Statistics and Probability Letters 83 (2013) 1854–1862

Contents lists available at SciVerse ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Association tests through combining *p*-values for case control genome-wide association studies

Zhongxue Chen *

Department of Epidemiology and Biostatistics, School of Public Health, Indiana University Bloomington, 1025 E. 7th Street, PH C104, Bloomington, IN, 47405, United States

ARTICLE INFO

Article history:
Received 20 December 2012
Received in revised form 11 April 2013
Accepted 19 April 2013
Available online 28 April 2013

Keywords:
Combining *p*-values
Genetic association
Single nucleotide polymorphism
Trend test

ABSTRACT

To detect single nucleotide polymorphisms (SNPs) that are associated with a common disease in a case control genome-wide association study (GWAS), powerful yet robust tests are desirable. Current available robust approaches in this area are mainly based on the optimal trend tests for some specific genetic models, such as recessive, additive, multiplicative, and dominant models. In this paper, we propose a class of robust association tests through combining *p*-values obtained by partitioning the 2 by 3 contingency table of the SNP data. Through simulation study and application to real data, we show that the proposed tests are powerful and robust. They provide alternative association tests for GWAS.

1. Introduction

For a diallele single nucleotide polymorphism (SNP) in a case control genome-wide association study (GWAS), there are three possible combinations (genotypes). Suppose the two alleles at a SNP locus are *A* and *a*, then the three genotypes are *AA*, *Aa*, and *aa*. Table 1 is a typical SNP data structure in GWAS.

To detect whether there is an association between the genotype and the disease status (case or control), a statistical test, such as the Pearson's chi-square test with 2 degrees of freedom (*df*), can be applied to the 2 by 3 table (rows 2–3 and columns 2–4 of Table 1). It is well known that the chi-square test is robust in the sense that it can detect all departures from the null hypothesis of no association between the row and column variables. However, for SNP data in GWAS, the chi-square test is not the best choice since it does not take the possible trend of relative risks into account. For example, if we assume *a* is the at-risk allele, then people having genotype *Aa* are more (less) likely to be affected by the disease than people with genotype *AA* (*aa*). In general, the Cochran–Armitage trend test (CATT) is more powerful than the chi-square test if the underlying genetic model and therefore the optimal scores are known (Armitage, 1955; Balding, 2006; Cochran, 1954; Joo et al., 2010; Zheng et al., 2006). Zheng et al. have shown that if the genetic models are recessive, additive/multiplicative (log additive), and dominant, the CATTs with optimal scores (0, 0, 1), (0, 0.5, 1), and (0, 1, 1), respectively, are more powerful than the chi-square test (Zheng et al., 2006, 2003).

However, if the underlying genetic model is unknown or misspecified, CATT may lose power dramatically. To overcome this shortcoming, several robust methods have been proposed (Chen, 2011a; Chen and Ng, 2012; Freidlin et al., 2002; González et al., 2008; Kwak et al., 2009; Freidlin et al., 1999; Sasieni, 1997; Slager and Schaid, 2001; Song and Elston, 2005; Wang and Sheffield, 2005; Zang et al., 2010; Zheng and Ng, 2008). Gastwirth (Gastwirth, 1966, 1985) proposed to use the maxmin efficiency robust test (MERT). Freidlin et al. (Freidlin et al., 2002) found that the maximum (MAX3) of the three above mentioned CATTs was robust. Zheng and Ng recently proposed a genetic model selection (GMS) method, with the

* Tel.: +1 812 855 1163; fax: +1 812 856 2488.
E-mail addresses: zc3@indiana.edu, zhongxue.chen@gmail.com.

0167-7152/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.spl.2013.04.021

Table 1
SNP data structure in case-control GWAS.

Genotype	*AA*	*Aa*	*aa*	Total
Case	r_1	r_2	r_3	r
Control	s_1	s_2	s_3	s
Total	n_1	n_2	n_3	n

first step selecting one of the three genetic models from data and then estimating the p-value in the second step by CATT with the optimal scores $(0, x, 1)$, where x is 0, 0.5, or 1 according to the selected model in the first step (Zheng and Ng, 2008).

In this paper, we take the trend of relative risks into account and propose a class of robust tests by partitioning the 2 by 3 contingency table of Table 1 to get two asymptotic independent test statistics and then combining two sets of p-values obtained from these two statistics based on two one-sided (left- and right-sided) tests. Through simulation study and real SNP data application, we show that the proposed tests are powerful and robust.

2. Method

For the 2 by 3 contingency table in Table 1, there are several ways to get a 2 by 2 table out of it. Our proposed tests use the following 2 by 2 subtables that are derived from the original 2 by 3 table: ST_1: from columns 1 and 2; ST_2: from columns 1+2 (i.e., collapsing columns 1 and 2), and column 3; ST_3: from columns 2 and 3; ST_4: from columns 2+3 (i.e., collapsing columns 2 and 3), and column 1.

For the above four 2 by 2 tables, a corresponding statistic for each of them can be constructed as follows:

$$T_1 = r_2 s_1 - r_1 s_2;$$
$$T_2 = r_3(s_1 + s_2) - (r_1 + r_2)s_3 = r_3 s - r s_3;$$
$$T_3 = r_3 s_2 - r_2 s_3;$$
$$T_4 = (r_2 + r_3)s_1 - r_1(s_2 + s_3) = r s_1 - r_1 s.$$

If a is the at-risk allele, the above four statistics are expected to be non-negative or positive. Suppose both cases and controls follow multinomial distributions, i.e. (r_1, r_2, r_3) and (s_1, s_2, s_3) are distributed as trinomials (r, p_1, p_2, p_3) and (s, q_1, q_2, q_3), respectively. Under the null hypothesis of no association between genotype and disease status, $(p_1, p_2, p_3) = (q_1, q_2, q_3)$. Some direct calculations give us the following results.

Theorem 2.1. (i) *Under the null hypothesis, the above random vector* $T = [T_1, T_2, T_3, T_4]'$ *has mean* 0 *and variance–covariance matrix:*

$$\Sigma_T = \begin{bmatrix} rsp_1p_2(n + (2-n)p_3) & 0 & rsp_1p_2p_3(2-n) & nrsp_1p_2 \\ 0 & nrsp_3(1-p_3) & nrsp_2p_3 & nrsp_1p_3 \\ rsp_1p_2p_3(2-n) & nrsp_2p_3 & rsp_2p_3(n + (2-n)p_1) & 0 \\ nrsp_1p_2 & nrsp_1p_3 & 0 & nrsp_1(1-p_1) \end{bmatrix}. \tag{1}$$

(ii) *Let* $U = [U_1, U_2, U_3, U_4]'$, *where* $U_i = T_i/SD(T_i)$ *and* $SD(T_i)$ *is the standard deviation of* $T_i (i = 1, 2, 3, 4)$, *then the variance–covariance matrix of random vector* U *is*

$$\Sigma_U = \begin{bmatrix} 1 & 0 & -\sqrt{\frac{(2-n)^2p_1p_3}{(n+(2-n)p_3)(n+(2-n)p_1)}} & \sqrt{\frac{np_2}{(n+(2-n)p_3)(1-p_1)}} \\ 0 & 1 & \sqrt{\frac{np_2}{(n+(2-n)p_1)(1-p_3)}} & \sqrt{\frac{p_1p_3}{(1-p_1)(1-p_3)}} \\ -\sqrt{\frac{(2-n)^2p_1p_3}{(n+(2-n)p_3)(n+(2-n)p_1)}} & \sqrt{\frac{np_2}{(n+(2-n)p_1)(1-p_3)}} & 1 & 0 \\ \sqrt{\frac{np_2}{(n+(2-n)p_3)(1-p_1)}} & \sqrt{\frac{p_1p_3}{(1-p_1)(1-p_3)}} & 0 & 1 \end{bmatrix}. \tag{2}$$

(iii) *Asymptotically, i.e., when n goes to infinity, the above variance–covariance matrix converges to:*

$$\begin{bmatrix} 1 & 0 & -\sqrt{\frac{p_1p_3}{(1-p_3)(1-p_1)}} & \sqrt{\frac{p_2}{(1-p_3)(1-p_1)}} \\ 0 & 1 & \sqrt{\frac{p_2}{(1-p_1)(1-p_3)}} & \sqrt{\frac{p_1p_3}{(1-p_1)(1-p_3)}} \\ -\sqrt{\frac{p_1p_3}{(1-p_3)(1-p_1)}} & \sqrt{\frac{p_2}{(1-p_3)(1-p_1)}} & 1 & 0 \\ \sqrt{\frac{p_2}{(1-p_3)(1-p_1)}} & \sqrt{\frac{p_1p_3}{(1-p_1)(1-p_3)}} & 0 & 1 \end{bmatrix}. \tag{3}$$

Since $p_i's$ are unknown, U_i is then modified by replacing p_i with its maximum likelihood estimator under H_0, n_i/n, and denote it by $Z_i(i = 1, 2, 3, 4)$ as follows:

$$Z_1 = \frac{r_2 s_1 - r_1 s_2}{\sqrt{rsn_1 n_2 (n + (2-n)n_3/n)/n^2}} \tag{4}$$

$$Z_2 = \frac{r_3 s - r s_3}{\sqrt{rsn_3 (n - n_3)/n}} \tag{5}$$

$$Z_3 = \frac{r_3 s_2 - r_2 s_3}{\sqrt{rsn_2 n_3 (n + (2-n)n_1/n)/n^2}} \tag{6}$$

$$Z_4 = \frac{r s_1 - r_1 s}{\sqrt{rsn_1 (n - n_1)/n}}. \tag{7}$$

Then, according to Theorem 2.1 and Slutsky's theorem, for these statistics, we have the following.

Theorem 2.2. *Under the null hypothesis, asymptotically, the above random vector* $Z = (Z_1, Z_2, Z_3, Z_4)'$ *is distributed as a multivariate normal distribution,* $\mathrm{MVN}(0, \Sigma_Z)$, *where*

$$\Sigma_Z = \begin{bmatrix} 1 & 0 & -\sqrt{\frac{p_1 p_3}{(1-p_3)(1-p_1)}} & \sqrt{\frac{p_2}{(1-p_3)(1-p_1)}} \\ 0 & 1 & \sqrt{\frac{p_2}{(1-p_1)(1-p_3)}} & \sqrt{\frac{p_1 p_3}{(1-p_1)(1-p_3)}} \\ -\sqrt{\frac{p_1 p_3}{(1-p_3)(1-p_1)}} & \sqrt{\frac{p_2}{(1-p_3)(1-p_1)}} & 1 & 0 \\ \sqrt{\frac{p_2}{(1-p_3)(1-p_1)}} & \sqrt{\frac{p_1 p_3}{(1-p_1)(1-p_3)}} & 0 & 1 \end{bmatrix}.$$

Our proposed tests are based on pairs of the above $Z_i's$. For a pair of Z_i and $Z_j(i, j = 1, 2, 3, 4, \text{and } i < j)$, we define

$$Z^{ij} = \begin{bmatrix} Z_1^{ij} \\ Z_2^{ij} \end{bmatrix} = \Sigma_{ij}^{-1/2} \begin{bmatrix} Z_i \\ Z_j \end{bmatrix}, \tag{8}$$

where $\Sigma_{ij} = \begin{bmatrix} 1 & \rho_{ij} \\ \rho_{ij} & 1 \end{bmatrix}$, and ρ_{ij} is the $(i, j)th$ element of Σ_Z with p_i being replaced by its MLE, n_i/n.

Then under the null hypothesis, the two statistics, Z_1^{ij} and Z_2^{ij} are asymptotically identically independently distributed as a standard normal $N(0, 1)$. If the at-risk allele in Table 1 is a, then Z_1^{ij} and Z_2^{ij} are expected to be nonnegative and at least one is positive. The two one-sided (right-sided) p-values from the two statistics are $p_{k,r}^{ij} = 1 - \Phi(z_k^{ij}) = \Phi(-z_k^{ij})$, $k = 1, 2$, where Φ is the cumulative distribution function (CDF) of the standard normal distribution, $N(0, 1)$. Under the null hypothesis of no association, the two p-values are asymptotically independent and uniformly distributed between 0 and 1. Then according to Fisher (Fisher, 1932), the null distribution of $W_r^{ij} = -2\ln(p_{1,r}^{ij} p_{2,r}^{ij})$ is a chi-square with 4 df. Therefore the overall p-value from the two tests Z_1^{ij} and Z_2^{ij} can be calculated by:

$$p_r^{ij} = 1 - F_{\chi_4^2}(w_r^{ij}) = 1 - F_{\chi_4^2}(-2\ln(p_{1,r}^{ij} p_{2,r}^{ij})), \tag{9}$$

where $F_{\chi_4^2}$ is the CDF of χ_4^2.

On the other hand, if the true at-risk allele is A rather than a, then the two one-sided (left-sided) p-values from Z_1^{ij} and Z_2^{ij} are calculated by $p_{k,l}^{ij} = \Phi(z_k^{ij})$, $k = 1, 2$ and $W_l^{ij} = -2\ln(p_{1,l}^{ij} p_{2,l}^{ij})$ is distributed as chi-square with 4 df under the null hypothesis. The overall p-value is then calculated by:

$$p_l^{ij} = 1 - F_{\chi_4^2}^{-1}(w_l^{ij}) = 1 - F_{\chi_4^2}^{-1}(-2\ln(p_{1,l}^{ij} p_{2,l}^{ij})). \tag{10}$$

With the uncertainty about the at-risk allele, we use the maximum of W_r^{ij} and W_l^{ij}:

$$W^{ij} = \max\{W_r^{ij}, W_l^{ij}\}. \tag{11}$$

Based on the concept of association of random variables (Esary et al., 1967), for statistic W^{ij}, we have the following result (Chen, 2011a; Chen et al., 2012a,b; Chen and Ng, 2012; Owen, 2009; Chen et al., 2013).

Theorem 2.3. *Under the null hypothesis of no association, the p-value of W^{ij} is bounded by*

$$2\beta - \beta^2 \le \Pr[W^{ij} > w] \le 2\beta, \tag{12}$$

where $\beta = 1 - F_{\chi_4^2}(w)$.

When the observed value w^{ij} of W^{ij} is large, $\beta = 1 - F_{\chi_4^2}(w^{ij})$ is small and β^2 is negligible. Therefore the p-value of W^{ij} can be approximated by the upper bound, $2(1 - F_{\chi_4^2}(w^{ij}))$. This approximation is very accurate when the true p-value is small.

For these four $Z_i's$ defined above, there are six different pairs and therefore we have six possible tests, i.e., W^{12}, W^{13}, W^{14}, W^{23}, W^{24} and W^{34}. However, as we will show in the next theorem, W^{14}, W^{23} are not robust and W^{13} and W^{24} are asymptotically equivalent. Therefore we propose three robust tests, W^{12}, W^{34}, and $W^{13}(W^{24})$ in this paper. For those tests, we have the following properties.

Theorem 2.4. (i) *W^{13} and W^{24} are asymptotically equivalent;*
(ii) *For W^{14} and W^{23}, $Z_1^{14} = -Z_2^{23}$, and $Z_2^{14} = Z_1^{23}$.*

Note Theorem 2.4(ii) says that tests W^{14} and W^{23} cannot be powerful simultaneously and therefore are not robust, although it is possible that one of them may have high power under some situations. This fact is also confirmed by the simulation study conducted in the next section. The proof of this theorem is given in the Appendix.

3. Simulation study

To assess the performances of the proposed tests, we conduct a simulation study to estimate their powers under different situations and compare them with other methods. For given relative risks λ_1, λ_2, defined as the relative risks of *Aa* and *aa* to *AA*, respectively, and genotypic frequencies $q = (q_1, q_2, q_3)$ for controls, the genotypic frequencies $p = (p_1, p_2, p_3)$ for cases can be calculated by the following formula (Chen, 2011a; Chen et al., 2012a; Chen and Ng, 2012; Chen et al., 2013):

$$\begin{aligned} p_1 &= \frac{q_1}{q_1 + \lambda_1 q_2 + \lambda_2 q_3} \\ p_2 &= \frac{\lambda_1 q_2}{q_1 + \lambda_1 q_2 + \lambda_2 q_3} \\ p_3 &= \frac{\lambda_2 q_3}{q_1 + \lambda_1 q_2 + \lambda_2 q_3}. \end{aligned} \tag{13}$$

In the simulation study, we consider both situations where Hardy–Weinberg equilibrium (HWE) holds and does not hold for controls. Under HWE, we choose two different values for minor allele frequency (maf): 0.3 and 0.5. When HWE does not hold, we use genotypic frequencies (0.2, 0.45, 0.35) and (0.35, 0.45, 0.2) for controls. In power comparison, we keep $\lambda_2 = 1.4$ and let λ_1 vary from 1 to 1.4 (i.e., 1, 1.1, 1.18, 1.2, 1.3, and 1.4). When $\lambda_1 = 1$, 1.18, 1.2 and 1.4, the corresponding genetic models are recessive, multiplicative, additive, and dominant, respectively. Under the null hypothesis, $\lambda_1 = \lambda_2 = 1$, which is used to estimate the type-I error rates in our simulation study.

We compare the new tests with existing methods, such as MERT, GMS, MAX3, Pearson's chi-square test and the default CATT with scores $x = 0.5$. To estimate the p-values from MERT, GMS, MAX3, the R package "Rassoc" with option "asy" was used (Zang et al., 2010). In our simulations, we assume there are 1000 cases and 1000 controls. The significance level is set to be 0.05 and 100,000 replicates are used to estimate the type I error rate and power. All the simulations are carried out by using the publicly available software R (http://www.r-project.org/).

Tables 2–5 list the estimated type I error rates (when $\lambda_1 = \lambda_2 = 1$) and powers (when $\lambda_2 = 1.4$) obtained from each method under four different settings with various relative risks. The simulation results show that all methods control type I error rate very well. Test W^{23} usually has the highest powers when the genetic models are recessive, but it loses power when the two relative risks are close to each other, especially under the dominant models. In contrast, W^{14} are more powerful than other methods for dominant models, but it is less powerful for recessive models. Neither W^{23} or W^{14} is robust, which is consistent with Theorem 2.4. Like the default CATT, W^{23} and W^{14} are not robust, and therefore are not recommended in practice if the underlying genetic model is unknown. The simulation results also confirm that W^{13} and W^{24} are asymptotically identical as they obtain almost identical powers for any situation considered. Furthermore, W^{13} and W^{24} are robust in the sense that they have reasonable powers for all situations.

The simulation study also shows that the default CATT is not robust although it is usually most powerful when the genetic models are multiplicative or additive. On the other hand, Pearson's chi-square test is robust but it usually has lower powers compared with other methods. Among those robust methods (chi-square test, MERT, GMS, MAX3, W^{12}, W^{34}, W^{13} (W^{24})), W^{12} is usually the most powerful test for dominant or close to dominant models; while W^{34} is more powerful than others for recessive or close to recessive models. The performance of W^{13} (W^{24}) is usually between those of W^{12} and W^{34}; and overall it is better than those of MERT, GMS, MAX3, and chi-square test.

Figs. 1–4 plot the estimated powers from all methods, except for W^{14}, W^{23}, and W^{24}, for the four situations considered in our simulations. They clearly show the robustness of our proposed methods.

Table 2
Estimated type I error rates and powers with significance level 0.05 and 100,000 replicates when HWE with maf = 0.5 holds for controls.

λ_1	1.00	1.00	1.10	1.18	1.20	1.30	1.40
λ_2	1.00	1.40	1.40	1.40	1.40	1.40	1.40
W^{12}	0.050	0.846	0.760	0.720	0.716	0.755	0.836
W^{34}	0.050	0.894	0.791	0.722	0.716	0.718	0.781
W^{14}	0.053	0.662	0.596	0.615	0.633	0.742	0.860
W^{23}	0.050	0.907	0.761	0.634	0.609	0.540	0.579
W^{13}	0.049	0.875	0.780	0.724	0.719	0.742	0.814
W^{24}	0.049	0.876	0.781	0.724	0.719	0.742	0.814
chi-sq	0.051	0.865	0.730	0.650	0.643	0.696	0.803
MAX3	0.051	0.879	0.766	0.701	0.696	0.732	0.822
GMS	0.051	0.866	0.758	0.700	0.693	0.719	0.808
CATT	0.051	0.808	0.781	0.751	0.751	0.721	0.702
MERT	0.052	0.797	0.775	0.751	0.752	0.729	0.714

Table 3
Estimated type I error rates and powers with significance level 0.05 and 100,000 replicates when HWE with maf=0.3 holds for controls.

λ_1	1.00	1.00	1.10	1.18	1.20	1.30	1.40
λ_2	1.00	1.40	1.40	1.40	1.40	1.40	1.40
W^{12}	0.048	0.914	0.764	0.633	0.595	0.482	0.453
W^{34}	0.047	0.944	0.805	0.657	0.609	0.463	0.396
W^{14}	0.047	0.767	0.570	0.479	0.456	0.435	0.484
W^{23}	0.050	0.952	0.801	0.600	0.548	0.327	0.239
W^{13}	0.049	0.933	0.787	0.650	0.606	0.476	0.428
W^{24}	0.049	0.933	0.787	0.650	0.606	0.476	0.428
chi-sq	0.048	0.928	0.753	0.579	0.536	0.407	0.417
MAX3	0.047	0.941	0.789	0.625	0.582	0.446	0.429
GMS	0.046	0.929	0.776	0.620	0.575	0.436	0.413
CATT	0.050	0.928	0.813	0.679	0.634	0.437	0.259
MERT	0.051	0.862	0.757	0.662	0.627	0.492	0.368

Table 4
Estimated type I error rates and powers with significance level 0.05 and 100,000 replicates when the genotypic frequencies are 0.20, 0.45, 0.35 for controls.

λ_1	1.00	1.00	1.10	1.18	1.20	1.30	1.40
λ_2	1.00	1.40	1.40	1.40	1.40	1.40	1.40
W^{12}	0.050	0.906	0.807	0.736	0.734	0.721	0.768
W^{34}	0.053	0.940	0.838	0.745	0.736	0.682	0.694
W^{14}	0.048	0.751	0.641	0.621	0.636	0.693	0.791
W^{23}	0.050	0.943	0.816	0.673	0.650	0.513	0.498
W^{13}	0.051	0.927	0.827	0.746	0.738	0.705	0.739
W^{24}	0.051	0.927	0.827	0.746	0.739	0.705	0.739
chi-sq	0.050	0.917	0.781	0.678	0.669	0.644	0.730
MAX3	0.050	0.931	0.818	0.725	0.718	0.692	0.746
GMS	0.050	0.914	0.807	0.722	0.715	0.681	0.738
CATT	0.052	0.898	0.833	0.773	0.766	0.679	0.600
MERT	0.052	0.868	0.817	0.768	0.767	0.706	0.646

Table 5
Estimated type I error rates and powers with significance level 0.05 and 100,000 replicates when the genotypic frequencies are 0.35, 0.45, 0.20 for controls.

λ_1	1.00	1.00	1.10	1.18	1.20	1.30	1.40
λ_2	1.00	1.40	1.40	1.40	1.40	1.40	1.40
W^{12}	0.052	0.792	0.738	0.756	0.763	0.8391	0.913
W^{34}	0.053	0.854	0.769	0.754	0.750	0.805	0.872
W^{14}	0.056	0.592	0.588	0.674	0.690	0.826	0.925
W^{23}	0.054	0.864	0.731	0.656	0.634	0.628	0.704
W^{13}	0.053	0.831	0.760	0.760	0.761	0.827	0.898
W^{24}	0.053	0.831	0.761	0.760	0.761	0.827	0.898
chi-sq	0.054	0.815	0.702	0.693	0.693	0.785	0.891
MAX3	0.054	0.837	0.746	0.741	0.742	0.819	0.905
GMS	0.050	0.824	0.734	0.736	0.735	0.806	0.892
CATT	0.050	0.727	0.756	0.787	0.787	0.821	0.840
MERT	0.051	0.750	0.766	0.784	0.785	0.807	0.817

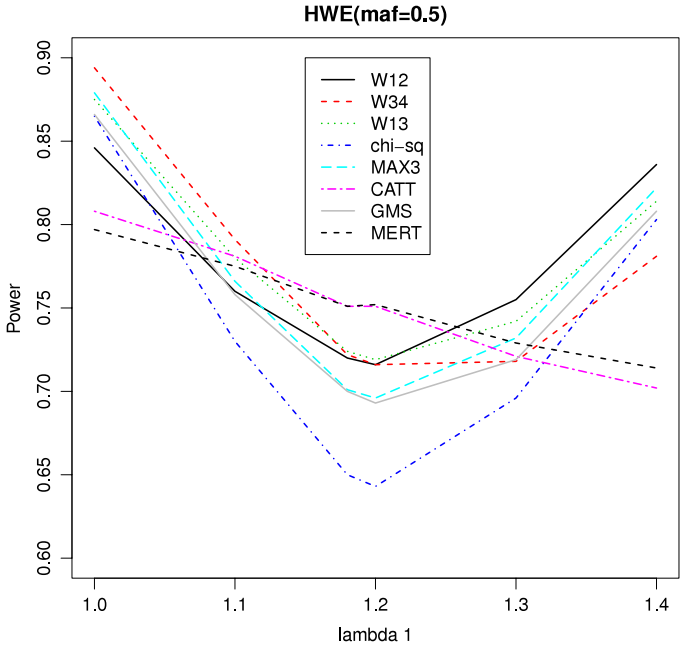

Fig. 1. Estimated powers from each method when HWE holds for controls with maf = 0.5.

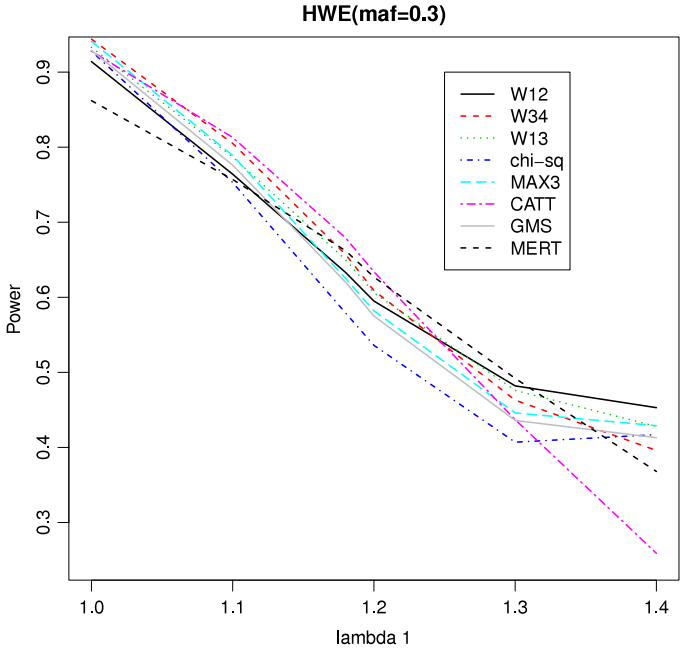

Fig. 2. Estimated powers from each method when HWE holds for controls with maf = 0.3.

4. Application to real data

We then applied the proposed methods to a SNP dataset from a real GWAS conducted by WTCCC (The Wellcome Trust Case Control Consortium and The Australo–Anglo–American Spondylitis Consortium (WTCCC), 2007) (Burton et al., 2007). About 14 000 SNPs were genotyped for about 1000 independent patients from each of the four common diseases: ankylosing spondylitis (AS), autoimmune thyroid disease (ATD), multiple sclerosis (MS) and breast cancer (BC), and a common control group of 1500 randomly selected healthy British individuals. About 12 000 SNPs which passed the quality controls were used in the final analyses.

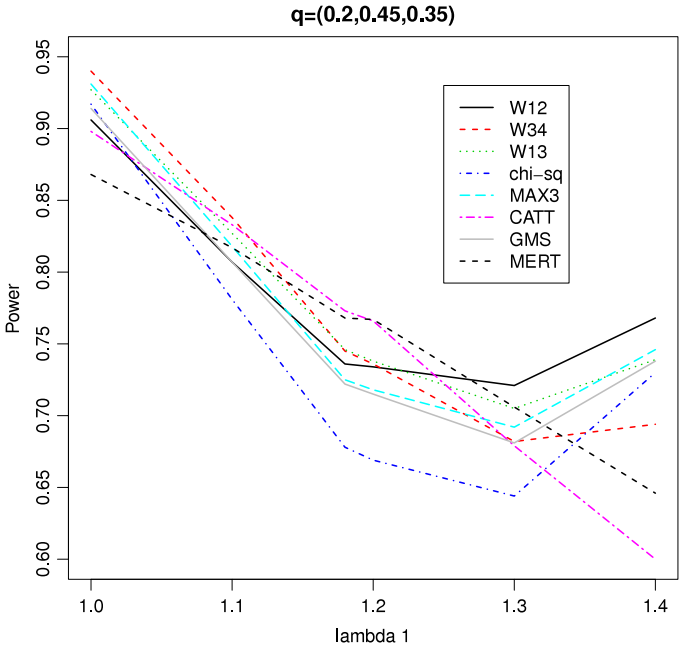

Fig. 3. Estimated powers from each method when HWE does not hold and the genotypic frequencies for controls are 0.20, 0.45, 0.35.

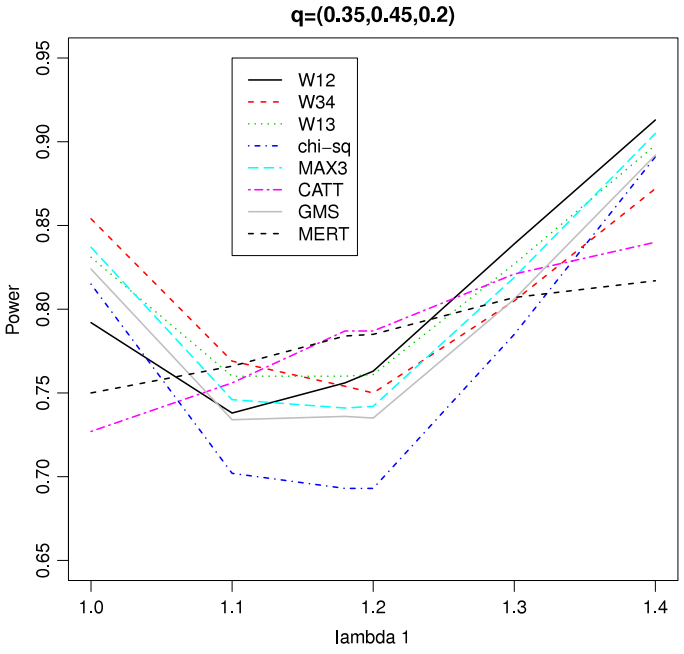

Fig. 4. Estimated powers from each method when HWE does not hold and the genotypic frequencies for controls are 0.35, 0.45, 0.20.

Table 6 lists the p-values obtained by our proposed robust methods, as well as Pearson's chi-square test, MAX3 and the default CATT, which was the test used by the original study, for SNPs with p-values less than 10^{-4} from at least one method. Except for rs27044, which was not genotyped in the follow-up study, all of the SNPs listed in Table 6 from disease AS were confirmed as associated with this disease by a follow-up study conducted by the same authors. The p-values from both studies and the overall p-values from the pooled data were listed in their Table 4 (Burton et al., 2007). From Table 6, we can see, in general, CATT generated larger p-values than MAX3 and our proposed tests did. For three SNPs, rs10050860, rs2287987, and rs17482078, the three proposed test (W^{12}, W^{34}, W^{13}) all obtained smaller p-values than those by other methods. It is also noticeable that the default CATT may lose power for some situations due to its unrobustness. For example, the p-value for SNP rs3733876 was 0.023 from CATT; while those from other methods were all less than 10^{-4}.

Table 6
p-values from various methods for SNPs of the WTCCC data with p-values less than 10^{-4} from at least one test.

Disease	SNP	W^{12}	W^{34}	W^{13}	chi-sq	MAX3	CATT
AS	rs27044	7.88E−7	3.20E−7	3.96E−7	6.19E−7	2.47E−7	1.02E−6
	rs30187	1.23E−6	4.83E−7	6.39E−7	9.53E−7	3.90E−7	2.99E−6
	rs2303138	3.28E−5	2.50E−5	2.57E−5	5.27E−5	2.45E−5	1.05E−5
	rs10050860	1.91E−6	3.77E−6	3.56E−6	6.11E−6	6.41E−6	1.15E−4
	rs2287987	2.18E−6	4.32E−6	4.08E−6	6.86E−6	6.62E−6	1.55E−4
	rs17482078	7.93E−6	1.42E−5	1.36E−5	2.43E−5	2.94E−5	2.32E−4
ATD	rs6427384	2.77E−5	1.20E−5	1.29E−5	1.64E−5	7.32E−6	1.33E−5
	rs3733876	1.23E−5	2.48E−5	2.23E−5	1.48E−5	6.60E−6	2.28E−2
	rs2012199	4.06E−4	1.45E−4	1.55E−4	1.27E−4	9.87E−5	2.83E−4
MS	rs1800437	1.31E−4	6.76E−5	7.13E−5	1.00E−4	4.54E−5	5.97E−5
	rs1132200	2.96E−4	2.07E−4	2.11E−4	4.16E−4	2.37E−4	9.59E−5
BC	rs2285374	6.36E−5	1.18E−4	9.16E−5	1.39E−4	7.55E−5	4.65E−4

5. Discussion

In GWAS, the underlying genetic models of associated SNPs vary and are usually unknown prior to the data analysis; therefore choosing a powerful yet robust statistical test to find the associated SNPs is desirable. Many of the current robust association tests are based on the optimal CATTs for the three special genetic models (recessive, additive, and dominant). They have been shown to be more powerful than Pearson's chi-square test and more robust than the default CATT. In this paper, we propose a class of association tests, W^{12}, W^{34}, W^{13} (W^{24}), which utilize the possible trend of relative risks of SNP data in GWAS. The new tests also incorporate the techniques for combining p-values obtained from independent tests.

There are many ways to combine independent p-values, for example, except for Fisher test used in this paper, we can use Z-test (Chen, 2011b). However, our simulation results (results using Z-test in combining p-values are not shown) indicate that in general Fisher test performs better than Z test. It is possible that under some situations other methods can be more powerful than Fisher test to combine two independent p-values obtained from SNP data in GWAS. This remains an open research topic in this area.

If the underlying genetic models are unknown, we recommend using W^{13} (W^{24}). When the information of the genetic model for a specific SNP is fully or partially known, we may choose other tests. For example, if the genetic models are exactly recessive, additive, multiplicative, or dominant, the CATTs with optimal scores can be used. However, when the genetic model is none of them, but close to dominant, for instance, W^{12} would be a better choice.

Another important issue for GWAS is to adjust for multiple comparisons. Given the large number of correlated SNPs to be tested, the traditional methods, such as Bonferroni correction, are not appropriate, although they are frequently used in GWAS. Many multiple comparison correction approaches based on different assumptions have been proposed for GWAS (Chen and Liu, 2011; Cheverud, 2001; Gao et al., 2009; Li and Ji, 2005), however, it is not clear in general which one is the best. Perhaps the optimal choice depends on the choice of the association test (Chen and Liu, 2011). Since all of those multiple comparison correction methods require subject-level data (i.e., genotypes of all SNPs for each subject), which were not available for the real data we used, we could not apply them to the real data application.

Acknowledgments

The author would like to thank the Associate Editor and two anonymous referees for their helpful comments which result in an improved presentation of this paper. The author also thanks the support from the faculty research funds awarded by the School of Public Health, Indiana University Bloomington.

Appendix. Proof of Theorem 2.4

In this Appendix, we provide the proof for Theorem 2.4. We only give the proof for (i) as the proof for (ii) is very similar. Here we assume n is a large number and random vector Z are defined in (4)–(7). We only need to show $Z_1^{13} = Z_2^{24}$, and $Z_2^{13} = Z_1^{24}$.

Recall $Z^{13} = \begin{bmatrix} Z_1^{13} \\ Z_2^{13} \end{bmatrix} = \Sigma_{13}^{-1/2} \begin{bmatrix} Z_1 \\ Z_3 \end{bmatrix} = \begin{bmatrix} 1 & \rho_{13} \\ \rho_{13} & 1 \end{bmatrix}^{-1/2} \begin{bmatrix} Z_1 \\ Z_3 \end{bmatrix}$, where $\rho_{13} = -\sqrt{\frac{p_1 p_3}{(1-p_1)(1-p_3)}}$. Since $\rho_{13} = -\rho_{24}$, $Z^{24} = \begin{bmatrix} Z_1^{24} \\ Z_2^{24} \end{bmatrix} = \Sigma_{24}^{-1/2} \begin{bmatrix} Z_2 \\ Z_4 \end{bmatrix} = \begin{bmatrix} 1 & \rho_{24} \\ \rho_{24} & 1 \end{bmatrix}^{-1/2} \begin{bmatrix} Z_2 \\ Z_4 \end{bmatrix} = \begin{bmatrix} 1 & -\rho_{13} \\ -\rho_{13} & 1 \end{bmatrix}^{-1/2} \begin{bmatrix} Z_2 \\ Z_4 \end{bmatrix}$. Let $Z_3 = c_{31}Z_1 + c_{32}Z_2$, $Z_4 = c_{41}Z_1 + c_{42}Z_2$, then $\mathrm{Cov}(Z_1, Z_3) = c_{31} = \rho_{13}$, $\mathrm{Cov}(Z_2, Z_3) = c_{32} = \rho_{23}$, $\mathrm{Cov}(Z_1, Z_4) = c_{41} = \rho_{14}$, $\mathrm{Cov}(Z_2, Z_4) = c_{42} = \rho_{24}$. Therefore, $\begin{bmatrix} Z_3 \\ Z_4 \end{bmatrix} = \begin{bmatrix} -\sqrt{\frac{p_1 p_3}{(1-p_1)(1-p_3)}} & \sqrt{\frac{p_2}{(1-p_1)(1-p_3)}} \\ \sqrt{\frac{p_2}{(1-p_1)(1-p_3)}} & \sqrt{\frac{p_1 p_3}{(1-p_1)(1-p_3)}} \end{bmatrix} \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}$. It can be shown that: $\Sigma_{13}^{-1/2} = \begin{bmatrix} a & b \\ b & a \end{bmatrix}$, where $a = (\frac{1}{\sqrt{1+\rho_{13}}} + \frac{1}{\sqrt{1-\rho_{13}}})/2$, $b =$

$(\frac{1}{\sqrt{1+\rho_{13}}} - \frac{1}{\sqrt{1-\rho_{13}}})/2$. Similarly, $\Sigma_{24}^{-1/2} = \begin{bmatrix} a & -b \\ -b & a \end{bmatrix}$. Therefore, $Z_1^{13} - Z_2^{24} = (a + b\rho_{13} - a\rho_{23})Z_1 + (b + b\rho_{23} + a\rho_{13})Z_2$. To show $Z_1^{13} = Z_2^{24}$, we only need to show $a + b\rho_{13} - a\rho_{23} = 0$ and $b + b\rho_{23} + a\rho_{13} = 0$, or $(a + b\rho_{13} - a\rho_{23})^2 = 0$ and $(b+b\rho_{23}+a\rho_{13})^2 = 0$. But $(a+b\rho_{13}-a\rho_{23})^2 = a^2+b^2+(a^2-b^2)\rho_{23}+2ab\rho_{13}-2ab\rho_{13}\rho_{23}-2a^2\rho_{23}$. Since $\rho_{13}^2+\rho_{23}^2 = 1$, we have $a^2 = (\frac{1}{\rho_{23}} + \frac{1}{\rho_{23}^2})/2$, $2ab = -\frac{\rho_{13}}{\rho_{23}^2}$, and $b^2 = (\frac{1}{\rho_{23}^2} - \frac{1}{\rho_{23}})/2$. Simple calculation gives us $(a + b\rho_{13} - a\rho_{23})^2 = 0$, therefore, $a + b\rho_{13} - a\rho_{23} = 0$. Similarly, we can show $b + b\rho_{23} + a\rho_{13} = 0$. Putting them together gives us $Z_1^{13} = Z_2^{24}$. $Z_2^{13} = Z_1^{24}$ can be proved in the same way. This proves (i).

References

Armitage, P., 1955. Tests for linear trends in proportions and frequencies. Biometrics 11 (3), 375–386.

Balding, D.J., 2006. A tutorial on statistical methods for population association studies. Nature Reviews Genetics 7 (10), 781–791.

Burton, P.R., Clayton, D.G., Cardon, L.R., Craddock, N., Deloukas, P., Duncanson, A., Kwiatkowski, D.P., McCarthy, M.I., Ouwehand, W.H., Samani, N.J., et al., 2007. Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. Nature Genetics 39 (11), 1329–1337.

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

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

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., 2013. Testing for association in case-control genome-wide association studies with shared controls. Statistical Methods in Medical Research http://dx.doi.org/10.1177/0962280212474061. Published online before print February 1 2013.

Chen, Z., Liu, Q., 2011. A new approach to account for the correlations among single nucleotide polymorphisms in genome-wide association studies. Human Heredity 72 (1), 1–9.

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

Cheverud, J.M., 2001. A simple correction for multiple comparisons in interval mapping genome scans. Heredity 87 (1), 52–58.

Cochran, W.G., 1954. Some methods for strengthening the common χ^2 tests. Biometrics 10 (4), 417–451.

Esary, J.D., Proschan, F., Walkup, D.W., 1967. Association of random variables, with applications. The Annals of Mathematical Statistics 38 (5), 1466–1474.

Fisher, R.A., 1932. Statistical Methods for Research Workers. Oliver and Boyd, Edinburgh.

Freidlin, B., Podgor, M.J., Gastwirth, J.L., 1999. Efficiency robust tests for survival or ordered categorical data. Biometrics 55 (3), 883–886.

Freidlin, B., Zheng, G., Li, Z., Gastwirth, J.L., 2002. Trend tests for case-control studies of genetic markers: power, sample size and robustness. Human Heredity 53 (3), 146–152.

Gao, X., Becker, L.C., Becker, D.M., Starmer, J.D., Province, M.A., 2009. Avoiding the high Bonferroni penalty in genome-wide association studies. Genetic Epidemiology 34 (1), 100–105.

Gastwirth, J.L., 1966. On robust procedures. Journal of the American Statistical Association 61 (316), 929–948.

Gastwirth, J.L., 1985. The use of maximin efficiency robust tests in combining contingency tables and survival analysis. Journal of the American Statistical Association 80 (390), 380–384.

González, J.R., Carrasco, J.L., Dudbridge, F., Armengol, L., Estivill, X., Moreno, V., 2008. Maximizing association statistics over genetic models. Genetic Epidemiology 32 (3), 246–254.

Joo, J., Kwak, M., Zheng, G., 2010. Improving power for testing genetic association in case-control studies by reducing the alternative space. Biometrics 66 (1), 266–276.

Kwak, M., Joo, J., Zheng, G., 2009. A robust test for two-stage design in genome-wide association studies. Biometrics 65 (4), 1288–1295.

Li, J., Ji, L., 2005. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. Heredity 95 (3), 221–227.

Owen, A.B., 2009. Karl Pearsons meta-analysis revisited. The Annals of Statistics 37 (6B), 3867–3892.

Sasieni, P.D., 1997. From genotypes to genes: doubling the sample size. Biometrics 53, 1253–1261.

Slager, S.L., Schaid, D.J., 2001. Case-control studies of genetic markers: power and sample size approximations for Armitages test for trend. Human Heredity 52 (3), 149–153.

Song, K., Elston, R.C., 2005. A powerful method of combining measures of association and Hardy–Weinberg disequilibrium for fine-mapping in case-control studies. Statistics in Medicine 25 (1), 105–126.

Wang, K., Sheffield, V.C., 2005. A constrained-likelihood approach to marker-trait association studies. The American Journal of Human Genetics 77 (5), 768–780.

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.

Zheng, G., Freidlin, B., Gastwirth, J.L., 2006. Comparison of robust tests for genetic association using case-control studies. IMS Lecture Notes—Monograph Series 49, 253–265.

Zheng, G., Freidlin, B., Li, Z., Gastwirth, J.L., 2003. Choice of scores in trend tests for case-control studies of candidate-gene associations. Biometrical Journal 45 (3), 335–348.

Zheng, G., Ng, H.K.T., 2008. Genetic model selection in two-phase analysis for case-control association studies. Biostatistics 9 (3), 391–399.