, X., Fuller, C. K., Song, Y., Meng, Q., Zhang, B., Yang, X. & Li, H. (2013), ‘Sherlock: detecting gene-disease associations by matching patterns of expression QTL and GWAS’, *The American Journal of Human Genetics* **92**(5), 667–680.
- Heller, R., Heller, Y., Kaufman, S. & Gorfine, M. (2014), ‘Consistent distribution-free tests of association between univariate random variables’, *arXiv preprint arXiv:1308.1559*.
- Hoeffding, W. (1948), ‘A non-parametric test of independence’, *The Annals of Mathematical Statistics* pp. 546–557.
- Ingster, Y. I. (1997), ‘Some problems of hypothesis testing leading to infinitely divisible distributions’, *Mathematical Methods of Statistics* **6**(1), 47–69.
- Ingster, Y. I. (2002*a*), ‘Adaptive detection of a signal of growing dimension, i’, *Mathematical Methods of Statistics* **10**, 395–421.
- Ingster, Y. I. (2002*b*), ‘Adaptive detection of a signal of growing dimension, ii’, *Mathematical Methods of Statistics* **11**(1), 37–68.
- Jager, L. & Wellner, J. A. (2007), ‘Goodness-of-fit tests via phi-divergences’, *The Annals of Statistics* **35**(5), 2018–2053.
- Ledwina, T. & Wyłupek, G. (2014), ‘Validation of positive quadrant dependence’, *Insurance: Mathematics and Economics* **56**, 38–47.

- Lehmann, E. E. L. & Romano, J. P. (2005), *Testing statistical hypotheses*, Springer Science+ Business Media.
- Nicolae, D. L., Gamazon, E., Zhang, W., Duan, S., Dolan, M. E. & Cox, N. J. (2010), ‘Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS’, *PLoS Genetics* **6**(4), e1000888.
- Phillips, D. & Ghosh, D. (2013), ‘Testing the disjunction hypothesis using voronoi diagrams with applications to genetics’, *arXiv preprint arXiv:1312.5782*.
- Phu, H. & Hoffmann, A. (1996), ‘Essential supremum and supremum of summable functions: Summable functions’, *Numerical functional analysis and optimization* **17**(1-2), 161–180.
- Reshef, D. N., Reshef, Y. A., Finucane, H. K., Grossman, S. R., McVean, G., Turnbaugh, P. J., Lander, E. S., Mitzenmacher, M. & Sabeti, P. C. (2011), ‘Detecting novel associations in large data sets’, *Science* **334**(6062), 1518–1524.
- Rhinn, H., Fujita, R., Qiang, L., Cheng, R., Lee, J. H. & Abeliovich, A. (2013), ‘Integrative genomics identifies apoe ϵ 4 effectors in alzheimer’s disease’, *Nature* **500**(7460), 45–50.
- Rivals, I., Personnaz, L., Taing, L. & Potier, M.-C. (2007), ‘Enrichment or depletion of a go category within a class of genes: which test?’, *Bioinformatics* **23**(4), 401–407.
- Scaillet, O. (2005), ‘A Kolmogorov-Smirnov type test for positive quadrant dependence’, *Canadian Journal of Statistics* **33**(3), 415–427.
- Storey, J. D. & Tibshirani, R. (2003), ‘Statistical significance for genomewide studies’, *Proceedings of the National Academy of Sciences* **100**(16), 9440–9445.
- Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S. et al. (2005), ‘Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles’, *Proceedings of the National Academy of Sciences of the United States of America* **102**(43), 15545–15550.

- Székely, G. J., Rizzo, M. L. et al. (2009), ‘Brownian distance covariance’, *Ann. Appl. Statist.* **3**(4), 1236–1265.
- Thas, O. & Ottoy, J.-P. (2004), ‘A nonparametric test for independence based on sample space partitions’, *Communications in Statistics-Simulation and Computation* **33**(3), 711–728.
- Ware, J., Petretto, E. & Cook, S. (2013), ‘Integrative genomics in cardiovascular medicine’, *Cardiovascular Research* . in press.
- Xie, J., Cai, T. T. & Li, H. (2011), ‘Sample size and power analysis for sparse signal recovery in genome-wide association studies’, *Biometrika* **98**(2), 273–290.
- Zhao, S. D., Cai, T. T. & Li, H. (2014), ‘Gene-disease associations via sparse simultaneous signal detection’, *Biometrika* . in revision.