Cauchy combination test: a powerful test with analytic p-value calculation under arbitrary dependency structures

Yaowu Liu* and Jun Xie†

Abstract

Combining individual p-values to aggregate multiple small effects has a long-standing interest in statistics, dating back to the classic Fisher's combination test. In modern large-scale data analysis, correlation and sparsity are common features and efficient computation is a necessary requirement for dealing with massive data. To overcome these challenges, we propose a new test that takes advantage of the Cauchy distribution. Our test statistic has a very simple form and is defined as a weighted sum of Cauchy transformation of individual p-values. We prove a non-asymptotic result that the tail of the null distribution of our proposed test statistic can be well approximated by a Cauchy distribution under arbitrary dependency structures. Based on this theoretical result, the p-value calculation of our proposed test is not only accurate, but also as simple as the classic z-test or t-test, making our test well suited for analyzing massive data. We further show that the power of the proposed test is asymptotically optimal in a strong sparsity setting. Extensive simulations demonstrate that the proposed test has both strong power against sparse alternatives and a good accuracy with respect to p-value calculations, especially for very small p-values. The proposed test has also been applied to a genome-wide association study of Crohn's disease and compared with several existing tests.

Keywords: Cauchy distribution; Correlation matrix; Non-asymptotic approximation; High dimensional data; Global hypothesis testing; Sparse alternative.

1 Introduction

Methods for combining individual p-values or test statistics are of historically substantial interest in statistics. A few well-known classical methods include the Fisher's combination test (Fisher, 1932) and the sum-of-squares type tests. However, in modern high-throughput data analysis where there is only a small fraction of significant effects, these traditional tests are ineffective and can have substantial power loss (Koziol and Perlman, 1978; Arias-Castro et al., 2011). For example, in genome-wide association studies (GWAS), massive amounts of genetic variants, e.g., single nucleotide polymorphism (SNP), are collected while only a small number of them are expected to be related to the phenotype of interest (e.g., a disease status). Various methods have been developed

*Postdoctoral fellow, Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA 02115
†Professor, Department of Statistics, Purdue University, West Lafayette, IN 47906. Research Sponsored by the National Institutes of Health Grant R21GM101504.

to improve power for detecting sparse alternatives in this situation. The Tippett's minimum p-value test (Tippett, 1931), the higher criticism test (Donoho and Jin, 2004), and the Berk-Jones test (Berk and Jones, 1979) are particularly popular and have received substantial attention in the literature. As all three tests combine individual p-values to aggregate multiple effects, we will also refer to them as combination tests hereafter.

In practice, the individual test statistics or p-values are often correlated. For instance, SNPs could be highly correlated due to linkage disequilibrium. In order to control the Type I error and draw valid statistical inferences, the correlation structure should be taken into account in the p-value calculation. Here and throughout this paper, by p-value calculation, we always mean the p-value of a combination test, rather than the individual p-values that can be readily calculated. To the best of our knowledge, no analytic methods are available for the p-value calculation of the Tippett's minimum p-value, higher criticism, and Berk-Jones tests under dependence structures. While permutation or other approaches based on numerical simulations (e.g., Liu and Xie, 2018) can be used to incorporate the correlation information, they are computationally burdensome or even at times intractable in the analysis of massive data, especially in the following situations. First, when a combination test needs to be performed numerous times, such as in large-scale multiple testing, it is too time consuming to use the permutation approach. Here and throughout this paper, by large-scale multiple testing, we always mean a large number of combination tests instead of individual tests. Second, when the p-value of a combination test is extremely small, permutation would require very intensive computation as a vast number of simulations are needed to stabilize the calculation. This situation is particularly important in large-scale multiple testing, where practitioners care about the validity of extremely small p-values. Third, when the number of individual p-values in a combination test, denoted by d, is very large (e.g., $d = 10^6$), the permutation approach is also impractical. Therefore, there is an increasing demand for developing tests whose p-values can be calculated analytically under dependence. Recently, Barnett et al. (2017) generalized the higher criticism test to incorporate the dependency structure and provided an analytic approximation method to compute the p-value of their new test. However, their analytic method is not accurate for extremely small p-values and still requires very intensive computation, even for a moderately large d (e.g., $d > 100$). In summary, due to a lack of computational efficiency and accuracy of p-values, none of the existing tests can handle large-scale data effectively.

Our main motivating examples of these challenging situations are also from GWAS, where the genetic data could contain hundreds of thousands of subjects and millions of SNPs and fast

computation is a necessary requirement for the analysis of such big data. One commonly used analysis approach in GWAS is to perform set-based analysis (Wu et al., 2010), which divides the SNPs into sets/groups (e.g., genes) based on some biological information and tests the association between each SNP-set and the phenotype one at a time. The combination tests are useful for testing the significance of each SNP-set by aggregating the p-values of individual SNPs. In each SNP-set, the number of SNPs is not very large and often in dozens. However, there are tens of thousands of sets which need to be tested, requiring a fast calculation of the p-value for each SNP-set. In addition, a stringent significance threshold needs to be applied to account for multiple testing and the significant SNP-sets, which are the primary interest of practitioners, have very small p-values (e.g., $< 10^{-6}$). Therefore, the p-value of a combination test needs to be accurate for exceedingly small p-values. Furthermore, one may be also interested in inferring whether there is any overall effect in the whole genome with millions of SNPs all together. This corresponds to the high-dimensional situation with a very large d.

In this article, we propose a new combination test based on the Cauchy distribution and refer to it as the Cauchy combination test. Similar to the Fisher's combination test, the new test statistic is defined as the weighted sum of transformed p-values (Xie et al., 2011; Xie and Singh, 2013), except that the p-values are transformed to follow a standard Cauchy distribution. We prove that the tail of the null distribution of this test statistic is approximately Cauchy under arbitrary correlation structures. According to this theoretical result, we then propose to calculate the p-value of the Cauchy combination test by the cumulative distribution function (c.d.f.) of a standard Cauchy distribution. Similar to the classic z-test or t-test, our test has low computational requirements for p-value calculation and therefore is (potentially) able to be used routinely in large-scale data analysis with a vast number of combination tests. We also establish similar theoretical result for the high-dimensional situation where the number of p-values d diverges. An extensive simulation study is carried out in Section 4, which shows that under general correlation structures, the analytic p-value approximation by the Cauchy distribution is very accurate, especially for extremely small p-values. In fact, the smaller the p-value, the more accurate the approximation. In addition, parallel to the optimality theory for the minimum p-value test shown in Arias-Castro et al. (2011), we prove that the power of our test is asymptotically optimal in a strong sparsity setting. In summary, the Cauchy combination test is well suited to deal with the challenges posed by sparsity, correlation, high-dimensionality and large scales, which, for example, are the situations we encountered in GWAS.

A related and more profound theory regarding the Cauchy distribution was also established in a recent work by Pillai and Meng (2016), which showed a remarkable result that the sum of some class of dependent Cauchy variables could be exactly Cauchy distributed. Our idea of using the Cauchy distribution was motivated from the strong need in GWAS for computationally scalable methods, and was originated from the observation that the sum of independent standard Cauchy variables follows the same distribution as the sum of perfectly dependent standard Cauchy variables. We provide a detailed discussion on the connections and differences between Pillai and Meng (2016)'s and our work in Section 2.2.

The rest of the paper is organized as follows. Section 2 presents our main theorems about the null distribution of the Cauchy combination test statistic. In Section 3, we establish the asymptotical optimality of the power of the new test in the strong sparsity setting. In Section 4, we conduct extensive simulations to evaluate the accuracy of the p-value calculation of the proposed test and compare its power with a few existing tests. We use an analysis of GWAS data to demonstrate the effectiveness of our test. Some concluding remarks and a discussion of future research are given in Section 5. All the technical proofs and additional simulation results are relegated to the supplementary material.

2 Null distribution

Let p_i be the individual p-value, for $i = 1, 2, \cdots, d$. We define the Cauchy combination test statistic as

$$T = \sum_{i=1}^{d} \omega_i \tan\{(0.5 - p_i)\pi\}, \tag{1}$$

where the weights ω_i's are nonnegative and $\sum_{i=1}^{d} \omega_i = 1$. Given that p_i is uniformly distributed between 0 and 1 under the null, the component $\tan\{(0.5 - p_i)\pi\}$ follows a standard Cauchy distribution.

When p_i's are independent or perfectly dependent (i.e., all the p_i's are equal), it is easy to see that the test statistic T has a standard Cauchy distribution under the null. This phenomenon results from the closeness of Cauchy distribution under convolution and is unique to our Cauchy combination test statistic. In fact, for the minimum p-value, higher criticism, Berk-Jones, and many other test statistics, the null distribution in the independent case is completely different from that in the perfectly dependent case. This simple observation indicates that correlation can have a substantial impact on the null distribution of these existing tests and should not be ignored. While

correlation also affects the null distribution of the Cauchy combination test statistic, we will show next that the impact on the tail is very limited.

2.1 Non-asymptotic approximation for the null distribution

To investigate the null distribution in the presence of correlation, we assume that the p-values are calculated from z-scores (i.e., test statistics that follow normal distributions). Specifically, let $\mathbf{X} = (X_1, X_2, \cdots, X_d)^T$, where X_i is a test statistic (or z-score) corresponding to the individual p-value p_i.

Denote $\mathrm{E}(\mathbf{X}) = \boldsymbol{\mu}$ and $\mathrm{Cov}(\mathbf{X}) = \boldsymbol{\Sigma}$. Because a test statistic must have a known null distribution to obtain its critical value or p-value, we can always rescale the individual test statistic X_i to have variance 1. Thus, without loss of generality, we assume that $\boldsymbol{\Sigma}$ is a correlation matrix and then the individual p-value is given by $2\{1-\Phi(|X_i|)\}$. For testing the global null hypothesis that $H_0 : \boldsymbol{\mu} = \mathbf{0}$, we can rewrite the Cauchy combination test statistic (1) with respect to $\mathbf{X}$ as

$$T(\mathbf{X}) = \sum_{i=1}^{d} \omega_i \tan[\{2\Phi(|X_i|) - 3/2\}\pi].$$

Large values of $T(\mathbf{X})$ provide evidence against the global null hypothesis H_0.

We assume the following condition about the test statistics X_i's.

(C.1) (*Bivariate normality*) For any $1 \le i < j \le d$, $(X_i, X_j)^T$ follows a bivariate normal distribution.

The bivariate normal condition (C.1) is mild and also assumed in Efron (2007), where the author studied a similar topic, i.e., controlling the false discovery rate under dependence.

The following theorem provides a non-asymptotic approximation to the null distribution of $T(\mathbf{X})$ under any arbitrary correlation structure $\boldsymbol{\Sigma}$.

Theorem 1. *Suppose that the bivariate normality condition (C.1) holds and* $E(\mathbf{X}) = \mathbf{0}$. *Then for any fixed* d *and any correlation matrix* $\boldsymbol{\Sigma} \ge 0$, *we have*

$$\lim_{t \to +\infty} \frac{P\{T(\mathbf{X}) > t\}}{P\{W_0 > t\}} = 1,$$

where W_0 *denotes a standard Cauchy random variable.*

Theorem 1 indicates that the test statistic $T(\mathbf{X})$ still has approximately a Cauchy tail under dependency structures. Note that $T(\mathbf{X})$ is defined as a weighted sum of "correlated" standard Cauchy variables. Roughly speaking, because of the heaviness of the Cauchy tail, the correlation structure $\boldsymbol{\Sigma}$ only has limited impact on the tail of $T(\mathbf{X})$.

As p-value corresponds to the tail probability of the null distribution, Theorem 1 suggests that we can use the standard Cauchy distribution to approximate the p-value of the test based on $T(\mathbf{X})$. Let t_α be the upper α-quantile of the standard Cauchy distribution, i.e., $P\{W_0 > t_\alpha\} = \alpha$. We define an α-level test as

$$R_\alpha(\mathbf{X}) = I\{T(\mathbf{X}) > t_\alpha\} \tag{2}$$

and refer to it as the Cauchy combination test, where $I(\cdot)$ is an indicator function.

Suppose that we observe $T(\mathbf{x}) = t_0$. From the c.d.f. of standard Cauchy distribution, the p-value of the test can be simply approximated by

$$\text{p-value} = 1/2 - (\arctan t_0)/\pi. \tag{3}$$

Therefore, given the observed test statistic, the computation cost of calculating p-value is almost negligible, making the Cauchy combination test $R_\alpha(\mathbf{X})$ well suited for analyzing massive data. Furthermore, Theorem 1 guarantees that the approximation should be particularly accurate for very small p-values, which are of primary interest in large-scale multiple testing but very difficult to be calculated accurately.

Note that $P\{T(\mathbf{X}) > t_\alpha\}$ represents the actual size, denoted by s_α, of the test $R_\alpha(\mathbf{X})$. Theorem 1 can be equivalently stated as the ratio of the size to significance level converges to 1 as the significance level tends to 0, i.e.,

$$\lim_{\alpha \to 0} \frac{s_\alpha}{\alpha} = 1.$$

Simulation studies in Section 4.1 show that when the significance level α is moderately small, this ratio would already be close to 1 under a variety of correlation matrices.

Theorem 1 can also be extended to the cases where the weights, w_i's, are random and independent of the test statistics:

Corollary 1. *If the weights, w_i's, are random variables and independent of $\mathbf{X}$, then Theorem 1 still holds.*

2.2 Why Cauchy distribution?

In statistical literature, the Cauchy distribution mainly serves as a counter example, such as the nonexistence of the mean and an exception to the Law of Large Number. In this sense, quoting Pillai and Meng (2016), "some introductory courses have given the Cauchy distribution the nickname Evil". Probably for these reasons, the Cauchy distribution has seldom been used in statistical

inference. Motivated by studying the large sample behavior of the Wald tests, Pillai and Meng (2016) recently revealed one of the angel aspects of Cauchy distribution and proved a surprising result that was originally conjectured by Drton and Xiao (2016). Specifically, let $\mathbf{Y} = (Y_1, \cdots, Y_d)^T$ and $\mathbf{Z} = (Z_1, \cdots, Z_d)^T$ be i.i.d. $N_d(\mathbf{0}, \mathbf{\Sigma})$. Note that Y_i/Z_i is Cauchy distributed. Pillai and Meng (2016) proved that for an arbitrary covariance matrix $\mathbf{\Sigma}$, $\sum_{i=1}^d \omega_i (Y_i/Z_i)$ still follows a standard Cauchy distribution, where $\sum_{i=1}^d \omega_i = 1$ and $\omega_i \geq 0$ for any $i = 1, \cdots d$.

Both their result and our Theorem 1 indicates that the Cauchy distribution could be insensitive to certain types of dependency structures. Specifically, their result shows that the weighted sum of a class of dependent Cauchy variables can still be a Cauchy variable, while our Theorem 1 considers another class of dependent Cauchy variables and indicates that the weighted sum of them still has a Cauchy tail. Therefore, we can use the Cauchy distribution to construct test statistics to deal with dependence structures, which are very challenging to be accounted for in general.

While sharing a similar interpretation with Pillai and Meng (2016), our result is substantially different and has unique contributions in terms of both methodology and theory. First and foremost, Pillai and Meng (2016)'s result is not a practical method and was motivated from a theoretical interest of studying the large-sample behaviour of the Wald test. After all, we rarely have a test statistic that is the ratio of two normal variables in practice. In contrast, the Cauchy combination test that we proposed maps the p-values to Cauchy variables and has a wide range of applications. Second, there is also fundamental distinctions between our and Pillai and Meng (2016)'s theories. Let $V_i = \tan[\{2\Phi(|X_i|) - 3/2\}\pi]$ denote the Cauchy variable in our theory and $W_i = Y_i/Z_i$ denote the Cauchy variable in theirs, where $i = 1, 2, \cdots, d$. While V_i and W_i have the same marginal distribution, the joint distribution of V_i's is completely different from that of W_i's. Therefore, the test statistics $\sum \omega_i V_i$ and $\sum \omega_i W_i$ might follow very distinct distributions. In fact, our proof strategy for Theorem 1 is also completely different from that in Pillai and Meng (2016). In addition, because the bivariate normality condition assumed in our Theorem 1 is much weaker than the joint normality assumption in Pillai and Meng (2016), our result allows for a broader class of dependent Cauchy variables than theirs. More discussions about the bivariate normality condition in high dimensions are provided in Section 2.3.

2.3 Asymptotic approximation for the null distribution in high dimensions

To establish the null distribution in the high dimensional situation where the number of p-values d is very large and diverging, we further assume the following conditions on the correlation matrix

$\mathbf{\Sigma}$.

(C.2) $\lambda_{max}(\mathbf{\Sigma}) \leq C_0$ for some constant $C_0 > 0$, where $\lambda_{max}(\mathbf{\Sigma})$ denotes the largest eigenvalue of $\mathbf{\Sigma}$.

(C.3) $\max_{1\leq i<j\leq d} \sigma_{ij}^2 \leq \sigma_{max}^2 < 1$ for some constant $0 < \sigma_{max}^2 < 1$, where σ_{ij} is the (i,j)-th element of $\mathbf{\Sigma}$.

Conditions (C.2) and (C.3) on the correlation matrix are mild and common assumptions in the high dimensional setting (see, e.g., Cai et al., 2014).

The following theorem shows that the Cauchy approximation for the null distribution is still valid when the dimension d diverges.

Theorem 2. *Suppose that conditions (C.1), (C.2) and (C.3) hold and* $E(\mathbf{X}) = \mathbf{0}$. *If* $d = o(t^c)$ *for any constant* $0 < c < 1/2$, *we have*

$$\lim_{t\to+\infty} \frac{P\{T(\mathbf{X}) > t\}}{P\{W_0 > t\}} = 1,$$

where W_0 *denotes a standard Cauchy random variable.*

In addition to the theoretical justification provided in Theorem 2, the Cauchy combination test also offers advantages that make it appealing in the high-dimensional situation from a practical point of view. We illustrate the challenges and the advantages of the Cauchy combination test in the high-dimensional situation by the following example. Let $\mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_2, \cdots, \mathbf{Z}_d)$ be an $n \times d$ fixed design matrix and $\mathbf{Y}$ be a vector of n i.i.d responses. Assume that $\mathbf{Y}$ and $\mathbf{Z}_i$'s are standardized to have mean 0 and variance 1. To test the marginal association between $\mathbf{Y}$ and $\mathbf{Z}_i$'s, we have the individual test statistics defined as $\mathbf{X} = \mathbf{Z}^T\mathbf{Y}/\sqrt{n}$. Many classic tests use X_i as the test statistic, such as the Cochran-Armitage trend test that is commonly used in GWAS for testing the association between a disease status and individual SNPs. When the p-values of all the SNPs in the genome are combined for a global significance, d is in the hundreds of thousands or even millions.

First, the correlation matrix $\mathbf{\Sigma}$ is highly singular in the high dimensional situation, with a rank less than the sample size n that could be much smaller than the dimension d. In the above example, $\mathbf{\Sigma} = \mathbf{Z}^T\mathbf{Z}/n$. For the minimum p-value, higher criticism, Berk-Jones, and many other tests, a highly singular correlation matrix would have a substantial impact on the null distribution and is very difficult to be accounted for. Note that perfect dependence is a special case of a highly singular correlation matrix, and that the Cauchy combination test statistic follows exactly a standard Cauchy distribution in this case. Thus, the Cauchy approximation should be particularly

accurate in the high-dimensional situation. Our simulation study in Section 4.1 also confirms this expectation. Moreover, as the p-value of our test is calculated by a simple explicit formula (3) without requiring the information of $\mathbf{\Sigma}$ that could be very big in the high-dimensional situation, there is no computational issue for the Cauchy combination test even when d is exceedingly large.

Second, the bivariate normality condition in Theorem 1 and 2 (also in Efron, 2007) is mild and appropriate in the high-dimensional setting. To see this, we compare it with the stronger condition of joint normality (i.e., $\mathbf{X} \sim N_d(\boldsymbol{\mu}, \mathbf{\Sigma})$), which is commonly assumed in the literature when dependency structures are considered (see, e.g., Hall and Jin, 2010; Fan et al., 2012; Fan and Han, 2016). Consider the aforementioned example when the response is not normally distributed, such as a binary response. Both the bivariate and joint normality conditions are on the basis of multivariate central limit theorem (CLT). However, due to the convergence rate of CLT, $\mathbf{X}$ may not jointly converge to a multivariate normal distribution in the high-dimensional scenario where d increases with n at a certain rate. See Chernozhukov et al. (2013) and Chernozhukov et al. (2014) for recent reviews on this topic. Therefore, it is not realistic to assume the joint normality of $\mathbf{X}$ when d is comparable with or even much larger than n. The bivariate normality, however, is a much weaker assumption and still reasonable in the high-dimensional setting.

2.4 Remarks

Remark 1. According to the test statistic (1) and p-value calculation (3), our test only requires the individual p-values (and the prespecified weights) as input. The bivariate normality condition and the correlation matrix $\mathbf{\Sigma}$ are only used to study the null distribution of the test statistic and are actually not needed for the application of our Cauchy combination test itself.

Remark 2. The weights ω_i's add flexibility to incorporate possible domain knowledge to boost power. For example, in GWAS, the biological information of genetic variants (e.g., annotation) can be integrated to improve the analysis power (Lee et al., 2014). In comparison, to the best of our knowledge, the minimum p-value, higher criticism and Berk-Jones tests do not allow for incorporation of weights, as all of them have a maximum-type test statistic. In the absence of prior knowledge, the equal weights (i.e., $\omega_i = 1/d$) can be employed.

Remark 3. Because the null distribution of $T(\mathbf{X})$ is symmetric, i.e., $P\{T(\mathbf{X}) > t\} = P\{T(\mathbf{X}) < -t\}$ for any $t \in \mathbb{R}$, it is trivial that Theorem 1 also holds for the lower tail of the distribution of $T(\mathbf{X})$, i.e., $\lim_{t\to-\infty} P\{T(\mathbf{X}) < t\}/P\{W_0 < t\} = 1$.

Remark 4. As mentioned in Remark 1, the proposed test only takes the individual p-values as

input. Therefore, our method can also be useful in applications where the original data is difficult to access and only summary statistics, such as the individual test statistics or p-values, are available. For example, in GWAS, the original data can be difficult to obtain due to various reasons including consent and privacy issues. In fact, statistical analysis based on summary statistics has emerged with an increasing demand. Recently developed methods includes Wen and Stephens (2010); Yang et al. (2012); Lee et al. (2014); Finucane et al. (2015).

Remark 5. If the data are discrete and certain exact tests are used (see, e.g., Liu et al., 2014), the individual p-values may not exactly follow the uniform distribution $U[0,1]$ under the null. In many applications such as GWAS with a binary outcome (Wu et al., 2010, 2011), the p-values are often conservative, i.e., stochastically smaller than $U[0,1]$. Let $\tilde{\mathbf{X}} = (\tilde{X}_1, \cdots, \tilde{X}_d)$, where $\tilde{X}_i$ is a test statistic corresponding to the p-value p_i and follows a normal distribution with mean 0 and variance less than 1. Then the p-value $p_i = 2\{1 - \Phi(|\tilde{X}_i|)\}$ is conservative. The following corollary shows that the Cauchy combination test can still protect the type I error and provide valid inference in the presence of conservative individual p-values.

Corollary 2. *Under the same assumptions of Theorem 1 (or Theorem 2) except that* $var(\tilde{X}_i) \leq 1$ *for any* $i = 1, 2, \cdots, d$, *we have*

$$\lim_{t \to +\infty} \frac{P\{T(\tilde{\mathbf{X}}) > t\}}{P\{W_0 > t\}} \leq 1.$$

3 Power

In this section, we study the asymptotic power of the proposed Cauchy combination test $R_\alpha(\mathbf{X})$ under sparse alternatives. Here asymptotics refers to d tending to infinity. We follow the theoretical setup of Donoho and Jin (2004). Assume that the individual test statistics $\mathbf{X} \sim N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is a banded correlation matrix, i.e., $\sigma_{ij} = 0$ for any $|i-j| > d_0$ for some positive constant $d_0 > 1$. Let μ_i denote the coordinates of $\boldsymbol{\mu}$ for $i = 1, 2, \cdots, d$. We are interested in testing the global null hypothesis that $H_0 : \boldsymbol{\mu} = \mathbf{0}$, against alternatives where only a small number of μ_i's are nonzero. Denote $S = \{1 \leq i \leq d : \mu_i \neq 0\}$ as the set of signals or nonzero effects. Suppose that the number of signals $|S| = d^\gamma$, where $|S|$ is the cardinality of S and the sparsity parameter $0 < \gamma < 1/2$. The signals are assumed to have the same magnitude, i.e., $|\mu_i| = \mu_0 > 0$ for all $i \in S$.

Theorem 3. *Suppose that* $\min_{1 \leq i \leq d} \omega_i \geq c_0/d$ *for some constant* $c_0 > 0$. *Let* $\mu_0 = \sqrt{2r \log d}$, *where* $r > 0$. *For any* $\alpha > 0$, $r > (1 - \sqrt{\gamma})^2$ *and* $0 < \gamma < 1/2$, *we have*

$$\lim_{d \to +\infty} P\{R_\alpha(\mathbf{X}) = 1\} = 1.$$

Theorem 3 states that the power of the Cauchy combination test converges to 1 for any significance level $\alpha > 0$, or equivalently, that the sum of Type I and II errors can vanish asymptotically, under sparse alternatives. Furthermore, Theorem 3 also indicates that the Cauchy combination test attains the optimal detection boundary defined in Donoho and Jin (2004) in the strong sparsity situation when $0 < \gamma < 1/4$.

The power of our proposed test has some similarity to that of the minimum p-value test. In fact, Arias-Castro et al. (2011) showed that the minimum p-value test also attains the optimal detection boundary when $0 < \gamma < 1/4$. Intuitively, the minimum p-value test has good power against sparse alternatives since it uses the smallest individual p-value to represent the overall significance of a set of variables. In contrast, the Cauchy combination test statistic (1) transforms individual p-values to standard Cauchy variables. It can be easily seen that small p-values correspond to very large values of a Cauchy variable and the sum in (1) is essentially dominated by a few of the smallest p-values (see a toy example in Table 1 in the supplementary material). Roughly speaking, the Cauchy combination test uses the few smallest p-values to represent the overall significance. Therefore, it is expected to also have strong power against sparse alternatives.

The theoretical analysis here only states the asymptotic power under banded correlation matrices. To examine the finite-sample power performance under general correlation structures, extensive simulation studies are carried out in Section 4.2 and show that the Cauchy combination test has very robust power across a range of correlation structures and sparsity levels compared with the existing tests. We also provide some discussions about the finite-sample power of the Cauchy combination test in Section 3 in the supplementary material.

4 Applications

We evaluate the p-value calculation accuracy of the Cauchy combination test under a variety of hypothetical and real-data-based correlation matrices. We also compare the power of our proposed test with three existing tests that have strong power against sparse alternatives, i.e., the minimum p-value (Tippett, 1931), higher criticism (Donoho and Jin, 2004) and Berk-Jones (Berk and Jones, 1979) tests. Throughout this section, the weights, ω_i's, in the Cauchy combination test statistic are chosen to be $1/d$ for all $i = 1, 2, \cdots, d$.

For both real-data analysis and parts of the simulations, we use the data of a Crohn's disease genome-wide association study (Duerr et al., 2006), which aims at identifying SNPs or genes that are associated with the inflammatory bowel disease. This data contains 1028 independent subjects

from the Non-Jewish population. After similar data quality control as in Duerr et al. (2006), the data set used in our analysis consists of 293,426 SNPs and 997 subjects, with 498 cases and 499 controls. SNPs are grouped into 15,279 genes on chromosomes 1–22 according to the Genome Build UCSC hg17 assembly. The gene size (number of SNPs) ranges from 1 to 705 and is highly skewed to the right.

4.1 Accuracy of p-value calculation

We use simulations to examine the accuracy of p-value calculation based on the Cauchy approximation under various correlation structures and different dimensions. The vector of individual test statistics $\mathbf{X}$ is generated from $N_d(0, \mathbf{\Sigma})$ under the null hypothesis. We consider six values of the dimension d, i.e., $d = 5, 20, 50, 100, 300, 500$, for each of the following correlation matrix $\mathbf{\Sigma} = (\sigma_{ij})$.

- Model 1 (AR(1) correlation): For each d, $\sigma_{ij} = \rho^{|i-j|}$ for $1 \le i, j \le d$, where $\rho = 0.2, 0.4, 0.6, 0.8, 0.99$. There are 30 conditions in total, corresponding to six dimension values and five correlation matrices.

- Model 2 (Polynomial decay): For each d, $\sigma_{ii} = 1$ and $\sigma_{ij} = \frac{1}{0.7+|i-j|^\rho}$ for $1 \le i \ne j \le d$, where $\rho = 0.5, 1.0, 1.5, 2.0, 2.5$. There are 30 conditions in total.

- Model 3 (Singular matrices): For each d, let $A = (a_{ij})$ be a $(d/5) \times d$ matrix, where $a_{ij} = \rho^{|i-j|}$ and $\rho = 0.2, 0.4, 0.6, 0.8, 0.99$. Further let $D = (d_{ij})$ be a diagonal matrix with diagonal elements $d_{ii} = (\tilde{a}_{ii})^{-1/2}$, where $\tilde{a}_{ii}$ is the i-th diagonal of $A^T A$. Then we take $\mathbf{\Sigma} = D^T A^T A D$. There are 30 conditions in total.

- Model 4 (Real genotypes): For each d, we randomly select 10 genes from the Crohn's disease data that have or approximately have d SNPs. Then we take $\mathbf{\Sigma}$ to be the sample correlation matrix of SNPs in a gene. There are 60 conditions in total.

Model 1 and 2 are commonly used in simulations. The singular matrices constructed in Model 3 aim to mimic the high-dimensional situation and contain many large and moderate correlation coefficients. We also consider realistic correlation structures in genetic data through Model 4. Since SNPs could be highly correlated due to linkage disequilibrium, the correlation matrices in Model 4 often contain very strong correlations (e.g., 0.99).

Recall that our Theorem 1 indicates that $\lim_{\alpha \to 0} s_\alpha / \alpha = 1$, where s_α denotes the size of the Cauchy combination test. For each correlation matrix $\mathbf{\Sigma}$ specified above, we generate 10^8 Monte

Carlo samples to evaluate the empirical size at significance levels $\alpha = 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}$, and use the ratio of empirical size to significance level to reflect the accuracy of p-value calculation.

The results are summarized by boxplots and shown in Figure 1. It can be seen that the Type I error of the Cauchy combination test is well controlled in general. As the significance level decreases, the Cauchy approximation becomes more accurate. For very small significance levels such as $\alpha = 10^{-5}$, the Monte Carlo errors are not negligible and are in fact the main cause of the variations in the boxplots. Given the total number of our simulation conditions (i.e., 150), the ratio of empirical size to significance level for $\alpha = 10^{-5}$ is not significantly different from 1, which indicates very good accuracy for extremely small p-values. Furthermore, under real correlation structures in Model 4 that contain very strong correlations, the Type I error is still well controlled. This is expected because the perfect dependency is also an extreme case of strong correlation. Moreover, it can be seen from the result of Model 3 that the accuracy is particulary good under singular correlation matrices, which agrees with our discussion in Section 2.3.

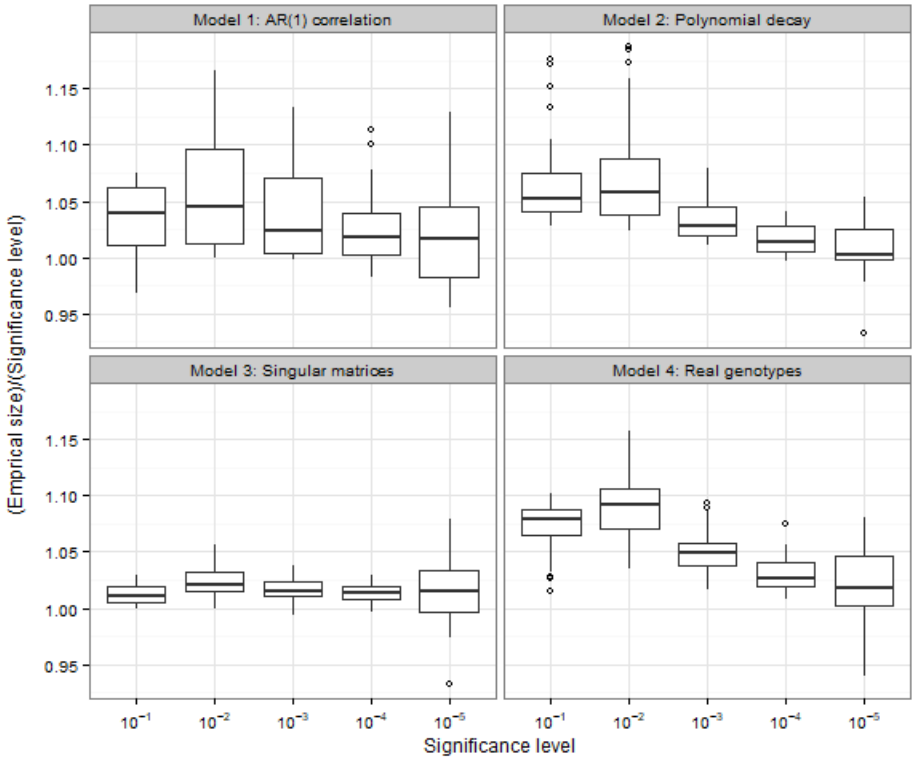

Figure 1: The ratio of empirical size to significance level for Model 1–4, summarized by boxplots. The x-axis is the significance level at $\alpha = 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}$.

To examine the accuracy of the p-value calculation in a more challenging high-dimensional

scenario with an exceedingly large d, we consider the following correlation matrix $\mathbf{\Sigma}$ based on real genomic data.

- Model 5 (High-dimensional singular matrix): We take $\mathbf{\Sigma}$ to be the sample correlation matrix of all the SNPs in the Crohn's disease data. Specifically, $\mathbf{\Sigma}$ is a highly singular matrix with dimension $d = 293,426$ and rank equal to the sample size $n = 997$.

Because the computation for a large $\mathbf{\Sigma}$ is very intensive, we use 10^6 Monte Carlo samples to calculate the empirical sizes. Figure 2 shows the simulation result and demonstrates that the p-value calculation is still accurate under high-dimensional singular correlation matrix.

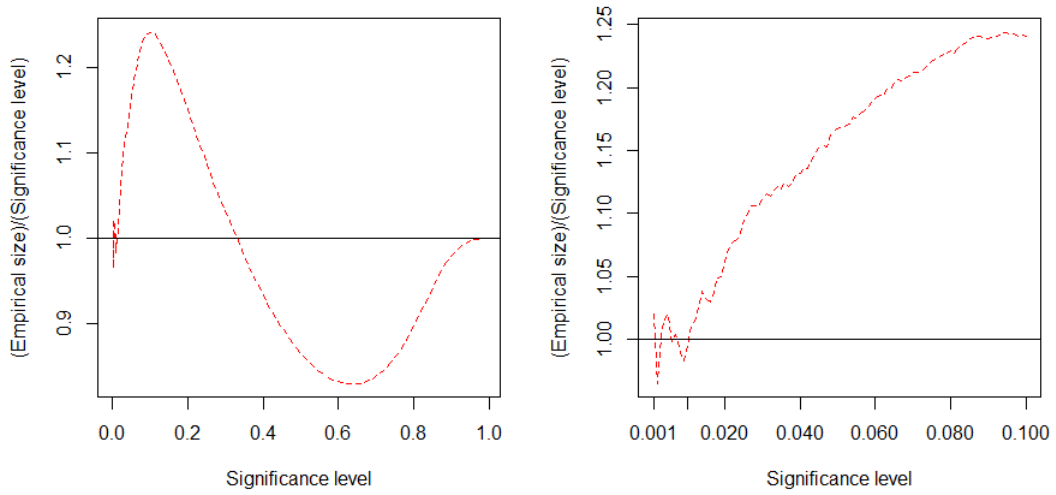

Figure 2: The ratio of empirical size to significance level (dashed lines) for Model 5. The straight line in each plot is the reference line. The plot on the right is a zoom-in image of the plot on the right. Note that the non-smoothness and fluctuation of the dashed curve in the right plot is due to the Monte Carlo errors.

We also investigate the accuracy of p-value calculation when the normality assumption is violated. The simulation setting is exactly the same as that of Figure 1, except that $\mathbf{X}$ is generated from a multivariate t distribution with 4 degrees of freedom, i.e, $\mathbf{X} \sim t_4(0, \mathbf{\Sigma})$. The result is presented in Figure 1 in the supplementary material and shows a similar phenomena as the Gaussian case.

4.2 Power comparison

We compare the power of the Cauchy combination, minimum p-value, higher criticism, and Berk-Jones tests, which are denoted by *CCT*, *MinP*, *HC* and *BJ*, respectively. More specifically, we

investigate how sparsity and correlation could influence the power of different tests. Under the alternative, the vector of individual test statistic $\mathbf{X}$ is generated from $N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu} = (\mu_i)$ and $\boldsymbol{\Sigma} = (\sigma_{ij})$. Three values of the dimension d are examined: $d = 20, 40, 60$. The percentage of signals (i.e., non-zero μ_i's) is set to be 5%, 10% and 20% for each d. All the signals have the same strength μ_0, which is chosen to be $\sqrt{3\log(d)}/s^{1/3}$ to make the power in different settings comparable, where s denotes the number of signals. The correlation matrix $\boldsymbol{\Sigma}$ is set to have an exchangeable structure, with $\sigma_{ij} = \rho$ for all $1 \leq i \neq j \leq d$, and a variety of correlation levels are considered with ρ chosen to be the nonnegative multiples of 0.05 between 0 and 0.4.

For each $\boldsymbol{\Sigma}$, we first draw 10^5 Monte Carlo samples to calculate the critical values of *CCT*, *MinP*, *HC* and *BJ* at the significance level 0.05. Here we also use simulation-based critical values for *CCT* in order to make a fair comparison. Then in each sparsity and correlation setting, 10^4 simulations are performed to calculate the power of the four tests.

The result is displayed in Figure 3. It demonstrates that *CCT* has very robust power across different sparsity and correlation levels compared with the other three tests. *MinP* is not sensitive to the magnitude of correlation. But when signals are not very sparse and weakly dependent, *MinP* has a considerable power loss and *BJ* is most advantageous in this situation. Both *HC* and *BJ* lose power substantially as correlation increases, even in the case of moderately sparse signals. One possible explanation for this is that both tests compare the ordered individual p-value $p_{(i)}$ with the reference value i/d, which is not a correct reference in the presence of correlation. In contrast, *CCT* has very robust power in every setting. It outperforms *MinP* when signals are not very sparse, and has higher power than *HC* and *BJ* in the case of moderate or strong correlations. In the absence of prior knowledge of sparsity and correlation, such as scanning genes in GWAS, *CCT* would be a robust choice and less likely to miss important signals. More importantly, the p-value of *CCT* can be computed accurately and analytically under general correlation structures, while the other three tests would require intensive computation and are not suitable to analyze large-scale data. We also present the result based on analytic critical values of *CCT* in Figure 2 in the supplementary material, which demonstrates a similar phenomena.

4.3 Real genetic data analysis

We apply our Cauchy combination test to the Crohn's disease genome-wide association study and compare it with the other three tests (i.e., *MinP*, *HC* and *BJ*) in terms of power and computation time. All the analyses are carried out on a computer node with 2.5 GHz quad-core Intel Xeon

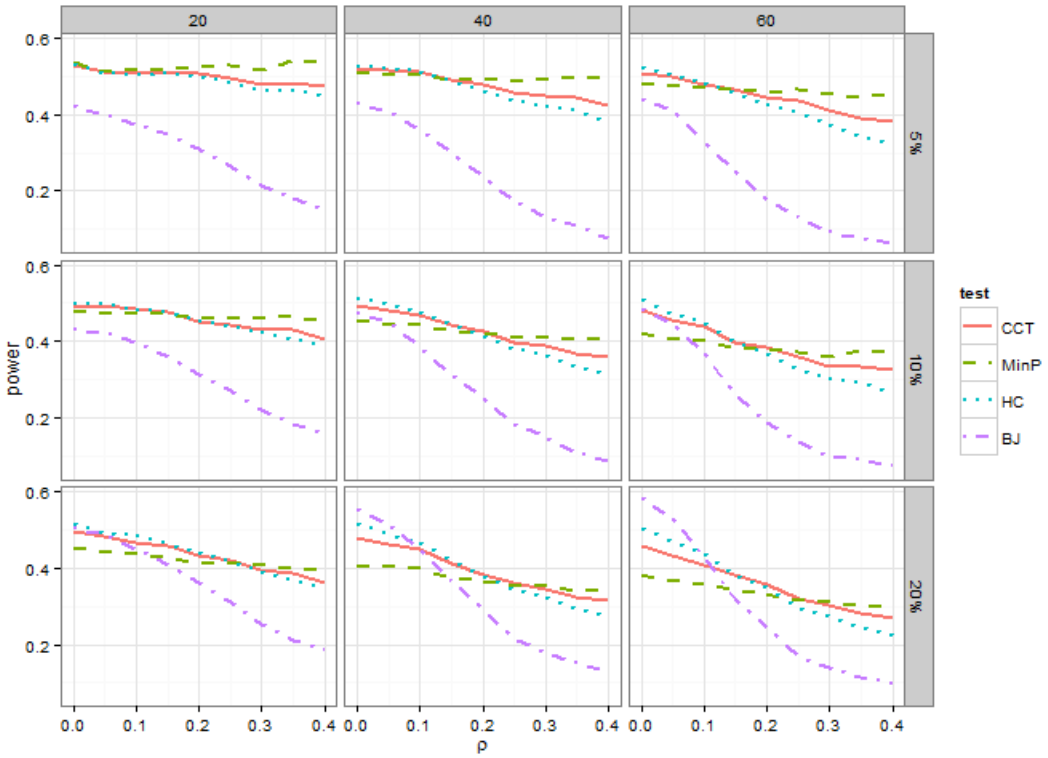

Figure 3: Power comparison of *CCT*, *MinP*, *HC* and *BJ*. The x-axis is the correlation strength ρ. The columns from left to right correspond to the dimension $d = 20, 40, 60$. The rows from top to bottom correspond to the signal percentage 5%, 10% and 20%. The signal strength is chosen to make the power in every setting comparable.

E3-1284 CPUs and 32 GB memory.

Firstly, we perform the single-SNP analysis. The individual p-values of the 293,426 SNPs in the study are obtained based on the Cochran–Armitage trend test for the association between the disease status and individual SNPs. Two SNPs are found to be significant at a level of 0.05 after the Bonferroni adjustment, with p-values of 2.8×10^{-8} and 7.5×10^{-8}. The analysis is performed using the standard software *Plink* (Purcell et al., 2007) for genome-wide association studies and takes about 3 minutes. Then we exclude these two SNPs and apply our proposed test to combine the individual p-values of the remaining 293,424 SNPs. The Cauchy combination test gives a p-value of $0.030 < 0.05$, which suggests that there still exists genetic information in the remaining SNPs. The computation of this step only takes about 1 second. Note that the Cochran–Armitage trend test statistics have a correlation matrix $\mathbf{\Sigma}$ equal to the sample correlation matrix of SNPs. The result for Model 5 (Figure 2) indicates that the p-value maintains satisfactory accuracy. In comparison,

for the other three tests, permutation is needed to incorporate the correlation structure $\mathbf{\Sigma}$ to obtain accurate p-values. Because of the high dimensionality in this situation, it is computationally very intensive to use permutation and we do not provide the results for the other three tests here.

We next perform a gene-based analysis, i.e., the p-values of individual SNPs in a gene are combined to test for an overall significance of the gene. We apply the Cauchy combination test and the other three tests (i.e., *MinP*, *HC* and *BJ*) to screen the 15,279 genes in the Crohn's disease data. The most significant genes based on the four tests are shown in Table 1. The p-values of our test are obtained according to (3) and permutation is employed for computing the p-values of the other three tests. In particular, we use 10^8 permutations for the genes listed in Table 1 and 10^6 permutations for the rest of genes. All four tests identify genes *IL23R* and *NOD2* as significant at a level of 0.05 after the Bonferroni adjustment. These two genes are also found to contain genetic variants associated with the Crohn's disease in the literature (Franke et al., 2010). The proposed Cauchy combination test only takes about 10 seconds to complete the analysis, i.e., computing the p-values of all the genes, compared with nearly 12 days for the other three tests based on permutation. In addition, our simulation result of Model 4 in Section 4.1 indicates that the p-values of the genes in Table 1 based on the Cauchy combination test should be very accurate, since these genes have very small p-values.

Table 1: P-values of the most significant genes in analysis of the Crohn's disease data using the four tests. The list is sorted in increasing order based on the smallest of the p-values of the Cauchy combination test (*CCT*).

Gene	d	MinP	HC	BJ	CCT
NOD2	8	$2.00 \cdot 10^{-7}$	$1.80 \cdot 10^{-7}$	$1.20 \cdot 10^{-7}$	$4.35 \cdot 10^{-7}$
IL23R	22	$1.20 \cdot 10^{-7}$	$1.60 \cdot 10^{-7}$	$3.10 \cdot 10^{-7}$	$5.84 \cdot 10^{-7}$
OR2AT4	5	$1.10 \cdot 10^{-4}$	$8.26 \cdot 10^{-5}$	$9.31 \cdot 10^{-5}$	$6.05 \cdot 10^{-5}$
RASD2	22	$6.86 \cdot 10^{-5}$	$8.95 \cdot 10^{-5}$	$9.21 \cdot 10^{-4}$	$8.45 \cdot 10^{-5}$
SLCO2B1	16	$1.53 \cdot 10^{-4}$	$6.96 \cdot 10^{-5}$	$1.42 \cdot 10^{-4}$	$1.22 \cdot 10^{-4}$
VSX2	5	$1.80 \cdot 10^{-4}$	$1.82 \cdot 10^{-4}$	$1.95 \cdot 10^{-3}$	$1.98 \cdot 10^{-4}$
TIFA	4	$1.66 \cdot 10^{-4}$	$2.38 \cdot 10^{-4}$	$7.27 \cdot 10^{-3}$	$2.00 \cdot 10^{-4}$
SLC44A4	6	$3.78 \cdot 10^{-4}$	$1.70 \cdot 10^{-4}$	$1.00 \cdot 10^{-3}$	$2.40 \cdot 10^{-4}$
GIMAP7	4	$5.65 \cdot 10^{-4}$	$7.70 \cdot 10^{-4}$	$5.23 \cdot 10^{-5}$	$5.49 \cdot 10^{-4}$
EVI5L	7	$2.35 \cdot 10^{-2}$	$3.18 \cdot 10^{-3}$	$7.01 \cdot 10^{-5}$	$8.96 \cdot 10^{-3}$

In summary, for either single-SNP or gene-based analysis, our method can be done within just a few seconds and provide reasonably accurate p-values, while the other three existing tests are computationally burdensome for the analysis of large genomic data.

5 Discussion

In this paper, we use the Cauchy distribution to construct a novel test that not only is powerful against sparse alternatives but also has accurate and efficient p-value calculations under arbitrary dependency structures. Our contributions are threefold. Firstly, the proposed Cauchy combination test fills the gap of testing against sparse alternatives. In the case of dense signals, with a variety of analytic p-value calculation methods, the classical sum-of-squares tests have been widely used in practice. However, in the case of sparse signals, none of the existing tests have efficient p-value calculations, which are of great importance for analyzing massive or big data. Second, the analytic method for computing the p-value of our proposed test maintains several notable properties, making the test particularly useful in modern large-scale and high-dimensional data analysis. Finally, besides the methodological contribution, our Theorem 1 also has interest on its own. It can be viewed as an extension of the closeness of Cauchy distribution under convolution from the independent case to a special dependent case. It is also established under very weak assumptions, essentially requiring only the bivariate normality of the individual test statistics.

The Cauchy combination test statistic T in (1) can be viewed as a special case of the general combination scheme based on the sum of transformed p-values (Xie et al., 2011; Xie and Singh, 2013), i.e., $\sum_{i=1}^{d} h(p_i)$, where $h(\cdot)$ can be any monotonically increasing function. Besides the advantages resulting from the special Cauchy transformation, this general combination scheme has many other advantages, for example, being able to make an exact inference in discrete data analysis and enhance finite sample efficiency (Liu et al., 2014). It is of great interest to explore other transformations and use this general combination scheme to develop tests that have other remarkable features.

While the bivariate normality assumption in Theorem 1 is appropriate in many cases, there are some applications where the individual p-values are calculated from test statistics that are not normally distributed and one can also apply the proposed test to combine the p-values. We observe through simulations that the Cauchy approximation is still quite accurate in these situations where the normality assumption is not satisfied, for example, the simulation under multivariate t distribution in Figure 1 in the supplementary material. Hence, it is interesting to generalize Theorem 1 to non-Gaussian individual test statistics.

The accuracy of the Cauchy approximation would certainly depend on the significance level and the correlation structure, which we have investigated empirically. Another interesting research

question is to derive the convergence rate or a non-asymptotic bound for Theorem 1.

Acknowledgments

The authors thank the associate editor and the two anonymous referees for their comments and suggestions that have helped greatly improve the paper. The first author would like to thank Xihong Lin for multiple inspiring discussions.

Supplementary material

Supplementary material includes the proofs of Theorem 1–3, Corollary 1–2 and technical lemmas, additional simulation results, as well as some further discussions about the finite-sample power of the proposed test.

References

Arias-Castro, E., E. J. Candès, and Y. Plan (2011). Global testing under sparse alternatives: Anova, multiple comparisons and the higher criticism. *The Annals of Statistics*, 2533–2556.

Barnett, I., R. Mukherjee, and X. Lin (2017). The generalized higher criticism for testing snp-set effects in genetic association studies. *Journal of the American Statistical Association 112*(517), 64–76.

Berk, R. H. and D. H. Jones (1979). Goodness-of-fit test statistics that dominate the kolmogorov statistics. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete 47*(1), 47–59.

Cai, T., W. Liu, and Y. Xia (2014). Two-sample test of high dimensional means under dependence. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 76*(2), 349–372.

Chernozhukov, V., D. Chetverikov, and K. Kato (2013). Supplemental material to "gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors". *The Annals of Statistics 41*(6), 2786–2819.

Chernozhukov, V., D. Chetverikov, and K. Kato (2014). Central limit theorems and bootstrap in high dimensions. *arXiv preprint arXiv:1412.3661*.

Donoho, D. and J. Jin (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Annals of Statistics*, 962–994.

Drton, M. and H. Xiao (2016). Wald tests of singular hypotheses. *Bernoulli 22*(1), 38–59.

Duerr, R. H., K. D. Taylor, S. R. Brant, J. D. Rioux, M. S. Silverberg, M. J. Daly, A. H. Steinhart, C. Abraham, M. Regueiro, A. Griffiths, et al. (2006). A genome-wide association study identifies il23r as an inflammatory bowel disease gene. *Science 314*(5804), 1461–1463.

Efron, B. (2007). Correlation and large-scale simultaneous significance testing. *Journal of the American Statistical Association 102*(477), 93–103.

Fan, J. and X. Han (2016). Estimation of the false discovery proportion with unknown dependence. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*.

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

Finucane, H. K., B. Bulik-Sullivan, A. Gusev, G. Trynka, Y. Reshef, P.-R. Loh, V. Anttila, H. Xu, C. Zang, K. Farh, et al. (2015). Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nature genetics 47*(11), 1228–1235.

Fisher, R. A. (1932). *Statistical methods for research workers, 4th edition*. Oliver and Boyd, London.

Franke, A., D. P. McGovern, J. C. Barrett, K. Wang, G. L. Radford-Smith, T. Ahmad, C. W. Lees, T. Balschun, J. Lee, R. Roberts, et al. (2010). Genome-wide meta-analysis increases to 71 the number of confirmed crohn's disease susceptibility loci. *Nature genetics 42*(12), 1118–1125.

Hall, P. and J. Jin (2010). Innovated higher criticism for detecting sparse signals in correlated noise. *The Annals of Statistics 38*(3), 1686–1732.

Koziol, J. A. and M. D. Perlman (1978). Combining independent chi-squared tests. *Journal of the American Statistical Association 73*(364), 753–763.

Lee, D., V. S. Williamson, T. B. Bigdeli, B. P. Riley, A. H. Fanous, V. I. Vladimirov, and S.-A. Bacanu (2014). Jepeg: a summary statistics based tool for gene-level joint testing of functional variants. *Bioinformatics*, btu816.

Lee, S., G. R. Abecasis, M. Boehnke, and X. Lin (2014). Rare-variant association analysis: study designs and statistical tests. *The American Journal of Human Genetics 95*(1), 5–23.

Liu, D., R. Y. Liu, and M.-g. Xie (2014). Exact meta-analysis approach for discrete data and its application to 2× 2 tables with rare events. *Journal of the American Statistical Association 109*(508), 1450–1465.

Liu, Y. and J. Xie (2018). Accurate and efficient p-value calculation via gaussian approximation: a novel monte-carlo method. *Journal of the American Statistical Association*, 1–9.

Pillai, N. S. and X.-L. Meng (2016). An unexpected encounter with cauchy and lévy. *The Annals of Statistics 44*(5), 2089–2097.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. De Bakker, M. J. Daly, et al. (2007). Plink: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics 81*(3), 559–575.

Tippett, L. H. C. (1931). The methods of statistics. *The Methods of Statistics.*.

Wen, X. and M. Stephens (2010). Using linear predictors to impute allele frequencies from summary or pooled genotype data. *The Annals of Applied Statistics 4*(3), 1158.

Wu, M. C., P. Kraft, M. P. Epstein, D. M. Taylor, S. J. Chanock, D. J. Hunter, and X. Lin (2010). Powerful snp-set analysis for case-control genome-wide association studies. *The American Journal of Human Genetics 86*(6), 929–942.

Wu, M. C., S. Lee, T. Cai, Y. Li, M. Boehnke, and X. Lin (2011). Rare-variant association testing for sequencing data with the sequence kernel association test. *The American Journal of Human Genetics 89*(1), 82–93.

Xie, M., K. Singh, and W. E. Strawderman (2011). Confidence distributions and a unifying framework for meta-analysis. *Journal of the American Statistical Association 106*(493), 320–333.

Xie, M.-g. and K. Singh (2013). Confidence distribution, the frequentist distribution estimator of a parameter: A review. *International Statistical Review 81*(1), 3–39.

Yang, J., T. Ferreira, A. P. Morris, S. E. Medland, P. A. Madden, A. C. Heath, N. G. Martin, G. W. Montgomery, M. N. Weedon, R. J. Loos, et al. (2012). Conditional and joint multiple-snp analysis of gwas summary statistics identifies additional variants influencing complex traits. *Nature genetics 44*(4), 369–375.