Restricted parameter space models for testing gene-gene interaction

Minsun Song[1] and **Dan L. Nicolae**[1,2]

Dan L. Nicolae: nicolae@galton.uchicago.edu

[1]Department of Statistics, The University of Chicago, 5734 S. University Ave., Chicago, IL 60637

[2]Department of Medicine, The University of Chicago, 5734 S. University Ave., Chicago, IL 60637

Abstract

There is a growing recognition that interactions (gene-gene and gene-environment) can play an important role in common disease etiology. The development of cost-effective genotyping technologies has made genome-wide association studies the preferred tool for searching for loci affecting disease risk. These studies are characterized by a large number of investigated SNPs, and efficient statistical methods are even more important than in classical association studies that are done with a small number of markers. In this paper we propose a novel gene-gene interaction test that is more powerful than classical methods. The increase in power is due to the fact that the proposed method incorporates reasonable constraints in the parameter space. The test for both association and interaction is based on a likelihood ratio statistic that has a chi-bar-squared distribution asymptotically. We also discuss the definitions used for "no interaction" and argue that tests for pure interaction are useful in genome-wide studies, especially when using two stage strategies where the analyses in the second stage are done on pairs of loci for which at least one is associated with the trait.

1 Introduction

Common complex diseases such as diabetes and asthma have been investigated for genetic risk factors for two decades, but the identification of disease susceptibility genes and the development of models that predict disease risk have been less successful than those for Mendelian traits. This is probably partly due to the interactions of genes with each other and with the environment. Recent human and animal studies of complex diseases have identified susceptibility genetic variants that are marginally associated to a minor extent only, but that interact significantly with each other; these are loci that can be found only when using interaction models. That is why there is a growing need for inference on models in which two or more susceptibility loci contribute to a common trait jointly. Genome-wide association studies (GWAS) where high-density SNP information is available, provide great potential for association mapping aiming to identify genetic variants that are associated marginally as well as interactively. It has been shown (Marchini *et al.*, 2005) that, even with a conservative penalty for multiple testing, analytic designs incorporating locus-locus interaction can be more powerful for GWAS than those performing only single-marker association tests.

Several approaches have been proposed for detecting gene-gene interactions and there is no unified definition for the null model they are testing. A common method for detecting two jointly associated loci is based on the logistic regression model composed of terms for the additive and dominance effects for each marker and the between-loci additive and dominance interactions (Marchini *et al.*, 2005; North *et al.*, 2005). The logistic regression is easy to implement, and the interaction test corresponds to testing if the four interaction parameters are equal to zero. An alternative measure of interaction between two unlinked loci is the deviance of the penetrance for a haplotype at two loci from product of the marginal penetrance of the individual alleles that span the haplotype, and this is motivated by linkage disequilibrium (LD) measures (Zhao *et al*, 2006). A Wald test for interaction between two unlinked loci (Zhao *et al*, 2006) can be used to investigate deviations from equilibrium. Also, various data-mining methods such as multidimensional reduction (MDR) (Ritchie *et al.*, 2001), the combinatorial partitioning method (CPM) (Nelson *et al.*, 2001), and the restricted partitioning method (RPM) (Culverhouse *et al.*, 2004) have been explored to detect gene-gene interaction. These methods assume no parametric model, and one of their limitation is their impracticability for large data sets because of the massive computation that is required.

Our focus is on developing methods that can be applied to large datasets such as those coming from GWA studies. Our belief is that two-stage strategies are most appropriate, where the first stage consists of single-marker association tests, and where the interaction analyses in the second stage are done only on a subset of markers. For simplicity in the interpretation of the results, we would like the gene-gene interaction tests to reflect evidence for association in addition to that offered by the marginal tests, so *pure interaction tests* that are independent of any marginal effect are necessary. Also, even in two-stage designs, the number of interaction tests that are performed can be very large, and efficient methods are needed. One way to improve efficiency is to investigate a smaller alternative space. This is similar to using allelic or trend tests in single-marker association for reducing the size of the alternative space. We argue in this paper that all natural interaction models have restrictions that can be utilized; as a parallel to marginal association tests, the restriction that is natural there is that the heterozygous risk is bounded by the risks corresponding to the two homozygous genotypes.

In this paper we focus on two locus models for gene-gene interaction. We start with a extensive discussion on the definition of "no interaction". Our main approach is based on imposing natural restrictions for the parameter spaces used to model interactions, and this is described next. We also investigate the type I error and power of our tests using simulations.

2 The null and alternative models

Let us introduce some notation before we address our model and test for interaction. We denote by G and H the markers of interest and we assume that the markers are biallelic and are not in linkage disequilibrium in the general population. The alleles of G are denoted by a and A and those of H by b and B. Where necessary, we let A and B be the disease predisposing alleles. We use $G_i(H_i)$, $i = 1, 2, 3$ to denote the genotypes aa, Aa, and AA (bb, Bb, and BB). We denote by p_{ij} the frequency of joint genotype (G_i, H_j) in cases and n_{ij} the

genotype counts for (G_i, H_j) in n cases. Finally, $p_{i.}$ and $p_{.j}$ denote the cases marginal genotype frequencies for G_i and H_j respectively.

The definition of the "no interaction" model is critically important since parametric statistical methods would be more powerful than non-parametric methods such as MDR as long as modeling assumptions are satisfied. In this paper we define " no interaction" between two unlinked loci as the model that satisfies,

$$p_{ij}=p_{i.}p_{.j}, \quad (1)$$

which is conditional independence of the two loci (i.e. independence conditioned on the subject having the disease, event denoted with D). This is the same hypothesis that would be tested by an independence test, such as Pearson's χ^2, in the 3x3 contingency table of genotype counts in affecteds. Properties and justification of this model for the null hypothesis are itemized below where we also compare it to nulls implied by other methods discussed in the paper. The most utilized method for testing for interaction is logistic regression where "no interaction" means no departure from the multiplicative odds ratio of disease.

The important factors to consider when assessing this null model we propose are:

- independence in cases, equation (1), is equivalent to multiplicative penetrances, and this implies multiplicative joint genotype relative risks (i.e. the ratio of penetrances); multiplicative penetrances have been used as a null model in linkage studies that focus on sharing among affecteds;

- under this model, the relative genotype risks for G are not changed when conditional on H, i.e

$$\frac{P(D|G_i, H_j)}{P(D|G_k, H_j)}=\frac{P(D|G_i)}{P(D|G_k)},$$

for all possible genotypes. This can be interpreted clearly as a no interaction model because genotypes at H do not affect the risk at G.

- the marginal counts for cases are sufficient statistics under this model. This is important as it shows that a cases-only design can be used to test this null model;

- the full log-likelihood ratio for the comparison of the fully parametrized model and the model under which neither G and H are associated can be decomposed as the sum of likelihoods for marginal association and pure interaction; this shows that the pure interaction test that results from testing this model is independent of any marginal association effect;

- testing this model requires no data and no assumptions on controls and this leads to a more robust approach than logistic regression, for example, where both cases and controls are needed. This is made clearer using an example described in the next paragraph.

- when control data is available and the markers are in linkage equilibrium in the controls (for example when the controls are "population controls"), this null model is equivalent to the null model tested by the logistic regression.

- for a rare disease, the magnitude of genotype relative risks is similar to that of odds ratios;

- finally, case-only analyses are in many situations at least as efficient (corresponding tests are more powerful) as having both cases and controls. This is true when the controls contain no information on the interaction part of the model, such as when the genotypes for *G* and *H* are independent in controls. This is similar to previous arguments, e.g. Yang *et al* (1997) and Yang *et al* (1999) which claimed that for identifying the relationship between gene and environment and between gene and gene in the view of interactions, fewer cases are needed in a case-only design than a case-control design, respectively. We will demonstrate this using simulations in the results section.

Our claim that cases-only studies are robust is based on the assumption that, in many studies, the controls contain no information on the model of pure interaction. That is why random deviations from independence in controls will lead to false positives for the interaction test. The real data example we use is from Hampe *et al.* (2007) where evidence for interaction between *ATG*16*L*1 SNP rs2241880 and *CARD*15/*NOD*2 disease-associated variants was found in a Crohn's disease dataset. Their argument is based on odds ratios which vary according to *CARD*15 genotypes, corresponding to a significant finding using the logistic regression model. Note that out of total nine parameters for a 3×3 table, one parameter models the overall penetrance and four parameters represent marginal effects. In the Crohn's dataset, the likelihood ratio test (LRT) for the logistic regression model involving 4 degrees of freedom (df) leads to a 0.05 p-value, whereas a LRT with 4 df for cases only (hence the proposed null) results in $P = 0.82$. In this case, the two tests lead to different conclusions. Subsequent analyses in Cummings *et al.* (2007) failed to replicate the interaction. Significant interaction is not found for both cases-only LRT (P=0.96 for the data from Cummings *et al.* (2007), P=0.93 for combined data Cummings *et al.* (2007); Hampe *et al.* (2007)) and logistic regression model (P=0.23 for the data from Cummings *et al.* (2007), P=0.26 for combined data Cummings *et al.* (2007); Hampe *et al.* (2007)). In addition to Cummings *et al.* (2007), there is another study which indicates that *CARD*15 and *ATG*16*L*1 contribute independently to Crohn's risk (Prescott *et al.*, 2007), where no significant interaction was found using logistic regression (P=0.98). Using LRT for those data we obtain a P value of 0.35. Given the lack of replication, it is likely that there is no interaction. In this situation, case-only LRT seems to lead to the correct conclusion as opposed to the logistic regression approach. The source of the signal implied by the logistic regression is easily detectable: there is a deviation from independence in controls (but not in cases). Although there are scenarios where (well selected) controls can display LD between unlinked markers, we believe that, in the Crohn's data, this is just a random deviation.

The deviance from (1) measures the dependence between two loci. It is easy to see that (1) is equivalent to the definition of "no interaction" proposed by Zhao *et al* (2006) when Hardy-Weinberg Equilibrium (HWE) holds in the disease population. Similarly to Zhao *et al*

(2006) it is natural to consider that interaction between two unlinked loci will result in deviation of the penetrance of the joint genotypes from independence of the marginal penetrance of each genotype. The main difference in the two definitions of "no interaction" is the use of genotypes in (1) versus haplotypes in Zhao *et al* (2006). We will also investigate the power of their method after the appropriate choice of alternative models.

As mentioned above, interaction can be modeled using four parameters leading to 4 df tests such as the likelihood ratio test, Pearson's test in the 3x3 table of genotype counts in cases, and the Wald test. There is a decrease in power associated with using a test with more df relative to a test with less df when the true model is close to the one specified by the test with fewer df, and the decrease tends to be severe when the significance thresholds are very small (such as those from genome-wide association studies where a large number of tests are performed). Therefore a reduction in the size of the alternative space can lead to a substantial increase in power. A good understanding of the two-locus models is necessary for choosing appropriate restrictions in the alternative parameter space. Two-locus models of disease have been classified and studied elsewhere (Neuman and Rice, 1992; Li and Reich, 2000; Hallgrímsdóttir and Yuster, 2008, e.g.). We introduce here several two-locus models which have been considered in other studies for gene-gene interaction (Neuman and Rice, 1992; Li and Reich, 2000; Zhao *et al*, 2006; Hallgrímsdóttir and Yuster, 2008). Table 1 displays the following two-locus models : the intersection of dominant and dominant (D∩D), the intersection of recessive and dominant (R∩D), the intersection of recessive and recessive (R∩R), the union of dominant and dominant (D∪D), the union of recessive and recessive (R∪R), and the union of dominant and recessive (D∪R). Among those models, D∩D, R∩D, and R∩R correspond to epistasis models and D∪D, D∪R, and R∪R correspond to heterogeneity models (or logical OR models). Heterogeneity models are a result of independent genetic mechanisms in which an individual manifests the phenotype by possessing a disease predisposing genotype at either locus. Therefore the penetrance of joint genotype is the union of two independent penetrances of each marginal genotype. The mathematical formulation of heterogeneity models corresponds to (Neuman and Rice, 1992; Hallgrímsdóttir and Yuster, 2008),

$$P(D|G_i, H_j) = P(D|G_i) + P(D|H_j) - P(D|G_i)P(D|H_j). \quad (2)$$

The heterogeneity models we consider (i.e. D∪D, R∪R, and D∪R) are approximations that correspond to the case where the marginal penetrance is (0,1,1) for a dominant trait and (0,0,1) for a recessive trait. Note that the models above span a wide range of genetic mechanisms for interactions. For example, to increase the risk in the D∩D model, disease predisposing alleles are needed at both loci while for D∪D model, a disease predisposing allele is needed at either locus. These models have been used to investigate traits which do not display marginal associations. The R∪R model has been used to explain prelingual deafness (Majumder *et al.*, 1989). In addition, other two locus models have been explored to describe the genetics of other phenotypes (Elandt-Johnson, 1971; Lerner, 1968; Vogel and Motulsky, 1986; Levy and Nagylaki, 1972). Assuming linkage equilibrium in the general population, models involving intersection, namely models of epistasis, show non-negative log local odds ratios (i.e. $\log \frac{p_{ij}p_{i+1j+1}}{p_{ij+1}p_{i+1j}}$ for $i, j=1,2$) and those related with union, namely

models of heterogeneity, lead to non-positive log local odds ratios. Also, models of heterogeneity from formula (2) show non-positive log local odds ratios as long as marginal penetrances are monotone. Note that (1) is equivalent to zero log local odds ratios. Although this is not a complete list of models, similarly to Kooperberg *et al.* (2008), we focus on plausible interactions where the effects are in the same direction in the number of disease alleles of both SNPs. The difference to the approach of Kooperberg *et al.* (2008) is that we will use inequality constraints to reduce dimensionality, as opposed to sharp constraints.

3 Inequality constrained penetrance test (ICPT)

3.1 Two-locus model

The inference developed in this section is based only on data in cases, and we use the same notation as in the previous section. The sufficient statistic for all investigated models consists of the genotype counts in cases for all nine combinations of pairs of genotypes at the two loci. We propose a LRT for which the alternative space is inequality-constrained. Using the Bayes rule and the assumption that the two markers are in linkage equilibrium in the general population, the local odds ratios satisfy

$$\frac{p_{ij}p_{i+1j+1}}{p_{i+1j}p_{ij+1}} = \frac{P(D|G_i, H_j)P(D|G_{i+1}, H_{j+1})}{P(D|G_i, H_{j+1})P(D|G_{i+1}, H_j)}, \quad (3)$$

so they are functions of penetrances. Thus the restriction on the sign of local odds ratio is equivalent to an inequality constraint on the penetrances of the joint genotypes, and that is why we use ICPT as the acronym for our test. The calculation of the likelihood is straightforward, and we will focus the discussion on the methods that are needed to find the maximum likelihood estimates. We consider first non-negative log local odds ratios as a restriction for an alternative parameter space. The test statistic for this model requires the maximization of the likelihood under the imposed constraint. The maximization problem can be formulated as follows,

$$\sup_{A} \prod_{i=1}^{3} \prod_{j=1}^{3} p_{ij}^{n_{ij}} \quad (4)$$

where $A = \{p; \log p \in \cap_{j=1}^{2} \cap_{i=1}^{2} K_{ij}\}$ and where $K_{ij} = \{x : x_{i,j} + x_{i+1,j+1} - x_{i+1,j} - x_{i,j+1} \geq 0\}$.

We will use I-projection(Robertson *et al.*, 1988) to solve this optimization problem. Suppose $p = (p_1, \ldots, p_k)$ and $r = (r_1, \ldots, r_k)$ are probability vectors (PV). The I-divergence of p with respect to r (also known as the Kullback-Leibler divergence), is given by

$$I(p\|r) = \sum_{i=1}^{k} p_i \log(p_i/r_i)$$

where $\frac{0}{0}$ is defined as 1. It is well known that $I(p\|r)$ can be interpreted as a "distance" between p and r, and it is natural to define the closest PV to r within a set of PVs E based on $I(p\|r)$. A solution to the problem is I-projection of r onto a set E, i.e. a vector $q \in E$ such that $I(q\|r) < \infty$ and $I(q\|r) = \min_{p \in E} I(p\|r)$.

Robertson *et al.* (1988) showed that maximizing $\prod_{i=1}^{k} p_i^{x_i}$ over this space, $\{p_i \geq 0, \forall i,$ $\sum_{i=1}^{k} p_i = 1, \log p \in M\}$, is equivalent to minimizing $I(p\|u)$ over $\{p_i \geq 0, \forall i, \sum_{i}^{k} p_i = 1, p \hat{-} p$ $\in M^*\}$, where $u = (\frac{1}{k}, \ldots, \frac{1}{k})$ is the uniform PV, $\hat{p} = (\sum_{1}^{k} x_i)^{-1} (x_1, \ldots, x_k)$, M is any closed convex cone containing the constant vectors, and $M^* = \{y; \sum_{1}^{k} x_i y_i \leq 0, \forall x \in M\}$. This duality allows obtaining maximum likelihood estimator (MLE) under inequality (or order) restrictions using numerical methods for I-projection problems.

The algorithm we use to calculate the MLE is based on Dykstra (1985) who introduced a method that can be used to solve the I-projection problem when the target set is expressed as a finite intersection of arbitrary closed convex sets. It can be shown (Shapiro, 1985) that the LRT under inequality constraints follows a chi-bar-squared distribution (i.e. weighted sum of chi-squared distributions) asymptotically, i.e.

$$P(LRT \geq t) \approx \sum_{l=1}^{4} w_l P(\chi_l^2 \geq t)$$

The weight w_l is equal to the probability that the projection of estimated log local odds ratios ($\log \frac{n_{i,j} n_{i+1,j+1}}{n_{i,j+1} n_{i+1,j}}$ for $i, j = 1,2$) on non-negative space takes on l positive values. There are analytical formulas for w_1, w_2 and w_3 (Shapiro, 1985) and an upper bound for w_4 (Kudo, 1963). Also, the weights are functions of the asymptotic variance matrix of log local odds ratios (see Appendix). Weights can be estimated using Monte-Carlo simulations where each weight w_l is calculated as the proportion of times that l positive estimated log local odds ratios occur after the projection on non-negative log local odds ratios.

3.2 One-locus model

A similar dimension reduction idea can be applied to the one-locus case. We present it here as it fits naturally in the context of this paper. We denote by p_{aa}, p_{Aa}, and p_{AA} (q_{aa}, q_{Aa}, and q_{AA}) the frequencies of genotype aa, Aa, and AA in cases (respectively, controls). Let r_i (s_i), $i = 0,1,2$ be the genotype counts for genotype aa, Aa, and AA in cases (controls). We use R and S to denote the number of cases and controls, respectively. Let $n_i = r_i + s_i$, $i = 0,1,2$ be the genotype counts for genotype aa, Aa, and AA for combined cases and controls and N to denote the number of combined cases and controls.

One-locus analyses are usually performed using a 1 df allelic test, a 2 df genotype test, or a 1 df trend test. When Hardy-Weinberg equilibrium holds, the allelic test is asymptotically equivalent to the trend test (Sasieni, 1997). In this paper we focus the discussion on the trend

test as it does not require additional assumptions. Note that the trend test detects a specific structure in the penetrance, and the test we propose is a generalization of this. To use the trend test, a set of scores $x = (x_1, x_2, x_3)$ should be assigned to genotypes *aa*, *Aa*, and *AA*. The test with $x = (0, 1, 1)$ is efficient if the underlying genetic model is dominant, and the test with $x=(0,0,1)$ is efficient when the genetic model is recessive. Also, the test with $x=(0,1,2)$ is optimal for an additive model. However, in practice, we do not know the underlying genetic model and most applications use the trend scores corresponding to additive effects because they are closest to the models discussed above. Previous studies (Freidlin *et al.*, 2002) have shown that there is a substantial loss of power when the score for the test is not optimal. A possible remedy is to develop a test which is robust to the specification of the trend scores. We achieve this by using a LRT for models with ordered penetrances (corresponding to ordered x's). The penetrances are ordered for all natural one locus models: recessive, dominant, multiplicative, and additive, and it is widely thought that the risk is monotone (Balding, 2006). Assuming *A* is a risk-increasing allele, our alternative model can be written as

$$\frac{p_{aa}}{q_{aa}} \le \frac{p_{Aa}}{q_{Aa}} \le \frac{p_{AA}}{q_{AA}}.$$

Then regardless of the fact that *q* is known or unknown (Dykstra *et al.*, 1995), the LRT statistic follows a chi-bar-squared distribution, i.e.

$$\lim_{R,S \to \infty} P(LRT \ge t) = \sum_{l=1}^{3} w_l P(\chi_l^2 \ge t)$$

Robertson *et al.* (1988) gives closed formula for w_1 and w_2 as follows :

$$w_1 = 0.5, w_2 = 0.25 + \frac{1}{2\pi} \sin^{-1} \rho_{12}$$

where $\rho_{12} = -\frac{p_1 p_3}{(p_1+p_2)(p_2+p_3)}$ and $p_i = n_i/N$.

In many cases including GWAS, we do not know the allele that increases risk, and we use a two-sided alternative. For the two-sided alternative, the distribution of the LRT statistic can be easily obtained from the distribution described above.

4 Results

4.1 Simulation studies

To evaluate the performance of ICPT, we perform simulation studies. Several scenarios are simulated including those of one locus and two locus models. For controls, we consider the case where controls are from general populations. Controls are randomly generated using a multinomial distribution with population genotype frequency. Cases are generated from multinomial distributions whose frequencies are calculated using Bayes rule from the

specified control genotype frequencies and penetrances. For two locus models, we assume that the markers are not in LD in controls so joint genotype frequencies in controls are generated from marginal genotype frequencies. We consider bi-allelic SNPs with genotype frequencies 0.03, 0.3, and 0.67 for one locus models and 0.3, 0.4, and 0.3 for both loci in the two locus models. All our simulations of the one locus models involve sample sizes of 1000 for both cases and controls whereas simulations for two locus models have sample sizes of 5000 for cases and controls. We perform 1000 simulations for one locus models and 10000 simulations for two-locus models. We consider five types of one-locus models and eight types of two-locus models including null models. We include a threshold model (Neuman and Rice, 1992; Li and Reich, 2000) which does not satisfy the inequality constraints required by our model (see Table 1) in order to investigate the robustness of our methodology against deviation from these assumptions. All the other alternative two-locus models satisfy the restrictions. For all two-locus models other than the null and the threshold model (Table 1), g is 0.08 and f is 0.1. For the threshold model, g is 0.09 and f is 0.1. For the null model, the gentotype relative risks are assumed to be in the 1.05–1.15 range. Two-sided alternatives for ICPT are considered for both one-locus and two-locus models.

4.2 Simulation results for assessing ICPT for one-locus models

The purpose of this set of simulations (results shown in Table 2) is to evaluate the performance of the proposed statistic for testing marginal association. The first column in the table represents one locus models, and the second column shows the penetrances used in the simulation. The third, fourth, and fifth columns present the power from the trend test using the score (0,1,2), ICPT, and the 2 df LRT. The power of the statistic obtained by taking minimum of p-values of trend test and the 2 df LRT is presented in the sixth column. The trend test is most powerful in the multiplicative and additive models. However, the power of the trend test drops when the dominant model is true which implies that the power of the trend test is sensitive to the choice of score. On the other hand, from the simulation studies, ICPT performs quite well compared to other tests in all the models. The restricted parameter space for the alternative makes the test powerful when compared to the 2 df test. It is important to notice that, even when the model underlying the trend test is satisfied, the power of ICPT is still comparable to the power of the trend test.

4.3 Simulation results for assessing ICPT of two-locus models

In this set of simulations, we consider the performance of ICPT for detecting gene-gene interaction. In Table 3 we show the results of the simulations when the two locus models are Null, D∪D, R∪R, D∪R, D∩D, R∩R, R∩D, and Threshold. Because for one of the weights in the distribution of the LRT we can specify only an upper bound, the p-values based on these weights are conservative. To illustrate the loss in power due to this, we also provide results with weights obtained by Monte Carlo simulation. To estimate the weights, we obtain the marginal distributions and calculate case genotype frequencies by assuming the two loci are independent. The frequencies are used to generate data based on which we project the probabilities of joint case genotype frequency (i.e. p_{ij} for i, j=1,2,3) onto the space on which log local odds ratios are non-negative. Each weight w_l is defined as the proportion of times that l positive log local odds ratios after the projection on non-negative log local odds ratios occur in the replication. To contrast, we also estimate the power of 4 df LRT without any

restrictions for the alternative. Because the Wald test and Pearson's test for the table of cases genotype counts are 4 df tests that have similar performance to the LRT (results not shown), we only show the power and the type 1 error for the LRT. We also include a maximum marginal trend test (Agresti, 1996) for the null (1). The choice of scores has a big effect on power so we try to select optimal scores. The algorithm for selecting the trend scores works as follows: for each locus, we select the score among (0,1,2), (0,0,1), and (0,1,1) that gives the largest marginal trend test statistic. We compare the empirical power for the 4 df LRT, the maximum marginal trend test, and two other interaction tests, logistic regression and the LD test (Zhao *et al*, 2006) with those for ICPT with analytical and empirical weights. The Zhao *et al* (2006) approach does not take into account the haplotype ambiguity, and we found from the simulation experiment that the corresponding test does not maintain the correct type 1 error rate (data now shown). That is why for the LD test (Zhao *et al*, 2006), we correct the variance matrix so that it reflects uncertainty in estimated haplotypes.

LRT for cases outperforms the other two methods, logistic regression method and the LD test significantly. Our results for case-only LRT and case-control logistic regression show that when the main interest is pure interaction, the controls just add noise that is not properly accounted for by logistic regression. The LD test is noticeably underpowered relative to LRT. This is likely due to increase in the variance from the uncertainty of haplotypic phase.

We next attempted to compare the power among tests having the null (1) (i.e. LRT, ICPT, and the maximum marginal trend test). Note that ICPT generally outperforms the 4 df tests except when using the threshold model. However, the decrease in power of ICPT compared with LRT for threshold model is not substantial even if the model does not satisfy the inequality constraints. For R∪R and D∩D model, the maximum marginal trend test is more powerful than ICPT because the strong marginal effect allows the selection of the optimal score. However, for other models, marginal effects are weaker and the maximum marginal trend tests for single locus do not select the optimal scores. Hence, the maximum marginal trend test for interaction is not powerful when the loci have weak marginal effects. From the power comparison between ICPT obtained using closed form of the weights and the ICPT with empirical weights, the power of ICPT with empirical weights is higher than that of ICPT based on closed form. Also, ICPT using closed form for weights tend to be conservative because w_4 is given as an upper bound. Power of ICPT with analytical form of weights is higher than 4 df LRT but lower than ICPT with empirical weights. Table 3 shows that the efficiency of both ICPT over 4 df tests would increase as type I error becomes smaller. For example, we can see that for D∪R, the ratios of power (i.e. $\frac{\text{Power of 4 df test}}{\text{Power of ICPT}}$) are approximately monotonically increasing in type I error.

To further evaluate the relative performance of ICPT and the 4 df LR test, the power was calculated for various levels of the type I error ranging from 10^{-2} to 10^{-5}. The ratio of LRT power to ICPT power is an almost monotone increasing function of the type I error for the range considered (data not shown). This implies that the power of the proposed test, ICPT, is much higher than that of the 4df LRT at small significance levels which are usually of interest for GWAS.

All the code that was used for the analysis of the simulated data is written in R.

5 Discussion

The vast majority of the GWA results published to date are based on single-SNP analysis of genotyped or imputed markers. This has led to findings of associations with relatively large genetic effects, and these strong signals are rare for complex traits such as asthma or type 2 diabetes. One possibility for finding associations with weaker marginal effects is to investigate gene-gene interactions. The multiple comparison adjustment when testing for gene-gene interaction can be even more severe than for single-SNP analysis (e.g. when all pairs of makers are tested) and this has motivated the search for more powerful strategies for testing gene-gene interactions.

We introduce a new test for gene-gene interaction that is based on a likelihood ratio statistic for a restricted parameter space. The test, ICPT, is a robust alternative to classical approaches as it is not based on a narrow underlying model. The imposed restrictions to the parameter space, namely ordered penetrances, is a plausible assumption. Simulations show that the power of two-locus ICPT is superior to other tests which do not restrict the parameter space. The difference in power between ICPT and 4 degree of freedom tests, such as the likelihood ratio test, increases with more stringent significance levels which are used in studies with a large number of markers such as GWAS.

In this paper, we only investigate the case of unlinked SNPs, i.e. those that are in linkage equilibrium in the population the data are sampled from. Extensions to linked SNPs will require control data as well, as opposed to the study design used in this paper that requires only data on cases. Like most methods for identifying interaction, the asymptotic approximations for the distribution of the proposed interaction test statistic are not accurate for sparse tables, i.e. tables with zero or low counts. Possible solutions for this problem include merging genotypes and using sampling methods for calculating p-values. Also extensions to testing for gene-gene interaction of imputed markers (Nicolae, 2006; Wen and Nicolae, 2008) are necessary for a complete investigation of the data from a GWA study. We will discuss these in future manuscripts.

Acknowledgments

References

Agresti, A. An introduction to categorical data analysis. John Wiley & sons; 1996.

Armitage P. Tests for linear trends in proportions and frequencies. Biometrics. 1955; 11:375–386.

Balding DJ. A tutorial on statistical methods for population association studies. Nature Reviews Genetics. 2006; 7:781–791.

Culverhouse R, Klein T, Shannon W. Detecting epistatic interactions contributing to quantitative traits. Genetic Epidemiology. 2004; 27:141–152. [PubMed: 15305330]

Cummings JRF, Cooney R, Pathan S, Anderson CA, Barrett JC, Beckly J, Geremia A, Hancock L, Guo C, Ahmad T, Cardon LR, Jewell DP. Confirmation of the role of ATG16L1 as a Crohn's disease susceptibility gene. Inflammatory Bowel Diseases. 2007; 13:941–946. [PubMed: 17455206]

Dykstra R. An iterative procedure for obtaining I-projections onto the intersection of convex sets. Ann Prob. 1985; 13:975–984.

Dykstra R, Kochar S, Robertson T. Inference for likelihood ratio ordering in the two-sample problem. Journal of the American Statistical Association. 1995; 90:1034–1040.

Elandt-Johnson, RC. Probability models and statistical methods in genetics. New York: John Wiley; 1971.

Freidlin B, Zheng G, Li Z, Gastwirth JL. Trend tests for case-control studies of genetic markers:power, sample size and robustness. Human Heredity. 2002; 53:146–152. [PubMed: 12145550]

Hampe J, Franke A, Rosenstiel P, Till A, Teuber M, Huse K, Albrecht M, Mayr G, De La Vega FM, Briggs J, Günther S, Prescott NJ, Onnie CM, Häsler R, Sipos B, Fölsch UR, Lengauer T, Platzer M, Mathew CG, Krawczak M, Schreiber S. A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. Nature Genetics. 2007; 39:207–211. [PubMed: 17200669]

Hallgrímsdóttir IB, Yuster DS. A complete classification of epistatic two-locus models. BMC Genetics. 2008; 9:17.10.1186/1471-2156-9-17 [PubMed: 18284682]

Kooperberg C, LeBlanc M. Increasing the power of identifying gene-gene interactions in genome-wide association studies. Genetic Epidemiology. 2008; 32:255–263. [PubMed: 18200600]

Kudo A. A multivariate analogue of the one-sided test. Biometrika. 1963; 50:403–418.

Lerner, IM. Heredity, evolution, and society. San Francisco: W.H. Freeman; 1968.

Levy J, Nagylaki T. A model for the genetics of handedness. Genetics. 1972; 72:117–128. [PubMed: 5073852]

Li W, Reich J. A complete enumeration and classification of two-locus disease models. Human Heredity. 2000; 50:334–349. [PubMed: 10899752]

Majumder PP, Ramesh A, Chinnappan D. On the genetics of prelingual deafness. American Journal of Human Genetics. 1989; 44:86–99. [PubMed: 2909169]

Marchini J, Donnelly P, Cardon LR. Genome-wide strategies for detecting multiple loci that influence complex diseases. Nature Genetics. 2005; 37:413–417. [PubMed: 15793588]

Nelson MR, Kardia SL, Ferrell RE, Sing CF. A combinatorial partitioning method to identify multilocus genotypic partitions that predict quantitative trait variation. Genome Res. 2001; 11:458–470. [PubMed: 11230170]

Neuman RJ, Rice JP. Two-locus models of disease. Genetic Epidemiology. 1992; 9:347–365. [PubMed: 1427023]

Nicolae DL. Testing untyped alleles (TUNA)-applications to genome-wide association studies. Genetic Epidemiology. 2006; 30(8):718–727. [PubMed: 16986160]

North BV, Curtis D, Sham PC. Application of logistic regression to case-control association studies involving two causative loci. Human Hered. 2005; 59:79–87.

Prescott NJ, Fisher SA, Franke A, Hampe J, Onnie CM, Soars D, Bagnall R, Mirza MM, Sanderson J, Forbes A, Mansfield JC, Lewis CM, Schreiber S, Mathew CG. A nonsynonymous SNP in ATG16L1 predisposes to Ileal Crohn's disease and is independent of CARD15 and IBD5. Gastroenterology. 2007; 132:1665–1671. [PubMed: 17484864]

Ritchie MD. Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. American Journal of Human Genetics. 2001; 69:138–147. [PubMed: 11404819]

Robertson, T.; Wright, FT.; Dykstra, RL. Order-restricted statistical inference. John Wiley & sons; 1988.

Sasieni PD. From genotypes to genes: doubling the sample size. Biometrics. 1997; 53:1253–1261. [PubMed: 9423247]

Shapiro A. Asymptotic distribution of test statistics in the analysis of moment structures under inequality constraints. Biometrika. 1985; 72:133–144.

Vogel, F.; Motulsky, AG. Human genetics: problems and approaches. 2. Berlin: Springer-Verlag; 1986.

Wen X, Nicolae DL. Association studies for untyped markers with TUNA. Bioinformatics. 2006; 24(3):435–437. [PubMed: 18057020]

Yang Q, Khoury MJ, Flanders WD. Sample size requirements in case-only designs to detect gene-environment interaction. Am J Epidemiology. 1997; 146:713–720.

Yang Q, Khoury MJ, Sun F, Flanders WD. Case-only design to measure gene-gene interaction. Epidemiology. 1999; 10:167–170. [PubMed: 10069253]

Zhao J, Jin L, Xiong M. Test for interaction between two unlinked loci. American Journal of Human Genetics. 2006; 79:831–845. [PubMed: 17033960]

7 Appendix

In this appendix we show the formulas for the weights in the chi-bar distribution that asymptotically approximates the distribution of the likelihood ratio statistic. It is clear that the estimates of the joint genotype frequencies $\hat{p_{ij}}$ are asymptotically distributed as a multivariate normal distribution $N(p, (1/n)\Sigma)$ where $p = (p_{ij})$ and $\Sigma = (\text{diag}(p_{ij}) - pp^T)$. Let $\beta = (\beta_1, \beta_2, \beta_3, \beta_4)^T$ be log local odds ratios and $\hat{\beta} = (\hat{\beta_1}, \hat{\beta_2}, \hat{\beta_3}, \hat{\beta_4})^T$ be the estimated log local odds ratios (i.e. $\log \hat{p_{ij}} + \log \hat{p_{i+1j+1}} - \log \hat{p_{ij+1}} - \log \hat{p_{i+1j}}$). Using the delta method, $\hat{\beta}$ is asymptotically distributed as $N(\beta, U)$, where $U = (1/n)AD\Sigma D^T A^T$, $D = \text{diag}(\frac{1}{p_{11}}, \frac{1}{p_{12}}, \ldots, \frac{1}{p_{33}})$, and

$$
A = \begin{pmatrix}
1 & -1 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\
0 & 1 & -1 & 0 & -1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & -1 & 0 & -1 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 & -1 & 0 & -1 & 1
\end{pmatrix}.
$$

Hence, we can show that

$$
AD\sum D^T A^T = \begin{pmatrix}
\frac{1}{p_{11}} + \frac{1}{p_{12}} + \frac{1}{p_{21}} + \frac{1}{p_{22}} & -\frac{1}{p_{12}} - \frac{1}{p_{22}} & -\frac{1}{p_{21}} - \frac{1}{p_{22}} & \frac{1}{p_{22}} \\
& \frac{1}{p_{12}} + \frac{1}{p_{13}} + \frac{1}{p_{22}} + \frac{1}{p_{23}} & \frac{1}{p_{22}} & -\frac{1}{p_{22}} - \frac{1}{p_{23}} \\
& & \frac{1}{p_{21}} + \frac{1}{p_{22}} + \frac{1}{p_{31}} + \frac{1}{p_{32}} & -\frac{1}{p_{22}} - \frac{1}{p_{32}} \\
& & & \frac{1}{p_{22}} + \frac{1}{p_{23}} + \frac{1}{p_{32}} + \frac{1}{p_{33}}
\end{pmatrix}.
$$

We denote with ρ_{ij} the (i, j)th element of the correlation matrix of $\hat{\beta}$, $\text{diag}(U)^{-1/2}U(\text{diag}(U))^{-1/2}$. Then the weights are given by (Shapiro, 1985; Kudo, 1963),

$$
\begin{aligned}
w_1 &= \tfrac{1}{8}\pi^{-1}(-4\pi + \sum_{i>j,i,j\neq k}\cos^{-1}\rho_{ij.k}), \\
w_2 &= \tfrac{1}{4}\pi^{-2}\sum_{i>j,k>l,k,l\neq i,j}\cos^{-1}\rho_{ij}(\pi - \cos^{-1}\rho_{kl.ij}), \\
w_3 &= \tfrac{1}{8}\pi^{-1}(8\pi + \sum_{i>j,i,j\neq k}\cos^{-1}\rho_{ij.k}), \\
w_4 &\leq \min(1 - w_1 - w_2 - w_3, \min_{(i,j,k)\subseteq(1,2,3,4)}\tfrac{1}{4\pi}(2\pi - \cos^{-1}\rho_{ij} - \cos^{-1}\rho_{ik} - \cos^{-1}\rho_{jk})),
\end{aligned}
$$

where $\rho_{ij.k}$ is the conditional correlation between $\hat{\beta_i}$ and $\hat{\beta_j}$ given $\hat{\beta_k}$ and $\rho_{kl.ij}$ denotes the conditional correlation between $\hat{\beta_k}$ and $\hat{\beta_l}$ given $\hat{\beta_i}$ and $\hat{\beta_j}$.

Table 1

Penetrance tables for two disease loci ($g < f$).

D∪D				R∪R				D∪R		
g	f	f		g	g	f		g	g	f
f	f	f		g	f	f		f	f	f
f	f	f		f	f	f		f	f	f

R∩D				R∩R				D∩D		
g	g	g		g	g	g		g	g	g
g	g	f		g	g	g		g	f	f
g	g	f		g	g	f		g	f	f

Threshold		
g	g	g
g	g	f
g	f	f

Table 2

Power for one-locus association tests. We used a setting with 1000 cases, 1000 controls, genotype frequencies for controls=(0.03,0.3,0.67), and significance level=0.05.

model	penetrance	trend (1df)	ICPT	2df	min p-value (1df,2df)
null	(0.01,0.01,0.01)	0.053	0.053	0.054	0.040
recessive	(0.01,0.01,0.015)	0.976	0.967	0.965	0.966
dominant	(0.01,0.025,0.025)	0.245	0.725	0.704	0.610
multiplicative	(0.01,0.015,0.0225)	0.991	0.987	0.983	0.985
additive	(0.01,0.011,0.012)	0.193	0.158	0.130	0.135

Table 3

Power of ICPT[1], ICPT[2], 4 df case-only LRT, maximum marginal trend test, LD test (Zhao *et al*, 2006), and LRT based on logistic regression model. ICPT[1] uses analytical formula for weights. ICPT[2] uses empirical weights. For each locus, the score maximizing trend test statistics for marginal effect is used for the score of the trend test for testing interaction. The simulations were done using 5000 cases, and 5000 controls and 5000 cases for the logistic regression test. The significance levels used are 0.05, 0.01, and 0.001.

Model	ICPT[1]			ICPT[2]			LRT			Trend test			LD test			Logistic regression		
	.05	.01	.001	.05	.01	.001	.05	.01	.001	.05	.01	.001	.05	.01	.001	.05	.01	.001
Null	.039	.008	.001	.050	.010	.001	.051	.009	.001	.051	.011	.001	.049	.009	.001	.050	.010	.001
D∪D	.746	.524	.268	.781	.564	.301	.723	.489	.238	.604	.431	.230	.610	.380	.158	.414	.203	.062
R∪R	.828	.641	.371	.850	.678	.404	.807	.602	.338	.893	.749	.498	.721	.483	.227	.444	.229	.073
D∪R	.770	.556	.296	.806	.604	.338	.752	.527	.271	.744	.575	.346	.641	.408	.177	.421	.209	.064
D∩D	.726	.507	.249	.762	.552	.278	.709	.477	.225	.836	.658	.396	.581	.344	.130	.419	.208	.064
R∩R	.815	.620	.343	.839	.656	.381	.798	.595	.322	.707	.546	.337	.706	.464	.212	.444	.226	.075
R∩D	.791	.582	.314	.820	.625	.352	.765	.547	.285	.771	.605	.363	.664	.424	.180	.432	.217	.072
Threshold	.213	.080	.015	.253	.096	.022	.321	.145	.036	.163	.055	.011	.192	.068	.015	.176	.056	.011