Statistics and Its Interface Volume 10 (2017) 267–277

Testing for gene-gene interaction in case-control GWAS

Zhongxue Chen

Detecting gene-gene interaction is an important but challenging task in genome-wide association studies (GWASs). To this end, many statistical methods have been proposed in the literature. However, powerful yet robust approaches are yet to be developed. In this paper we study the gene-gene interaction tests for case-control GWASs. A number of powerful tests can be constructed for given situations. We also discuss some tests for the main effects and the overall tests for association between genotype and phenotype. A simulation study is conducted to compare some of the proposed tests with existing methods. A real data application is also conducted to illustrate the use of the proposed tests.

Keywords and phrases: Asymptotically independent, Combining p-values, Odds ratio, Single nucleotide polymorphism.

1. INTRODUCTION

In the past decade, hundreds of genome-wide association studies (GWASs) have been successfully conducted; thousands of single nucleotide polymorphisms (SNPs) which are significantly associated with common complex diseases have been identified [Hindorff, et al.]. However, those genetic markers can only explain a small portion of the variabilities in phonotypes, indicating the missing heritability is yet to be uncovered [Manolio, et al. 2009]. A partial solution of finding the missing heritability is to identify gene-gene interaction in GWASs. In the literature, a large number of statistical tests have been proposed to detect the gene-gene interaction effects [Barhdadi and Dubé 2010; Hu, et al. 2014; Jiao, et al. 2012; Song and Nicolae 2009; Ueki 2014; VanderWeele and Laird 2011; Wu, et al. 2008; Yang, et al. 2009; Yu, et al. 2015].

Many current statistical tests for gene-gene interaction were constructed based on certain assumptions. For example, the fast-epistasis test in PLINK [Purcell, et al. 2007], which collapses the 3×3 genotype counts tables for case and control into 2×2 tables, respectively, assumes additive main and interaction effects. Ueki and Cordell have shown that the variance formula in that test underestimates the variance and, therefore, results in inflated type I error rate [Ueki and Cordell 2012]. Ueki and Cordell further proposed a joint test for detecting gene-gene interaction. However, their test is only valid when at most one SNP has main effect [Yu, et al. 2015].

In general, for case-control GWAS data, the likelihood ratio test (LRT) based on the logistic regression performs well and is recommended [Hu, et al. 2014; Yu, et al. 2015]. Although LRT is a robust test, under some situations, its power can be very low due to its large number (4) of degrees of freedom (df) [Song and Nicolae 2009]; several tests with only 1 df have been proposed in the literature [Barhdadi and Dubé 2010; Jiao, et al. 2012; VanderWeele and Laird 2011]. Another disadvange of the LRT test is that it doesn't have a closed form and may require a large amount of computing time. To overcome this limitation, recently, some Wald test based methods have been developed [Yu, et al. 2015]. Furthermore, under the assumption of additive interaction, a more powerful Wald test with 1 df can be constructed [Yu, et al. 2015].

Studies have shown that there is no uniformly most powerful gene-gene interaction test [Hu, et al. 2014]. Some methods may work better than others under some assumptions. However, if the assumptions are not met, those methods may perform poorly. Therefore, it is important to choose an appropriate method for a given situation. On the other hand, if the prior information about the genetic models is unavailable, robust methods are preferred.

In this paper, we propose some gene-gene interaction tests, which have closed forms and are easy to be computed. The test statistics are obtained through combining information from four asymptotically independent test statistics. Therefore, it is easy to incorporate the prior information about the genetic models to construct a powerful test. Without prior information, robust tests will be obtained. We also propose some statistical tests for the main effects. We show that the interaction tests and the main effect tests proposed in this paper are asymptotically independent. Based on this fact, an overall test for the association between the genotype and phenotype is developed via the technique of combining p-values. We compare the proposed interaction tests and the overall test with some commonly used methods through a simulation study and a real data application.

2. METHODS

In this section, we will review some existing gene-gene interaction tests and then describe the proposed ones. In this paper, we use A, a, and B, b to denote the two alleles for the

Table 1. Date structure (cases, controls) of a pair of SNPs in a case-control GWAS

Genotype		SNP 2		
		BB (h_1)	Bb (h_2)	bb (h_3)
	AA (g_1)	r_1, s_1	r_2, s_2	r_3, s_3
SNP 1	Aa (g_2)	r_4, s_4	r_5, s_5	r_6, s_6
	aa (g_3)	r_7, s_7	r_8, s_8	r_9, s_9

two SNPs, respectively. We also use g_i and h_j $(i, j = 1, 2, 3)$ to denote the genotypes of the two SNPs. For instance, we assume $g_1 = AA$, $h_2 = Bb$. The data from a pair of SNPs can be summarized in the above Table 1, where r_k and s_k are the counts of cases and controls for genotype $g_i h_j$, where $k = 3(i-1) + j$.

There are two different null hypotheses that are of interest in GWASs. The first one is about the interaction between the two SNPs, H_{01}: there is not interaction effect; and the second one is on the overall association, H_{02}: there is no association between the disease status and any of the two SNPs. The alternatives of those two null hypotheses are: the respective null hypothesis is not true. It should be pointed out that many existing gene-gene interaction tests are actually based on the null hypothesis of H_{02}, instead of H_{01}.

2.1 Some existing methods

For the data in Table 1 (a 2×9 table), the commonly used Pearson's overall chi-square test can be applied to test the null hypothesis of H_{02}. To test the null hypothesis of H_{01}, Yang et al. proposed a chi-square partition based method [Yang, et al. 2009]. However, it has been shown that under the null hypothesis of H_{01} their test statistic may not have a chi-square distribution with 4 df as claimed by the authors [Plackett 1962; Yu, et al. 2015]. Therefore, it may have inflated type I error rates under some conditions [Hu, et al. 2014].

Another commonly used test for H_{01} with the data in Table 1 is the LRT test, which is based on the following logistic regression models. We use g and h to denote the genotype for SNP 1 and SNP 2, respectively; and for each SNP, we code the common homozygote as 1, the heterozygote as 2, and the rare homozygote as 3. We consider the full model, M_1:

(1)
$$\begin{aligned}\operatorname{logit}(\pi_{gh}) = \mu &+ \alpha_1 I(g=2) + \alpha_2 I(g=3) + \alpha_3 I(h=2)\\ &+ \alpha_4 I(h=3) + \beta_1 I(g=2) I(h=2)\\ &+ \beta_2 I(g=2) I(h=3) + \beta_3 I(g=3) I(h=2)\\ &+ \beta_4 I(g=3) I(h=3).\end{aligned}$$

And the reduced model M_0:

(2)
$$\begin{aligned}\operatorname{logit}(\pi_{gh}) = \mu &+ \alpha_1 I(g=2)\\ &+ \alpha_2 I(g=3) + \alpha_3 I(h=2) + \alpha_4 I(h=3).\end{aligned}$$

Table 2. Frequency distributions of two SNPs in a case-control GWAS

Genotype	AA BB	AA Bb	AA bb	Aa BB	Aa Bb	Aa bb	aa BB	aa Bb	aa bb	Total
Case	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8	p_9	r
control	q_1	q_2	q_3	q_4	q_5	q_6	q_7	q_8	q_9	s

Denote L_{M1} and L_{M0} the $\log_e$ (i.e., ln) of the estimated maximum likelihood values from the models M_1 and M_0, respectively. The test statistic is defined as:

(3)
$$T_L = -2(L_{M0} - L_{M1}).$$

Under the null hypothesis of H_{01}, T_L has an asymptotic chi-square distribution with 4 df.

Denote $\theta = (\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4)^T$, where $\hat{\beta}_i$ is the MLE for β_i $(i = 1, 2, 3, 4)$ in the model M_1, and Σ_θ the estimated covariance matrix for θ, then under the null hypothesis of H_{01}, the following Wald test statistic, denoted by W, has an asymptotic chi-square distribution with 4 df [Plackett 1962; Yu, et al. 2015].

(4)
$$W = \theta^T \Sigma_\theta^{-1} \theta.$$

If we know the relationship among those β_i's, a more powerful modified Wald test with 1 degree of freedom can be obtained. For instance, if we assume $\theta = \lambda A \mathbf{1}$ with known matrix A, where $\mathbf{1} = (1, 1, 1, 1)^T$, then we can define the following modified Wald test with 1 df [Yu, et al. 2015]:

(5)
$$W_1 = \frac{(\mathbf{1}^T A \Sigma_\theta^{-1} \theta)^2}{\mathbf{1}^T A \Sigma_\theta^{-1} A \mathbf{1}}.$$

For additive interaction, $A = \operatorname{diag}(1, 2, 2, 4)$ [Yu, et al. 2015].

2.2 The proposed methods

For the data in Table 1, we assume both cases and controls follow independent multinomial distributions with probabilities described in Table 2. In other words, Let $R = (R_1, R_2, R_3, R_4, R_5, R_6, R_7, R_8, R_9)^T$ the random numbers of subjects with genotypes $AABB$, $AABb$, …, $aabb$, out of r cases, then R follows a multinomial distribution: $R \sim MN(r, p = (p_1, p_2, p_3, p_4, p_5, p_6, p_7, p_8, p_9))$. Similarly, for $S = (S_1, S_2, S_3, S_4, S_5, S_6, S_7, S_8, S_9)^T$, the random numbers of subjects with genotypes $AABB$, $AABb$, …, $aabb$, out of s controls, then $S \sim MN(s, q = (q_1, q_2, q_3, q_4, q_5, q_6, q_7, q_8, q_9))$.

For the main effect of SNP 1, we define the following two statistics:

$$\begin{aligned}T_1 = &(r_4 + r_5 + r_6)(s_1 + s_2 + s_3)\\ &- (r_1 + r_2 + r_3)(s_4 + s_5 + s_6)\end{aligned}$$

$$\begin{aligned}T_2 &= (r_7+r_8+r_9)\,(s_1+s_2+s_3+s_4+s_5+s_6)\\ &\quad -(r_1+r_2+r_3+r_4+r_5+r_6)\,(s_7+s_8+s_9)\,.\end{aligned}$$

For the main effect of SNP 2, we define the other two statistics:

$$\begin{aligned}T_3 &= (r_2+r_5+r_8)\,(s_1+s_4+s_7)\\ &\quad -(r_1+r_4+r_7)\,(s_2+s_5+s_8)\\ T_4 &= (r_3+r_6+r_9)\,(s_1+s_4+s_7+s_2+s_5+s_8)\\ &\quad -(r_1+r_4+r_7+r_2+r_5+r_8)\,(s_3+s_6+s_9)\,.\end{aligned}$$

Under the null hypothesis of no main effects, it can be shown (see Appendix A for more details) that $E(T_i)=0$ for $i=1,2,3,4$.

Under the null hypothesis of H_{02}, it can be shown that asymptotically the following equalities hold

$$\begin{cases} E\Big[\ln\Big(\frac{r_1r_5s_2s_4}{r_2r_4s_1s_5}\Big)\Big]=0 & (i)\\ E\Big[\ln\Big(\frac{(r_1+r_2)r_6s_3(s_4+s_5)}{r_3(r_4+r_5)(s_1+s_2)s_6}\Big)\Big]=0 & (ii)\\ E\Big[\ln\Big(\frac{(r_1+r_4)r_8(s_2+s_5)s_7}{(r_2+r_5)r_7(s_1+s_4)s_8}\Big)\Big]=0 & (iii)\\ E\Big[\ln\Big(\frac{(r_1+r_2+r_4+r_5)r_9(s_3+s_6)(s_7+s_8)}{(r_3+r_6)(r_7+r_8)(s_1+s_2+s_4+s_5)s_9}\Big)\Big]=0 & (iv)\end{cases}. \tag{6}$$

Similarly, we have the following results.

$$\begin{cases} E[r_1r_5s_2s_4-r_2r_4s_1s_5]=0 & (i)\\ E[(r_1+r_2)r_6s_3(s_4+s_5) & \\ \quad -r_3(r_4+r_5)(s_1+s_2)s_6]=0 & (ii)\\ E[(r_1+r_4)r_8(s_2+s_5)s_7 & \\ \quad -(r_2+r_5)r_7(s_1+s_4)s_8]=0 & (iii)\\ E[(r_1+r_2+r_4+r_5)r_9(s_3+s_6)(s_7+s_8) & \\ \quad -(r_3+r_6)(r_7+r_8)(s_1+s_2+s_4+s_5)s_9]=0 & (iv)\end{cases}. \tag{7}$$

The test statistics for interaction effects can be based on either (6) or (7). For example, based on (7), we define the following statistics:

$$\begin{cases} T_5=r_1r_5s_2s_4-r_2r_4s_1s_5\\ T_6=(r_1+r_2)r_6s_3(s_4+s_5)-r_3(r_4+r_5)(s_1+s_2)s_6\\ T_7=(r_1+r_4)r_8(s_2+s_5)s_7-(r_2+r_5)r_7(s_1+s_4)s_8\\ T_8=(r_1+r_2+r_4+r_5)r_9(s_3+s_6)(s_7+s_8)\\ \qquad -(r_3+r_6)(r_7+r_8)(s_1+s_2+s_4+s_5)s_9\end{cases}. \tag{8}$$

Under the null hypothesis of H_{02}, it is easy to see that $E(T_i)=0$ for $i=5,6,7,8$ (use the facts listed in Appendix A). It is not difficult to obtain the variance-covariance matrix of the above four statistics, T_5, T_6, T_7, and T_8 (see Appendix A). We define the following test statistics:

$$\begin{cases} Z_5=\frac{T_5}{\sqrt{\hat{v}_5}}\\ Z_6=\frac{T_6}{\sqrt{\hat{v}_6}}\\ Z_7=\frac{T_7}{\sqrt{\hat{v}_7}}\\ Z_8=\frac{T_8}{\sqrt{\hat{v}_8}}\end{cases}, \tag{9}$$

where $\hat{v}_i\,(i=5,6,7,8)$ is the estimated variance for T_i, and

$$\begin{aligned}\hat{v}_5 &= r^{(4)}s^{(3)}\left[\hat{p}_1^2\hat{p}_5^2\hat{q}_2\hat{q}_4\,(\hat{q}_2+\hat{q}_4)+\hat{p}_2^2\hat{p}_4^2\hat{q}_1\hat{q}_5\,(\hat{q}_1+\hat{q}_5)\right]\\ &\quad + r^{(3)}s^{(4)}\left[\hat{p}_1\hat{p}_5\,(\hat{p}_1+\hat{p}_5)\,\hat{q}_2^2\hat{q}_4^2+\hat{p}_2\hat{p}_4\,(\hat{p}_2+\hat{p}_4)\,\hat{q}_1^2\hat{q}_5^2\right],\\ \hat{v}_6 &= r^{(4)}s^{(3)}\big[(\hat{p}_1+\hat{p}_2)^2\hat{p}_6^2\hat{q}_3\,(\hat{q}_4+\hat{q}_5)\,(\hat{q}_3+\hat{q}_4+\hat{q}_5)\\ &\quad +\hat{p}_3^2(\hat{p}_4+\hat{p}_5)^2\,(\hat{q}_1+\hat{q}_2)\,\hat{q}_6\,(\hat{q}_1+\hat{q}_2+\hat{q}_6)\,\big]\\ &\quad + r^{(3)}s^{(4)}\big[(\hat{p}_1+\hat{p}_2)\hat{p}_6\,(\hat{p}_1+\hat{p}_2+\hat{p}_6)\,\hat{q}_3^2(\hat{q}_4+\hat{q}_5)^2\\ &\quad +\hat{p}_3(\hat{p}_4+\hat{p}_5)\,(\hat{p}_3+\hat{p}_4+\hat{p}_5)\,(\hat{q}_1+\hat{q}_2)^2\hat{q}_6^2\big],\\ \hat{v}_7 &= r^{(4)}s^{(3)}\big[(\hat{p}_1+\hat{p}_4)^2\hat{p}_8^2\hat{q}_2\hat{q}_7\,(\hat{q}_2+\hat{q}_5+\hat{q}_7)\\ &\quad +(\hat{p}_2+\hat{p}_5)^2\hat{p}_7^2(\hat{q}_1+\hat{q}_4)\hat{q}_8\,(\hat{q}_1+\hat{q}_4+\hat{q}_8)\,\big]\\ &\quad + r^{(3)}s^{(4)}\big[(\hat{p}_1+\hat{p}_4)\hat{p}_8\,(\hat{p}_1+\hat{p}_4+\hat{p}_8)\,(\hat{q}_2+\hat{q}_5)^2\hat{q}_7^2\\ &\quad +(\hat{p}_2+\hat{p}_5)\hat{p}_7\,(\hat{p}_2+\hat{p}_5+\hat{p}_7)\,(\hat{q}_1+\hat{q}_4)^2\hat{q}_8^2\big],\\ \hat{v}_8 &= r^{(4)}s^{(3)}\big[(\hat{p}_1+\hat{p}_2+\hat{p}_4+\hat{p}_5)^2\hat{p}_9^2(\hat{q}_3+\hat{q}_6)(\hat{q}_7+\hat{q}_8)\\ &\quad \times(\hat{q}_3+\hat{q}_6+\hat{q}_7+\hat{q}_8)+(\hat{p}_3+\hat{p}_6)^2(\hat{p}_7+\hat{p}_8)^2\\ &\quad \times(\hat{q}_1+\hat{q}_2+\hat{q}_4+\hat{q}_5)\,\hat{q}_9\,(\hat{q}_1+\hat{q}_2+\hat{q}_4+\hat{q}_5+\hat{q}_9)\,\big]\\ &\quad + r^{(3)}s^{(4)}\big[(\hat{p}_1+\hat{p}_2+\hat{p}_4+\hat{p}_5)\hat{p}_9\\ &\quad \times(\hat{p}_1+\hat{p}_2+\hat{p}_4+\hat{p}_5+\hat{p}_9)\,(\hat{q}_3+\hat{q}_6)^2(\hat{q}_7+\hat{q}_8)^2\\ &\quad +(\hat{p}_3+\hat{p}_6)(\hat{p}_7+\hat{p}_8)\,(\hat{p}_3+\hat{p}_6+\hat{p}_7+\hat{p}_8)\\ &\quad \times(\hat{q}_1+\hat{q}_2+\hat{q}_4+\hat{q}_5)^2\,\hat{q}_9^2\big].\end{aligned}$$

Here, $\hat{p}_i=\frac{r_i}{r}$, $\hat{q}_i=\frac{s_i}{s}$ for $i=1,2,\ldots,9$, $r^{(k)}=r\,(r-1)\ldots(r-k+1)$, $s^{(k)}=s\,(s-1)\ldots(s-k+1)$ for $k=3,4$.

The above test statistics have the following properties.

Theorem 1. *Under the null hypothesis of H_{02}, asymptotically, $Z_I=(Z_5,Z_6,Z_7,Z_8)^T$ follows a multivariate normal distribution, $Z_I\sim MVN(0,I_4)$, where I_4 is the 4×4 identity matrix.*

The proof of Theorem 1 is given in the Appendix A. Theorem 1 indicates that under the null hypothesis of H_{02}, Z_5,Z_6,Z_7,Z_8 are asymptotically independent. Many test statistics for detecting gene-gene interaction effects can be constructed based on this fact. Some possible tests are discussed as follows.

2.2.1 Robust test without any assumption about the structure of the four interaction terms

If we don't know the relationship among the 4 interaction effects, we can use a chi-square test to combine the information obtained from the four asymptotically independent tests statistics Z_5,Z_6,Z_7,Z_8. We define the following test [Chen and Nadarajah 2014]:

$$\chi_4^2=Z_5^2+Z_6^2+Z_7^2+Z_8^2. \tag{10}$$

It is easily seen that the above test χ_4^2 has an asymptotic chi-square distribution with 4 df under the null hypothesis of H_{02} as the four terms in the right side of (8) are asymptotically independently and identically distributed as a chi-square distribution with 1 df.

2.2.2 Weighted Z test when only the directions of the 4 interaction terms are known

If we know the signs (positive or negative) of the interaction effects, we can construct a potentially more powerful test based on the following weighted Z test.

$$Z^D = (I_{\beta_1>0}Z_5 + I_{\beta_2>0}Z_6 + I_{\beta_3>0}Z_7 + I_{\beta_4>0}Z_8)/2, \tag{11}$$

where $I_{\beta_i>0}$ equals 1 if $\beta_i > 0$ and -1 otherwise ($i = 1, 2, 3, 4$). The right-sided p-value of Z^D can be calculated based on the standard normal distribution, which is the asymptotic null sampling distribution of Z^D under the null hypothesis of H_{02}.

2.2.3 Robust test if all the 4 interaction effects have the same but unknown direction

If the four interaction coefficients have the same but unknown direction, we can use a chi-square test with 1 df.

$$\chi_1^2 = \left(\frac{Z_5 + Z_6 + Z_7 + Z_8}{2}\right)^2. \tag{12}$$

Under the null hypothesis of H_{02}, χ_1^2 has an asymptotic chi-square distribution with 1 df.

Alternatively, a more robust test can be constructed based on Fisher's method of combining independent p-values [Fisher 1932; Owen 2009]:

$$W = \max(W_1, W_2), \tag{13}$$

where $W_1 = -2\ln(\Phi(Z_5)\Phi(Z_6)\Phi(Z_7)\Phi(Z_8))$, $W_2 = -2\ln(\Phi(-Z_5)\Phi(-Z_6)\Phi(-Z_7)\Phi(-Z_8))$. The p-value of W can be approximated by [Chen 2013; Chen 2014; Chen, et al. 2013a; Chen, et al. 2014a; Chen, et al. 2012a; Chen, et al. 2013b; Chen, et al. 2014b; Chen, et al. 2012b; Chen and Ng 2012; Chen, et al. 2014c; Owen 2009] $\min(1, 2\chi_8^2(W))$, where $\chi_8^2()$ is the cumulative density distribution (CDF) of the chi-square distribution with 8 df. The test W in (13) is more robust than the test χ_1^2 in (12), especially when some of the Z_i's have small effects.

2.2.4 Powerful test when both the directions and magnitudes of the 4 interaction effects are known

If the directions and magnitudes of all of the 4 interaction coefficients in model (1) are known, a powerful weighted Z test with weights equal to the effect sizes can be constructed for this ideal situation. For instance, under the assumption of additive interaction effects [VanderWeele and Laird 2011; Yu, et al. 2015], we can construct the following test statistic:

$$Z^A = \frac{Z_5 + 2Z_6 + 2Z_7 + 4Z_8}{5}. \tag{14}$$

Under the null hypothesis of H_{02}, asymptotically, Z^A follows the standard normal distribution. Furthermore, if the signs of β_i's are known (e.g., all positive or all negative), an even more powerful one-sided test can be used to calculate the p-value; otherwise, the equivalent chi-square test of $(Z^A)^2$ will be used to calculate the two-sided p-value for Z^A in (14).

In general, under the ideal situation that the signs and the relative magnitudes of the four interaction terms in model (1) are known, a powerful test statistic can be obtained. Suppose $\beta_1 : \beta_2 : \beta_3 : \beta_4 = 1 : \gamma_1 : \gamma_2 : \gamma_3$, the test statistic is constructed as follows:

$$Z^I = \frac{I_{\beta_1>0}Z_5 + I_{\beta_2>0}\gamma_1 Z_6 + I_{\beta_3>0}\gamma_2 Z_7 + I_{\beta_4>0}\gamma_3 Z_8}{\sqrt{1+\gamma_1^2+\gamma_2^2+\gamma_3^2}}. \tag{15}$$

Under the null hypothesis of H_{01}, Z^I asymptotically follows the standard normal distribution, and its p-value can be calculated based on the right-sided test. It is easily seen that when $\gamma_1 = \gamma_2 = \gamma_3 = 1$, (15) becomes (11).

2.3 Interaction tests based on LOR

The above interaction tests are based on (7), the differences of the product of proportions. Alternatively, those tests can be constructed based on (6), the log odds ratios (LOR).

$$\begin{cases} T_5^{LOR} = \ln\left(\frac{r_1 r_5 s_2 s_4}{r_2 r_4 s_1 s_5}\right) \\ T_6^{LOR} = \ln\left(\frac{(r_1+r_2)r_6 s_3(s_4+s_5)}{r_3(r_4+r_5)(s_1+s_2)s_6}\right) \\ T_7^{LOR} = \ln\left(\frac{(r_1+r_4)r_8(s_2+s_5)s_7}{(r_2+r_5)r_7(s_1+s_4)s_8}\right) \\ T_8^{LOR} = \ln\left(\frac{(r_1+r_2+r_4+r_5)r_9(s_3+s_6)(s_7+s_8)}{(r_3+r_6)(r_7+r_8)(s_1+s_2+s_4+s_5)s_9}\right) \end{cases}. \tag{16}$$

The estimate of the variance-covariance matrix of the above statistics can be obtained using the delta method. We define the following test statistics based on LOR.

$$\begin{cases} Z_5^{LOR} = \frac{T_5^{LOR}}{\sqrt{\hat{v}_5^{LOR}}} \\ Z_6^{LOR} = \frac{T_6^{LOR}}{\sqrt{\hat{v}_6^{LOR}}} \\ Z_7^{LOR} = \frac{T_7^{LOR}}{\sqrt{\hat{v}_7^{LOR}}} \\ Z_8^{LOR} = \frac{T_8^{LOR}}{\sqrt{\hat{v}_8^{LOR}}} \end{cases}, \tag{17}$$

where $\hat{v}_5^{LOR} = \frac{1}{r_1} + \frac{1}{r_2} + \frac{1}{r_4} + \frac{1}{r_5} + \frac{1}{s_1} + \frac{1}{s_2} + \frac{1}{s_4} + \frac{1}{s_5}$, $\hat{v}_6^{LOR} = \frac{1}{r_1+r_2} + \frac{1}{r_3} + \frac{1}{r_4+r_5} + \frac{1}{r_6} + \frac{1}{s_1+s_2} + \frac{1}{s_3} + \frac{1}{s_4+s_5} + \frac{1}{s_6}$, $\hat{v}_7^{LOR} = \frac{1}{r_1+r_4} + \frac{1}{r_2+r_5} + \frac{1}{r_7} + \frac{1}{r_8} + \frac{1}{s_1+s_4} + \frac{1}{s_2+s_5} + \frac{1}{s_7} + \frac{1}{s_8}$, and $\hat{v}_8^{LOR} = \frac{1}{r_1+r_2+r_4+r_5} + \frac{1}{r_3+r_6} + \frac{1}{r_7+r_8} + \frac{1}{r_9} + \frac{1}{s_1+s_2+s_4+s_5} + \frac{1}{s_3+s_6} + \frac{1}{s_7+s_8} + \frac{1}{s_9}$.

For the above statistics, we have the following results.

Theorem 2. *Under the null hypothesis of H_{02}, asymptotically the random vector $Z_I^{LOR} = (Z_5^{LOR}, Z_6^{LOR}, Z_7^{LOR}, Z_8^{LOR})^T$ follows a multivariate normal distribution, e.g., $Z_I^{LOR} \sim MVN(0, I_4)$.*

The proof of Theorem 2 is given in the Appendix B.

Many test statistics discussed before can be constructed based on Z_I^{LOR}, instead of Z_I. In fact, $Z_I^{LOR}$and Z_I are highly correlated. We have the following results.

Theorem 3. $Z_I^{LOR} \approx Z_I$ *if* β_i $(i = 1, 2, 3, 4)$ *in model (1) are small.*

The proof of Theorem 3 is given in the Appendix C. When β_i are large, or the sample sizes are small, the variance estimates for Z_I are more accurate than those for Z_I^{LOR}, therefore, in this paper we focus on the interaction tests based on Z_I only.

2.4 Testing for the main effects

For a single SNP in a case-control GWAS, many robust association tests have been proposed in the literature [Chen 2011b; Chen and Ng 2012; Zang, et al. 2010; Zheng and Ng 2008]. Recently, we proposed a robust association test based on the generalized genetic model (GGM) [Chen and Ng 2012], which includes the commonly assumed dominant, recessive, and additive models as special cases. GGM assumes that, under the alternative, the relative risk of g_2 to g_1 is between one and the relative risk of g_3 to g_1. This implies that under the alternative hypothesis of main effects present, T_1 and T_2 have the same sign and so for T_3 and T_4.

The statistics, T_1, T_2, T_3, and T_4, defined in the subsection of the proposed methods are asymptotically independent under the null hypothesis of H_{02}. Their variances v_i $(i = 1, 2, 3, 4)$ can be estimated and we define the following test statistics.

$$(18) \qquad \begin{cases} Z_1 = \frac{T_1}{\sqrt{\hat{v}_1}} \\ Z_2 = \frac{T_2}{\sqrt{\hat{v}_2}} \\ Z_3 = \frac{T_3}{\sqrt{\hat{v}_3}} \\ Z_4 = \frac{T_4}{\sqrt{\hat{v}_4}} \end{cases} .$$

For the above defined test statistics, we have the following properties.

Theorem 4. *Under the null hypothesis of* H_{02}*, asymptotically, the random vector* $Z_M = (Z_1, Z_2, Z_3, Z_4)^T$ *follows a multivariate normal distribution, e.g.,* $Z_M \sim MVN(0, I_4)$*, where*

$$\begin{aligned} \hat{v}_1 &= rs\left(\hat{\bar{p}}_1 + \hat{\bar{p}}_2 + \hat{\bar{p}}_3\right)\left(\hat{\bar{p}}_4 + \hat{\bar{p}}_5 + \hat{\bar{p}}_6\right) \\ &\quad \times \left[(n-2)\left(\hat{\bar{p}}_1 + \hat{\bar{p}}_2 + \hat{\bar{p}}_3 + \hat{\bar{p}}_4 + \hat{\bar{p}}_5 + \hat{\bar{p}}_6\right) + 2\right], \\ \hat{v}_2 &= nrs\big(\hat{\bar{p}}_1 + \hat{\bar{p}}_2 + \hat{\bar{p}}_3 + \hat{\bar{p}}_4 + \hat{\bar{p}}_5 + \hat{\bar{p}}_6\big)\big(\hat{\bar{p}}_7 + \hat{\bar{p}}_8 + \hat{\bar{p}}_9\big), \\ \hat{v}_3 &= rs\left(\hat{\bar{p}}_1 + \hat{\bar{p}}_4 + \hat{\bar{p}}_7\right)\left(\hat{\bar{p}}_2 + \hat{\bar{p}}_5 + \hat{\bar{p}}_8\right) \\ &\quad \times \left[(n-2)\left(\hat{\bar{p}}_1 + \hat{\bar{p}}_4 + \hat{\bar{p}}_7 + \hat{\bar{p}}_2 + \hat{\bar{p}}_5 + \hat{\bar{p}}_8\right) + 2\right], \\ \hat{v}_4 &= nrs\big(\hat{\bar{p}}_1 + \hat{\bar{p}}_4 + \hat{\bar{p}}_7 + \hat{\bar{p}}_2 + \hat{\bar{p}}_5 + \hat{\bar{p}}_8\big)\big(\hat{\bar{p}}_3 + \hat{\bar{p}}_6 + \hat{\bar{p}}_9\big), \end{aligned}$$

$\hat{\bar{p}}_i = \frac{r_i + s_i}{r + s}$ *is the estimate of* $\bar{p}_i = \frac{rp_i + sq_i}{r + s}$.

The proof of Theorem 4 is given in the Appendix D. To test for the main effect of SNP 1, we assume GGM and denote $W_1 = \max(W_{11}, W_{12})$, where $W_{11} = (\chi_1^2)^{-1}(\Phi(Z_1)) + (\chi_1^2)^{-1}(\Phi(Z_2))$, $W_{12} = (\chi_1^2)^{-1}(\Phi(-Z_1)) + (\chi_1^2)^{-1}(\Phi(-Z_2))$, $(\chi_1^2)^{-1}()$ is the inverse of the CDF of the chi-square distribution with 1 df. The p-value for W_1 can be approximated by $P_1 = 2\chi_2^2(W_1)$ [Chen and Nadarajah 2014; Chen and Ng 2012; Owen 2009]. Similarly, the p-value for testing the main effect of SNP 2 can be approximated by $P_2 = 2\chi_2^2(W_2)$, where $W_2 = \max(W_{21}, W_{22})$, and $W_{21} = (\chi_1^2)^{-1}(\Phi(Z_3)) + (\chi_1^2)^{-1}(\Phi(Z_4))$, $W_{22} = (\chi_1^2)^{-1}(\Phi(-Z_3)) + (\chi_1^2)^{-1}(\Phi(-Z_4))$. It should be pointed out that, if the GGM assumption is slightly violated, the main effect tests mentioned above are robust and still have reasonable powers.

2.5 The relationship between the main and the interaction tests and the overall association test

It can be shown that $\mathrm{Cov}(T_1, T_5) = (p_1p_5q_2q_4 - p_2p_4q_1q_5)[r^{(2)}s^{(3)}((q_1+q_2+q_3)-(q_4+q_5+q_6))-r^{(3)}s^{(2)}((p_1+p_2+p_3)-(p_4+p_5+p_6))]$. If the null hypothesis of no interaction, i.e., H_{01}, is assumed, then $\mathrm{Cov}(T_1, T_5) = 0$. In general, under the null hypothesis of no interaction, Z_I and Z_M are independent. Therefore, the above interaction tests based on Z_I and the main effects tests based on Z_M are asymptotically independent under the null hypothesis of H_{02}.

Since the current gene-gene interaction tests are based on different definitions of interaction effects, it is difficult to directly compare their performances. In addition, many times, with data in Table 1, we are interested in testing for the overall association, i.e., for the null hypothesis of H_{02}. We can use the above proposed interaction tests and the main effect tests to obtain an overall test through the techniques of combining independent p-values [Chen 2011a; Chen and Nadarajah 2014; Chen, et al. 2014d; Fisher 1932; Owen 2009]. Suppose the p-value from the interaction test (any one from section 2.2) is P_3. Since those two p-values, P_1, P_2, obtained from the two main effect tests, and P_3 are asymptotically independent under the null hypothesis of H_{02}, many techniques of combining independent p-values can be applied. If any information about the main effects and the interaction effect is available, it should be used. In general, the following robust chi-square test can be used:

$$(19) \quad W = \left(\chi^2_{df1}\right)^{-1}(P_1) + \left(\chi^2_{df2}\right)^{-1}(P_2) + \left(\chi^2_{df3}\right)^{-1}(P_3).$$

Under the null hypothesis of H_{02}, W can be approximated by a chi-square distribution with df $df = df1 + df2 + df3$; therefore, its p-value can be approximated by $p_o = \chi^2_{df}(w)$. Since in GWAS, the effects of interaction are usually relatively small, it is preferable to assign a relatively small number for $df3$. Without prior information, we may choose $df1 = df2 = df3 = 1$.

3. SIMULATION STUDY

In this section, we conduct a simulation study to compare the proposed tests with some existing methods. As

Table 3. Frequencies for genotypes under the assumptions of HWE and LE

Genotype	AA	Aa	aa
BB	$p_A^2 p_B^2$	$2p_A(1-p_A)p_B^2$	$(1-p_A)^2 p_B^2$
Bb	$2p_A^2 p_B(1-p_B)$	$4p_A(1-p_A)p_B(1-p_B)$	$2(1-p_A)^2 p_B(1-p_B)$
bb	$p_A^2(1-p_B)^2$	$2p_A(1-p_A)(1-p_B)^2$	$(1-p_A)^2(1-p_B)^2$

mentioned early, under different assumptions, test statistics with more power can be constructed accordingly. However, in most situation, we don't know the truth and robust tests are preferred. In this simulation study, we only consider robust tests of our proposed methods and some commonly used robust methods recommended in the literature [Hu, et al. 2014]. Specifically, for the proposed tests, we include the interaction test (IT) in (10), the overall test (OT) in (19) with $df1 = df2 = df3 = 1$; for existing tests, we consider the LRT for interaction, or the logistic regression based interaction test (LI), in (3), the overall chi-square test (CS), and the LRT for testing for the overall association (LO), which compares the model (1) and the null model logit $(\pi_{gh}) = \mu$.

We denote $p_A =$ Pr (allele A for SNP 1), the probability of having the allele A for the first SNP 1, $p_B =$ Pr (allele B for SNP 2), the frequencies for the 9 combinations of genotypes are determined as shown in Table 3 when Hardy-Weinberg equilibrium (HWE) and linkage equilibrium (LE) are assumed.

In the simulation study, we assume HWE and LE hold for controls, therefore the frequencies for controls will be determined by p_A, and p_B. As mentioned in the subsection of the proposed methods, we assume the random vectors of cases, R, and controls, S, both follow multinomial distribution, $R \sim MN(r, p = (p_1, p_2, p_3, p_4, p_5, p_6, p_7, p_8, p_9))$, and $S \sim MN(s, q = (q_1, q_2, q_3, q_4, q_5, q_6, q_7, q_8, q_9))$.

We denote relative risks $r_1 = \frac{\Pr(case|Aa)}{\Pr(case|AA)}$, $r_2 = \frac{\Pr(case|aa)}{\Pr(case|AA)}$, $r_{11} = \frac{\Pr((case|Bb)|AA)}{\Pr((case|BB)|AA)}$, $r_{12} = \frac{\Pr((case|bb)|AA)}{\Pr((case|BB)|AA)}$, $r_{21} = \frac{\Pr((case|Bb)|Aa)}{\Pr((case|BB)|Aa)}$, $r_{22} = \frac{\Pr((case|bb)|Aa)}{\Pr((case|BB)|Aa)}$, $r_{31} = \frac{\Pr((case|Bb)|aa)}{\Pr((case|BB)|aa)}$, and $r_{32} = \frac{\Pr((case|bb)|aa)}{\Pr((case|BB)|aa)}$. Note that r_1 and r_2 are the marginal relative risks of genotypes Aa and aa to AA; r_{i1} and r_{i2} $(i = 1, 2, 3)$ are the conditional relative risks of genotypes Bb and bb to BB when the genotype for SNP 1 is AA, Aa, and aa, respectively. Given the above relative risks and the frequencies for controls, the frequencies for cases can be determined.

Denote $p_{01} = p_1 + p_2 + p_3$, $p_{02} = p_4 + p_5 + p_6$, $p_{03} = p_7 + p_8 + p_9$, $q_{01} = q_1 + q_2 + q_3$, $q_{02} = q_4 + q_5 + q_6$, $q_{03} = q_7 + q_8 + q_9$, then we have [Chen and Ng 2012]

$$\begin{cases} p_{01} = \frac{q_{01}}{q_{01}+r_1 q_{02}+r_2 q_{03}} \\ p_{02} = \frac{r_1 q_{02}}{q_{01}+r_1 q_{02}+r_2 q_{03}} \\ p_{03} = \frac{r_2 q_{03}}{q_{01}+r_1 q_{02}+r_2 q_{03}} \end{cases}.$$

Further, denote

$$\begin{cases} p_{11} = \frac{p_1}{p_{01}} \\ p_{12} = \frac{p_2}{p_{01}} \\ p_{13} = \frac{p_3}{p_{01}} \end{cases},$$

and

$$\begin{cases} q_{11} = \frac{q_1}{q_{01}} \\ q_{12} = \frac{q_2}{q_{01}} \\ q_{13} = \frac{q_3}{q_{01}} \end{cases},$$

then we have

$$\begin{cases} p_{11} = \frac{q_{11}}{q_{11}+r_{11}q_{12}+r_{12}q_{13}} \\ p_{12} = \frac{r_{11}q_{12}}{q_{11}+r_{11}q_{12}+r_{12}q_{13}} \\ p_{13} = \frac{r_{12}q_{13}}{q_{11}+r_{11}q_{12}+r_{12}q_{13}} \end{cases},$$

or

$$\begin{cases} p_1 = p_{01}\frac{q_{11}}{q_{11}+r_{11}q_{12}+r_{12}q_{13}} = \frac{q_{01}}{q_{01}+r_1 q_{02}+r_2 q_{03}} \frac{q_1}{q_1+r_{11}q_2+r_{12}q_3} \\ p_2 = p_{01}\frac{r_{11}q_{12}}{q_{11}+r_{11}q_{12}+r_{12}q_{13}} = \frac{q_{01}}{q_{01}+r_1 q_{02}+r_2 q_{03}} \frac{r_{11}q_2}{q_1+r_{11}q_2+r_{12}q_3} \\ p_3 = p_{01}\frac{r_{12}q_{13}}{q_{11}+r_{11}q_{12}+r_{12}q_{13}} = \frac{q_{01}}{q_{01}+r_1 q_{02}+r_2 q_{03}} \frac{r_{12}q_3}{q_1+r_{11}q_2+r_{12}q_3} \end{cases}.$$

Similarly, we have the following results

$$\begin{cases} p_4 = p_{02}\frac{q_{21}}{q_{21}+r_{21}q_{22}+r_{22}q_{23}} = \frac{r_1 q_{02}}{q_{01}+r_1 q_{02}+r_2 q_{03}} \frac{q_4}{q_4+r_{21}q_5+r_{22}q_6} \\ p_5 = p_{02}\frac{r_{21}q_{22}}{q_{21}+r_{21}q_{22}+r_{22}q_{23}} = \frac{r_1 q_{02}}{q_{01}+r_1 q_{02}+r_2 q_{03}} \frac{r_{21}q_5}{q_4+r_{21}q_5+r_{22}q_6} \\ p_6 = p_{02}\frac{r_{22}q_{23}}{q_{21}+r_{21}q_{22}+r_{22}q_{23}} = \frac{r_1 q_{02}}{q_{01}+r_1 q_{02}+r_2 q_{03}} \frac{r_{22}q_6}{q_4+r_{21}q_5+r_{22}q_6} \end{cases},$$

and

$$\begin{cases} p_7 = p_{03}\frac{q_{31}}{q_{31}+r_{31}q_{32}+r_{32}q_{33}} = \frac{r_2 q_{03}}{q_{01}+r_1 q_{02}+r_2 q_{03}} \frac{q_7}{q_7+r_{31}q_8+r_{32}q_9} \\ p_8 = p_{03}\frac{r_{31}q_{32}}{q_{31}+r_{31}q_{32}+r_{32}q_{33}} = \frac{r_2 q_{03}}{q_{01}+r_1 q_{02}+r_2 q_{03}} \frac{r_{21}8}{q_7+r_{31}q_8+r_{32}q_9} \\ p_9 = p_{03}\frac{r_{32}q_{33}}{q_{31}+r_{31}q_{32}+r_{32}q_{33}} = \frac{r_2 q_{03}}{q_{01}+r_1 q_{02}+r_2 q_{03}} \frac{r_{22}q_9}{q_7+r_{31}q_8+r_{32}q_9} \end{cases}.$$

In the simulation study, we assume HWE hold in controls for both SNPs with minor allele frequency (MAF) equals 0.3 and 0.5. For the relative risks of SNP 1, we assume $r_2 = 1.4$ and r_1 takes values 1, 1.1, 1.2, 1.3, 1.4. For the relative risks of SNP 2 at each genotype of SNP 1, we assume $r_{i2} = 1.4$, and $r_{i1} = 1$, 1.2, or 1.4, for $i = 1$, 2, 3. We simulate 5000 cases and 5000 controls and estimate the empirical type I error rate using 1000 replicates and significance level of 0.05. To make the comparisons appreciable, when estimate the empirical powers of the interaction tests and the overall association tests, we use significance level of 0.05 and 10^{-16}, respectively.

Table 4 reports the empirical type I error rate for each of the methods compared when the null hypothesis of no

Table 4. Empirical type I error rates for the interaction and the overall association tests under the null hypothesis of H_{02} based on 1000 replicates with 5000 cases and 5000 controls, and the nominal significance level of 0.05

p_A, p_B	IT	LI	OT	CS	LO
0.3, 0.3	0.054	0.054	0.061	0.052	0.053
0.3, 0.5	0.046	0.048	0.047	0.047	0.047
0.5, 0.5	0.059	0.058	0.053	0.054	0.054

Table 5. Empirical type I error rates of the interaction tests (use significance level of 0.05) and powers of the overall association tests (use significance level of 1e-16) when no interaction but only one main effect presents with $r_1 = 1, 1.2, 1.4$, and $r_2 = 1.4$

p_A, p_B	r_1, r_2	IT	LI	OT	CS	LO
0.3, 0.3	1, 1.4	0.045	0.047	0.552	0.180	0.187
0.3, 0.3	1.2, 1.4	0.048	0.052	0.000	0.000	0.000
0.3, 0.3	1.4, 1.4	0.046	0.050	0.000	0.000	0.000
0.3, 0.5	1, 1.4	0.044	0.046	0.225	0.045	0.047
0.3, 0.5	1.2, 1.4	0.043	0.046	0.001	0.000	0.000
0.3, 0.5	1.4, 1.4	0.042	0.046	0.056	0.010	0.012
0.5, 0.5	1, 1.4	0.062	0.064	0.233	0.050	0.053
0.5, 0.5	1.2, 1.4	0.055	0.057	0.001	0.000	0.001
0.5, 0.5	1.4, 1.4	0.041	0.046	0.059	0.024	0.024

Table 6. Empirical type I error rates of the interaction tests (use significance level of 0.05) and powers of the overall association tests (use significance level of 1e-16) when no interaction but two main effect present with $p_A = 0.5$, $p_B = 0.5$, $r_1 = 1, 1.2, 1.4$, $r_2 = 1.4$, $r_{i1} = 1, 1.2, 1.4$, and $r_{i2} = 1.4$

r_1, r_2	r_{i1}, r_{i2}	IT	LI	OT	CS	LO
1, 1.4	1, 1.4	0.039	0.044	0.940	0.900	0.904
1, 1.4	1.2, 1.4	0.035	0.039	0.716	0.599	0.612
1, 1.4	1.4, 1.4	0.049	0.056	0.916	0.835	0.836
1.2, 1.4	1, 1.4	0.051	0.059	0.739	0.630	0.638
1.2, 1.4	1.2, 1.4	0.050	0.050	0.332	0.212	0.218
1.2, 1.4	1.4, 1.4	0.041	0.042	0.554	0.414	0.417
1.4, 1.4	1, 1.4	0.048	0.051	0.882	0.822	0.828
1.4, 1.4	1.2, 1.4	0.056	0.058	0.587	0.454	0.457
1.4, 1.4	1.4, 1.4	0.042	0.043	0.805	0.673	0.682

overall association, H_{02}, is assumed. It shows that under this assumption all methods control type I error rate.

When only the main effect of SNP 1 (Table 5), or both main effects (Table 6), but no interaction effects present, the empirical type I error rates for the interaction tests (IT, LI) and the empirical powers for the overall association tests (OT, CS, and LO) are listed in Table 5 and Table 6, respectively. It can be seen that both interaction tests control type I error rate quite well. Furthermore, in these situations, the proposed overall association test is usually more powerful than the other two overall association tests.

Table 7. Empirical powers of the interaction tests (use significance level of 0.05) and the overall association tests (use significance level of 1e-16) when interaction effects present with $p_A = 0.5$, $p_B = 0.5$, and $r_{i2} = 1.4$

r_1, r_2	r_{11}, r_{21}, r_{31}	IT	LI	OT	CS	LO
1, 1	1, 1.2, 1.4	0.657	0.652	0.007	0.004	0.004
1, 1.4	1, 1.2, 1.4	0.660	0.667	0.827	0.757	0.766
1.2, 1.4	1, 1.2, 1.4	0.630	0.646	0.415	0.342	0.346
1.4, 1.4	1, 1.2, 1.4	0.614	0.626	0.666	0.597	0.604

Table 8. (Number of cases)/(Number of controls) in each of the two-locus genotypes, data were from Sha et al. [Sha, et al. 2009]

Genotype		SNP 1		
		TT	TC	CC
SNP 2	AA	11/23	14/37	3/7
	AG	29/50	73/56	29/11
	GG	23/45	65/24	28/16
SNP 3	CC	33/95	95/89	37/30
	CA	29/20	52/25	22/4
	AA	1/3	5/3	1/0

Finally, Table 7 lists the empirical powers for all of the tests considered in the simulation study when interaction effects present and the main effect of SNP 1 either presents ($r_1 = r_2 = 1$) or not ($r_2 = 1.4$). It clearly shows that the performance of the two interaction tests are very similar. In addition, the proposed overall association test has larger empirical powers than the other two overall association tests. We also simulated data when the genetic models various (i.e., different values for the marginal and conditional relative risks), we saw similar patterns.

4. REAL DATA APPLICATION

We apply the proposed tests to the data presented in Table 2 of the paper by Sha et al. [Sha, et al. 2009], which studied the interaction effects between pairs of SNPs using the GWAS data set of sporadic Amyotrophic lateral sclerosis (ALS) [Schymick, et al. 2007]. Three previously reported associated SNPs were found to have possible interactions. They were rs4363506 (SNP 1), rs3733242 (SNP 2), and rs16984239 (SNP 3), among which SNP 1 was thought to have interaction with both SNP 2 and SNP 3. The data were summarized in Table 8.

Table 9 lists the p-value obtained from each method. Although the overall association tests each has a small p-value, indicating the existence of the overall association between each pair of SNPs and the disease, the interaction tests (IT and LI) only identified the significant interaction between SNP 1 and SNP 2, but not SNP 1 and SNP 3, at the significance level of 0.05. The eight statistics, z_1, z_2, $\ldots$, z_8, are listed in Table 10. For the four interaction test statistics

Table 9. P-value obtained by each of the method for the data

Pair of SNPs	IT	LI	OT	CS	LO
SNP 1 & SNP 2	3.07e-02	1.08e-02	9.55e-11	2.53e-10	1.12e-10
SNP 1 & SNP 3	3.41e-01	3.29e-01	1.19e-10	1.82e-09	4.29e-10

Table 10. Test statistics obtained by each of the method for the data

Pair of SNPs	z_1	z_2	z_3	z_4	z_5	z_6	z_7	z_8
SNP 1 & SNP 2	4.51	2.83	3.87	2.56	1.59	2.37	1.18	-1.07
SNP 1 & SNP 3	4.51	2.83	5.05	0.24	-1.57	0.57	0.94	0.93

(z_5, z_6, z_7, and z_8), their absolute values were not large for each of the two pairs of SNPs, indicating the interaction effects are either small or zero. Furthermore, for each of the pairs of statistics, z_1 and z_2, z_3 and z_4, they have the same direction (both positive), indicating the GGM is valid for both main effects. For the pair of SNP 1 and SNP 2, suppose we know the signs of β_1, β_2, β_3, and β_4 as the ones listed in Table 10 (i.e., $+,+,+,-$) before we see the data, we then apply the test Z^D defined in (11) to test for the interaction effect. We obtain the test statistic $z^D = 3.1$, and a much smaller one-sided p-value 0.00097.

5. DISCUSSION AND CONCLUSION

In this paper we proposed a number of interaction tests based on the four asymptotically independent statistics. Many powerful tests under certain situations can be developed based on those statistics. However, without any prior information about the genetic model, robust methods are recommended.

When both the counts of cases and controls are zero for a given genotype, the test statistics should be corrected accordingly. For instance, if in Table 1 $r_1 = s_1 = 0$, then $Z_5 = 0$, the robust interaction test in (10) should base on Z_6, Z_7, and Z_8 only, and the df should be 3, instead of 4. Similarly, the Pearson's chi-square test will have 7, instead of 8, df in this situation.

Although many gene-gene interaction tests have been proposed for GWASs in the literature, they were constructed based on different definitions of interaction, making direct comparison of their performances difficult. In this case, it might be more appropriate to compare the performance of the overall tests. Furthermore, the findings of current gene-gene interactions are statistical method-biased. More robust gene-gene interaction tests are desired. On the other hand, powerful tests with specific assumptions may be preferable when the information about the genetic models is available from prior studies. To this end, the proposed statistical tests in the paper may provide useful tools.

ACKNOWLEDGEMENTS

The author is grateful to the editor, the associate editor and referees for their helpful comments and suggestions that lead to improvement of the paper. The author also acknowledges the support from the internal research funds awarded by Indiana University School of Public Health-Bloomington.

APPENDIX A. PROOF OF THEOREM 1

Suppose $W = (W_1, W_2, \cdots, W_k)^T$ is a k-dimensional multinomial variable with parameters $p = (p_1, p_2, \cdots, p_k)^T$ and cluster size m. Denote $x^{(a)} = x(x-1)\ldots(x-a+1)$, we have the following results [Mosimann 1962]:

$$(i)\ E\left(W_i\right) = mp_i,\ i = 1, 2, \ldots, k.$$

$$(ii)\ E\left(W_iW_j\right) = m^{(2)}p_ip_j,\ i \neq j.$$

$$(iii)\ E\left(W_i^2\right) = m^{(2)}p_i^2 + mp_i,\ i = 1, 2, \ldots, k.$$

$$(iv)\ E\left(W_iW_jW_l\right) = m^{(3)}p_ip_jp_l,\ i \neq j \neq l.$$

$$(v)\ E\left(W_i^2W_j\right) = m^{(3)}p_i^2p_j + m^{(2)}p_ip_j,\ i \neq j.$$

$$(vi)\ E\left(W_i^3\right) = m^{(3)}p_i^3 + 3m^{(2)}p_i^2 + mp_i,\ i = 1, 2, \ldots, k.$$

$$(vii)\ E\left(W_iW_jW_lW_r\right) = m^{(4)}p_ip_jp_lp_r,\ i \neq j \neq l \neq r.$$

$$(viii)\ E\left(W_i^2W_jW_l\right) = m^{(4)}p_i^2p_jp_l + m^{(3)}p_ip_jp_l,\ i \neq j \neq l.$$

$$\begin{aligned}(ix)\ E\left(W_i^2W_j^2\right) &= m^{(4)}p_i^2p_j^2 + m^{(3)}(p_i^2p_j + p_ip_j^2)\\ &\quad + m^{(2)}p_ip_j,\ i \neq j.\end{aligned}$$

Based on the above results, let $v_i = \text{var}\left(T_i\right)$ $(i = 5, 6, 7, 8)$, where T_i's are defined in (8), it is easy to show that $E\left(T_i\right) = 0$. Therefore $v_i = \text{var}\left(T_i\right) = E(T_i^2)$, and $\text{cov}\left(T_i, T_j\right) = E\left(T_iT_j\right)$. Some algebra show the following:

$$\text{(a)}\ v_5 = a_{51} + a_{52} + a_{53} + a_{54} + a_{55} + a_{56} + a_{57} + a_{58},$$

where

$$\begin{aligned}
a_{51} &= r^{(4)}s^{(3)}\left[p_1^2p_5^2q_2q_4(q_2+q_4) + p_2^2p_4^2q_1q_5(q_1+q_5)\right],\\
a_{52} &= r^{(4)}s^{(2)}\left[p_1^2p_5^2q_2q_4 + p_2^2p_4^2q_1q_5\right],\\
a_{53} &= r^{(3)}s^{(4)}\left[p_1p_5\left(p_1+p_5\right)q_2^2q_4^2 + p_2p_4\left(p_2+p_4\right)q_1^2q_5^2\right],\\
a_{54} &= r^{(3)}s^{(3)}\left[p_1p_5\left(p_1+p_5\right)q_2q_4\left(q_2+q_4\right)\right.\\
&\qquad \left. + p_2p_4\left(p_2+p_4\right)q_1q_5\left(q_1+q_5\right)\right],\\
a_{55} &= r^{(3)}s^{(2)}\left[p_1p_5\left(p_1+p_5\right)q_2q_4 + p_2p_4\left(p_2+p_4\right)q_1q_5\right],\\
a_{56} &= r^{(2)}s^{(4)}\left[p_1p_5q_2^2q_4^2 + p_2p_4q_1^2q_5^2\right],
\end{aligned}$$

$$\begin{aligned}
a_{57} &= r^{(2)}s^{(3)}\left[p_1p_5q_2q_4\left(q_2+q_4\right)+p_2p_4q_1q_5\left(q_1+q_5\right)\right],\\
a_{58} &= r^{(2)}s^{(2)}\left[p_1p_5q_2q_4+p_2p_4q_1q_5\right].
\end{aligned}$$

(*b*) $v_6 = a_{61}+a_{62}+a_{63}+a_{64}+a_{65}+a_{66}+a_{67}+a_{68}$, where

$$\begin{aligned}
a_{61} &= r^{(4)}s^{(3)}\bigl[(p_1+p_2)^2p_6^2q_3(q_4+q_5)(q_3+q_4+q_5)\\
&\quad + p_3^2(p_4+p_5)^2(q_1+q_2)q_6(q_1+q_2+q_6)\bigr],\\
a_{62} &= r^{(4)}s^{(2)}\bigl[(p_1+p_2)^2p_6^2q_3\left(q_4+q_5\right)\\
&\quad + p_3^2(p_4+p_5)^2\left(q_1+q_2\right)q_6\bigr],\\
a_{63} &= r^{(3)}s^{(4)}\bigl[(p_1+p_2)p_6\left(p_1+p_2+p_6\right)q_3^2(q_4+q_5)^2\\
&\quad + p_3(p_4+p_5)\left(p_3+p_4+p_5\right)(q_1+q_2)^2q_6^2\bigr],\\
a_{64} &= r^{(3)}s^{(3)}\bigl[(p_1+p_2)p_6\left(p_1+p_2+p_6\right)q_3(q_4+q_5)\\
&\quad \times\left(q_3+q_4+q_5\right)+p_3(p_4+p_5)\left(p_3+p_4+p_5\right)\\
&\quad \times\left(q_1+q_2\right)q_6\left(q_1+q_2+q_6\right)\bigr],\\
a_{65} &= r^{(3)}s^{(2)}\bigl[(p_1+p_2)p_6\left(p_1+p_2+p_6\right)q_3(q_4+q_5)\\
&\quad + p_3(p_4+p_5)\left(p_3+p_4+p_5\right)\left(q_1+q_2\right)q_6\bigr],\\
a_{66} &= r^{(2)}s^{(4)}\bigl[(p_1+p_2)p_6q_3^2(q_4+q_5)^2\\
&\quad + p_3(p_4+p_5)(q_1+q_2)^2q_6^2\bigr],\\
a_{67} &= r^{(2)}s^{(3)}\bigl[(p_1+p_2)p_6q_3\left(q_4+q_5\right)\left(q_3+q_4+q_5\right)\\
&\quad + p_3(p_4+p_5)(q_1+q_2)q_6\left(q_1+q_2+q_6\right)\bigr],\\
a_{68} &= r^{(2)}s^{(2)}\bigl[(p_1+p_2)p_6q_3\left(q_4+q_5\right)\\
&\quad + p_3(p_4+p_5)(q_1+q_2)q_6\left(q_1+q_2+q_6\right)\bigr].
\end{aligned}$$

(*c*) $v_7 = a_{71}+a_{72}+a_{73}+a_{74}+a_{75}+a_{76}+a_{77}+a_{78}$, where

$$\begin{aligned}
a_{71} &= r^{(4)}s^{(3)}\bigl[(p_1+p_4)^2p_8^2q_2q_7(q_2+q_5+q_7)\\
&\quad + (p_2+p_5)^2p_7^2(q_1+q_4)q_8(q_1+q_4+q_8)\bigr],\\
a_{72} &= r^{(4)}s^{(2)}\bigl[(p_1+p_4)^2p_8^2(q_2+q_5)q_7\\
&\quad + (p_2+p_5)^2p_7^2(q_1+q_4)q_8\bigr],\\
a_{73} &= r^{(3)}s^{(4)}\bigl[(p_1+p_4)p_8(p_1+p_4+p_8)(q_2+q_5)^2q_7^2\\
&\quad + (p_2+p_5)p_7(p_2+p_5+p_7)(q_1+q_4)^2q_8^2\bigr],\\
a_{74} &= r^{(3)}s^{(3)}\bigl[(p_1+p_4)p_8(p_1+p_4+p_8)(q_2+q_5)\\
&\quad \times q_7(q_2+q_5+q_7)\\
&\quad + (p_2+p_5)p_7(p_2+p_5+p_7)q_1q_8(q_1+q_4+q_8)\bigr],\\
a_{75} &= r^{(3)}s^{(2)}\bigl[(p_1+p_4)p_8(p_1+p_4+p_8)(q_2+q_5)q_7\\
&\quad + (p_2+p_5)p_7(p_2+p_5+p_7)(q_1+q_4)q_8\bigr],\\
a_{76} &= r^{(2)}s^{(4)}\bigl[(p_1+p_4)p_8(q_2+q_5)^2q_7^2\\
&\quad + (p_2+p_5)p_7(q_1+q_4)^2q_8^2\bigr],\\
a_{77} &= r^{(2)}s^{(3)}\bigl[(p_1+p_4)p_8(q_2+q_5)q_7(q_2+q_5+q_7)\\
&\quad + (p_2+p_5)p_7(q_1+q_4)q_8(q_1+q_4+q_8)\bigr],\\
a_{78} &= r^{(2)}s^{(2)}\bigl[(p_1+p_4)p_8(q_2+q_5)q_7\\
&\quad + (p_2+p_5)p_7(q_1+q_4)q_8(q_1+q_4+q_8)\bigr].
\end{aligned}$$

(d) $v_8 = a_{81}+a_{82}+a_{83}+a_{84}+a_{85}+a_{86}+a_{87}+a_{88}$, where

$$\begin{aligned}
a_{81} &= r^{(4)}s^{(3)}[(p_1+p_2+p_4+p_5)^2p_9^2(q_3+q_6)(q_7+q_8)\\
&\quad \times(q_3+q_6+q_7+q_8)+(p_3+p_6)^2(p_7+p_8)^2\\
&\quad \times(q_1+q_2+q_4+q_5)q_9(q_1+q_2+q_4+q_5+q_9)],\\
a_{82} &= r^{(4)}s^{(2)}\bigl[(p_1+p_2+p_4+p_5)^2p_9^2(q_3+q_6)(q_7+q_8)\\
&\quad + (p_3+p_6)^2(p_7+p_8)^2(q_1+q_2+q_4+q_5)q_9\bigr],\\
a_{83} &= r^{(3)}s^{(4)}\bigl[(p_1+p_2+p_4+p_5)p_9(p_1+p_2+p_4+p_5+p_9)\\
&\quad \times(q_3+q_6)^2(q_7+q_8)^2+(p_3+p_6)(p_7+p_8)\\
&\quad \times(p_3+p_6+p_7+p_8)(q_1+q_2+q_4+q_5)^2q_9^2\bigr],\\
a_{84} &= r^{(3)}s^{(3)}\bigl[(p_1+p_2+p_4+p_5)p_9(p_1+p_2+p_4+p_5+p_9)\\
&\quad \times(q_3+q_6)(p_7+p_8)(q_3+q_6+p_7+p_8)\\
&\quad + (p_3+p_6)(p_7+p_8)(p_3+p_6+p_7+p_8)\\
&\quad \times(q_1+q_2+q_4+q_5)q_9(q_1+q_2+q_4+q_5+q_9)\bigr],\\
a_{85} &= r^{(3)}s^{(2)}\bigl[(p_1+p_2+p_4+p_5)p_9(p_1+p_2+p_4+p_5+p_9)\\
&\quad \times(q_3+q_6)(q_7+q_8)+(p_3+p_6)(p_7+p_8)\\
&\quad \times(p_3+p_6+p_7+p_8)(q_1+q_2+q_4+q_5)q_9\bigr],\\
a_{86} &= r^{(2)}s^{(4)}\bigl[(p_1+p_2+p_4+p_5)p_9(q_3+q_6)^2(q_7+q_8)^2\\
&\quad + (p_3+p_6)(p_7+p_8)(q_1+q_2+q_4+q_5)^2q_9^2\bigr],\\
a_{87} &= r^{(2)}s^{(3)}\bigl[(p_1+p_2+p_4+p_5)p_9(q_3+q_6)(p_7+p_8)\\
&\quad \times(q_3+q_6+p_7+p_8)+(p_3+p_6)(p_7+p_8)\\
&\quad \times(q_1+q_2+q_4+q_5)q_9(q_1+q_2+q_4+q_5+q_9)\bigr],\\
a_{88} &= r^{(2)}s^{(2)}\bigl[(p_1+p_2+p_4+p_5)p_9(q_3+q_6)(p_7+p_8)\\
&\quad + (p_3+p_6)(p_7+p_8)(q_1+q_2+q_4+q_5)\\
&\quad \times q_9(q_1+q_2+q_4+q_5+q_9)\bigr].
\end{aligned}$$

(e) $\operatorname{cov}(T_i, T_j) = 0,\ i \neq j.$

From the above results, the variances can be estimated by the estimates of the leading terms. For example, $\mathrm{Var}(T_5)$, i.e., v_5, can be approximated by $\hat{v}_5 = \hat{a}_{51}+\hat{a}_{53}$, the estimate of $a_{51}+a_{53}$, with the p_i and q_i being replaced by their MLEs $\hat{p}_i = \frac{r_i}{r}$, and $\hat{q}_i = \frac{s_i}{s}$. This is because $\frac{a_{5i}}{a_{51}+a_{53}} \to 0$ for $i = 2, 4, 5, \ldots, 8$ when r and $s \to \infty$.

APPENDIX B. PROOF OF THEOREM 2

We only need to show that $\mathrm{var}(T_5^{LOR}) \approx \frac{1}{r_1}+\frac{1}{r_2}+\frac{1}{r_4}+\frac{1}{r_5}+\frac{1}{s_1}+\frac{1}{s_2}+\frac{1}{s_4}+\frac{1}{s_5}$, and $\mathrm{cov}(T_5^{LOR}, T_6^{LOR}) = 0$ as the other results can be proven in the same way. Since $T_5^{LOR} = \ln(\frac{r_1r_5s_2s_4}{r_2r_4s_1s_5}) = \ln(\frac{\hat{p}_1\hat{p}_5\hat{q}_2\hat{q}_4}{\hat{p}_2\hat{p}_4\hat{q}_1\hat{q}_5})$, let $f(p_1, p_2, p_4, p_5, q_1, q_2, q_4, q_5) = \ln(p_1) - \ln(p_2) - \ln(p_4) +$

$\ln(p_5) - \ln(q_1) + \ln(q_2) + \ln(q_4) - \ln(q_5)$, then the gradient $\nabla f = (\frac{1}{p_1}, -\frac{1}{p_2}, -\frac{1}{p_4}, \frac{1}{p_5}, -\frac{1}{q_1}, \frac{1}{q_2}, \frac{1}{q_4}, -\frac{1}{q_5})^T$. The variance-covariance matrix, denoted by Σ, of vector $(\hat{p}_1, \hat{p}_2, \hat{p}_4, \hat{p}_5, \hat{q}_1, \hat{q}_2, \hat{q}_4, \hat{q}_5)^T$ can be found based on the properties of multinomial distribution and the fact that cases and controls are independent. $\Sigma = \begin{bmatrix} \Sigma_P & 0 \\ 0 & \Sigma_Q \end{bmatrix}$, where

$$\Sigma_P = \frac{1}{r}\begin{bmatrix} p_1(1-p_1) & -p_{11}p_2 & -p_1p_4 & -p_1p_5 \\ -p_1p_2 & p_2(1-p_2) & -p_2p_4 & -p_2p_5 \\ -p_1p_4 & -p_2p_4 & p_4(1-p_4) & -p_4p_5 \\ -p_1\hat{p}_5 & -p_2p_5 & -p_4p_5 & p_5(1-p_5) \end{bmatrix},$$

and

$$\Sigma_Q = \frac{1}{s}\begin{bmatrix} q_1(1-q_1) & -q_1q_2 & -q_1q_4 & -q_1q_5 \\ -q_1q_2 & q_2(1-q_2) & -q_2q_4 & -q_2q_5 \\ -q_1q_4 & -q_2q_4 & q_4(1-q_4) & -q_4q_5 \\ -q_1q_5 & -q_2q_5 & -q_4q_5 & q_5(1-q_5) \end{bmatrix}.$$

The variance of T_5^{LOR} can then be approximated by $(\nabla f)^T\Sigma(\nabla f)$. It is not difficulty to show that $(\nabla f)^T\Sigma(\nabla f) = \frac{1}{r_1} + \frac{1}{r_2} + \frac{1}{r_4} + \frac{1}{r_5} + \frac{1}{s_1} + \frac{1}{s_2} + \frac{1}{s_4} + \frac{1}{s_5}$.

Next, we show that $\mathrm{cov}(T_5^{LOR}, T_6^{LOR}) \approx 0$. Let $f(p_1, p_2, p_4, p_5, p_7, p_8, q_1, q_2, q_4, q_5, q_7, q_8) = \ln(p_1) - \ln(p_2) - \ln(p_4) + \ln(p_5) - \ln(q_1) + \ln(q_2) + \ln(q_4) - \ln(q_5)$ as defined above and $g(p_1, p_2, p_4, p_5, p_7, p_8, q_1, q_2, q_4, q_5, q_7, q_8) = \ln(p_1 + p_2) - \ln(p_3 + p_4) - \ln(p_5) + \ln(p_6) - \ln(q_1 + q_2) + \ln(q_3 + q_4) + \ln(q_5) - \ln(q_6)$.

Then $\nabla f = (\frac{1}{p_1}, -\frac{1}{p_2}, -\frac{1}{p_4}, \frac{1}{p_5}, 0, 0, -\frac{1}{q_1}, \frac{1}{q_2}, \frac{1}{q_4}, -\frac{1}{q_5}, 0, 0)^T$, and $\nabla g = (\frac{1}{p_1+p_2}, \frac{1}{p_1+p_2}, -\frac{1}{p_3+p_4}, -\frac{1}{p_3+p_4}, -\frac{1}{p_5}, \frac{1}{p_6}, -\frac{1}{q_1+q_2}, -\frac{1}{q_1+q_2}, \frac{1}{q_3+q_4}, \frac{1}{q_3+q_4}, \frac{1}{q_5}, \frac{1}{q_6})^T$. The variance-covariance matrix of random vector $(p_1, p_2, p_4, p_5, p_7, p_8, q_1, q_2, q_4, q_5, q_7, q_8)^T$can be found as $\Sigma = \begin{bmatrix} \Sigma_P & 0 \\ 0 & \Sigma_Q \end{bmatrix}$, with

$$\Sigma_P = \\ \frac{1}{r}\begin{bmatrix} p_1(1-p_1) & -p_1p_2 & -p_1p_3 & -p_14 & -p_1p_5 & -p_1p_6 \\ -p_1p_2 & p_2(1-p_2) & -p_2p_3 & -p_2p_4 & -p_2p_5 & -p_2p_6 \\ -p_1p_3 & -p_2p_3 & p_3(1-p_3) & -p_3p_4 & -p_3p_5 & -p_3p_6 \\ -p_1p_4 & -p_2p_4 & -p_3p_4 & p_4(1-p_4) & -p_4p_5 & -p_4p_6 \\ -p_1p_5 & -p_2p_5 & -p_3p_5 & -p_4p_5 & p_5(1-p_5) & -p_5p_6 \\ -p_1p_6 & -p_2p_6 & -p_3p_6 & -p_4p_6 & -p_5p_6 & p_6(1-p_6) \end{bmatrix},$$

and

$$\Sigma_Q = \\ \frac{1}{s}\begin{bmatrix} q_1(1-q_1) & -q_1q_2 & -q_1q_3 & -q_1q_4 & -q_1q_5 & -q_1q_6 \\ -q_1q_2 & q_2(1-q_2) & -q_2q_3 & -q_2q_4 & -q_2q_5 & -q_2q_6 \\ -q_1q_3 & -q_2q_3 & q_3(1-q_3) & -q_3q_4 & -q_3q_5 & -q_3q_6 \\ -q_1q_4 & -q_2q_4 & -q_3q_4 & q_4(1-q_4) & -q_4q_5 & -q_4q_6 \\ -q_1q_5 & -q_2q_5 & -q_3q_5 & -q_4q_5 & q_5(1-q_5) & -q_5q_6 \\ -q_1q_6 & -q_2q_6 & -q_3q_6 & -q_4q_6 & -q_5q_6 & q_6(1-q_6) \end{bmatrix}.$$

The $\mathrm{cov}(T_5^{LOR}, T_6^{LOR})$ can be approximated by $(\nabla f)^T\Sigma(\nabla g)$, which can be shown equal to 0.

APPENDIX C. PROOF OF THEOREM 3

We only show that $Z_5^{LOR} \approx Z_5$ when β_1 is small, the others can be proved in the same way. Since $\beta_1 = \ln(\frac{r_1r_5s_2s_4}{r_2r_4s_1s_5})$, if β_1 is a small, which is usually true in GWAS, $r_1r_5s_2s_4 = e^{\beta_1}r_2r_4s_1s_5 \approx (1+\beta_1)r_2r_4s_1s_5$, or $p_1p_5q_2q_4 \approx (1+\beta_1)p_2p_4q_1q_5$. Because $Z_5 = \frac{\beta_1r_5s_2s_4 - r_2r_4s_1s_5}{\sqrt{\hat{a}_{51}+\hat{a}_{53}}} \approx \frac{\beta_1r_5s_2s_4 - r_2r_4s_1s_5}{\sqrt{a_{51}+a_{53}}}$, where

$$\begin{aligned} a_{51} &= r^{(4)}s^{(3)}\left[p_1^2p_5^2q_2q_4(q_2+q_4) + p_2^2p_4^2q_1q_5(q_1+q_5)\right] \\ &= r^{(4)}s^{(3)}p_2^2p_4^2q_1^2q_5^2\left[\left(1+\beta_1\right)^2\left(\frac{1}{q_2}+\frac{1}{q_4}\right)+\left(\frac{1}{q_1}+\frac{1}{q_5}\right)\right] \\ &\approx r^{(4)}s^{(3)}p_2^2p_4^2q_1^2q_5^2\left[\frac{1}{q_2}+\frac{1}{q_4}+\frac{1}{q_1}+\frac{1}{q_5}\right], \end{aligned}$$

and

$$\begin{aligned} a_{53} &= r^{(3)}s^{(4)}\left[p_1p_5(p_1+p_5)q_2^2q_4^2 + p_2p_4(p_2+p_4)q_1^2q_5^2\right] \\ &= r^{(3)}s^{(4)}p_2^2p_4^2q_1^2q_5^2\left[\left(1+\beta_1\right)^2\left(\frac{1}{p_1}+\frac{1}{p_5}\right)+\left(\frac{1}{p_2}+\frac{1}{p_4}\right)\right] \\ &\approx r^{(3)}s^{(4)}p_2^2p_4^2q_1^2q_5^2\left[\frac{1}{p_1}+\frac{1}{p_5}+\frac{1}{p_2}+\frac{1}{p_4}\right]. \end{aligned}$$

Therefore,

$$\begin{aligned} Z_5 &\approx \left\{r^2s^2\beta_1p_2p_4q_1q_5\right\} \\ &\quad\times\left\{r^{(4)}s^{(3)}p_2^2p_4^2q_1^2q_5^2\left[\frac{1}{q_2}+\frac{1}{q_4}+\frac{1}{q_1}+\frac{1}{q_5}\right]\right. \\ &\quad\left.+\,r^{(3)}s^{(4)}p_2^2p_4^2q_1^2q_5^2\left[\frac{1}{p_1}+\frac{1}{p_5}+\frac{1}{p_2}+\frac{1}{p_4}\right]\right\}^{-\frac{1}{2}} \\ &\approx \frac{\beta_1}{\sqrt{\frac{1}{s}\left[\frac{1}{q_2}+\frac{1}{q_4}+\frac{1}{q_1}+\frac{1}{q_5}\right]+\frac{1}{r}\left[\frac{1}{p_1}+\frac{1}{p_5}+\frac{1}{p_2}+\frac{1}{p_4}\right]}} \\ &= Z_5^{LOR}. \end{aligned}$$

APPENDIX D. PROOF OF THEOREM 4

For the statistics T_i's, denote their variances by v_i ($i = 1, 2, 3, 4$). Similar as in the proof of theorem 1, under the null hypothesis of H_{02} we can show that

(a) $v_1 = rs\left(\overline{p}_1+\overline{p}_2+\overline{p}_3\right)\left(\overline{p}_4+\overline{p}_5+\overline{p}_6\right)$
$\times\left[(n-2)\left(\overline{p}_1+\overline{p}_2+\overline{p}_3+\overline{p}_4+\overline{p}_5+\overline{p}_6\right)+2\right]$,

(b) $v_2 = nrs(\overline{p}_1+\overline{p}_2+\overline{p}_3+\overline{p}_4+\overline{p}_5+\overline{p}_6)(\overline{p}_7+\overline{p}_8+\overline{p}_9)$,

(c) $v_3 = rs\left(\overline{p}_1+\overline{p}_4+\overline{p}_7\right)\left(\overline{p}_2+\overline{p}_5+\overline{p}_8\right)$
$\times\left[(n-2)\left(\overline{p}_1+\overline{p}_4+\overline{p}_7+\overline{p}_2+\overline{p}_5+\overline{p}_8\right)+2\right]$,

(d) $v_4 = nrs(\overline{p}_1+\overline{p}_4+\overline{p}_7+\overline{p}_2+\overline{p}_5+\overline{p}_8)(\overline{p}_3+\overline{p}_6+\overline{p}_9)$,

and

(e) $\mathrm{cov}(T_i, T_j) = 0,\ i \neq j$.

In the above formula, $\overline{p}_i = \frac{rp_i+sq_i}{r+s}$. The variances v_i can be estimated by $\hat{v}_i$ for which $\overline{p}_i$ in v_i is replaced by its estimate $\hat{\overline{p}}_i = \frac{r_i+s_i}{r+s}$.

Received 22 August 2015

REFERENCES

Barhdadi A., Dubé M.-P. (2010). Testing for gene-gene interaction with AMMI models. *Statistical Applications in Genetics and Molecular Biology* **9**(1). MR2594941

Chen Z. (2011a). Is the weighted z-test the best method for combining probabilities from independent tests? *Journal of Evolutionary Biology* **24**(4):926–930.

Chen Z. (2011b). A new association test based on chi-square partition for case-control GWA studies. *Genetic Epidemiology* **35**(7): 658–663.

Chen Z. (2013). Association tests through combining p-values for case control genome–wide association studies. *Statistics and Probability Letters* **83**(8):1854–1862. MR3069889

Chen Z. (2014). A new association test based on disease allele selection for case-control genome-wide association studies. *BMC Genomics* **15**:358.

Chen Z., Huang H., Liu J., Ng H. K. T., Nadarajah S., Huang X., Deng Y. (2013a). Detecting differentially methylated loci for Illumina array methylation data based on human ovarian cancer data. *BMC Medical Genomics* **6**(Suppl 1):S9.

Chen Z., Huang H., Liu Q. (2014a). Detecting differentially methylated loci for multiple treatments based on high-throughput methylation data. *BMC Bioinformatics* **15**:142.

Chen Z., Huang H., Ng H. K. T. (2012a). Design and analysis of multiple diseases genome-wide association studies without controls. *GENE* **510**(1):87–92.

Chen Z., Huang H., Ng H. K. T. (2013b). Testing for association in case-control genome-wide association studies with shared controls. *Statistical Methods in Medical Research*, Published online before print February 1, 2013, doi:10.1177/0962280212474061.

Chen Z., Huang H., Ng H. K. T. (2014b). An Improved Robust Association Test for GWAS with Multiple Diseases. *Statistics & Probability Letters* **91**:153–161. MR3208129

Chen Z., Liu Q., Nadarajah S. (2012b). A new statistical approach to detecting differentially methylated loci for case control Illumina array methylation data. *Bioinformatics* **28**(8):1109–1113.

Chen Z., Nadarajah S. (2014). On the optimally weighted z-test for combining probabilities from independent studies. *Computational Statistics & Data Analysis* **70**:387–394. MR3125501

Chen Z., Ng H. K. T. (2012). A robust method for testing association in genome-wide association studies. *Human Heredity* **73**(1):26–34.

Chen Z., Ng H. K. T., Li J., Liu Q., Huang H. (2014c). Detecting associated single-nucleotide polymorphisms on the X chromosome in case control genome-wide association studies. *Statistical Methods in Medical Research*, Published online before print September 24, 2014, doi:10.1177/0962280214551815.

Chen Z., Yang W., Liu Q., Yang J. Y., Li J., Yang M. Q. (2014d). A new statistical approach to combining p-values using gamma distribution and its application to genome-wide association study. *BMC Bioinformatics* **15**(Suppl 17):S3.

Fisher R. A., editor. (1932). *Statistical Methods for Research Workers*. Edinburgh: Oliver and Boyd.

Hindorff L., MacArthur J., Morales J., Junkins H. A., Hall P. N., Klemm A., Manolio T. A. A Catalog of Published Genome-Wide Association Studies. Available at: www.genome.gov/gwastudies.

Hu J. K., Wang X., Wang P. (2014). Testing gene–gene interactions in genome wide association studies. *Genetic Epidemiology* **38**(2):123–134.

Jiao S., Hsu L., Berndt S., Bézieau S., Brenner H., Buchanan D., Caan B. J., Campbell P. T., Carlson C. S., Casey G. (2012). Genome-wide search for gene-gene interactions in colorectal cancer. *PloS One* **7**(12):e52535.

Manolio T. A., Collins F. S., Cox N. J., Goldstein D. B., Hindorff L. A., Hunter D. J., McCarthy M. I., Ramos E. M., Cardon L. R., Chakravarti A. (2009). Finding the missing heritability of complex diseases. *Nature* **461**(7265):747–753.

Mosimann J. E. (1962). On the compound multinomial distribution, the multivariate β-distribution, and correlations among proportions. *Biometrika* 65–82. MR0143299

Owen A. B. (2009). Karl Pearson's meta-analysis revisited. *Ann. Statist.* **37**(6B):3867–3892. MR2572446

Plackett R. (1962). A note on interactions in contingency tables. *Journal of the Royal Statistical Society. Series B (Methodological)* 162–166.

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

Schymick J. C., Scholz S. W., Fung H.-C., Britton A., Arepalli S., Gibbs J. R., Lombardo F., Matarin M., Kasperaviciute D., Hernandez D. G. (2007). Genome-wide genotyping in amyotrophic lateral sclerosis and neurologically normal controls: first stage analysis and public release of data. *The Lancet Neurology* **6**(4): 322–328.

Sha Q., Zhang Z., Schymick J. C., Traynor B. J., Zhang S. (2009). Genome-wide association reveals three SNPs associated with sporadic amyotrophic lateral sclerosis through a two-locus analysis. *BMC Medical Genetics* **10**(1):86.

Song M., Nicolae D. L. (2009). Restricted parameter space models for testing gene-gene interaction. *Genetic Epidemiology* **33**(5): 386.

Ueki M. (2014). On the choice of degrees of freedom for testing gene–gene interactions. *Statistics in Medicine* **33**(28):4934–4948. MR3276510

Ueki M., Cordell H. J. (2012). Improved statistics for genome-wide interaction analysis. *PLoS Genet.* **8**(4):e1002625.

VanderWeele T. J., Laird N. M. (2011). Tests for compositional epistasis under single interaction-parameter models. *Annals of Human Genetics* **75**(1):146–156.

Wu X., Jin L., Xiong M. (2008). Composite measure of linkage disequilibrium for testing interaction between unlinked loci. *European Journal of Human Genetics* **16**(5):644–651.

Yang Y., He C., Ott J. (2009). Testing association with interactions by partitioning chi-squares. *Annals of Human Genetics* **73**(1):109–117.

Yu Z., Demetriou M., Gillen D. L. (2015). Genome-wide analysis of gene-gene and gene-environment interactions using closed-form wald tests. *Genetic Epidemiology.*

Zang Y., Fung W. K., Zheng G. (2010). Simple algorithms to calculate the asymptotic null distributions of robust tests in case-control genetic association studies in R. *Journal of Statistical Software* **33**(8):1–24.

Zheng G., Ng H. K. T. (2008). Genetic model selection in two-phase analysis for case-control association studies. *Biostatistics* **9**(3):391–9.

Zhongxue Chen
Department of Epidemiology and Biostatistics
School of Public Health
Indiana University Bloomington
1025 E. 7th street
Bloomington, IN 47405
USA
Phone: 1-812-855-1163
E-mail address: zc3@indiana.edu