mvgS Summary

Irene Zhang

March 21, 2020

1 Introduction

Conceptually, complex traits are thought to result from genetic variation at multiple genes or their regulators, and how this genetic profile interacts with behavioral and environmental factors. In case-control or cohort studies, testing for gene-environment interactions is often performed to identify genetic variants whose impact seems to be modified by different environmental exposures [**thomas2010gene**], although the power of such tests is often low [**cordell2009detecting**]. Traditionally, the phenotipic effect of a gene-environment interaction is estimated by fitting a regression model for the outcome of interest(Y) including both the main effects of a genetic variant(G) and an environmental factor (E), as well as the product between G and E. For a known environmental exposure, genome wide association studies (GWAS) for interaction effects usually scan through millions of single nucleotide polymorphisms (SNPs) using this model; this approach requires adjustment for multiple comparison [**he2017set**]. Results from such single SNP analyses have not proven to be reproducible. Statistical power is low for tests of interaction. Also, the environmental exposure is often measured with error or misclassified, which can reduce the power[**thomas2010gene**, **tchetgen2011robustness**, **cornelis2012gene**, **boonstra2016tests**]. To alleviate penalties from multiple testing, filtering strategies have been adopted to reduce the number of candidate SNPs considered for interaction effects [**kooperberg2008increasing**, **thomas2010gene**] by setting a threshold p-value for significance of the main effects and only considering significant subset of SNPs for interaction analysis.

These approaches do not explicitly consider or use any dependence among genetic variants and can lose power from not capturing those variants whose effects are only manifest through interaction. In an attempt to improve power, it has become popular to aggregate or jointly modelling the effects of all SNPs in a pre-defined set, such as a gene, pathway or specific genomic region. Some previously-proposed methods for testing GxE interaction effects are extension of joint tests of multiple-SNP main effects, for example, by aggregating SNPs in the same gene and testing the interaction between the environment and the aggregated measures, or by aggregating the test statistics of each individual SNP and drawing conclusion on the overall significance [**jiao2013sberia**, **wang2015powerful**, **basu2011comparison**]. // However, these gene based tests have been shown to be biased and have inflated type I error, particularly when the genes and environment are not independent or when the sizes of individual interaction effects are not of comparable magnitude ([**lin2016test**]). In contrast, kernel based methods, including the gene-environment set association test(GESAT)[**lin2013test**], and the interaction sequence kernel association test (iSKAT)[**lin2016test**] have been proposed by extending the sequence kernel association test (SKAT)[**wu2011rare**]. By assuming the individual interaction effects follows a zero-mean distribution, kernel based methods for interaction can test if the variance of the individual effects is different from 0, a one-degree-of-freedom test. We are going to review GESAT and iSKAT in secition 2.3.

All these direct tests assume that the environmental factor is observed and measured without error. However, this is not always the case.. Robust interaction tests are needed to cope with inaccurately-measured exposure as well as the case where the environment factor is not observed at all. Pare et al. [**pare2010use**] noticed that when a SNP is involved in an interaction, with either another SNP or environmental exposure, quantitative trait variance can differ across the SNP genotypes. They therefore proposed a SNP prioritization method based on the variances, using Levene's test to assess the variance heterogeneity. SNPs with insignificant heterogeneity signals are filtered out of further interaction analysis so to reduce the number of tests performed. For a single SNP analyses, Levene's test can be directly applied to a categorical genotype, AA, AB, or BB. However, implementation of this concept when jointly testing multiple SNPs is not straightforward. Simply enumerating haplotype pairs would lead to an exponential increase in the number of categories, and Levene's test would have little power. //

Soave and Sun (2017) proposed a two-stage regression framework that generalized the classic Levene's test to allow for non-categorical genotypes,such as those found, for example, when missing genotypes are imputed. Their method is called the generalized Levene's Scale test (gS). The gS test first fits a main effect model and obtains the residuals, then tests if the variance of the residual vary between genotype groups; this latter is by modeling the absolute residuals versus the genotype. To be precise, the gS using ordinary least squares in the first stage and an

F statistic in the second stage is equivalent to classic Levene's test; and gS with least absolute deviance in the first stage corresponds to the Brown-Forsythe test. Although there are some other popular variance tests, such as the likelihood ratio test(LRT) ([**cao2014versatile**]), they have been shown to be less robust than the Levene's test. Therefore, here we generalize the gS test to multi-SNP version to create a region based test for variance-genotype dependence. As for the gS test, our multi-SNP generalized Levene's scale test (mvgS) does not require a measured environmental factor and hence is robust to missing or poorly measured environmental factors. Also, mvgS can be used for imputed (non-categorical) genotype data.//

In section 2, we first review two direct interaction tests (GESAT and iSKAT) together with the underpinning sequence kernel association test (SKAT). Then we introduce the multi-SNP generalized Levene's scale test (mvgS) and together with the relevant model assumptions, choice of residuals, and the choice of second stage test statistics. Simulations are performed in section 3 to compare the performance of variants of mvgS containing different assumptions. First we estimate type 1 error of all versions of the mvgS tests under a range of error distributions, using a permutation-based empirical simulation design. Then we compare the power of all all versions tests in different settings and select the best combinations of residual type and second stage test statistic. At last, we compare the power of all tests, where the environmental factor is binary or continuous, and with or without measurement error.

Results of this type 1 error simulation study are presented in section 4, and based on the result, we decide to use (xxx residual and xxx as second stage test) in mvgS. We then applied the mvgS test in UK10K data (need more details) and result is presented in section 5. Overall, our proposed multi-variable generalized Levene's scale test can be used as a filter of regions demonstrating possible interaction effects when the interacting factor is not observed; The test has been shown to be robust to measurement error when compared to existing direct interaction tests; and exhibits good power when using the xxx second stage approach.

2 Method

2.1 Review of Direct Interaction Tests

In this section, we are going to introduce the kernel based tests in detail. Both GESAT and iSKAT are extensions of SKAT, originally developed by Wu et al. [**wu2011rare**] to test the main phenotypic effects of a region of SNPs.Suppose the data consist of n independent samples with phenotype Y_i, environment exposure E_i and J candidate SNPs $\mathbf{G}_i = (G_{i1}, \cdots, G_{iJ})^T$. GESAT and iSKAT consider a generalized linear model with interaction effect:

$$g(\mathbb{E}(Y_i)) = \beta_0 + \sum_{j=1}^{J} \beta_{1j} G_{ij} + \beta_2 E_i = \sum_{j=1}^{J} \beta_{3j} G_{ij} E_i \tag{1}$$

and are interested in testing $\mathcal{H}_0 : \beta_{3j} = 0$ for $j = 1, \cdots, J$. GESAT assumes that the interaction effects $\{\beta_{3j}\}$ are independent, following an arbitrary distribution with mean 0 and variance τ^2. The above null hypothesis is then equivalent to $\mathcal{H}_0 : \tau^2 = 0$. Following the idea of SKAT, the test statistics for the variance component τ^2 is:

$$Q_s = (\mathbf{Y} - \widehat{\mu})^{\mathbf{T}} \mathbf{S} \mathbf{S}^{\mathbf{T}} (\mathbf{Y} - \widehat{\mu}) \tag{2}$$

where $S_i = (G_{i1}E_i, G_{i2}E_i, \cdots, G_{iJ}E_i)^T$ and matrix $\mathbf{S} = (S_1, \cdots, S_n)^T$ containing the gene-environment interaction terms. $\widehat{\mu}$ is estimated from a main effect model via ridge regression in case number of SNPs J is large and candidate SNPs are of high linkage disequilibrium. GESAT has advantages over burden type of test where the target region has many noncausal variants or when causal variants have different directions of association. However, if the target region has a high proportion of causal variants with the effects in the same direction, burden tests can be more powerful than GESAT. Those biological knowledge is often unknown prior to the test and an unified optimal test is in needed for the whole genome screening. iSKAT is then proposed as an combination of burden type test and variance component test. For a particular ratio of combination $\rho \in [0, 1]$, iSKAT define the combined statistics:

$$Q_\rho = (\mathbf{Y} - \widehat{\mu})^{\mathbf{T}} \mathbf{S} \mathbf{W} \big((1 - \rho) \mathbf{I} + \rho \mathbf{1} \mathbf{1}^{\mathbf{T}} \big) \mathbf{W}^{\mathbf{T}} \mathbf{S}^{\mathbf{T}} (\mathbf{Y} - \widehat{\mu}) \tag{3}$$

where $\mathbf{W} = diag(w_1, w_2, \cdots, w_j)$ is the diagnoal kernel matrix. Weights w_i may depends on the minor allele frequencies. As ρ is unknown in practice, iSKAT search for the minimum p-value of Q_ρ over the support of ρ and developed the optimal statistics:

$$Q_{iSKAT} = \min_{0 \leq \rho \leq 1} p_\rho \tag{4}$$

The unified test is equivalent to a generalized GESAT test that no longer assumes independent β_{3j}s. It assumes that β_{3j} follows an arbitraty distribution with mean 0, variance τ and pairwise correlation ρ. Theoretically, iSKAT with flat weight, and $\rho = 0$ is equivalent to GESAT.

2.2 Proposed Multi-variable Generalized Levene's Scale Test

Assuming that the true data generating model involves J SNPs with either mean effect, interaction effect, or both:

$$Y_i = \beta_0 + \sum_{j=1}^{J} \beta_{1j} X_{ij,1} + \sum_{j=1}^{J} \beta_{2j} X_{ij,2} + \beta_3 E_i + \sum_{j=1}^{J} \beta_{4j} X_{ij,1} \times E_i + \sum_{j=1}^{J} \beta_{5j} X_{ij,2} \times E_i + \varepsilon_i \tag{5}$$

where $X_{ij,k} = \mathbb{1}(G_{ij} = k)$ for $k = 0, 1, 2$. Assumptions of the avgS test includes: normal error $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ independent from the region $\{G_{ij}\}$. And the environmental factor E_i is independent from the region$\{G_{ij}\}_{j=1}^{J}$.

Here we have the conditional expectation and variance:

$$\begin{aligned} \mathbb{E}\big[Y_i | \{G_{ij}\}\big] &= \beta_0 + \beta_3 \mathbb{E}(E_i) + \sum_{j=1}^{J} \gamma_{1j} X_{ij,1} + \sum_{j=1}^{J} \gamma_2 X_{ij,2} \\ Var\big[Y_i | \{G_{ij}\}\big] &= \Big(\beta_3 + \sum_{j=1}^{J} \beta_{4j} X_{ij,1} + \sum_{j=1}^{J} \beta_{5j} X_{ij,2}\Big)^2 \cdot \sigma_E^2 + \sigma^2 \end{aligned} \tag{6}$$

where $\gamma_{1j} = \beta_{1j} + \beta_{4j}\mathbb{E}(E_i)$, $\gamma_{2j} = \beta_{2j} + \beta_{5j}\mathbb{E}(E_i)$. The regression framework of gS test can be easily generalized to multi-variable version:

(I). Fit the regression model:

$$Y_i = b_0 + \sum_{j=1}^{J} b_{1j} X_{ij,1} + \sum_{j=1}^{J} b_{2j} X_{ij,2}$$

with LAD, and obtain the residuals $e_i = Y_i - \widehat{Y}_i = Y_i - X_i^T \widehat{\beta}$. Under the alternative model (5), the residual follows $e_i \sim \mathcal{N}(0, \sigma_i^2)$ where $\sigma_i^2 = Var[Y_i | \{G_{ij}\}]$. Here we consider the absolute residual following gS test, then the distribution of $d_i = |e_i|$ is a folded normal with expectation associated with G_{ij}:

$$\mathbb{E}[d_i] \propto \sqrt{Var[e_i]} = \sqrt{\Big(\beta_3 + \sum_{j=1}^{J} \beta_{4j} X_{ij,1} + \sum_{j=1}^{J} \beta_{5j} X_{ij,2}\Big)^2 \cdot \sigma_E^2 + \sigma^2} \tag{7}$$

If we are willing to assume $\sigma^2 \ll \Big(\beta_3 + \sum_{j=1}^{J} \beta_{4j} X_{ij,1} + \sum_{j=1}^{J} \beta_{5j} X_{ij,2}\Big)^2 \cdot \sigma_E^2$, an approximate linear relationship is obtained :

$$\mathbb{E}[d_i] \propto \beta_3 \sigma_E + \sum_{j=1}^{J} \beta_{4j} \sigma_E X_{ij,1} + \sum_{j=1}^{J} \beta_{5j} \sigma_E X_{ij,2} \tag{8}$$

If the assumption is not feasible, we consider the squared residule where:

$$\begin{aligned} \mathbb{E}(e_i^2) = Var(e_i) =& \Big(\beta_3 + \sum_{j=1}^{J} \beta_{4j} X_{ij,1} + \sum_{j=1}^{J} \beta_{5j} X_{ij,2}\Big)^2 \cdot \sigma_E^2 + \sigma^2 \\ \text{(simplified to be)} =& \gamma_0 + \sum_{j=1}^{J} \gamma_{1j} X_{ij,1}^2 + \sum_{j=1}^{J} \gamma_{2j} X_{ij,2}^2 \\ & + \sum_{j<k} \theta_{jk}^{1,1} X_{ij,1} X_{ik,1} + \theta^{2,2} X_{ij,2} X_{ik,2} + \sum_{j \neq k} \theta_{j,k}^{1,2} X_{ij,1} X_{ik,2} \end{aligned} \tag{9}$$

where $\gamma_{1j} = \beta_{4j}^2$ and $\gamma_{2j} = \beta_{5j}^2$. And $\theta_{j,k}^{1,1} = \beta_{4j} \cdot \beta_{4k}, \theta_{j,k}^{2,2} = \beta_{5j} \cdot \beta_{5k}, \theta_{j,k}^{1,2} = \beta_{4j} \cdot \beta_{5k}$ for $j, k = 1, \cdots, J$. If all genotypes are categorical, naturally $X_{ij,1} \cdot X_{ij,2} = \mathbb{1}(G_{ij} = 1) \cdot \mathbb{1}(G_{ij} = 2) = 0$ and $X_{ij,1}^2 = X_{ij,1}$, $X_{ij,2}^2 = X_{ij,2}$. Hence we can safely dropping the terms $\sum_{j,k} \theta_{j,k}^{1,2} X_{ij,1} X_{ik,2}$ and reduce the expectation to:

$$\mathbb{E}(e_i^2) \propto \gamma_0 + \sum_j \gamma_{1j} X_{ij,1} + \sum_j \gamma_{2j} X_{ij,2} + \sum_{j<k} \theta_{j,k}^{1,1} X_{ij,1} X_{ik,1} + \sum_{j<l} \theta_{j,k}^{2,2} X_{ij,2} X_{ik,2} \tag{10}$$

(II) For the absolute residual, fit the regression model:

$$d_i = c_0 + \sum_{j=1}^{J} c_{1j} X_{ij,1} + \sum_{j=1}^{J} c_{2j} X_{ij,2}$$

and jointly test:

$$\mathcal{H}_0 : c_{1j} = c_{2j} = 0, \forall j \in [J]$$

which is equivalent to test that no interaction effect exist in any SNPs in the sequence: $\mathcal{H}_0 : \beta_{4j} = \beta_{5j} = 0 \ \forall j \in [J]$.

For the squared residual, simply take $d_i = e_i^2$, fit the same regression model and jointly test: $\mathcal{H}_0 : c_{1j} = c_{2j} = 0, \forall j \in [J]$, is equivalent to test that no interaction effect exist in any SNPs in the sequence: $\mathcal{H}_0 : \beta_{4j}^2 = \beta_{5j}^2 = 0 \ \forall j \in [J]$.

In the original paper (Soave and Sun 2017), the two options of first stage regression and how they equivalent to Levene's test have been addressed. We follow the original paper and use the least absolute deviance (LAD) regression in the first stage. When the number of SNPs $J = 1$, the mvgS test reduced to the original gS test, where are more considerations in this multi-variable generalized scale test:

(1) Choice of residuals in the second stage: squared residual or absolute value. The approximation linearity in absolute residuals is subject to the assumption of negligible σ^2. If it is violated, mvgS would suffer from power loss. Although the squared residuals does not require such assumption, it is sensitive to normality assumption. Loh [**loh1987some**] compared several modification of Levene's test, among which, the power transformation with $p = 2$ is equivalent to using squared residuals in [illegible] stage II. Power transformation has unstable type 1 error when $\varepsilon_i \sim t_{(df)}$, i.e. with heavier tail. Applying normalization on phenotype before analysis can control the type 1 error at a cost of power loss. We consider squared residual as one option and demonstrate the T1E problem in section 3 simulation study.

(2) Choice of joint test in the second stage. No matter which type of residual the test uses in stage 2, it requires a statistical test for $2J$ coefficients jointly. Candidate test statistics include ANOVA F-test as adopted in the original gS test, SKAT and SKAT-O.

(3) Choice of genotype coding: the original gS test uses genotypic coding, and so is our proposed mvgS test. However, if the true data generating model is additive, with F-test in stage 2, mvgS will suffer from power loss due to double the degree of freedom if we chose to use additive regression model.

$$\begin{aligned} &\text{stage I: } Y_i = b_0 + \sum_{j=1}^{J} b_{1j} G_{ij} + b_2 E_i + \sum_{j=1}^{J} b_{3j} G_{ij} \times E_i \\ &\text{stage II: } d_i = c_0 + \sum_{j=1}^{J} c_{1j} G_{ij} \end{aligned} \tag{11}$$

Hence we consider both coding and examine the power difference.

3 Simulation Study

In the simulation study, we use the real genotype data from a small subset of the UK10K data, which contains 1004 independent individuals and the number of copies of minor allele at 10901 SNPs, and simulate the environment and phenotype data. About 10.87% of the SNPs has minor allele frequency above 0.05, which considered as common variants. Majority of the SNPs are rare.

We first take sample size $n = 1004$ (all individuals from the given dateset), and number of the SNPs $J = 11$, following the iSKAT paper. The genotypes $\{G_j\}_{j=1}^{J}$ are from a randomly sampled region from the UK10K data set, with restriction $\mathbf{G}^T\mathbf{G}$ semi-positive definite. Environmental factor is simulated independently from SNPs. We consider binary environmental factor:

$$E_i \sim Bernoulli(0.3) \tag{12}$$

Assumed additive effect, phenotype is generated from:

$$Y_i = \beta_0 + \sum_{j=1}^{J} \beta_{1j} G_{ij} + \beta_2 E_i + \sum_{j=1}^{J} \beta_{3j} G_{ij} \times E_i + \varepsilon_i, \varepsilon_i \sim \mathcal{N}(0, \sigma^2) \tag{13}$$

where $\sigma = 0.17$, β_1 =(0, 0.0343, 0.0131, 0, 0.0017, 0.0169, 0.0305, 0, -0.001, 0.0115, 0), $\beta_2 = 0.15$, and β_3 =(0, 0.0652, 0.1702, 0.0913, 0, -0.0634, 0.2301, 0, 0, 0, -0.0182).

In addition to the regular setting, we consider environmental factor with measurement error, too. For each generated dataset $\{Y_i, E_i, G_{i1}, \cdots, G_{iJ}\}$, we assume some measurement error occurs (at different level) and E_i^{obs} is collected instead of E_i. $E_i^{obs} = E_i$ when there is no measurement error.

Type 1 error is evaluated from a permutation based empirical simulation design. With the generated dataset $\{Y_i, E_i, G_{i1}, \cdots, G_{iJ}\}$, where Y_i is dependent on $\{G_{ij} \times E_i\}$ set, we first randomly permute (Y_i, E_i) pairs and obtain the new set $\{Y_i^{perm}, E_i^{perm}, G_{i1}, \cdots, G_{iJ}\}$ where $Y_i^{perm} \perp\!\!\!\perp G_{ij} \times E_i^{perm}$ for all $j = 1, \cdots, J$. Jointly permuting (Y_i, E_i) pair can maintain the association between Y_i and E_i, and the linkage pattern in $\{G_{ij}\}_{j=1}^{J}$, while breaks the association between Y_i and $\{G_{ij} \times E_i\}_{j=1}^{J}$. And in GWAS study of interaction effects, environment exposure E is always fixed, as the phenotype of interest Y, while sets of SNPs are scanned for signal. Permuting (Y, E) jointly can well approximate the GWAS setting. In each permutation, we evaluate the mvgS with candidate combinations and direct methods GESAT and iSKAT. From 10^5 permutations, we can obtain an empirical null distribution and type 1 error of each test statistics with respect to the particular set of coefficients β and the data generating model 13.

For one generated dataset, we evaluate all candidates/methods once on the original dataset $\{Y_i, E_i, G_{i1}, \cdots, G_{iJ}\}$ and use the quantile of the test statistic with respect to its empirical null distribution as the p-value. Then we repeat the whole process for 1000 times to obtain the empirical power. The simulation design can be summarized below:

1. Generate dataset $\{Y_i, E_i, G_{i1}, \cdots, G_{iJ}\}$;

2. Generate E_i^{obs} from E_i with or without measurement error;

    i For each observed dataset $\{Y_i, E_i^{obs}, G_{i1}, \cdots, G_{iJ}\}$, permute (Y_i, E_i^{obs}) jointly;

    ii Evaluate all candidate mvgS tests and direct tests on each $\{Y_i^{perm}, E_i^{obs,perm}, G_{i1}, \cdots, G_{iJ}\}$, collect the test statistics;

    iii Repeat permutation 10^5 times to obtain the empirical null distributions.

3. Evaluate all 5 methods on $\{Y_i, E_i^{obs}, G_{i1}, \cdots, G_{iJ}\}$, calculate p-value as the quantile of the statistics in its empirical null distribution.

4. Repeat the whole process 1000 times to obtain the empirical power.

We also consider changing parts of the settings to include more possible cases in the simulation study so to first identify the best combination of residual type and stage II statistic within mvgS.

(a) T1E control under $\varepsilon_i \sim t_{(5)}$ in the data generating model (13)

(b) Power evaluation when mvgS uses an additive model, to understand the power loss due to model mis-specification.

(c) Power evaluation when $\varepsilon_i \sim \mathcal{N}(0, 2 * 0.27^2)$, to assess how sensitive mvgS (with absolute residual) is to the assumption of negligible σ^2.

(d) Power evaluation when $\beta_2 = 0.05$, to assess how sensitive mvgS is to the magnitude of unobserved environment effect.

Then to compare the proposed mvgS test with direct interaction test GESAT and iSKAT, we consider 2 type of environment exposure variable: continuous and binary, matched with 5 different situations where we have no measurement error, symmetric measurement errors and asymmetric measurement errors.

(e) Power evaluation for continuous environment exposure $E_i \sim \mathcal{N}(0.3, 0.46)$

  (i) no measurement error: $E_i^{obs} = E_i$

  (ii) 10% measurement error: $E_i^{obs} = E_i + \delta_i, \delta_i \sim \mathcal{N}()$

  (iii) 20% measurement error: $E_i^{obs} = E_i + \delta_i, \delta_i \sim \mathcal{N}()$

  (iv) 10% measurement error, skewed: $E_i^{obs} = E_i + \delta_i, \delta_i \sim \chi^2()$

  (v) 20% measurement error, skewed: $E_i^{obs} = E_i + \delta_i, \delta_i \sim \chi^2()$

(f) Power evluation for binary $E_i \sim Bin(0.3)$

  (i) no measurement error: $E_i^{obs} = E_i$

  (ii) 10% measurement error: $E_i^{obs} = (1 - \alpha_i)E_i + \alpha_i(1 - E_i), \alpha_i \sim Ber(0.1)$

  (iii) 20% measurement error: $E_i^{obs} = (1 - \alpha_i)E_i + \alpha_i(1 - E_i), \alpha_i \sim Ber(0.2)$

  (iv) 10% measurement error, skewed: if $E_i = 0$, $E_i^{obs} \sim Ber(1/3)$, else $E_i^{obs} = E_i$

  (v) 20% measurement error, skewed: if $E_i = 0$, $E_i^{obs} \sim Ber(2/3)$, else $E_i^{obs} = E_i$

4 Result

Table 1 shows the empirical type 1 error evaluation of mvgS test with 6 different combinations from 10^5 replicates. At a 5% nominal level, any empirical T1E within (0.0494, 0506) are considered as correct type 1 error control. Under normal error model, absolute residual with F-test in second stage is the only combination within this range. Absolute residual with SKAT and squared residual with SKAT are considered acceptable since their T1E are slightly below the nominal level, which will give conservative conclusion when apply to real data. SKATO has significant T1E inflation with both residual types. Under t distributed error model, absolute residual with all three candidate test statistics give lower T1E than under normal error model, while squared residual give higher T1E than under normal error model.

	absolute residual $d_i = \|e_i\|$			squared residual $d_i = e_i^2$		
	F-test	SKAT	SKATO	F	SKAT	SKATO
Original	0.0505	0.0478	0.1155	0.0200	0.0469	0.0521
(a) t distributed error	0.0300	0.0476	0.0908	0.0277	0.0539	0.0588

Table 1: Empirical type 1 error simulation for 6 candidate combinations of mvgS test under normal error model and under setting (a) t distributed error.

Within mvgS test, we evaluate the power of each combinations under the genotypic regression model and the additive model. Comparing the performance of two differently coded mvgS, there is around 49% to 67% power loss due to model mis-specification(Table 2). In the same table when we increase the variance of error term in data generating model (setting(b)), the power of mvgS decrease due to the variance explained by observed variables are lessened. Among mvgS, those with absolute residual are affected more than mvgS with squared residual. In setting (d) where β_2 increase from 0.015 to 0.1, the power loss of mvgS with absolute residual varies from 13% to 17%, and from 5% to 9% for mvgS with squared residual.

	absolute residual $d_i = |e_i|$			squared residual $d_i = e_i^2$		
	F-test	SKAT	SKATO	F	SKAT	SKATO
Original	0.259	0.212	0.250	0.232	0.193	0.211
(b) Additive coding	0.673	0.413	0.499	0.695	0.445	0.442
(c) $\sigma^2 = 2 \times 0.27^2$	0.120	0.161	0.194	0.196	0.160	0.171
(d) $\beta_2 = 0.1$	0.214	0.183	0.211	0.212	0.184	0.198

Table 2: Within mvgS power evaluation of 6 combinations of different residual types and stage 2 statistics for original mvgS and for setting (b) additive coded mvgS, (c) larger variance of ϵ_i, and (d) larger environment effect β_2.

Overall, the power of mvgS with squared residual is higher than the original mvgS with absolute residual when the error variance gets larger. As addressed in many previous literature ([**loh1987some**] and xxx), using squared error instead of absolute error in Levene's test gives unstable type 1 error control under non-normal error distribution. We do not carry squared residuals in the following power comparison with direct tests, GESAT and iSKAT.

5 Discussion

- Another powerful method: adaptive combination of Bayes factors method(ADABF) has been proposed in 2018 (Lin et al., 2018) that allows prior information in the test. Others propose mixed effect model instead of GLM in GxE interaction test.

- A summary of permutation testing in regression can be found in [**anderson2001permutation**]. The article summarises permutation testing in models with one and two main effects, and notes that in a model with two main effects and an interaction term there is no exact spermutation method for testing the interaction term. Consider the same data generating model equation

  $$Y_i = \beta_0 + \beta_1 G_i + \beta_2 E_i + \beta_3 G_i E_i + \varepsilon_i, \varepsilon_i \sim \mathcal{N}(0, \sigma^2) \tag{14}$$

  a simple permutation based method would fix G, E and permutes outcome Y to give Y^{perm}, independent of G and E. However, Y is not independent of G and E in model 15. In our simulation design, jointly permutated phenotype Y^{perm} is associated with E^{perm} but still independent from G. Anderson claims that unless the main effect β_1 equals to zero or its exact value being known, joint permutation only approximate a restrictive empirical null distribution. Comparing our design to the real GWAS setting, where we keep the environmental factor E and phenotype Y as fixed, and search through all candidate SNPs for the one with interaction effect. Most of the SNPs have no interaction effect, that is, under the null hypothesis, but with possible main effect. Our proposed permutation cannot honestly keep the main effect of G, thus is not perfect under the case. (need to rephrase this part to show: although not perfect, permuting YE pair is the closest to a real GWAS study.

- Contribution and limitation

Supplement Materials

A short proof of the independence between Y_i^{perm} and $G_i \times E_i^{perm}$ is stated below:

proof:

Consider a single SNP case where the data generating model is

$$Y_i = G_i + E_i + G_i \times E_i + \varepsilon_i \tag{15}$$

Imaging we permute all variables (Y, E, G, ε) together, we can write the permuted phenotype as:

$$Y_i^{perm} = G_i^{perm} + E_i^{perm} + G_i^{perm} E_i^{perm} + \varepsilon_i^{perm}$$

Although G_i is never permuted and G_i^{perm} is not available in application, we use the representation of Y_i^{perm} for the independence derivation. Naturally, the permuted G_i^{perm} is independent of G_i, and ε_i^{perm} is independent of ε_i. From the model assumption, we further have $G_i \perp\!\!\!\perp E_i$, $G_i \perp\!\!\!\perp E_i^{perm}$ and $G_i^{perm} \perp\!\!\!\perp E_i^{perm}$. To evaluate the type 1 error of a particular test of interaction, we need to make sure we are under the null hypothesis that $Y_i^{perm} \perp\!\!\!\perp G_i \times E_i^{perm}$.

$$\begin{aligned}
&Cov(Y_i^{perm}, G_i \times E_i^{perm}) = Cov(G_i^{perm} + E_i^{perm} + G_i^{perm} E_i^{perm} + \varepsilon_i^{perm}, G_i E_i^{perm}) \\
=&Cov(G_i^{perm}, G_i E^{perm}) + Cov(E_i^{perm}, G_i E_i^{perm}) + Cov(G_i^{perm} E^{perm}, G_i E^{perm}) \\
=&\mathbb{E}(E_i^{perm} G_i E_i^{perm}) - \mathbb{E}(E_i^{perm})\mathbb{E}(G_i)\mathbb{E}(E_i^{perm}) \\
&\quad + \mathbb{E}(G_i^{perm} E^{perm} G E^{perm}) - \mathbb{E}(G_i^{perm})\mathbb{E}(E_i^{perm})\mathbb{E}(G_i)\mathbb{E}(E^{perm}) \\
=&\big(1 + \mathbb{E}(G_i)\big) \cdot \mathbb{E}(G_i) \cdot Var(E_i) \\
=&0 \quad \text{(given that } G \text{ is centered and scaled)}
\end{aligned} \tag{16}$$