Bootstrap Restricted Likelihood Ratio Test for the Detection of Rare Variants

Ping Zeng1,* and Ting Wang2

1Department of Epidemiology and Biostatistics, and Center of Medical Statistics and Data Analysis, School of Public Health, Xuzhou Medical College, Xuzhou, Jiangsu, 221004, P. R. China; 2Department of Epidemiology and Biostatistics, School of Public Health, Xuzhou Medical College, Xuzhou, Jiangsu, 221004, P. R. China

Abstract: In this paper the detection of rare variants association with continuous phenotypes of interest is investigated via the likelihood-ratio based variance component test under the framework of linear mixed models. The hypothesis testing is challenging and nonstandard, since under the null the variance component is located on the boundary of its parameter space. In this situation the usual asymptotic chi-square distribution of the likelihood ratio statistic does not necessarily hold. To circumvent the derivation of the null distribution we resort to the bootstrap method due to its generic applicability and being easy to implement. Both parametric and nonparametric bootstrap likelihood ratio tests are studied. Numerical studies are implemented to evaluate the performance of the proposed bootstrap likelihood ratio test and compare to some existing methods for the identification of rare variants. To reduce the computational time of the bootstrap likelihood ratio test we propose an effective approximation mixture for the bootstrap null distribution. The GAW17 data is used to illustrate the proposed test.

Keywords: Rare variants association study, Variance component, Likelihood ratio test, Linear mixed model, Bootstrap test, Distribution approximation.

1. BACKGROUND

Over the past few years due to rapid advances in next-generation sequencing technologies, the detection of rare variants (i.e., minor allele frequency MAF less than 0.01) association with phenotypes of interest has drawn much attention in the literature. However, because of their rarity the frequently-used methods for common variants (MAF greater than 0.01) have remarkably limited power for rare variants; thus a variety of approaches especially for rare variants association analysis have been developed recently [1-5], including the collapsing-based burden test [6-8], the variable allele-frequency threshold test [9], the data-adaptive sum test [10] and the sequence kernel association test (SKAT) [11, 12], and many others.

However, many existing methods for the association of rare variants are subject to a serious power loss when both harmful and protective effects of causal rare variants within a gene are present [12, 13]. It has been shown that an efficient way for handling the directionality problem is to treat the effects of rare variants to be random with a common variance component and then test the significance based on this variance component [12, 14, 15]. By doing so the detection of rare variants is immediately translated into the variance component test under the context of linear mixed effects model [16].

Assume for the moment the variance component for multiple rare variants to be $r^2$; then $H_0: r^2 = 0$ indicates no association exists between rare variants and phenotypes. The hypothesis testing is nonstandard since $r^2$ should be nonnegative and is located on the boundary of its parameter space under the null. In this situation the usual asymptotic $\chi^2$ distribution of the test statistic does not necessarily hold. Under some regularity conditions Self and Liang [17] proved that the likelihood ratio statistic asymptotically followed a 50:50 mixture of $\chi^2_0$ and $\chi^2_1$, where $\chi^2_0$ is a point mass at zero and $\chi^2_1$ is the $\chi^2$ distribution with one degree of freedom; see also Stram and Lee [18] and Liang and Self [19]. However, some authors pointed out that this equal-weight mixture may be not reasonable in practice and may lead to conservative results [20]. To adjust for the conservatism, Pinheiro and Bates [21] advised to use the 65:35 mixture. In fact, determining a proper weight for the mixture of the likelihood ratio statistic is not straightforward [22].

In the literature under complex hypothesis-testing scenarios a natural way to resolve the problem is to resort to the bootstrap technique because of its generality and conceptual simplicity [23-25]. Faraway [26] and Samuh et al. [27] adopted parametric bootstrap methods for the likelihood ratio variance component test in linear mixed models. To examine the variance component in generalized linear mixed...
models, Sinha [28] proposed a score test and conducted it by
means of parametric bootstrap. They have shown the boot-
strap test can maintain the type I error rate at the nominal
level and is more powerful than the method that uses the
usual asymptotic 50:50 mixture. The main advantage of the
bootstrap test is that it avoids the derivation of the complex
null distribution analytically and is easy to implement.

Motivated by these results, in this paper we attempt to
apply the bootstrap test to detect the association of rare vari-
ants in sequencing data. Both parametric and nonparametric
bootstrap likelihood ratio tests are studied. Since the restrict-
ed maximum likelihood estimation generally offers less un-
bias estimates for variance components than the maximum
likelihood estimation [20, 29, 30], thus instead of the likeli-
hood ratio test, we use the restricted likelihood ratio test
(ReLRT) in the paper.

This paper has the following organizations. After intro-
ducing the association of rare variants briefly and defining
the ReLRT statistic, we present the algorithm for the para-
metric and nonparametric bootstrap tests. We implement
extensive numerical studies to evaluate the performance of
the proposed bootstrap likelihood ratio test, and compare it
with other existing methods for the identification of rare vari-
ants. We finally apply the proposed ReLRT to the GAW17
data and give some discussions.

2. MATERIALS AND METHODS

2.1. Linear Mixed Model

Denote \( X = [x_1, x_2, \ldots, x_p] \) covariates, such as smoking
and age; \( Z = [z_1, z_2, \ldots, z_q] \) genotypes of rare variants defined
within a gene, where \( z \) is equal to 0, 1, and 2 representing
the number of minor alleles; the phenotype is \( y \); and the sample
size is \( n \). We use the following linear mixed model [16, 21]
to characterize the relationship between \( y, X \) and \( Z \)

\[
M_i: y_i = X_i \beta + Z_i b_i + e_i,
\]

\[
b_i \sim N(0, \sigma^2_w I),
\]

\[
e_i \sim N(0, \sigma^2),
\]

where \( \beta = [\beta_1, \beta_2, \ldots, \beta_p] \) are fixed effects, \( b = [b_1, b_2, \ldots, b_q] \)
are random effects with variance component \( \tau^2 \), where \( \sigma^2_w \)
is the given weight for rare variant \( j \) and is typically calculated
in light of MAF \([8, 12]\). Following Wu et al. [12], we select-
net \( w = Beta(MAF; 1, 25) \) in the paper, where \( Beta \) is the
beta probability density function with parameters 1 and 25.
This selection implies that rarer variants have more weights.
Clearly, \( \tau^2 = 0 \) suggests that \( b = 0 \) and there is no statistical
association between the genotypes \( Z \) and phenotype \( y \).

Note that model (1) is different from the model consider-
in Fitzmaurice et al. [22] and Sinha [28], where the mixed
effects model was built for the dependent clustered data and
included only one random effect representing the cluster
effect; while in model (1) the phenotype \( y \) is assumed to be
unrelated although the random effects can induce a marginal
correlation among \( y \). In addition, multiple random effects are
contained in model (1).

In model (1) the problem of the identification of rare vari-
ants has been converted into the test of variance component
\( \tau^2 \). This transformation results in at least two useful conse-
quences: (i) it circumvents the directionality problem en-
countered by the burden test and other collapsing-based
methods, accordingly the variance-component based test
generally has a better power when the effects of rare variants
are heterogeneous [12, 31]; (ii) it is a single parameter test
and thus avoids performing the multivariate hypothesis test-
ing by directly deeming \( b \) as fixed effects, which may lead to
a serious loss of degrees of freedom if a large number of rare
variants are included and hence is often less powerful.

Note that the null hypothesis \( H_0: \tau^2 = 0 \) is nonstandard;
the usual asymptotic \( \chi^2 \) distribution does not hold [22, 28].
Under \( H_0 \), model (1) reduces to the general linear model

\[
M_0: y_i = X_i \beta_i + e_{i0},
\]

\[
e_{i0} \sim N(0, \sigma^2).
\]

2.2. Restricted Likelihood Ratio Test

The methods of maximum likelihood (ML) and restricted
maximum likelihood (REML) are two commonly-used ap-
proaches for fitting model (1). It has been proved that REML
provides less biased estimates for variance components
than ML [29, 30]. So we only consider ReLRT for the prob-
lem of testing \( H_0: \tau^2 = 0 \) versus \( H_1: \tau^2 > 0 \). By replacing \( \beta \) and \( \sigma^2 \) with their restricted maximum likelihood estimators,
we can obtain the restricted profile log-likelihood function for
model (1) up to an independent constant

\[
1(\lambda) = \frac{1}{2} \left[ (n-p) \log \left\{ (w \beta')' \left( \lambda Z Z' + I_n \right)^{-1} w \beta \right\} + \log \left\{ \left( \lambda Z Z' + I_n \right)^{-1} w \right\} \right]
\]

where \( \lambda = \lambda Z W W' + I_n \) with \( \lambda = \tau^2 / \sigma^2 \), in which \( I_n \) is the
identity matrix with order \( n \) and \( W \) is a diagonal matrix with
elements \( w \); \( P = I_n - X (X' V X)^{-1} X' V \); where \( Y \), \( X \) and \( Z \)
are respectively the vector of \( y \) and the matrices of \( X \) and \( Z \).
It is easy to see that testing \( \lambda = 0 \) is equivalent to testing \( \tau^2 = 0 \).

The ReLRT statistic is accordingly defined as

\[
T = 2 \left\{ \sup \left[ 1_{M_0}(\lambda) - 1_{M_0}(\hat{\lambda}) \right] \right\}
\]

Noted that \( 1_{M_0} \) in Equation (4) is actually a constant in-
dependent of \( \lambda \). Since \( T \) has a complicated expression, deriv-
ing its null distribution analytically is typically not feasible.
To overcome this difficulty we next turn to the bootstrap
procedure [23, 24].

2.3. Bootstrap Procedure

The bootstrap procedure is conceptually simple and easy
to conduct as long as one can repeatedly obtain bootstrap
samples. First we fit the null model \( M_0 \) in (2), and yield \( \hat{M}_0 \)
with \( \hat{\beta}_0 \), \( \hat{\sigma}_0^2 \) and \( \hat{e}_{i0} \) \( (i = 1, 2, \ldots, n) \), here the subscript 0 indica-
tes these quantities are estimated under \( M_0 \). Sampling on
\( \hat{e}_{i0} \) in the bootstrap procedure is a natural choice; however,
also, as Davison and Hinkley [23] stated that a better strategy is
on the modified residuals

\[
\hat{e}_{i0} = \hat{e}_{i0} / \sqrt{1 - \hat{h}_{i0}},
\]

(5)
where $h_{0ii}$ is the leverage value calculated as the diagonal element of hat matrix $X(X'X)^{-1}X'$. Then $\hat{e}_0$ has a constant variance while $\hat{e}_0$ is heteroscedastic. Under these settings, the general bootstrap procedure can be described as follows.

**Algorithm 1.** The general bootstrap procedure for ReLRT

(i) Calculate the observed value of $T$ according to $M_1$ and $M_0$ using the original data, denote as $T^{obs}$,

(ii) Generate bootstrap samples many times from $M_0$, say $B$, and calculate $T$ for each bootstrap sample using Equation (4), obtain the bootstrap value of $T$, say $T^b, b = 1, 2, \ldots, B$;

(iii) The bootstrap null distribution of $T$ consists of $T^b$, and the Monte Carlo p value of ReLRT is simply estimated as the proportion of $T^b$ equal to or greater than $T^{obs}$.

In the second step if we directly generate bootstrap samples using the fitted null model $M_0$, i.e., if the bootstrap phenotype is created as $Y' \sim N(X\hat{\beta}_0, \sigma^2_0)$, then we are conducting the parametric bootstrap test [26]; on the other hand, if we generate bootstrap samples by resampling the modified residuals $\hat{e}_0$ with replacement, then we are conducting the nonparametric bootstrap test [23]. More specifically, the nonparametric bootstrap test is carried out by creating the bootstrap phenotype $Y^* = X\hat{\beta}_0 + \hat{e}_0^*$, where $\hat{e}_0^*$ is produced through resampling $\hat{e}_0$ with replacement. Note that in the nonparametric bootstrap test if we create $\hat{e}_0^*$ via resampling $\hat{e}_0$ without replacement, then actually we are performing the permutation test [23, 25], and they typically produce similar results.

2.4. Approximate Mixture of the Bootstrap Null Distribution

To reduce the computation time of the proposed bootstrap test, we propose a mixture to approximate the bootstrap null distribution of ReLRT. Following Greven et al. [32], we specify the approximate mixture as

$$T \sim \pi X_0' + (1-\pi) a X_1',$$  \hspace{1cm} (6)

where $\pi$ and $a$ are respectively the proportion and scale parameters, and both of them can be estimated via the method of moment. Let $T^i, i = 1, 2, \ldots, L$ is the vector of the bootstrap value of $T$ via the parametric or nonparametric bootstrap procedure described in Algorithm 1. Then the moment estimate for $a$ is $\hat{a} = T / (1-\hat{a})$, where $T$ is the average of $T^i$ and $\hat{a}$ is the estimate of $a$ via the spectral representation [20, 33]

$$\hat{a} = \operatorname{prob} \left\{ \sum_{j=1}^{k} \mu_j u_j^2 < \left( \frac{1}{n-p} \sum_{j=1}^{k} \mu_j \right)^2 \sum_{j=1}^{n-p} u_j^2 \right\}, \hspace{1cm} (7)$$

where $\mu$'s are the eigenvalues of $WZP_0ZW$ with $P_0 = I_{n-p} X(X'X)^{-1}X'$, $u$'s are independently standard normal random variables. Note that this estimation of $\pi$ is different from the original moment estimation in Greven et al. [32]. The estimate of $\pi$ in (7) is independent of the choice of $L$; hence it has the advantage of numerical stability over the moment estimation.

3. NUMERICAL STUDY AND REAL DATA APPLICATION

3.1. Settings of Numerical Study

The genotypes of rare variants $Z$ were generated via the package COSI based on a coalescent model for European population [12, 34]. The covariates $X = [1, x_1, x_2]$, where $x_1$ and $x_2$ are the standard normal variable and the binary variable with a probability of 0.5, respectively. The phenotype was sampled from the normal distribution

$$y \sim N(-1.0 - 0.5x_1 - 0.5x_2 + \sum_{j=1}^{k} z_j h_j, 1),$$  \hspace{1cm} (8)

where $k_0$ is the number of causal rare variants; i.e., the variants that have nonzero effects on the phenotype. The sample size was set to 200, 400, 600 and 800; in our simulations the average number of rare variants (i.e., $k$) corresponding to these sample sizes are 32, 46, 54 and 61, respectively.

In the bootstrap ReLRT algorithm, $B$ was set to 2000, and for the approximation mixture in (6) we selected $L = 2000, 1500, 1000, 800, 500, 300$ and $100$. We also implemented the simulation-based algorithm for the finite sample null distribution of ReLRT [20, 35, 36], and the number of runs in this algorithm is set to 10000. Besides ReLRT, the burden test, the optimal SKAT (SKAT-O) [37, 38], SKAT [12], the genetic random field (GenRF) model [39, 40] and the mixed effects score test (MiST) [41] were conducted together for comparisons. For type I error simulations the number of replicates was 2000; for power simulations the number of replicates was 1000. The estimated type I error rate and power were calculated as the proportion of $p$ values equal to or less than 0.01.

3.2. Numerical Study 1: Type I Error Control

In model (8) $k_0$ was set to 0. The results are given in (Tables 1-2). In the tables sim corresponds to results of ReLRT computed via the simulation-based algorithm [20], mix0 indicates results of ReLRT computed via the bootstrap test with $B = 2000$, and mix1-mix7 represent results of ReLRT computed via the approximate mixture with $L = 2000, 1500, 1000, 800, 500, 300$ and $100$, respectively.

3.3. Numerical Study 2: Statistical Power Evaluation

For power simulations, only 30% rare variants were causal to reflect the fact that not all the rare variants are related to the phenotype. Like in Wu et al. [12], the effect size $|b|$ was specified to $0.3|\log_{10} MAF|$, which results in an effect size of 1.20 for MAF = 0.0001 and an effect size of 0.60 for MAF = 0.01. To study the influence of the direction of effects of rare variants on the statistical power, we considered two situations: (A) all of the causal rare variants had positive effects, i.e., $b = 0.3|\log_{10} MAF|$; (B) half of the causal rare variants had positive effects, i.e., $b = 0.3|\log_{10} MAF|$ and the rest half had negative effects, i.e., $b = -0.3|\log_{10} MAF|$. The results are given in (Fig. 1-5).
Table 1. Type I error rate for the 50:50 and 65:35 mixtures and other existing methods.

| n   | 50:50 | 65:35 | Burden | SKAT-O | SKAT   | GenRF  | MiST  |
|-----|-------|-------|--------|--------|--------|--------|-------|
| 200 | 0.007 | 0.013 | 0.008  | 0.008  | 0.007  | 0.009  | 0.008 |
| 400 | 0.007 | 0.012 | 0.011  | 0.010  | 0.007  | 0.009  | 0.009 |
| 600 | 0.007 | 0.013 | 0.010  | 0.009  | 0.009  | 0.011  | 0.008 |
| 800 | 0.008 | 0.013 | 0.011  | 0.010  | 0.008  | 0.008  | 0.009 |

Note: here 50:50 and 65:35 correspond to results of ReLRT obtained by using the 50:50 and 65:35 null mixtures.

Table 2. Type I error rate for ReLRT and its approximation with the mixture.

| n   | sim  | mix0 | mix1 | mix2 | mix3 | mix4 | mix5 | mix6 | mix7 |
|-----|------|------|------|------|------|------|------|------|------|
|     |      |      |      |      |      |      |      |      |      |
|     | parametric bootstrap | | | | | | | | |
| 200 | 0.011 | 0.012 | 0.012 | 0.012 | 0.011 | 0.011 | 0.012 | 0.009 | 0.013 |
| 400 | 0.010 | 0.008 | 0.009 | 0.009 | 0.009 | 0.010 | 0.010 | 0.011 | 0.011 |
| 600 | 0.011 | 0.011 | 0.011 | 0.011 | 0.012 | 0.012 | 0.013 | 0.013 | 0.013 |
| 800 | 0.010 | 0.009 | 0.011 | 0.011 | 0.009 | 0.010 | 0.011 | 0.013 | 0.013 |
|     | nonparametric bootstrap | | | | | | | | |
| 200 | 0.009 | 0.011 | 0.011 | 0.011 | 0.010 | 0.011 | 0.011 | 0.012 | 0.016 |
| 400 | 0.009 | 0.009 | 0.011 | 0.011 | 0.011 | 0.011 | 0.009 | 0.009 | 0.013 |
| 600 | 0.010 | 0.010 | 0.011 | 0.010 | 0.011 | 0.011 | 0.011 | 0.012 | 0.016 |
| 800 | 0.007 | 0.010 | 0.011 | 0.010 | 0.010 | 0.010 | 0.010 | 0.012 | 0.014 |

Note: here sim corresponds to results of ReLRT computed via the simulation-based algorithm described by Crainiceanu and Ruppert (2004); mix0 indicates results of ReLRT computed via the bootstrap test with $B = 2000$; and mix1-mix7 represent results of ReLRT computed via the mixture distribution described in Equation (6) with $L = 2000, 1500, 1000, 800, 500, 300$ and 100, respectively.

Fig. (1). Empirical power with the sample size $n = 200$. The first line is for situation $A$ (i.e., all the effects of the causal rare variants are in the same direction) and the second line is for situation $B$ (i.e., both positive and negative effects are present). In each line the first is for the burden test, SKAT-O, SKAT, GenRF and MiST, the second is for the parametric bootstrap ReLRT, and the third is for the nonparametric bootstrap ReLRT. In the figure sim corresponds to results of ReLRT computed via the simulation-based algorithm described by Crainiceanu and Ruppert (2004); mix0 indicates results of ReLRT computed via the bootstrap test with $B = 2000$; and mix1-mix7 represent results of ReLRT computed via the mixture distribution described in Equation (6) with $L = 2000, 1500, 1000, 800, 500, 300$ and 100, respectively.
Fig. (2). Empirical power with the sample size $n = 400$. The first line is for situation $A$ (i.e., all the effects of the causal rare variants are in the same direction) and the second line is for situation $B$ (i.e., both positive and negative effects are present). In each line the first is for the burden test, SKAT-O, SKAT, GenRF and MiST, the second is for the parametric bootstrap ReLRT, and the third is for the nonparametric bootstrap ReLRT. In the figure sim corresponds to results of ReLRT computed via the simulation-based algorithm described by Crainiceanu and Ruppert (2004); mix0 indicates results of ReLRT computed via the bootstrap test with $B = 2000$; and mix1-mix7 represent results of ReLRT computed via the mixture distribution described in Equation (6) with $L = 2000, 1500, 1000, 800, 500, 300$ and $100$, respectively.

Fig. (3). Empirical power with the sample size $n = 600$. The first line is for situation $A$ (i.e., all the effects of the causal rare variants are in the same direction) and the second line is for situation $B$ (i.e., both positive and negative effects are present). In each line the first is for the burden test, SKAT-O, SKAT, GenRF and MiST, the second is for the parametric bootstrap ReLRT, and the third is for the nonparametric bootstrap ReLRT. In the figure sim corresponds to results of ReLRT computed via the simulation-based algorithm described by Crainiceanu and Ruppert (2004); mix0 indicates results of ReLRT computed via the bootstrap test with $B = 2000$; and mix1-mix7 represent results of ReLRT computed via the mixture distribution described in Equation (6) with $L = 2000, 1500, 1000, 800, 500, 300$ and $100$, respectively.

3.4. Data Analysis

We applied the proposed bootstrap ReLRT to the genetic analysis workshop (GAW) 17 data [42, 43]. The GAW17 data has 24487 single nucleotide polymorphisms (SNPs) on 3205 genes across 697 individuals. The MAF ranges from 0.07% to 25.8% and most of the SNPs are rare. The covariates include age, sex and smoking. We used the quantitative trait Q1 as the phenotype in our analysis. Three genes (i.e., $ARNT$, $FLT1$ and $KDR$) listed in (Table 1) in Almasy et al. [43] were analyzed here. The results are given in (Tables 3 and 4).

4. RESULTS

4.1. Results for Numerical Study

The 50:50 and 65:35 mixture distributions for ReLRT typically lead to incorrect type I error rates (Table 1). Specifically,
Fig. (4). **Empirical power with the sample size** $n = 800$. The first line is for situation $A$ (i.e., all the effects of the causal rare variants are in the same direction) and the second line is for situation $B$ (i.e., both positive and negative effects are present). In each line the first is for the burden test, SKAT-O, SKAT, GenRF and MiST, the second is for the parametric bootstrap ReLRT, and the third is for the nonparametric bootstrap ReLRT. In the figure sim corresponds to results of ReLRT computed via the simulation-based algorithm described by Crainiceanu and Ruppert (2004); mix0 indicates results of ReLRT computed via the bootstrap test with $B = 2000$; and mix1-mix7 represent results of ReLRT computed via the mixture distribution described in Equation (6) with $L = 2000, 1500, 1000, 800, 500, 300$ and $100$, respectively.

Fig. (5). **Power loss between situations** $A$ and $B$ for the burden test, SKAT-O, SKAT, GenRF, MiST and ReLRT. Here the power loss is calculated as the difference of power between situations $A$ and $B$ for the corresponding methods. Since the powers of ReLRT calculated via the parametric and nonparametric bootstrap tests as well as their mixture distribution were almost the same, thus here only the power for ReLRT calculated from the parametric bootstrap test with $B = 2000$ was used. From top left to bottom right these four plots respectively correspond to the sample size $n = 200, 400, 600$ and $800$. 
Table 3. The \( p \) values for the three genes in the GWA17 data via the burden test, SKAT-O, SKAT, GenRF, MiST and ReLRT.

|         | Burden | SKAT-O | SKAT  | GenRF | MiST  | ReLRT |
|---------|--------|--------|-------|-------|-------|-------|
| ARNT    | 0.409  | 0.533  | 0.341 | 0.720 | 0.483 | 0.343 |
| FLT1    | 2.286E-3 | 1.034E-4 | 8.266E-5 | 2.589E-2 | 5.320E-5 | 2.000E-4 |
| KDR     | 1.080E-6 | 2.685E-6 | 8.425E-5 | 2.135E-1 | 6.156E-8 | 9.999E-5 |

Note: here the \( p \) value of ReLRT is computed through the simulation-based algorithm of Crainiceanu and Ruppert (2004).

Table 4. The \( p \) values for the three genes in the GWA17 data via ReLRT computed by using the parametric and nonparametric bootstrap procedures.

| Mixture | KDR      | FLT1      | ARNT      |
|---------|----------|-----------|-----------|
|         | pboot    | nboot     | pboot     | nboot     | pboot     | nboot     |
| mix0    | 0.356    | 0.344     | 5.00E-4   | 5.00E-4   | 5.00E-4   | 5.00E-4   |
| mix1    | 0.352    | 0.352     | 4.72E-5   | 1.27E-4   | 5.50E-7   | 9.01E-6   |
| mix2    | 0.353    | 0.352     | 4.86E-5   | 1.92E-4   | 5.54E-7   | 6.78E-6   |
| mix3    | 0.353    | 0.351     | 5.30E-5   | 1.20E-4   | 3.15E-7   | 6.81E-6   |
| mix4    | 0.354    | 0.350     | 5.11E-5   | 1.06E-4   | 7.12E-7   | 3.01E-6   |
| mix5    | 0.356    | 0.350     | 2.93E-5   | 7.27E-5   | 5.51E-7   | 1.45E-6   |
| mix6    | 0.356    | 0.347     | 3.13E-5   | 2.07E-5   | 8.74E-7   | 4.55E-6   |
| mix7    | 0.351    | 0.343     | 1.26E-4   | 2.55E-5   | 4.17E-10  | 6.76E-6   |

Note: pboot and nboot respectively represent the parametric and nonparametric bootstrap procedures; mix0 indicates results of ReLRT computed via the bootstrap test with \( B = 2000 \); and mix1-mix7 represent results of ReLRT computed via the mixture distribution described in Equation (6) with \( L = 2000, 1500, 1000, 800, 500, 300 \) and 100, respectively.

in our simulations the 50:50 mixture usually produces conservative results, i.e., the type I error rate is less than the given nominal level, this is similar to that of Fitzmaurice et al. [22]; in contrast, the 65:35 mixture usually produces liberal results, i.e., the type I error rate is larger than the given nominal level. These results suggest that fixing the proportion parameter in the asymptotic null mixture for ReLRT is inappropriate. In fact, Claeskens [33] and Crainiceanu and Ruppert [20] have found that the mixture proportion parameter is case-specific instead of a fixed constant. For fair comparisons, in the following for power evaluations we do not consider the 50:50 and 65:35 mixtures.

(Table 1) shows that the burden test, SKAT-O, SKAT, GenRF and MiST can correctly control the type I error rate, except that SKAT is slightly conservative when the sample size is small (e.g., \( n = 200 \) and 400). SKAT is a score-based variance component test [44]; its relatively conservative nature has been also observed previously [12, 40].

It is seen from (Table 2) that ReLRT based on the simulation algorithm of Crainiceanu and Ruppert [20] can protect the type I error rate. (Table 2) also shows that both the parametric and nonparametric bootstrap ReLRT methods can effectively control the type I error, except when a very small value of \( L \) (e.g., \( L = 100 \)) is used in the approximation mixture, under which slightly liberal results are obtained.

From (Fig. 1 to Fig. 4), on the whole we can see that when all the causal rare variants are in the same direction (situation \( A \)), the burden test, SKAT-O and MiST have a relatively higher power, and MiST has the highest power; while when both negative and positive effects are present (situation \( B \)), SKAT and ReLRT have better power compared with the other competing methods.

Additionally, it can be seen from (Fig. 1-4) that: (i) the performance of the parametric and nonparametric bootstrap procedures for ReLRT is comparable, little difference is observed; (ii) and the performance of ReLRT based on various methods (e.g., the simulation-based algorithm, the parametric and nonparametric bootstrap procedures and mixture distribution) is largely analogous; (iii) under both situations \( A \) and \( B \), ReLRT uniformly outperforms SKAT; (iv) in our simulations GenRF has the lowest power; (v) when the effects of the causal rare variants are heterogeneous (situation \( B \)), all the methods are subject to the power loss (Fig. 5); however, SKAT and ReLRT suffer less reduction than the others, implying they are relatively robust to the effect heterogeneity.

4.2. Results for the GAW 17 Data

From (Tables 3 and 4) the genes FLT1 and KDR can be viewed as to be statistically associated with the phenotype at the level of 0.01, while ARNT cannot be. From (Table 4) it is observed empirically that when the \( p \) value is large, the proposed approximate mixture leads to the very similar results regardless of the choice of \( L \); when the \( p \) value is small, the mixture may lead to different results. But as seen from the
simulations for a long run the approximate mixture with different choice of \(L\) generally produces comparable results.

5. DISCUSSION

Since its introduction in 1979 by Efron [45], the bootstrap method has become a simple and powerful means for statistical inference and has been widely employed in various scientific problems [23-25]. In this paper we propose the bootstrap-based restricted likelihood ratio test for the identification of rare variants in sequencing data. The simulation results have shown that both the parametric and nonparametric bootstrap ReLRT can correctly maintain the type I error at the nominal significant level, while the usual asymptotic mixture cannot. The two bootstrap procedures typically lead to almost the same results.

Several competing methods for the identification of rare variants, including the burden test, SKAT-O, SKAT and MiST, are also considered for comparisons; actually, these methods are essentially score tests based only on the null model [12, 41]. An attractive feature of these tests is that they are computationally fast [12]. However, tests based only the null model explicitly ignore useful information about the possible alternative model. In contrast, ReLRT needs to fit both the null and alternative models; as a result, it is less computationally efficient. But compared to these existing methods, the proposed ReLRT based on bootstrap resampling has a comparable power or behave better. More importantly, the simulations show that ReLRT is robust in the sense that it suffers from less power loss when the effects of causal rare variants are in mixed directions. This is a satisfactory feature since in practice the true underlying genetic model of diseases is generally not known and tests that are robust against the effect directionality are more preferred.

Despite its generic applicability, the bootstrap test is a computer-intensive statistical method. This problem will be more pronounced for the likelihood ratio test. Varieties of strategies are developed to decrease the computational cost of bootstrap; for example, reducing the number of replicates required by the bootstrap test, subsampling and the \(m\) out of \(n\) bootstrap. Kleiner et al. [46] recently presented an overview of these strategies, and pointed out that none of them are perfect and each has different degrees of shortcomings.

To reduce the computational cost in this paper we employ an approximate mixture null distribution for the bootstrap ReLRT. The simulation results have shown that this mixture performs very well even for a relatively small value of \(L\) (e.g., \(L = 100 - 300\)). The importance is that the computation time of the bootstrap ReLRT is reduced significantly compared with the original bootstrap procedure. For example, compared to the bootstrap procedure with 2000 replicates, the proposed mixture with \(L = 200\) can decrease the computation burden about 10 times while maintaining its desired performance. The performance with various choices of \(L\) has little difference. This property is desired since we are typically not clear about how the magnitude \(L\) should be, and tend to select a relatively small value.

To improve the speed, the modern parallel computing via clusters and multicore processor can be utilized [46, 47]. Recently, a computationally efficient method, called bag of little bootstraps, has been proposed by Kleiner et al. [46]. This method has been shown to have favorable statistical performance for massive data; however, this method cannot be applicable to our setting, because it was initially designed for assessing the uncertainty of the estimates. Extending it to the likelihood ratio test may be an interesting problem needing further investigation.

LIST OF ABBREVIATIONS

| Abbreviation | Description |
|--------------|-------------|
| MAF | Minor allele frequency |
| SKAT | Sequence kernel association test |
| ReLRT | Restricted likelihood ratio test |
| GenRF | Genetic random field |
| SKAT-O | Optimal sequence kernel association test |
| MiST | Mixed effects score test |
| GAW | Genetic analysis workshop |

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

We are grateful to reviewers for their helpful comments and suggestions which substantially improve the manuscript, we are also grateful to supports from the editors. We thank the computation support from the high performance computing center at the School of Public Health of Nanjing Medical University. We would like to thank the Genetic Analysis Workshop 17 (GAW17) committee for providing the GAW17 data. Preparation of the Genetic Analysis Workshop 17 Simulated Exome Data Set was supported in part by NIH grant R01 MH059490 and used sequencing data from the 1000 Genomes Project (www.1000genomes.org). The Genetic Analysis Workshop is supported by NIH grant R01 GM031575. This work was supported by the National Natural Science Foundation of P. R. China (81402765), the Statistical Science Research Project from National Bureau of Statistics of P. R. China (2014LY112), the College Philosophy and Social Science Foundation from Education Department of Jiangsu Province of P. R. China (No. 2013SJB790059, 2013SJD790032), and the Research Foundation from Xuzhou Medical College. The work was also supported by the Cultivation Foundation of Excellent Doctoral Dissertation from the School of Public Health of Nanjing Medical University.

REFERENCES

[1] Moutsianas, L.; Morris, A. P. Methodology for the analysis of rare genetic variation in genome-wide association and re-sequencing studies of complex human traits. Br. Funct. Genom., 2014,13(5), 362-370.
[2] Lee, S.; Abecasis, Gonçalo R.; Boehnke, M.; Lin, X. Rare-Variant Association Analysis: Study Designs and Statistical Tests. Am. J. Hum. Genet., 2014, 95(1), 5-23.
[3] Lange, K.; Papp, J. C.; Sinsheimer, J. S.; Sobel, E. M. Next-Generation Statistical Genetics: Modeling, Penalization, and Optimization in High-Dimensional Data. Annu. Rev. Stat. Appl., 2014, 1(1), 279-300.
[4] Derkach, A.; Lawless, J. F.; Sun, L. Pooled Association Tests for Rare Genetic Variants: A Review and Some New Results. Stat. Sci.,
