1 Introduction

In nearly all areas of high-dimensional statistics distinguishing between the important and unimportant predictors is of intrest. In genetics many candidate snips and their association with some outcome is of interest. Many studies take measurements on a large number of biomarkers to predict medical outcomes such as the presence or progression of a disease. While there has been much focus on finding important predictors, there has been less interest in knowing if any predictors are actually important. The vast number of potential predictors creates new statistical challenges to answering this less obvious question.

In the univariate case, once a measure of association is selected, standard approaches exist to construct asymptotically valid (and sometimes optimal) tests [Neyman Jerzy et al.]. In this setting, a test is considered optimal if it achieves the largest power for every alternative compared to every other test with the same type one error. However, scientists are often interested if any of a multitude of covariate are associated with the outcome of interest. Even in the case with two covariates, a test with the above optimal property no longer exists and the alternative plays a larger role in which tests are more or less powerful.

To illustrate this, consider two tests with equal type one error rates. Both test the null that both of two association measures ψ_1, ψ_2 are zero. The first test focuses only on the ψ_1 and the second test considers both ψ_1 and ψ_2. When the ψ_1 is non-zero ψ_2 is zero the first test can outperform all other tests. Conversely when only ψ_2 is non-zero the first test will be outperformed by the second test. Additionally in this setting, the power of these tests will be tied to the covariance between the covariates (a problem that did not exist for a single covariate). These two factors make it difficult to define a test that performs well across all possible alternatives.

Work with Simultaneous hypothesis testing began with Tukey in 1953 [Miller]. Previous work by Bonferroni was used by [Dunn, a,b] to come up with some of the first multiple hypothesis testing procedures. Further improvements were proposed by [Hochberg, Holm, S. Holland and DiPonzio Copenhaver]. Bonferroni-based correction procedures have the advantage of being easy to apply to already existing tests, while guaranteeing family-wise error control. However, these tests suffer from low power, especially in cases where the probability of rejecting each hypothesis is highly correlated.

Newer procedures [Donoho and Jin] and add other papers here attain improved power compared to Bonferroni-based methods, but often rely on asymptotics to obtain these results, and don't account for the irregularity of the estimator on which their test is based. Subsequent work [McKeague and Qian, Pan et al., Xu et al.] addressed these concerns by accounting for the adaptive nature of the considered tests. Still, these newer tests are restricted to testing a particular hypothesis, must make assumptions about the data-generating mechanism to obtain theoretical guarantees.

In this article a test is proposed that works across a wide variety of data generating mechanisms and parameters of interest, but also achieves comparable power to tailor made procedures. Section 2 describes the data generating mechanisms that are considered to evaluate the performance of the test and the competing test made for the given data generating mechanism. Section 3 proposes the testing procedure. Add other sections here

2 Working Examples

Let $X_1, \ldots, X_n$ be independent identically distributed draws from some distribution P, and let $\boldsymbol{X} = \{X_1, \ldots, X_n\}$. Let $X_i = (Y_i, W_{1i}, \ldots, W_{di}), i \in \{1, \ldots n\}$ where Y is the outcome of interest, and each W is a covariate. Let $\psi_1 = \Psi_1(P), \ldots, \psi_a = \Psi_a(P)$ be measures of association between Y and some combination of the W_j's. While the results found in this article are valid for any integer a, for the remainder of this article assume $a = d$, the number of covariates. Also, let ψ_j correspond to a measure of association between Y and W_j. The null hypothesis for our test will be the strong null:

$$H_0 : \psi_1 = \psi_2 = \cdots = \psi_a = 0 \quad \text{versus} \quad H_1 : \psi_j \neq 0 \text{ for some } j \in \{1, \ldots, a\}.$$

Last, let $\mathcal{M}$ denote the set of all possible distributions. Define $\mathcal{M}_0 = \{P \in \mathcal{M} : H_0 \text{ holds}\}$

2.1 Correlation Parameter

We will compare our method to both a simple bonferroni correction method, and the method described in [Zhang and Laber]. The settings considered will be the same as the firs setting in [McKeague and Qian]. The parameter of interest, $\psi_j(P)$ will be the correlation between the outcome of interest and the j'th covariate.

The vector of covariates in this setting will be generated from a normal distribution with mean zero and a variance covariance of Σ with Σ_{ij} equal to ρ when $i \neq j$ and equal to 1 when $i = j$, and . Three different models are considered. Three different models for the outcome of interest (Y) will be considered. Letting $\varepsilon \sim N(0, 1)$ and be independent of X, the first model let $Y = \varepsilon$, the second has $Y = X_1/4$, and third has $Y = \sum_{k=1}^{1} 0\beta_k X_k + \varepsilon$ where $\beta_k = 0.15$ for $k = \{1, \ldots, 5\}$, and $\beta_k = -0.1$ for $k = \{6 \ldots 10\}$. in which Sample sizes of 100 and 200, dimensions of 10, 50, 100, 150, and 200, and ρ of $0, 0.5$, or 0.8 will be considered,

2.2 Missing Data Example

In the second example, Y is binary, Δ is a missingness indicator, and each W_j is a covariate of interest. When $\Delta = 0$ we don't observe Y. The identifying assumption is $\Delta \perp\!\!\!\perp Y|W$. The parameter of interest is the risk ratio,

$$\Psi_j\left(P^{\text{full}}\right) = \frac{\text{Cov}\left(\log\left(Pr\left(Y=1|W_j\right)\right), W_j\right)}{\text{Var}(W_j)}.$$

Using the identifying assumption, the observed data parameter is:

$$\tilde{\Psi}_j\left(P^{\text{obs}}\right) = \frac{\text{Cov}\left(\log\left(E\left[Pr\left(\Delta Y=1|\Delta=1, W=W\right)|W_j\right]\right), W_j\right)}{\text{Var}(W_j)}$$

Right now, it seems that the time for a single test is quite a bit larger for this test than it is for the correlation test (which is reasonable), though it may speed up if Superlearner can be improved. I am thinking of using similar simulation settings as was used before. There will be a single mechanism for the missingness, but it will be variable if the mechanism is known. There will be three settings for between W correlation. All W will be equally correlated, with $\rho = 0, 0.3$, or 0.7. Dimension will (depending on computation time) be $10, 50, 100$ and possibly 200.

There will be three settings for the β's in the data generating model:

$$\log(Pr(Y=1|W)) = \beta_0 + W^\top \beta$$

In one setting $\beta = 0$. In the second only one or two entries of β will be non-zero. In the last setting, many of the entries of β will be non-zero (possibly 70 or 80 percent of the entries).

2.3 Marginal Structural Model

For this marginal structural model, we are interested in if the average treatment is modified by any covariates. The marginal structural model for each W_j is defined by

$$\text{logit}\left(Pr(Y^{(a)}=1|w)\right) = \beta_0 + \beta_1 a + \beta_2 w_j + \beta_3 w_j a$$

$$(\beta_0^*, \beta_1^*, \beta_2^*, \beta_3^*) = \text{argmin}_{\beta_0,\beta_1,\beta_2,\beta_3} \int \left(\text{logit}\left(Pr(y^{(a)}=1\right) - (\beta_0 + \beta_1 a + \beta_2 w_j + \beta_3 w_j a)\right)^2 dP(y^{(a)}, w, a)$$

The parameter of interest is β_3^* The other W_j's are included in the analysis and are marginalized over, but are not part of the defined model. I am planning on using the same simulation settings as above, but also including non-changing effects for a, but am open to other ideas.

3 Proposed testing procedure

Denote the estimator of the $\boldsymbol{\psi}$ by $\hat{\boldsymbol{\Psi}}(\boldsymbol{X})$ and let $\hat{\boldsymbol{\psi}} \equiv \hat{\boldsymbol{\Psi}}(x)$ be an estimate of $\boldsymbol{\psi}$. Suppose that $\sqrt{n}\hat{\boldsymbol{\psi}}$ converges in law to some distribution $Q(P)$ when $P \in \mathcal{M}_0$. Any test of $P \in \mathcal{M}_0$ can be characterized by an acceptance region $\Theta_0(P) \subset \mathbb{R}^d$. The region $\Theta_0(P)$ can be chosen so the probability of rejection under the null is controlled asymptotically:

$$Pr(Z \in \Theta_0(P)) = 1 - \alpha \text{ for every } P \in M_0 \text{ where } Z \sim Q(P). \tag{1}$$

While there are infinitely many regions satisfying (1), we focus on a class of regions defined using ℓ_p norms. For simplicity, first consider regions defined using an ℓ_2 norm:

$$\Theta_0(r) = \{\omega : ||\omega||_2 \leq r\}$$

A region satisfying (1) has a radius:

$$r_\alpha(P) = \min\left\{r : Pr_{Q(P)}(||Z||_2 \leq r) \geq 1 - \alpha\right\}.$$

By constraining the possible regions we consider, we can now define our test in a simple algebraic form:

$$\text{reject } H_0 \text{ if } ||\sqrt{n}\hat{\boldsymbol{\psi}}||_2 \geq r_\alpha.$$

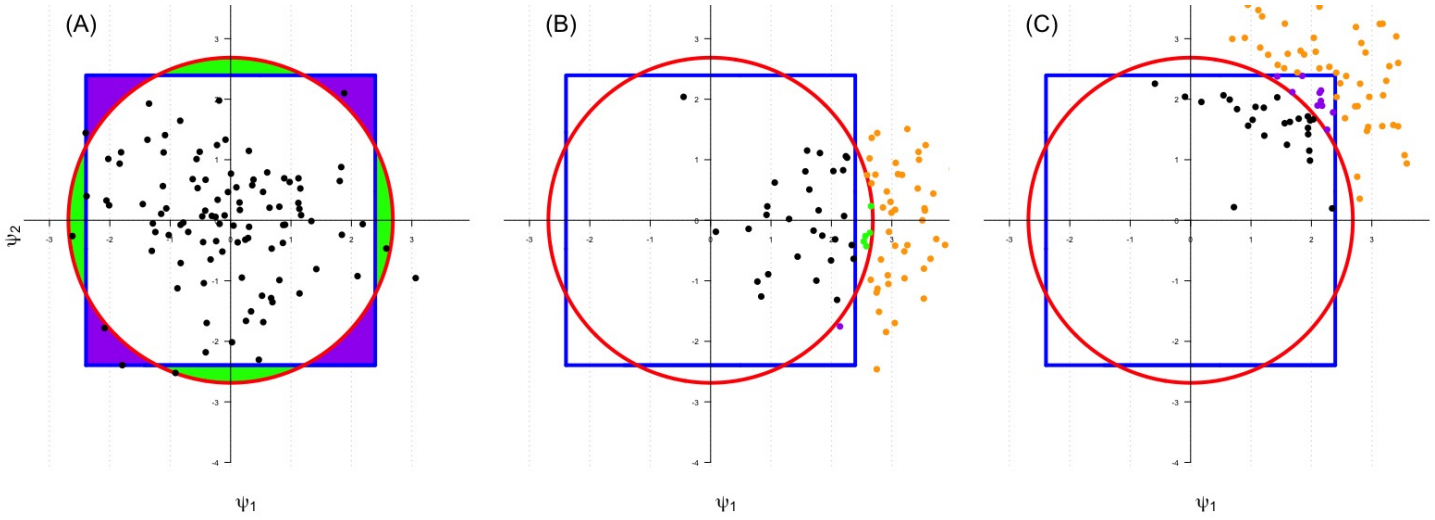

Figure 1: Plots of 100 observations from a limiting distribution of a hypothetical vector of parameter estimators in $\mathbb{R}^2$ (A) under the null, (B) under an alternative with $\psi_1 = 0, \psi_2 \neq 0$, and (C) under an alternative with $\psi_1, \psi_2 \neq 0$. The 95% quantiles for the data based on the max (blue) and ℓ_2 (red) norms under the null are given in all three panels. If a test statistic fell within the purple regions the test would fail to reject H_0 if the ℓ_∞ norm was used, but would reject H_0 if the ℓ_2 norm was used. The converse is true for the green regions. Depending on the alternative, the ℓ_∞ norm (B) or the ℓ_2 norm(C) will achieve higher power.

Similarly, p-values can now be simply defined as $Pr(||Z||_2 \geq ||\sqrt{n}\hat{\boldsymbol{\psi}}||_2)$.

Above we used a norm to constrain the considered rejection regions and allow for a test based on the value of $\sqrt{n}\hat{\boldsymbol{\psi}}$ under the norm and the distribution of Q under the norm. However, this approach will also work with more complicated functions. Consider instead of norm, some function $\Gamma_Q(x)$ that maps from a value and distribution in $\mathbb{R}^d$ to $\mathbb{R}$. The distribution of $\Gamma_Q(Z)$ can be compared to $\Gamma_Q(\sqrt{n}\hat{\boldsymbol{\psi}})$ to obtain p-values and define a test. Consider, for example:

$$\Gamma_Q(x) = Pr_Q(||Z^* + x|| > c_{0.8}) \text{ where } c_{0.8} \equiv \min_c\{c : Pr_Q(||Z|| < c) \geq 0.8\} \text{ and } Z^* \stackrel{d}{=} Z.$$

In this scenario, the test can be carried out by comparing our test statistic, $\Gamma_Q(\sqrt{n}\hat{\boldsymbol{\psi}}, ||\cdot||)$ to $\Gamma_Q(Z, ||\cdot||)$.

While this method for defining tests allows for many different specifications, it often requires users chooses a norm as part of their definition of Γ. Figure 1 shows that the power a test achieves can depend on the norm chosen. Consider the case when $d = 2$ and 100 draws are taken from Z. Cutoffs can be obtained by transforming these data (using the l_p norm) and taking the 95% quantile of this transformed data. The perimeter of the acceptance region when using the ℓ_2 and ℓ_∞ norm are given in red and blue respectively. If $\sqrt{n}\hat{\boldsymbol{\psi}}$ falls within any of the purple regions, using the ℓ_∞ norm would result in failing to reject H_0 and using the ℓ_2 norm would result in rejecting H_0. The converse is true for the green regions.

3.1 Adaptive selection of a norm

In the previous section it was shown that a test can be defined by choosing a summarizing function and norm. However, the choice of norm influences the power of the test. In most scenarios it will not be clear a priori which norm will have maximal power for the true alternative. The procedure proposed in this section adaptively selects a norm and it will be shown that this procedure achieves greater power than a test with a fixed norm, while maintaining type 1 error control.

To achieve this, first define $\Gamma_{1,Q}(x), \ldots, \Gamma_{p,Q}(x)$ as collection of functions in which the functions only differ by the norm used in their definition. Next, define

$$\Gamma^*_Q(x) = \max\left\{\Gamma_{1,Q}(x), \Gamma_{2,Q}(x), \ldots, \Gamma_{p,Q}(x)\right\}.$$

This will could be the min if smaller values indicate larger distances away from the norm (such as p-values).

While it may be difficult to obtain the exact distribution of $\Gamma^*_Q(Z)$ in practice, the distribution is a function of Q, so obtaining very good approximations of $\Gamma^*_Q(Z)$ is possible by taking many draws from Z. This technique can be used both to estimate the function Γ and to estimate $\Gamma^*_Q(Z)$.

3.2 Obtaining the null distribution

The above procedure requires knowledge of the limiting distribution of $\sqrt{n}\hat{\boldsymbol{\psi}}$ when $P \in \mathcal{M}_0$. To obtain an estimate of this limiting distribution, assume the vector of parameter estimates $\hat{\boldsymbol{\psi}}$, converges to a normal distribution with an estimable

variance covariance matrix when properly centered and normalized. Also assume each of the estimators $\hat{\boldsymbol{\psi}}_1, \ldots, \hat{\boldsymbol{\psi}}_d$ of $\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_d$ is asymptotically linear. That is for each $j \in \{1, \ldots, d\}$:

$$\hat{\psi}_j = \psi_j + \frac{1}{n}\sum_{i=1}^{n} D_j(\boldsymbol{x}_i) + o_p(1/\sqrt{n}) \text{ for some function } D_j$$

When there is a fixed number of covariates, the Cramer-Wold device can be used to show that the vector of parameter estimates is asymptotically normal with mean zero, and variance covariance matrix given by $\Sigma = E_{P_0}\left[D(X)^\top D(X)\right]$:

$$\sqrt{n}\left(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}\right) \xrightarrow{d} Z \sim N\left(0, \Sigma\right)$$

Under H_0, $\sqrt{n}\hat{\boldsymbol{\psi}}$ converges to Z, and Σ can be approximates with $\widehat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n} D(\boldsymbol{x}_i)^\top D(\boldsymbol{x}_i)$. Thus, in practice we will use an estimate of Q, $\hat{Q}$ which is distributed $N(0, \widehat{\Sigma})$ in place of Q, and the random variable $\hat{Z}$ with distribution $\hat{Q}$ in place of Z. Thus, the test statistic will be $\Gamma_{\hat{Q}}(\sqrt{n}\hat{\boldsymbol{\psi}})$ and will be compared to $\Gamma_{\hat{Q}}(\hat{Z})$.

3.3 Using a permutation test for the test statistic

While the above approach will work asymptotically, there can be issues for small sample sizes. To avoid inflated type one error, a permutation based test can be used. Here, $\hat{Q}^{\#}$ is used to define Γ, and the test will compare $\Gamma_{\hat{Q}}(\sqrt{n}\hat{\boldsymbol{\psi}})$ to $\Gamma_{\hat{Q}}(Z^{\#})$. To determine $\hat{Q}^{\#}$, the Y's from the observed data are permuted before calculating $\widehat{\Sigma}$. Draws from $Z^{\#}$ are taken by permuting all of the Y's of the observed data. There are a few more complications here that need to be figured out.

3.4 Summary of how simulations are run

4 Simulation Study

4.1 Correlation

1. Two-Phase sampling Scheme of correlation
2. Marginal structural model example.

All three examples will come with simulation results in the form of figures. Not sure how many figures to use or what sample size(s) to use.

5 Data Application

1. Data from Peter Gilbert

6 Discussion

7 Appendix

7.1 Test Consistency

Theorem 1. *Assume the performance metric of choice is $\hat{r}_{\alpha,p}(a)$, and each the norms g_a considered have the following properties:*

$$\text{for } x \in \mathbb{R}^d, \text{ and } s, l > 0, s \cdot \max(\boldsymbol{x}) \leq g_a(\boldsymbol{x}) \leq l \cdot d \cdot \max(\boldsymbol{x}) \tag{2}$$

$$\text{for } s \leq 1, l \geq 1, g_a(s \cdot \boldsymbol{x}) \leq g_a(\boldsymbol{x}) \leq g_a(l \cdot \boldsymbol{x}) \tag{3}$$

Then for all $P \notin \mathcal{M}_0$

$$Pr\left(\frac{1}{B}\sum_{k=1}^{B} I\left\{\hat{T} \leq \hat{T}_k^{\#}\right\} < 0.05\right) \xrightarrow{p} 1 \text{ as } n \to \infty$$

or Equivalently

$$Pr\left(\hat{T} \leq F^{-1}_{\hat{T}^{\#}}(0.05)\right) \xrightarrow{p} 1 \text{ as } n \to \infty$$

Proof. This proof will consist of three parts. We will first show that as $n \to \infty, \hat{T}_a^{\#}$ becomes bounded away from 0 for any valid norm. Next we will show $\hat{T}_a$ converges to 0 in probability for any valid norm. Last we will show the two previous findings imply theorem 1.

Let $P_X^{\#}$ denote the distribution of the randomly permuted observations. Because $Pr(Y_i^{\#} \perp\!\!\!\perp \boldsymbol{W}_i^{\#}) \to 0$, $\boldsymbol{\Psi}(P_X^{\#}) = \mathbf{0}$. Additionally, $\sqrt{n}\left(\hat{\boldsymbol{\psi}}^{\#} - \boldsymbol{\psi}^{\#}\right) \xrightarrow{d} Z^{\#} \sim N(\mathbf{0}, \Sigma_{\text{perm}})$. Define $\hat{T}_a^{\#} = \min_s\{s : g_a(s \cdot \sqrt{n}\hat{\boldsymbol{\psi}}^{\#}) \geq C^{\#}_{0.95,a}\}$ and $C^{\#}_{0.95,a}$ is $F^{-1}_{Z^{\#}}(0.95)$. Since we know that $\psi^{\#} = \mathbf{0}$, it follows that $\sqrt{n}\hat{\boldsymbol{\psi}}^{\#} \overset{d}{\approx} Z^{\#}$.

Now, consider:

$$\begin{aligned}
\Pr\left(\hat{T}_a^{\#} > \epsilon\right) &= \Pr\left(g_a\left(\epsilon \cdot \sqrt{n}\hat{\boldsymbol{\psi}}^{\#}\right) \leq C_{0.95,a}\right) \\
&\geq \Pr\left(\epsilon \cdot d \cdot \max\left(\sqrt{n}\hat{\boldsymbol{\psi}}^{\#}\right) \leq C_{0.95,a}\right) \\
&= \Pr\left(\max\left(\sqrt{n}\hat{\boldsymbol{\psi}}^{\#}\right) \leq C_{0.95,a}/(\epsilon \cdot d)\right)
\end{aligned}$$

Because $\max\left(\left|\sqrt{n}\hat{\boldsymbol{\psi}}^{\#}\right|\right)$ converges to a well defined, positive distribution as a result of the continuous mapping theorem, for each constant $c < 1$, we know there exists an ϵ_c such that $\Pr\left(\hat{T}_a^{\#} > \epsilon_c\right) \geq c$.

Now, shifting our focus to $\hat{T}_a$, under alternatives, $\boldsymbol{\psi} \neq \mathbf{0}$. Define $\psi_{\max} = \max(\psi_1, \ldots, \psi_d)$. Using this knowledge, and (2), note that

$$\begin{aligned}
\Pr\left(\hat{T}_a < \epsilon\right) &= \Pr\left(g_a\left(\epsilon \cdot \sqrt{n}\hat{\boldsymbol{\psi}}\right) \geq C_{0.95,a}\right) \\
&\geq \Pr\left(\epsilon \cdot \sqrt{n}\max(\hat{\psi}_1, \ldots, \hat{\psi}_d) \geq C_{0.95,a}\right) \\
&= \Pr\left(\max(\hat{\psi}_1, \ldots, \hat{\psi}_d) \geq C_{0.95,a}/\left(\epsilon \cdot \sqrt{n}\right)\right) \\
&\geq \Pr\left(\max(\hat{\psi}_1, \ldots, \hat{\psi}_d) \geq \psi_{\max}/2\right) \Pr\left(\psi_{\max}/2 \geq C_{0.95,a}/\left(\epsilon \cdot \sqrt{n}\right)\right) \qquad (4)
\end{aligned}$$

The first factor of the product in (4) will converge to 1 as $n \to \infty$ from the consistency of $\hat{\boldsymbol{\psi}}$. The second quantity will be equal to 1 for sufficiently large n. Thus $\hat{T}_a \xrightarrow{p} 0$ under any alternative.

It was shown that for each a that $\hat{T}_a \xrightarrow{p} 0$. This means that our adaptive estimator $\hat{T} \xrightarrow{p} 0$ as well. Now, let $c = 0.05/k$ and ϵ_c be small enough that $\Pr\left(\hat{T}_a^{\#} > \epsilon_c\right) \geq 1 - (0.05/k)$. The permutation version of the adaptive estimator $\hat{T}^{\#}$ has the property that

$$\Pr\left(\hat{T}^{\#} < \epsilon_c\right) \leq \Pr(\hat{T}_1^{\#} < \epsilon_c) + \cdots + \Pr(\hat{T}_k^{\#} < \epsilon_c) \leq 0.05,$$

and the theorem's conclusion follows. □

7.2 Unbiasedness at local alternatives

Show that the test will reject the null with a probability greater than α for an α level test Proof Outline:

- Under local alternatives, $\sqrt{n}\hat{\boldsymbol{\psi}} \xrightarrow{d} N(\boldsymbol{c}, \Sigma)$
- The permutation test statistic will converge in distribution to a standard normal. This paper should help: [Omelka and Pauly]
- For each norm selected, [Gupta et al.] states that the power will be non-decreasing as long as the rejection region is convex (This should be true most of our rejection regions), and the probability density is decreasing away from the mean (which is true of a normal distribution).
- Show that for n large, the norm is selected to give the best power. Thus since each norm obtains local power, the adaptive test will also obtain local power for n large.

7.3 Consistency of norm selection

This proof would likely be difficult and require some slower than $\sqrt{n}$ covergence rates of the local alternative considered. Under fixed alternatives, all norms perform equally perfectly

7.4 Type 1 error control

Proof is so short, I am not sure if it is worth including Under the null, $Y \perp\!\!\!\perp X$. Thus the test statistic will be taken from the same distribution as all of the permutation based test statistics used to estimate the distirbution of the test statistic under the null. Therefore, as B grows, $Pr(T_n \geq F_{T^{\#}_{k,n}|X_n}(0.95)) \Rightarrow 0.05$

References

David Donoho and Jiashun Jin. Higher criticism for detecting sparse heterogeneous mixtures. 32(3):962–994. ISSN 0090-5364, 2168-8966. doi: 10.1214/009053604000000265. URL https://projecteuclid.org/euclid.aos/1085408492.

Olive Jean Dunn. Estimation of the medians for dependent variables. 30(1):192–197, a. ISSN 0003-4851, 2168-8990. doi: 10.1214/aoms/1177706374. URL https://projecteuclid.org/euclid.aoms/1177706374.

Olive Jean Dunn. Multiple comparisons among means. 56(293):52–64, b. ISSN 0162-1459. doi: 10.1080/01621459.1961.10482090. URL https://www.tandfonline.com/doi/abs/10.1080/01621459.1961.10482090.

S. Das Gupta, M. L. Eaton, I. Olkin, M. Perlman, L. J. Savage, and M. Sobel. Inequalitites on the probability content of convex regions for elliptically contoured distributions. The Regents of the University of California. URL https://projecteuclid.org/euclid.bsmsp/1200514222.

Yosef Hochberg. A sharper bonferroni procedure for multiple tests of significance. 75(4):800–802. ISSN 0006-3444. doi: 10.1093/biomet/75.4.800. URL https://academic.oup.com/biomet/article/75/4/800/423177.

Sture Holm. A simple sequentially rejective multiple test procedure. 6:65–70. doi: 10.2307/4615733.

Ian W. McKeague and Min Qian. An adaptive resampling test for detecting the presence of significant predictors. 110(512):1422–1433. ISSN 0162-1459. doi: 10.1080/01621459.2015.1095099. URL https://amstat.tandfonline.com/doi/abs/10.1080/01621459.2015.1095099.

Rupert G. Jr Miller. *Simultaneous Statistical Inference*. Springer Series in Statistics. Springer-Verlag, 2 edition. ISBN 978-1-4613-8124-2. URL https://www.springer.com/la/book/9781461381242.

Neyman Jerzy, Pearson Egon Sharpe, and Pearson Karl. IX. on the problem of the most efficient tests of statistical hypotheses. 231(694):289–337. doi: 10.1098/rsta.1933.0009. URL https://royalsocietypublishing.org/doi/abs/10.1098/rsta.1933.0009.

M. Omelka and M. Pauly. Testing equality of correlation coefficients in two populations via permutation methods. 142(6):1396–1406. ISSN 0378-3758. doi: 10.1016/j.jspi.2011.12.018. URL http://www.sciencedirect.com/science/article/pii/S0378375811004630.

Wei Pan, Junghi Kim, Yiwei Zhang, Xiaotong Shen, and Peng Wei. A powerful and adaptive association test for rare variants. 197(4):1081–1095. ISSN 0016-6731, 1943-2631. doi: 10.1534/genetics.114.165035. URL https://www.genetics.org/content/197/4/1081.

Burt S. Holland and Margaret DiPonzio Copenhaver. Improved bonferroni-type multiple testing procedures. 104:145–149. doi: 10.1037/0033-2909.104.1.145.

Gongjun Xu, Lifeng Lin, Peng Wei, and Wei Pan. An adaptive two-sample test for high-dimensional means. 103(3):609–624. ISSN 0006-3444. doi: 10.1093/biomet/asw029. URL https://academic.oup.com/biomet/article/103/3/609/1744173.

Yichi Zhang and Eric B. Laber. Comment. 110(512):1451–1454. ISSN 0162-1459. doi: 10.1080/01621459.2015.1106403. URL https://amstat.tandfonline.com/doi/full/10.1080/01621459.2015.1106403.