HHS Public Access

Author manuscript

J Am Stat Assoc. Author manuscript; available in PMC 2019 May 22.

Published in final edited form as:

J Am Stat Assoc. 2019 ; 114(525): 384–392. doi:10.1080/01621459.2017.1407776.

Accurate and Efficient *P*-value Calculation via Gaussian Approximation: a Novel Monte-Carlo Method

Yaowu Liu and

Department of Biostatistics, Harvard School of Public Health

Jun Xie

Department of Statistics, Purdue University

Abstract

It is of fundamental interest in statistics to test the significance of a set of covariates. For example, in genome-wide association studies, a joint null hypothesis of no genetic effect is tested for a set of multiple genetic variants. The minimum *p*-value method, higher criticism, and Berk–Jones tests are particularly effective when the covariates with nonzero effects are sparse. However, the correlations among covariates and the non-Gaussian distribution of the response pose a great challenge towards the *p*-value calculation of the three tests. In practice, permutation is commonly used to obtain accurate *p*-values, but it is computationally very intensive, especially when we need to conduct a large amount of hypothesis testing. In this paper, we propose a Gaussian approximation method based on a Monte Carlo scheme, which is computationally more efficient than permutation while still achieving similar accuracy. We derive non-asymptotic approximation error bounds that could vanish in the limit even if the number of covariates is much larger than the sample size. Through real-genotype-based simulations and data analysis of a genome-wide association study of Crohn's disease, we compare the accuracy and computation cost of our proposed method, of permutation, and of the method based on asymptotic distribution.

Keywords

Berk-Jones test; Genome-wide association study; Higher criticism; High dimensionality; Monte Carlo method

1 Introduction

Testing whether a set of covariates have any effect on a response is commonly encountered in practice and a fundamental statistical problem. In many applications, only a small fraction of covariates are expected to be related with the response, i.e., the covariates with nonzero effects in the set are sparse. For example, in typical genome-wide association studies, a

Correspondence to: Jun Xie.

*This work is supported by the National Institutes of Health Grant R21GM101504.

Supplementary material

Supplementary material includes the proofs of Theorem 1–3 and technical lemmas, the justification of Condition (A.3), the demonstration of the accuracy of our theorems, the formulae of asymptotical critical values, as well as additional simulation results of a variety of error distributions. The R code for the simulations is available upon request from the corresponding author.

sample of subjects is collected with their phenotypes and genetic information that may contain millions of genetic variants, e.g., single nucleotide polymorphism (SNP). It is often of interest to jointly test the existence of any genetic effect within a set of SNPs, such as a gene, pathway, or other functional genetic segment. One would expect that most SNPs have no effect on the phenotype (see, e.g., Wu et al., 2010). Therefore, there is an increasing demand for tests that are particularly powerful against sparse alternatives. Among these tests, the minimum p-value (Tippett, 1931), higher criticism (Donoho and Jin, 2004), and Berk-Jones (Berk and Jones, 1979) tests have received substantial interests in the literature. Specifically, they have been shown to have strong power in sparse settings (Arias-Castro et al., 2011; Li et al., 2015; Moscovich et al., 2016), and been adapted to genome-wide association studies to scan the whole genome for significant genes (Chen et al., 2006; Ballard et al., 2010; Wu et al., 2014). All the three tests can be viewed as approaches of combining marginal test statistics of individual covariates to aggregate individual effects.

To apply statistical tests in practice, it is important to obtain accurate p-values in order to make valid inference. However, the p-value calculation of the aforementioned three tests could be very challenging for various reasons, including correlations among covariates, non-Gaussian responses, and the large scale of the data. Using the example of genome-wide association study again, the genotypes of SNPs are possibly highly correlated due to linkage disequilibrium, and the phenotype of interest may be a binary disease status or follow a skewed distribution. In the literature, the majority of p-value calculation methods for the three tests are derived under the independence and normality assumptions of marginal test statistics, such as methods based on the asymptotic null distributions of the test statistics and analytic (approximation) methods including Noé (1972); Barnett and Lin (2014); Li et al. (2015). However, when the two assumptions are violated, there is no guarantee that these methods can provide accurate p-values for practical uses. As an alternative strategy, the permutation method has been widely adopted for p-value calculation, as it naturally incorporates the dependency structure and is robust to the normality assumption. For example, permutation was employed to compute p-values by Ballard et al. (2010) for the minimum p-value method and Wu et al. (2014) for the higher criticism test. Nevertheless, in a large-scale analysis that involves an enormous number of tests, simulating the null distributions of test statistics by permutation is computationally very intensive. For instance, tens of thousands of genes need to be tested in genome-wide association studies, making permutation computationally expensive (see also Barnett and Lin, 2014).

In this paper, we aim to provide a p-value calculation method that is computationally more efficient than permutation and also maintains reasonable accuracy under general distributions and dependency structures. We prove that the null distributions of the three test statistics can be well approximated by replacing the original marginal statistics with a Gaussian vector that has the same covariance matrix. Based on this theoretical implication, we propose to compute the p-values of the three tests by simulating correlated Gaussian variables. Similar to Barnett and Lin (2014), our proposed method is computationally advantageous over permutation when the number of covariates, denoted by d, is not large. More importantly, our method can be considered as an approach that Efron (2014) referred to as “a combination of a little mathematics with a lot of computation”, achieving a good balance between accuracy and computational efficiency. In comparison, the permutation

method and the methods derived under the independence and normality assumptions are solely based on numerical simulations or theoretical approximation, respectively. Finally, although the idea of Gaussian approximation is not new, to the best of our knowledge, it has not been used to calculate p-values of the three tests that have particularly strong power in sparse settings.

In addition to the methodological contribution, our theoretical development has its own interest and is based on a recent theory of high-dimensional Gaussian approximation developed in Chernozhukov et al. (2013). The theory of Chernozhukov et al. (2013) provides a non-asymptotic bound for the Gaussian approximation errors that could converge to 0 even if d is much larger than the sample size n under arbitrary covariance structures. Specifically, d can be as large as $O(\exp(Cn^c))$ for some constants $C > 0$ and $1 > c > 0$. In addition, the non-asymptotic bound is considered more advanced than the traditional asymptotic result as it specifies how the approximation errors depend on n and d explicitly. However, the theory of Chernozhukov et al. (2013) only applies to a type of maximum test statistics, which essentially are the minimum p-value test statistic considered here. We extend their remarkable result to the higher criticism and Berk-Jones tests, which complement the minimum p-value test and can be more powerful under a wide range of sparsity levels. Our extension is nontrivial since the higher criticism and Berk–Jones test statistics involve a sequence of order statistics and have a much more complicated form than the minimum p-value test statistic. In addition, we also extend to allow unknown error variance.

The paper is organized as follows. In Section 2, we establish the non-asymptotic error bounds of Gaussian approximation for the three tests. In Section 3, we compare the computation procedures of the Gaussian approximation and permutation methods, and discuss their efficiency. In Section 4, we evaluate the accuracy of p-values computed based on our proposed method, permutation, and asymptotic null distribution using real genotype-based simulation, and demonstrate the effectiveness of our method on data from a genome-wide association study of Crohn's disease. A discussion is given in Section 5. The technical proofs of the theorems and additional simulation results are provided in the supplementary material.

2 Theory of Gaussian approximation

2.1 Gaussian approximation

For a set of d covariates with n samples, we consider a regression model:

$$\mathbf{Y} = \alpha_0 \mathbf{1}_n + \mathbf{X}^T \boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (1)$$

with a response vector $\mathbf{Y} \in \mathbb{R}^n$, an intercept α_0, a vector of coefficients $\boldsymbol{\beta} \in \mathbb{R}^d$, an error vector $\boldsymbol{\varepsilon} = (\varepsilon_1, \cdots, \varepsilon_n)^T \in \mathbb{R}^n$, and a fixed design matrix $\mathbf{X} = (\mathbf{X}_1, \cdots \mathbf{X}_d) \in \mathbb{R}^{n \times d}$, where $\mathbf{1}_n \in \mathbb{R}^n$ is a vector of ones. The error terms ε_k's are assumed to be independent and identically distributed with mean 0 and variance $\sigma^2 > 0$. The problem of interest is to test the joint null hypothesis H_0: $\boldsymbol{\beta} = 0$ against a sparse alternative that only a small fraction of regression coefficients are nonzero. Throughout the paper, we assume $n, d \geq 2$.

Let $\widetilde{\mathbf{X}} = (\widetilde{\mathbf{X}}_1, \cdots \widetilde{\mathbf{X}}_d) \in \mathbb{R}^{n \times d}$ be a matrix with standardized columns, namely, $\mathbf{1}_n^T \widetilde{\mathbf{X}}_i = 0$ and $\widetilde{\mathbf{X}}_i^T \widetilde{\mathbf{X}}_i = n$ for $1 \le i \le d$. Define the marginal statistic as

$$r_i = \frac{1}{\sqrt{n}} \frac{\widetilde{\mathbf{X}}_i^T \mathbf{Y}}{s_y},$$

where s_y is the sample standard deviation of the response $\mathbf{Y}$. For the regression model (1), we take $\widetilde{\mathbf{X}}_i$ to be the standardized ith covariate $\mathbf{X}_i$, then $r_i/\sqrt{n}$ is the sample correlation coefficient between the ith covariate and the response variable.

In the case of known error variance σ^2, let

$$r_i^{\sigma} = \frac{1}{\sqrt{n}} \frac{\widetilde{\mathbf{X}}_i^T \mathbf{Y}}{\sigma}.$$

Write $\mathbf{r} = (r_1, \cdots, r_d)^T$ and $\mathbf{r}^{\sigma} = (r_1^{\sigma}, \cdots, r_d^{\sigma})^T$, and then we have $\mathbf{r} = (\sigma/s_y)\mathbf{r}^{\sigma}$. Since each $\widetilde{\mathbf{X}}_i$ is centered, the response vector $\mathbf{Y}$ in r_i and r_i^{σ} can be replaced by the error vector $\boldsymbol{\varepsilon}$ under the null hypothesis. Note that we do not require the error terms to be Gaussian variables. The marginal statistics $\mathbf{r}$ and $\mathbf{r}^{\sigma}$ can have general multivariate distributions under the null.

We further define

$$v_i = \frac{1}{\sqrt{n}} \widetilde{\mathbf{X}}_i^T \mathbf{e}$$

and $\mathbf{v} = (v_1, \cdots, v_d)^T$, where $\mathbf{e} = (e_1, \cdots, e_n)^T$ is a vector of independent standard Gaussian variables. Under the null hypothesis, the Gaussian vector $\mathbf{v}$ has the same mean 0 and covariance matrix $\widetilde{\mathbf{X}}^T \widetilde{\mathbf{X}}/n$ as the marginal statistics $\mathbf{r}^{\sigma}$.

Suppose $T(\mathbf{r})$ is a test statistic for the null hypothesis H_0, which summarizes the marginal statistics or mathematically is a function of $\mathbf{r}$. For instance, the minimum p-value, higher criticism and Berk-Jones test statistics, which we will introduce later, are such statistics. We refer to $T(\mathbf{v})$ as the Gaussian approximation of $T(\mathbf{r})$. Without a strong restriction on the correlation structure of $\mathbf{r}$ and the normality assumption of the error $\boldsymbol{\varepsilon}$, the null distribution of $T(\mathbf{r})$ is often theoretically intractable. But the distributions of $T(\mathbf{r})$ and its Gaussian approximation $T(\mathbf{v})$ could be very close in the sense of Kolmogorov–Smirnov distance. Therefore, $T(\mathbf{v})$ can be utilized to approximate the p-value of the test based on $T(\mathbf{r})$. In general, the accuracy of this approximation depends on the form of the test statistic, i.e., $T(\cdot)$, and can be poor. But for the test statistics considered in this paper, we will show that the approximation error converges to 0 even when d is much larger than n.

In the example of genome-wide association studies, we use the regression model (1) for testing the joint null hypothesis of no genetic effect within a set of d SNPs, where $\mathbf{Y}$ denotes

a vector of quantitative phenotypes of n subjects and $\mathbf{X}_i$ represents the genotypes of the ith SNP in the set. The SNP genotype is coded as 0, 1, 2 representing the copy number of the minor alleles. The magnitude of the marginal statistic r_i reflects the individual effect of the ith SNP and the covariance matrix $\widetilde{\mathbf{X}}^T\widetilde{\mathbf{X}}/n$ characterizes the patterns of correlation among SNPs. Since the genotypes of SNPs can be highly correlated due to linkage disequilibrium and the correlation patterns vary among different genes, it is desirable to study the approximation error under general dependency structures.

2.2 Non-asymptotic bounds of approximation errors

We next establish the non-asymptotic bounds of the Gaussian approximation errors for the minimum p-value, higher criticism, and Berk–Jones tests, which are particularly powerful for testing the joint null hypothesis H_0 against sparse alternatives.

The minimum p-value method corresponds to a maximum test statistic

$$T_{\text{MinP}}(\mathbf{r}) = \max_{1 \le i \le d} |r_i|,$$

of which large values reject the null hypothesis. Its Gaussian approximation is given by

$$T_{\text{MinP}}(\mathbf{v}) = \max_{1 \le i \le d} |v_i|.$$

The minimum p-value method summarizes the marginal statistics by their maximum, and is therefore powerful when the effects of individual covariates are sparse and strong.

Write $a \preceq b$ if a is smaller than or equal to b up to multiplying some positive constant independent of n and d. Assume the following conditions are satisfied:

(A.1) $E(\varepsilon_k^4) \le C$, where C is some positive constant;

(A.2) $|\widetilde{x}_{ki}| \le B_n$ for any $1 \le k \le n$ and $1 \le i \le d$, where $\widetilde{x}_{ki}$'s are the entries of $\widetilde{\mathbf{X}}$ and $B_n \ge 1$ is a sequence of constants, possibly growing to infinity as $n \to \infty$.

Theorem 1—Suppose that Conditions (A.1) and (A.2) are satisfied. Under the null hypothesis, we have

$$\sup_{t \in \mathbb{R}} |P\{T_{MinP}(\mathbf{r}) \le t\} - P\{T_{MinP}(\mathbf{v}) \le t\}| \preceq \frac{B_n^{2/3} \log^{7/6}(dn)}{n^{1/6}}.$$

Theorem 1 essentially is a direct consequence of the main theoretical result of Chernozhukov et al. (2013), except that we extend it to allow for unknown variance σ^2. More specifically, our Conditions (A.1) and (A.2) follow one of the conditions of Corollary 2.1 in Chernozhukov et al. (2013) for a fixed design. Note that the left-hand side of the inequality above is the Kolmogorov–Smirnov distance between the null distributions of

$T_{\text{MinP}}(\mathbf{r})$ and $T_{\text{MinP}}(\mathbf{v})$. Theorem 1 indicates that the approximation error is uniformly bounded at any significance level.

The Condition (A.2) requires that the entries of the standardized design matrix are uniformly bounded by B_n. In genome-wide association studies, SNPs with a minor allele frequency less than a given threshold (e.g. 0.01) are often excluded. Given this fact and that the value of the SNP genotype is between 0 and 2, $\tilde{x}_{ki}$'s are uniformly bounded by a constant and thus $B_n = O(1)$. In general situations where covariates are generated from sub-Gaussian distributions, we can expect that $B_n = O(\sqrt{\log nd})$. For both cases, the non-asymptotic bound in Theorem 1 implies that the approximation error increases very slowly as d grows and converges to 0 even if d is much larger than n. This result is established under arbitrary correlation structures. Moreover, only a bounded fourth moment is required for the error terms in Condition (A.1), which allows a broad range of distributions.

As the minimum p-value, higher criticism and Berk-Jones tests all are powerful against sparse alternatives, intuitively, their critical regions should be similar and the Gaussian approximation should be also accurate for the other two tests. However, the higher criticism and Berk-Jones test statistics are much more complicated than the minimum p-value test statistic, which makes the extension of Theorem 1 not straightforward.

We introduce some notations for the higher criticism test statistic first. For $x > 0$, define

$$\psi_i(x) = \frac{\sqrt{d}[i/d - \pi(x)]}{\sqrt{\pi(x)[1 - \pi(x)]}},$$

where $\pi(x) = 2[1 - \Phi(x)]$ and $\Phi(\cdot)$ is the cumulative distribution function of standard Gaussian distribution. Let $r_{(i)}$ and $v_{(i)}$ be the ith largest absolute value of r_i's and v_i's, respectively. For example, $r_{(1)} = T_{\text{MinP}}(\mathbf{r})$. The higher criticism test statistic and its Gaussian approximation are given by

$$T_{\text{HC}}(\mathbf{r}) = \max_{1 \le i \le d} \psi_i\left(r_{(i)}\right) \text{ and } T_{\text{HC}}(\mathbf{v}) = \max_{1 \le i \le d} \psi_i\left(v_{(i)}\right).$$

Note that $\pi(r_{(i)})$ can be viewed as the ordered ith smallest marginal p-value. The higher criticism uses the maximum of standardized ordered marginal p-values as a summary statistic and is particularly effective in the case of rare and weak effects (Donoho and Jin, 2015).

To facilitate the analysis, and similar to Arias-Castro et al. (2011), we search for the maximum over $c_0 \log d$ terms, where $c_0 \ge 1$ is a fixed constant, and define

$$T^*_{\text{HC}}(\mathbf{r}) = \max_{1 \le i \le c_0 \log d} \psi_i\left(r_{(i)}\right), \quad T^*_{\text{HC}}(\mathbf{v}) = \max_{1 \le i \le c_0 \log d} \psi_i\left(v_{(i)}\right)$$

In comparison, Arias-Castro et al. (2011) searches for the maximum over at most $\sqrt{5 \log d}$ terms. Further, assume that

(A.3) The density of $v_{(i)}$ is bounded by $\log d$ up to some multiplicative positive constant for any $1 \le i \le c_0 \log d$.

The following theorem shows that a similar bound of the Gaussian approximation error holds for the higher criticism test.

Theorem 2—Suppose that Conditions (A.1), (A.2) and (A.3) are satisfied. Under the null hypothesis, we have

$$\sup_{t \in \mathbb{R}} |P\{T^*_{HC}(\mathbf{r}) \le t\} - P\{T^*_{HC}(\mathbf{v}) \le t\}| \le \frac{B_n^{3/2}(\log d)^{5/2}}{n^{1/8}}.$$

Remark 1—Our Condition (A.3) is motivated from the following observations. Recall that $v_{(i)}$'s are the order statistics of the standard Gaussian variables v_i's. Firstly, when v_i's are independent, it can be easily shown that Condition (A.3) holds. A proof is given in Lemma 8 of the supplementary material. Secondly, when v_i's are correlated, we can use simulations to support the validity of (A.3). For instance, in Figure 1 of the supplementary material, a variety of correlation matrices are examined. The results show that the maximum density values are no larger than that of the independent case. Therefore, the density of $v_{(i)}$ is also bounded by $\log d$ (up to some multiplicative positive constant) under these correlation matrices. We anticipate that the phenomena observed in these examples would be true for general correlation structures of v_i's. Lastly, for the maximum order statistic $v_{(1)}$, Theorem 3 of Chernozhukov et al. (2015) implies that the upper bound in (A.3) holds uniformly for any correlation structures of v_i's.

We next consider the Berk–Jones statistic proposed by Berk and Jones (1979). Let

$$\phi_i(x) = d\left\{(i/d)\log\frac{i/d}{\pi(x)} + (1 - i/d)\log\frac{1 - i/d}{1 - \pi(x)}\right\},$$

for $x > 0$. The Berk–Jones statistic and its Gaussian approximation are

$$T_{\mathrm{BJ}}(\mathbf{r}) = \max_{1 \le i \le d} \phi_i\left(r_{(i)}\right) \text{ and } T_{\mathrm{BJ}}(\mathbf{v}) = \max_{1 \le i \le d} \phi_i\left(v_{(i)}\right),$$

respectively. The Berk–Jones test is motivated by considering the Kullback–Leibler distance between two Bernoulli distributions, one with a success probability i/d and the other with $\pi(r_{(i)})$. It also has strong power against sparse alternatives (Li et al., 2015).

In analogy to the higher criticism statistic, we consider the maximum over the first $c_0 \log d$ terms to facilitate the analysis and define

$$T^*_{\mathrm{BJ}}(\mathbf{r}) = \max_{1 \le i \le c_0 \log d} \phi_i\left(r_{(i)}\right), \quad T^*_{\mathrm{BJ}}(\mathbf{v}) = \max_{1 \le i \le c_0 \log d} \phi_i\left(v_{(i)}\right),$$

A similar non-asymptotic bound is obtained for the Berk–Jones test in the following result.

Theorem 3—Suppose that Conditions (A.1), (A.2) and (A.3) are satisfied. Under the null hypothesis, we have

$$\sup_{t \in \mathbb{R}} |P\{T^*_{BJ}(\mathbf{r}) \le t\} - P\{T^*_{BJ}(\mathbf{v}) \le t\}| \preceq \frac{B_n^{3/2}(\log d)^{5/2}}{n^{1/8}}.$$

To demonstrate the accuracy of Theorem 1–3, we carry out simulations and use p-p plots to compare the distributions of $\mathcal{T}(\mathbf{r})$ and its Gaussian approximation $\mathcal{T}(\mathbf{v})$ for the three tests, respectively. The result is displayed in Figure 2 in the supplementary material. It shows that the distributions of the test statistic and its Gaussian approximation are close to each other for all the three tests.

Note that our primary interest is to use $\mathcal{T}(\mathbf{v})$ for p-value approximation. As p-values that indicate significance correspond to the tail probabilities, more extensive simulations are performed in Section 4.1 to examine the Gaussian approximation accuracy at a range of stringent significance levels.

2.3 Binary phenotype

The regression model (1) only applies to problems with continuous responses. In case-control genome-wide association studies, the phenotype of interest is a binary disease status. Recall that the SNP genotypes (covariates) can only take three values (i.e., 0, 1 and 2). If the Cochran-Armitage trend test is employed to test the association between each SNP and the disease status, then the marginal test statistic has exactly the same form as r_i and our theorems in Section 2.2 can be directly applied.

In addition, for balanced case-control studies, Zuo et al. (2006) proposed another Z-statistic:

$$r_i^b = \sqrt{n}\frac{\hat{p}_{\text{case}} - \hat{p}_{\text{control}}}{\sqrt{2\hat{p}_{\text{all}}(1-\hat{p}_{\text{all}})}},$$

where $\hat{p}_{\text{case}}$, $\hat{p}_{\text{control}}$ and $\hat{p}_{\text{all}}$ are the estimated minor allele frequency of the ith SNP in cases, controls and all subjects, respectively. We adopt $\mathbf{r}^b$ as the marginal statistics in a balanced case-control study, where $\mathbf{r}^b = (r_1^b, \cdots, r_d^b)^{\mathrm{T}}$. In this case, we take $\widetilde{\mathbf{X}}_i = (\mathbf{X}_i - 2\hat{p}_{\text{all}})/\sqrt{2\hat{p}_{\text{all}}(1-\hat{p}_{\text{all}})}$ in the definitions of r_i, r_i^σ and v_i for any $1 \le i \le d$, then the approximation error bounds in Theorem 1–3 also hold. Furthermore, some straightforward algebra leads to $\mathbf{r} = (2s_y)\mathbf{r}^b$. As $2s_y$ converges to 1 at the rate of $\sqrt{n}$, the marginal statistics $\mathbf{r}$ and $\mathbf{r}^b$ are very close to each other with a high probability. Hence, the Gaussian approximation $\mathcal{T}(\mathbf{v})$ can also be applied for the test statistic $\mathcal{T}(\mathbf{r}^b)$.

Remark 2—In the marginal test statistics $\mathbf{r}^b$, note that $n \le \widetilde{\mathbf{X}}_i^{\mathrm{T}}\widetilde{\mathbf{X}}_i \le 2n$ instead of $\widetilde{\mathbf{X}}_i^{\mathrm{T}}\widetilde{\mathbf{X}}_i = n$. With some minor modifications, Theorem 1–3 can be established in a general situation where $c_1 n \le \widetilde{\mathbf{X}}_i^{\mathrm{T}}\widetilde{\mathbf{X}}_i \le c_2 n$ for some fixed positive constants $c_1 \le c_2$. By the relationship

between $\mathbf{r}$ and $\mathbf{r}^b$, it is straightforward to further generalize Theorem 1–3 for the approximation errors between the null distributions of $T(\mathbf{r}^b)$ and $T(\mathbf{v})$. We omit the proofs.

3 Computation procedures and their efficiency

In practice, permutation has been commonly used for p-value calculation. Let $\mathbf{Y}^p$ denote a random permutation sample of $\mathbf{Y}$ and define

$$\mathbf{r}^p = \frac{1}{\sqrt{n}} \frac{\tilde{\mathbf{X}}^T \mathbf{Y}^p}{s_y},$$

which represents the marginal statistics under the permuted sample. We refer to the null distribution of $T(\mathbf{r})$, $T(\mathbf{v})$, and $T(\mathbf{r}^p)$ as the true, Gaussian approximation, and permutation null distribution, respectively. The true null distribution is unknown in practice, and the other two null distributions are used to approximate it. We have derived the non-asymptotic bounds for the Kolmogorov–Smirnov distance between the true and Gaussian approximation null distributions in Section 2. In this section, we study the computational efficiency of the Gaussian approximation and permutation methods.

Let T_{obs} denote the observed test statistic calculated from a given data set. The p-values based on the Gaussian approximation and permutation methods are given by

$$P\left\{T(\mathbf{v}) \geq T_{\text{obs}}\right\} \text{ and } P\left\{T(\mathbf{r}^p) \geq T_{\text{obs}}\right\},$$

where the probability is with respect to the Gaussian approximation and permutation null distribution, respectively. The analytic forms of the two null distributions are barely available, but we can simulate independent Monte Carlo samples from them to obtain the empirical p-value, which is simply the proportion of samples greater than the observed test statistic T_{obs}.

For either Gaussian approximation or permutation, it consists of three steps to generate M independent Monte Carlo samples of the test statistic under the corresponding null distribution. Note that $\mathbf{v}$ follows a multivariate Gaussian distribution with mean 0 and covariance matrix $\tilde{\mathbf{X}}^T\tilde{\mathbf{X}}/n$. When $d < n$, as a pre-step, we calculate the Cholesky decomposition of the covariance matrix, namely, $\mathbf{Q}^T\mathbf{Q} = \tilde{\mathbf{X}}^T\tilde{\mathbf{X}}/n$. The upper triangular matrix $\mathbf{Q}$ has $d \times d$ dimensions and is used to speed up the computation for the Gaussian approximation method.

Step 1: Randomly generate a matrix. For permutation, generate a matrix $\mathbf{G}_{n\times M}$ where the columns are independent permuted samples of $\mathbf{Y}/(\sqrt{n}s_y)$. For Gaussian approximation, generate a matrix $\mathbf{E}_{d\times M}$ when $d < n$ or $\mathbf{E}_{n\times M}$ otherwise, where the entries are independent standard Gaussian samples.

Step 2: Compute the marginal statistics. For permutation, $\mathbf{R}_{d\times M} = \tilde{\mathbf{X}}^T_{d\times n} \cdot \mathbf{G}_{n\times M}$. For Gaussian approximation, $\mathbf{V}_{d\times M} = \mathbf{Q}^T_{d\times d} \cdot \mathbf{E}_{d\times M}$ when $d < n$ or $\mathbf{V}_{d\times M} = \tilde{\mathbf{X}}^T_{d\times n} \cdot \mathbf{E}_{n\times M}$ otherwise.

Step 3: Compute test statistics based on each column of $\mathbf{R}_{d\times M}$ or $\mathbf{V}_{d\times M}$.

We analyze the computation cost of the two methods step by step. We consider the case of $d < n$ at first. In Step 1, the ratio of the matrix sizes of **G** and **E** is n/d, and thus a larger matrix needs to be generated for permutation than for Gaussian approximation. Step 2 involves matrix multiplication, where the computation complexity is $O(n \times dM)$ for permutation and $O(d \times dM)$ for Gaussian approximation. Step 3 is the same for both methods in terms of computation cost. In addition, the Gaussian approximation method requires less memory than permutation in the first two steps. The Cholesky decomposition in the pre-step for Gaussian approximation is computationally cheap and almost negligible compared to other steps, since it only needs to be computed once. Therefore, given a fixed d, the computation time of Gaussian approximation remains almost constant for any sample size n. When $d \geq n$, the Cholesky decomposition is not performed and hence the two methods require a similar amount of computation and memory. To conclude, our method is computationally more efficient than permutation in the situation where $d < n$, and the computation saving becomes more dramatic as the ratio n/d increases. In practice with large-scale hypothesis testing, there may be a wide range of d. The computation savings would be substantial when a big portion of the tests have $d < n$. This is clearly demonstrated by a genome-wide association study in Section 4.

Remark 3

In the case of an orthogonal design matrix **X**, the Gaussian approximation can be implemented straightforwardly without Step 2. Therefore, its computation is further reduced and is much faster than that of permutation. When **X** is not orthogonal, one may consider to render **X** orthogonal (e.g., by the Gram-Schmidt transformation) as a preprocessing step to speed up the computation. On the other hand, in some situations where the signals are sparse and the correlation between covariates is moderate or strong, the de-correlation transformation may dampen the signals and result in power loss (see, e.g., Barnett and Lin, 2014). Since these situations are expected in genome-wide association studies, we do not perform the orthogonal transformation in our analysis.

4 Applications

We evaluate the accuracy and computation cost of the Gaussian approximation method, permutation, and the method based on asymptotic distribution, which are denoted by *GA*, *Permu*, and *Asym*, respectively. The original test statistics $T_{\mathrm{HC}}(\cdot)$ and $T_{\mathrm{BJ}}(\cdot)$ are adopted for the higher criticism and Berk–Jones tests. For all three tests, we use the marginal statistics **r** or $\mathbf{r}^b$, depending on whether the response is quantitative or binary.

For both simulation and real-data analysis, we use the data of the Crohn's disease genome-wide association study (Duerr et al., 2006), which aims at identifying genes that are

associated with the inflammatory bowel disease. This data consists of a total of 1760 independent subjects from Jewish and non-Jewish populations. Following the data quality control in Duerr et al. (2006), we exclude subjects with overall SNP call rates less than 94%, and remove SNPs with minor allele frequencies less than 1%, call rates less than 95%, or Hardy-Weinberg equilibrium p-values less than 0.01. The final data set consists of 293,426 SNPs and a total of 1719 subjects. SNPs are grouped into 15,279 genes on chromosomes 1–22 according to Genome Build UCSC hg 17 assembly. The gene size (number of SNPs) ranges from 1 to 705 and is highly skewed to the right. The first quartile, median and third quartile are 3, 7 and 17, respectively.

In practice, one may want to control for clinical covariates, which can be easily incorporated for quantitative phenotypes. Denote the clinical covariates by $\mathbf{Z}_1, \mathbf{Z}_2, \cdots, \mathbf{Z}_q$, where fixed constant $q \leq n - 2$. Let $\mathbf{Z} = (\mathbf{1}_n, \mathbf{Z}_1, \cdots, \mathbf{Z}_q)$ and $\mathbf{P}_Z = \mathbf{Z}(\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T$ be the orthogonal projection matrix of $\mathbf{Z}$. Then we take

$$\tilde{\mathbf{X}}_i = (\mathbf{I} - \mathbf{P}_Z)\mathbf{X}_i / \left\{ \mathbf{X}_i^T (\mathbf{I} - \mathbf{P}_Z)\mathbf{X}_i / (n - q) \right\}^{1/2}$$

and $s_y = \{\mathbf{Y}^T(\mathbf{I} - \mathbf{P}_Z)\mathbf{Y}/(n - q)\}^{1/2}$, where $\mathbf{I}$ is an $n \times n$ identity matrix. The marginal statistic and its Gaussian approximation are

$$r_i = \frac{1}{\sqrt{n-q}} \frac{\tilde{\mathbf{X}}_i^T \mathbf{Y}}{s_y} \text{ and } v_i = \frac{1}{\sqrt{n-q}} \tilde{\mathbf{X}}_i^T \mathbf{e},$$

respectively.

4.1 Simulation based on real genotypes and simulated phenotypes

To examine the accuracy of p-value calculation methods under realistic dependency structures, namely, the real patterns of correlation among SNPs, we use real genotypes from the Crohn's disease data and simulate phenotypes. Six settings of gene size are considered: d = 5, 20, 50, 100, 300, 500. We randomly select 10 genes containing (exactly or around) d SNPs for each gene size and a total of 60 genes from the Crohn's disease data. For simulating binary phenotypes, $\mathbf{Y}$ is independently generated from Bernoulli(p), where p is the probability parameter and three values of p are examined (p = 1/8, 1/4, 1/2). For simulating quantitative phenotypes, we also consider three covariates in the null model: gender and two principal components for population stratification (Price et al., 2006). Denote the three covariates by Z_1, Z_2 and Z_3. The response variable is simulated according to the null model $Y = a_0 + a_1 Z_1 + \cdots + a_3 Z_3 + \varepsilon$, where the coefficients a_i's are independently generated from $N(0, 0.4)$ for $i = 0, 1, \ldots, 3$. Three distributions of the error term ε are examined: Unif(0, 1), t(4) and Gamma(10, 1), which represent bounded, heavy-tailed, and skewed distributions. These distributions are standardized to have mean 0 and variance 1. We consider two sample sizes according to the Crohn's disease data: the non-Jewish population with n = 997 and the entire data with n = 1719.

The empirical sizes (or type I errors) of the three p-value calculation methods are compared over a range of significance levels: $\alpha = 0.01, 0.001, 0.0001$. For each gene, we first draw 10^6 independent Monte Carlo samples from the Gaussian approximation and permutation null distributions of each statistic, respectively, to calculate their critical values at significant level α. The asymptotic critical values for the minimum p-value method, higher criticism and Berk–Jones tests are computed according to the formulas in the literature (Cai et al., 2014; Donoho and Jin, 2015; Wellner and Koltchinskii, 2003), which are also listed in the supplementary Table 1. Next, we generate 10^6 independent Monte Carlo samples from the true null distribution. Then the empirical size of each method is the proportion of samples greater than the corresponding critical value obtained above. For each value of d, we average the empirical sizes over the 10 genes with the same size.

The results for t distribution are summarized in Table 1 and 2. Additional results for the uniform, gamma and Bernoulli distributions are given in the supplementary Tables 2–11. An empirical size that is closer to the significance level α indicates better performance of the corresponding method. It can be seen that (i) the Gaussian approximation error drops when sample size n increases and/or gene size d decreases; (ii) the asymptotic p-values are wildly inaccurate, especially for the higher criticism and Berk–Jones tests; (iii) the Gaussian approximation error increases very slowly with respect to d, as indicated by the bounds in Theorem 1–3 that depend on d only at the logarithmic rate; (iv) the Gaussian approximation method performs similarly for each of the three test statistics; (v) in general, the Gaussian approximation method is slightly less accurate than permutation, but still provides reasonably accurate p-values for practical uses, even in the situation of small p-values.

In Table 3, we demonstrate the computation time in seconds based on our implementation. The computation was carried out on a computer node with 2.5 GHz quad-core Intel Xeon E3-1284 CPUs and 32 GB memory. Since the computation time does not depend on the specific numbers in a genotype matrix, we use simulated genotypes independently generated from Binomial(2, 0.3) to investigate a broader range of sample size: $n = 1000, 2000, 4000$. Table 3 shows that the computation of Gaussian approximation is much less intensive than permutation, especially for small or moderate d. The computation savings of Gaussian approximation over permutation increases along with the ratio of n/d. For a fixed d, the computation time of Gaussian approximation remains almost the same for different sample sizes, while the time of permutation increases roughly linearly with the sample size n.

To study the computation time in large-scale hypothesis testing, we apply the Gaussian approximation and permutation methods to screen the whole genome in the Crohn's disease data, where 77% of genes have $d \leq 20$ SNPs and the sample size is 1719. Specifically, the p-value of each gene is calculated by simulating 10^6 Monte Carlo samples. The Gaussian approximation method requires hours to complete the computation over the genome, with 3.8, 9.3 and 12.3 hours for the minimum p-value method, higher criticism and Berk-Jones tests, respectively. On the other hand, the permutation method can only complete screening a fraction of the genome within a day. Based on the proportion of genes being processed in one day, we estimate that the permutation method would take 12.0, 12.3 and 12.4 days for the three tests.

To summarize, our simulation study demonstrates the tradeoff between computation efficiency and accuracy. The p-value calculation based on asymptotic distribution requires negligible computation compared to the other two methods, while its precision is very poor. Our proposed Gaussian approximation method slightly sacrifices the accuracy but substantially speeds up the computation in comparison with permutation.

4.2 Real-data analysis

We apply the three tests and p-value calculation methods to analyze a subpopulation of the Crohn's disease data, where the phenotype of interest is the Crohn's disease and the subset of samples are from the non-Jewish population. After quality control, the final data consists of 498 cases and 499 controls.

We specifically study 12 genes that are found to be functionally interesting or associated with Crohn's disease in the literature (Franke et al., 2010). In particular, IL23R and NOD2 are identified as the most significant two genes by all the three tests in this analysis. We use 10^8 Monte Carlo samples for these two genes and 10^6 for the rest to compute the p-values based on permutation and the Gaussian approximation method. The results are summarized in Table 4. It can be seen that the p-values computed by the Gaussian approximation method and permutation are close in general, while the asymptotic p-values are widely off, especially for the higher criticism and Berk–Jones tests. According to both permutation and Gaussian approximation p-values, IL23R and NOD2 are identified as significant genes at a level of 0.05 with Bonferroni correction.

5 Discussion

As can be seen from the simulation, our proposed Gaussian approximation method is particularly accurate and computationally much more efficient than permutation when the sample size n is large and the number of covariates d is small or moderate. In the application of genome-wide association studies, a small number of covariates is the case for the vast majority of genes. There may be a few genes with d comparable to or even larger than n. The computational advantage of Gaussian approximation is not substantial in this situation. Thus, a mixture of both methods might lead to an overall faster computation and accurate p-values. For example, one may consider using the Gaussian approximation method for genes with n/d larger than 2 and permutation for the other genes.

As d grows, the approximation errors for the three tests increase very slowly, more specifically, at a rate of $(\log d)^c$ for some constant $c > 0$. This nice property implies the good performance of Gaussian approximation and is mainly owing to the maximum form of the three test statistics. For other types of test statistics that are functions of the marginal statistics, the Gaussian approximation method can be directly applied, but the performance may be quite poor. For instance, we observe through simulations that the accuracy of Gaussian approximation for the Fisher's combination test (Fisher, 1925) is much worse than that for the three tests considered in this paper.

This research is motivated by large-scale genome-wide data analysis, which requires massive hypothesis testing and hence it is tricky to calculate p-values for powerful tests.

Despite of its motivating example, the proposed Gaussian approximation method is generally applicable to broader statistical problems and is very easy to implement. It is convenient to use the method in many applications of modern high-throughput data analysis, for example, differential gene expression from next generation sequencing and signal detection in engineering.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank the associate editor and the two anonymous referees for their comments and suggestions that have helped greatly improve the paper.

References

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

Ballard DH, Cho J, Zhao H. 2010; Comparisons of multi-marker association methods to detect association between a candidate region and disease. Genetic epidemiology. 34(3):201–212. [PubMed: 19810024]

Barnett IJ, Lin X. 2014; Analytical p-value calculation for the higher criticism test in finite-d problems. Biometrika. 101(4):964–970. [PubMed: 25750439]

Berk RH, Jones DH. 1979; Goodness-of-fit test statistics that dominate the kolmogorov statistics. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete. 47(1):47–59.

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

Chen BE, Sakoda LC, Hsing AW, Rosenberg PS. 2006; Resampling-based multiple hypothesis testing procedures for genetic case-control association studies. Genetic epidemiology. 30(6):495–507. [PubMed: 16755536]

Chernozhukov V, Chetverikov D, Kato K. 2013; Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. The Annals of Statistics. 41(6):2786–2819.

Chernozhukov V, Chetverikov D, Kato K. 2015; Comparison and anti-concentration bounds for maxima of gaussian random vectors. Probability Theory and Related Fields. 162(1–2):47–70.

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

Donoho D, Jin J. 2015; Higher criticism for large-scale inference, especially for rare and weak effects. Statistical Science. 30(1):1–25.

Duerr RH, Taylor KD, Brant SR, Rioux JD, Silverberg MS, Daly MJ, Steinhart AH, Abraham C, Regueiro M, Griffiths A, et al. 2006; A genome-wide association study identifies il23r as an inflammatory bowel disease gene. science. 314(5804):1461–1463. [PubMed: 17068223]

Efron B. 2014; Rejoinder: Estimation and accuracy after model selection. Journal of the American Statistical Association. 109(507):1021–1022.

Fisher, RA. Statistical methods for research workers. Genesis Publishing Pvt Ltd; 1925.

Franke A, McGovern DP, Barrett JC, Wang K, Radford-Smith GL, Ahmad T, Lees CW, Balschun T, Lee J, Roberts R, 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. [PubMed: 21102463]

Li J, Siegmund D, et al. 2015; Higher criticism: *p*-values and criticism. The Annals of Statistics. 43(3): 1323–1350.

Moscovich A, Nadler B, Spiegelman C, et al. 2016; On the exact berk-jones statistics and their *p*-value calculation. Electronic Journal of Statistics. 10(2):2329–2354.

Noé M. 1972The calculation of distributions of two-sided kolmogorov-smirnov type statistics. The Annals of Mathematical Statistics. :58–64.

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. 2006; Principal components analysis corrects for stratification in genome-wide association studies. Nature genetics. 38(8):904–909. [PubMed: 16862161]

Tippett LHC. 1931The methods of statistics. The Methods of Statistics.

Wellner, JA, Koltchinskii, V. High Dimensional Probability III. Springer; 2003. A note on the asymptotic distribution of berk–jones type statistics under the null hypothesis; 321–332.

Wu MC, Kraft P, Epstein MP, Taylor DM, Chanock SJ, Hunter DJ, Lin X. 2010; Powerful snp-set analysis for case-control genome-wide association studies. The American Journal of Human Genetics. 86(6):929–942. [PubMed: 20560208]

Wu Z, Sun Y, He S, Cho J, Zhao H, Jin J. 2014; Detection boundary and higher criticism approach for rare and weak genetic effects. The Annals of Applied Statistics. 8(2):824–851.

Zuo Y, Zou G, Zhao H. 2006; Two-stage designs in case–control association analysis. Genetics. 173(3):1747–1760. [PubMed: 16624925]

J Am Stat Assoc. Author manuscript; available in PMC 2019 May 22.

Table 1

Empirical sizes at $-\log_{10}$ scale over a range of significance levels α. Sample size $n = 997$. The error term $\boldsymbol{e}$ follows standardized t(4).

d	$-\log_{10}(\alpha)$	HC			MinP			BJ		
		GA	Permu	Asym	GA	Permu	Asym	GA	Permu	Asym
5	2	2.00	2.02	1.07	1.99	2.01	2.35	2.00	2.02	1.17
	3	2.98	3.02	1.47	2.97	3.01	3.37	3.01	3.02	1.68
	4	3.98	3.99	1.75	3.97	3.98	4.34	3.99	4.03	2.13
20	2	2.01	2.00	0.92	2.01	2.00	2.28	2.01	2.00	1.02
	3	3.01	3.00	1.24	3.00	3.00	3.31	3.03	3.01	1.41
	4	3.99	3.99	1.49	3.99	3.98	4.34	4.01	4.00	1.75
50	2	2.00	2.01	0.83	2.00	2.01	2.25	2.02	2.01	0.82
	3	2.98	3.00	1.13	2.98	2.99	3.25	3.03	3.01	1.14
	4	3.93	3.98	1.37	3.92	3.99	4.24	4.03	4.00	1.43
100	2	2.00	2.00	0.84	1.99	1.99	2.20	2.03	2.01	0.88
	3	2.98	2.98	1.15	2.97	2.97	3.20	3.05	3.01	1.24
	4	3.92	3.93	1.40	3.91	3.94	4.12	4.05	3.99	1.58
300	2	1.99	2.00	0.74	1.98	1.99	2.17	2.07	2.01	0.68
	3	2.93	2.96	1.04	2.92	2.95	3.13	3.09	3.01	0.97
	4	3.80	3.83	1.30	3.79	3.82	4.00	4.09	4.01	1.24
500	2	1.98	2.00	0.73	1.97	2.00	2.15	2.12	2.01	0.66
	3	2.92	2.94	1.03	2.91	2.94	3.10	3.16	3.02	0.93
	4	3.78	3.80	1.29	3.74	3.78	3.93	4.19	4.01	1.19

Table 2

Empirical sizes at $-\log_{10}$ scale over a range of significance levels α. Sample size $n = 1719$. The error term $\boldsymbol{\varepsilon}$ follows standardized t(4).

d	$-\log_{10}(\alpha)$	HC			MinP			BJ		
		GA	Permu	Asym	GA	Permu	Asym	GA	Permu	Asym
5	2	2.00	2.02	1.07	2.00	2.02	2.36	2.00	2.02	1.17
	3	2.98	3.02	1.47	2.98	3.02	3.39	3.00	3.03	1.67
	4	3.98	3.99	1.75	3.98	3.98	4.36	4.01	4.02	2.13
20	2	2.01	2.03	0.92	2.00	2.03	2.28	2.01	2.03	1.03
	3	3.01	3.04	1.25	3.01	3.05	3.32	3.01	3.03	1.42
	4	3.97	4.01	1.49	3.98	4.00	4.33	4.00	4.03	1.77
50	2	2.00	2.02	0.84	1.99	2.02	2.24	2.01	2.02	0.83
	3	2.98	3.03	1.14	2.98	3.02	3.25	3.03	3.03	1.16
	4	3.93	4.00	1.38	3.91	3.98	4.22	4.02	4.04	1.46
100	2	1.99	2.03	0.84	1.99	2.02	2.20	2.01	2.04	0.89
	3	2.99	3.02	1.15	2.98	3.01	3.21	3.04	3.06	1.25
	4	3.94	3.99	1.40	3.92	3.98	4.21	4.05	4.07	1.59
300	2	1.99	2.02	0.75	1.98	2.01	2.18	2.04	2.04	0.68
	3	2.95	2.99	1.05	2.94	2.99	3.15	3.06	3.06	0.97
	4	3.85	3.88	1.30	3.83	3.85	4.03	4.10	4.08	1.24
500	2	1.99	2.02	0.73	1.98	2.01	2.16	2.07	2.04	0.65
	3	2.94	2.98	1.03	2.93	2.97	3.12	3.10	3.08	0.93
	4	3.81	3.87	1.29	3.80	3.85	3.98	4.11	4.08	1.19

J Am Stat Assoc. Author manuscript; available in PMC 2019 May 22.

Table 3

Computation time in seconds of generating 10^5 independent Monte Carlo samples from the Gaussian approximation or permutation null distribution for a gene with size *d*. The computation time is an average over 10 replications. The columns "Ratio" show the ratio of computation time of permutation to Gaussian approximation.

n	*d*	HC			MinP			BJ		
		GA	Permu	Ratio	GA	Permu	Ratio	GA	Permu	Ratio
1000	5	0.05	5.70	110.59	0.02	4.58	188.30	0.08	5.45	64.29
	20	0.27	5.95	22.12	0.09	5.31	57.18	0.32	5.44	16.76
	50	0.65	5.95	9.12	0.27	5.70	21.28	0.95	6.26	6.58
	100	1.36	6.67	4.90	0.40	6.88	17.19	1.89	7.65	4.04
	300	4.79	10.75	2.24	1.67	7.50	4.48	6.53	12.84	1.97
	500	8.66	13.54	1.56	3.06	9.58	3.13	11.95	16.90	1.41
2000	5	0.05	10.31	214.85	0.02	9.74	421.84	0.07	10.19	136.04
	20	0.28	10.14	35.71	0.13	10.98	85.50	0.34	10.17	29.52
	50	0.58	11.52	19.72	0.27	10.56	38.99	0.79	11.34	14.31
	100	1.44	11.93	8.28	0.48	11.31	23.60	1.82	13.27	7.29
	300	4.45	16.43	3.69	1.72	13.53	7.85	7.35	18.29	2.49
	500	8.48	20.90	2.46	3.10	16.37	5.28	11.88	23.63	1.99
4000	5	0.06	18.73	304.09	0.02	19.06	787.55	0.09	19.05	204.45
	20	0.27	19.37	70.72	0.14	19.05	140.91	0.30	19.29	64.84
	50	0.68	19.01	27.97	0.31	20.21	65.58	0.93	20.72	22.35
	100	1.30	22.71	17.44	0.43	21.34	49.55	1.88	22.59	11.99
	300	4.70	28.33	6.03	1.46	25.02	17.14	6.94	29.96	4.32
	500	8.55	33.69	3.94	3.26	28.99	8.90	11.70	37.02	3.16

Table 4

P-values of 12 genes at the $-log_{10}$ scale. The first two columns are the gene name and the number of SNPs in the gene, respectively.

Genes	SNPs	HC			MinP			BJ		
		GA	Permu	Asym	GA	Permu	Asym	GA	Permu	Asym
IL23R	22	6.03	6.43	+∞	6.11	6.57	5.70	5.84	5.27	+∞
NOD2	8	6.68	7.22	+∞	6.59	7.15	6.63	5.97	6.51	+∞
SMAD3	48	0.10	0.10	0.20	0.27	0.26	0.17	0.15	0.14	0.26
ERAP2	11	0.13	0.13	0.44	0.04	0.04	0.05	0.14	0.13	0.25
IL10	4	0.24	0.24	0.94	0.17	0.16	0.16	0.24	0.24	0.60
IL2RA	23	0.23	0.23	0.50	0.25	0.25	0.15	0.12	0.11	0.25
TYK2	6	0.93	0.93	2.02	0.78	0.78	0.56	0.88	0.88	1.85
BACH2	81	0.34	0.34	0.69	0.86	0.87	0.74	0.44	0.43	0.79
TAGAP	9	0.05	0.06	0.37	0.11	0.11	0.12	0.04	0.04	0.15
FUT2	5	1.71	1.70	5.48	1.59	1.62	1.05	1.42	1.43	5.34
DENND1B	35	0.96	0.96	2.24	0.31	0.31	0.21	1.21	1.19	2.58
DNMT3A	17	0.05	0.05	0.21	0.04	0.04	0.04	0.15	0.15	0.26

J Am Stat Assoc. Author manuscript; available in PMC 2019 May 22.