Published in final edited form as:

Ann Hum Genet. 2009 March ; 73(2): 133–140. doi:10.1111/j.1469-1809.2008.00500.x.

Pearson's test, trend test, and MAX are all trend tests with different types of scores

Gang Zheng^{a,1}, Jungnam Joo^a, and Yaning Yang^b

^aOffice of Biostatistics Research, DPPS, National Heart, Lung and Blood Institute, USA

^bDepartment of Statistics and Finance, University of Science and Technology of China, China

Summary

Pearson's test is one of the most commonly used statistics for testing genetic association of case-control data. The trend test is another one which assumes a dose-response model between the risk of the disease and genotypes. To apply the trend test, a set of ordered scores is assigned *a priori* based on the underlying genetic model. Pearson's test is model-free and robust, but is less powerful for common genetic models. MAX is another robust test statistic, which takes the maximum of the trend tests over a family of scientifically plausible genetic models. We show that the three test statistics are all trend tests but with different types of scores; whether the scores are prespecified or data-driven, or whether the scores are ordered (restricted) or not ordered (unrestricted). We then provide insights into power performance of the three tests when the underlying genetic model is unknown and discuss which test to use for the analyses of case-control genetic association studies.

Keywords

Case-control design; genetic models; MAX; Pearson's test; robust tests; scores; trend tests

INTRODUCTION

A case-control study is a useful design to test for association between a genetic marker and a disease. Case-control samples for a single marker can be summarized in a 2 by 3 contingency table, with the rows corresponding to the case and control groups and the columns to three genotypes. Basic test statistics for genetic association using case-control samples are reviewed in Balding (2006). Pearson's Chi-square test (Pearson's test for short) is one of the most commonly used statistics to test for genetic association. The Cochran-Armitage trend test (trend test for short) has also been proposed to test for genetic association (Sasieni 1997), which assumes a dose-response effect between the risk of having the disease and genotypes, i.e. the risk of having the disease increases with the number of risk alleles in the genotype. Three trend tests are often used corresponding to the recessive, additive (multiplicative) and dominant models (Sasieni 1997; Freidlin *et al.* 2002; Lettre *et al.* 2007).

For common and complex diseases, the underlying genetic model is usually unknown and is not even necessarily restricted to the above four common genetic models. Choosing a single trend test is not robust when the genetic model is unknown. Therefore, the maximum of the three trend tests across the recessive, additive (multiplicative) and dominant models is

¹Corresponding author: Dr. Gang Zheng, Office of Biostatistics Research, 6701 Rockledge Drive, Bethesda, MD 20892-7913. Fax: 1(301)480-1862. E-mail: zhengg@nhlbi.nih.gov.

proposed (Freidlin *et al.* 2002), denoted by MAX3. Gonzalez *et al.* (2008) proposed a method similar to MAX3. MAX3 is included in SAS JMP Genomics Software (SAS 2008) and has been frequently applied to genome-wide association studies (Sladek *et al.* 2007; Li *et al.* 2008a; 2008b). We consider here a generalization of MAX3 based on (Davies 1977, 1987). We take the maximum of the trend tests over all possible genetic models rather than only the three genetic models. We denote this maximum-type test by MAX. From Zheng & Chen (2005), it is expected that MAX and MAX3 would have similar power performance.

We only focus on Pearson's test, the trend test and MAX, because our goal is to find insightful relationship among them. Other tests, including the allelic test and restricted likelihood ratio test (Wang & Sheffield 2005), are also used in practice. But their power performance is similar to that of the three test statistics we consider here (Sasieni 1997; Wang & Sheffield 2005; Guedj *et al.* 2008; Knapp 2008). Empirical studies have shown that Pearson's test is more robust than the trend test, and that the maximum-type tests are more powerful than Pearson's test (Freidlin *et al.* 2002; Zheng *et al.* 2006; Gonzalez *et al.* 2008). Why and when one test is better than the other is not specifically studied. We will show that the link among the three test statistics is that they are all trend tests with different types of scores, either prespecified or data-driven; either restricted or unrestricted. Our results can then be used to explain the power performance of the three tests and provide insight into the choices of appropriate testing method for case-control genetic association studies.

METHODS

Notation and Model

Denote the genotype counts for cases by (r_0, r_1, r_2) and for controls by (s_0, s_1, s_2) . Let $r = r_0 + r_1 + r_2$ and $s = s_0 + s_1 + s_2$ be the numbers of cases and controls, and $n_i = r_i + s_i$ for $i = 0, 1, 2$. Let $n = r + s$. Penetrances are denoted by $f_i = \text{pr}(\text{disease}|G_i)$, $i = 0, 1, 2$, where $(G_0, G_1, G_2) = (AA, AB, BB)$ are three genotypes. Genotype relative risks (GRRs) are given by $\lambda_1 = f_1/f_0$ and $\lambda_2 = f_2/f_0$. The null hypothesis H_0 is expressed as $H_0: \lambda_1 = \lambda_2 = 1$. If allele B is the risk allele under the alternative hypothesis H_1 , we expect under the dose-response model that $\lambda_2 \geq \lambda_1 \geq 1$ and $\lambda_2 > 1$. The recessive, additive, multiplicative and dominant models correspond to $\lambda_1 = 1$, $\lambda_1 = (1 + \lambda_2)/2$, $\lambda_1 = \lambda_2^{1/2}$, and $\lambda_1 = \lambda_2$ respectively.

To illustrate the three models, we consider a reparameterization with $\lambda_1 = 1 + \lambda \cos \theta$ and $\lambda_2 = 1 + \lambda \sin \theta$, where $\theta \in [0, 2\pi]$ and $\lambda \geq 0$ (Figure 1). Under the alternative hypothesis H_1 , (λ_1, λ_2) and (λ, θ) are one-to-one correspondent. Under H_0 , $(\lambda_1, \lambda_2) = (1, 1)$ is equivalent to $H_0: \lambda = 0$, under which the parameter θ vanishes and is not defined. Under H_1 , the parameter θ corresponds to a genetic model. For examples, the recessive, additive and dominant models correspond to $\theta = \pi/2$, $\arctan(2)$, and $\pi/4$, respectively. How $\theta = \arctan(2) = \tan^{-1}(2)$ is obtained under the additive model is as follows. Under the additive model, we have $2\lambda_1 = 1 + \lambda_2$, which is equivalent to $2\lambda \cos \theta = \lambda \sin \theta$. Since $\lambda \neq 0$ under H_1 , it is equivalent to $\tan \theta = \sin \theta / \cos \theta = 2$, i.e. $\theta = \arctan 2$. Using θ can make it easier to plot the powers across all possible genetic models. However, we will use both parameterizations, one using GRRs (λ_1, λ_2) and the other using the new coordinates (λ, θ) , for the comparison of powers. When $(\lambda_1, \lambda_2) \approx (1, 1)$, the additive model approximates the multiplicative model. If we restrict our attention to a family of genetic models ranging from the recessive to the dominant models (including the additive and multiplicative models), then θ is contained in $[\pi/4, \pi/2]$ when B is the risk allele or in $[5\pi/4, 3\pi/2]$ when A is the risk allele, where the interval $[5\pi/4, 3\pi/2]$ is obtained by adding π to $[\pi/4, \pi/2]$. Adding π , $\cos(\theta+\pi) = -\cos \theta$ and $\sin(\theta+\pi) = -\sin \theta$. Thus, by adding π , the genetic model does not change, but the order of $1 \leq \lambda_1 \leq \lambda_2$ is reversed to $\lambda_1 \geq \lambda_2 \geq 1$. If we do not put restrictions on the underlying genetic model, then $\theta \in [0, 2\pi]$.

Cochran-Armitage Trend Test and Scores

To obtain the Cochran-Armitage trend test (CATT), a set of increasing scores is specified, denoted by (x_0, x_1, x_2) such that $x_0 \leq x_1 \leq x_2$ and $x_0 < x_2$. Later we show that the scores can be decreasing too ($x_0 \geq x_1 \geq x_2$ and $x_0 > x_2$). Then the trend test can be written as (Sasieni 1997; Freidlin *et al.* 2002)

$$T_{\text{CATT}}(x_0, x_1, x_2) = \frac{\left\{ \sum_{j=0}^2 x_j ((1-\phi)r_j - \phi s_j) \right\}^2}{\left[n\phi(1-\phi) \left\{ \sum_{j=0}^2 x_j^2 n_j/n - (\sum_{j=0}^2 x_j n_j/n)^2 \right\} \right]}, \quad (1)$$

where $\phi = r/n$. The trend test $T_{\text{CATT}}(x_0, x_1, x_2)$ is invariant to a linear transformation of the scores. That is, $T_{\text{CATT}}(x_0, x_1, x_2) \equiv T_{\text{CATT}}(0, x, 1)$ where $x = (x_1 - x_0)/(x_2 - x_0)$. Under H_0 , $T_{\text{CATT}}(0, x, 1)$ asymptotically follows a Chi-square distribution with 1 degree of freedom (DF) for a fixed x . For the recessive, additive (or multiplicative) and dominant models, $x = 0, 1/2$ and 1, respectively. These choices of scores correspond to $\theta = \pi/2, \arctan(2)$, and $\pi/4$, respectively.

Note that the scores (x_0, x_1, x_2) increase when $x = (x_1 - x_0)/(x_2 - x_0) \in [0, 1]$. The scores (x_0, x_1, x_2) can be decreasing, under which linear transformations can change the decreasing scores to increasing scores. For example, a linear transformation of the decreasing scores (x_0, x_1, x_2) is given by $(1, (x_1 - x_2)/(x_0 - x_2), 0)$ followed by another linear transformation to $(0, 1 - (x_1 - x_2)/(x_0 - x_2), 1) = (0, (x_0 - x_1)/(x_0 - x_2), 1) = (0, x, 1)$, which are increasing scores. Therefore we only require the scores in the trend test in (1) be ordered or $x \in [0, 1]$. When the possible genetic model ranges from the recessive model to the dominant model, we have $x \in [0, 1]$ with increasing scores, which corresponds to $\theta \in [\pi/4, \pi/2]$. ($\theta \in [5\pi/4, 3\pi/2]$ corresponds to the decreasing scores.) For the underlying genetic models outside of this range, $x \notin [0, 1]$, e.g. the overdominant model ($x > 1$) or underdominant model ($x < 0$).

In practice, when the GRRs (λ_1, λ_2) are ordered, $\lambda_2 \geq \lambda_1 \geq 1$ and $\lambda_2 > 1$ or $\lambda_2 \leq \lambda_1 \leq 1$ and $\lambda_2 < 1$, the trend test is a powerful statistic if it incorporates the correct orders in the scores. However, the trend test may not be robust when the orders are misspecified.

Pearson's Test

Pearson's test, denoted by $T_{\chi^2_2}$, is given by

$$T_{\chi^2_2} = \sum_{i=0}^2 \frac{(r_i - n_i r/n)^2}{n_i r/n} + \sum_{i=0}^2 \frac{(s_i - n_i s/n)^2}{n_i s/n}. \quad (2)$$

Under H_0 , $T_{\chi^2_2}$ asymptotically follows a Chi-square distribution with 2 DF.

From (2), unlike the trend test $T_{\text{CATT}}(0, x, 1)$, $T_{\chi^2_2}$ does not depend on the scores or the underlying genetic model. Thus, it is more robust than $T_{\text{CATT}}(0, x, 1)$, in particular when the score x is misspecified. Analyses of ordered categorical data using Pearson's test and its decomposition studied in Haberman (1974) and Goodman (1979) indicated that Pearson's test is related to the trend tests for ordered categorical data. One such relationship is obtained here in the following result (its proof is given in the Appendix).

RESULT 1—Define scores $(x_0^*, x_1^*, x_2^*) = (r_0/n_0, r_1/n_1, r_2/n_2)$. Let $x^* = (x_1^* - x_0^*)/(x_2^* - x_0^*)$. Then

$$T_{\chi^2_2} \equiv T_{\text{CATT}}(0, x^*, 1). \quad (3)$$

This result shows that Pearson's test is also the trend test with the scores $(0, x^*, 1)$. For the retrospective case-control design, the score x^* is not determined a priori. It is data-driven.

One insight of these scores can be seen from a prospective case-control study, in which the scores (x_0^*, x_1^*, x_2^*) are the maximum likelihood estimates of the penetrances (f_0, f_1, f_2) . For the retrospective case-control study, however, (x_0^*, x_1^*, x_2^*) are the proportions of cases with the given genotypes, which are biased as estimates for the penetrances. As a contrast, the trend test, $T_{\text{CATT}}(0, x, 1)$ with $x \in [0, 1]$, uses increasing scores. Thus, even when the true GRRs are not ordered, the scores for the trend test are still ordered, while the scores $(0, x^*, 1)$ for Pearson's test are not necessarily ordered. The data-driven scores $(0, x^*, 1)$ may capture more information about whether or not the true GRRs are ordered than just using arbitrary ordered scores. Result (3) provides insight into the performance between Pearson's test and the trend test. The robustness of Pearson's test comes from its use of the data-driven scores which is not necessarily ordered. On the other hand, the trend test is more powerful when the scores are correctly specified in terms of the orders. In addition, because the scores for Pearson's test are random, Pearson's test needs higher DF than the trend test that uses fixed scores. Therefore, when the true GRRs are ordered, the trend test with smaller DF is more powerful than Pearson's test with higher DF.

MAX

The MAX statistic discussed in Freidlin *et al.* (2002) can be written as

$$\text{MAX3} = \max \left(T_{\text{CATT}}^{1/2}(0, 0, 1), T_{\text{CATT}}^{1/2}(0, 1/2, 1), T_{\text{CATT}}^{1/2}(0, 1, 1) \right).$$

Under H_0 , the distribution of MAX3 can be obtained by simulation (Freidlin *et al.* 2002) or multiple integrations (Conneely & Boehnke 2007; Gonzalez *et al.* 2008; Li *et al.* 2008). We consider the modified MAX-type test studied by Davies (1977, 1987), given by

$$\text{MAX} = \max_{x \in [0, 1]} T_{\text{CATT}}(0, x, 1). \quad (4)$$

The MAX statistic given in (4) not only covers all three genetic models considered in MAX3 but also includes all other models between the recessive model ($x = 0$) and the dominant model ($x = 1$).

Following the isotonic regression and pool adjacent violator algorithm (Robertson *et al.* 1988; Kimeldorf *et al.* 1992; Zheng, 2003), we show in Appendix that MAX is attained at $T_{\text{CATT}}(0, x^*, 1)$ if the x^* given in Result 1 lies in interval $[0, 1]$ and, otherwise, MAX is reached at $\max(T_{\text{CATT}}(0, 0, 1), T_{\text{CATT}}(0, 1, 1))$. This can be summarized as follows

$$\begin{aligned} \text{MAX} &= T_{\text{CATT}}(0, x^*, 1) \text{ if } x^* \in [0, 1]; \\ &= \max(T_{\text{CATT}}(0, 0, 1), T_{\text{CATT}}(0, 1, 1)) \text{ if } x^* \notin [0, 1]. \end{aligned}$$

From the above argument, we obtain

RESULT 2—When the data-driven score $x^* \in [0, 1]$, we have $\text{MAX} \equiv T_{\chi^2_2}$.

This result shows that, when the score x^* is restricted to 0 and 1 due to the orders corresponding to the underlying genetic model from the recessive model to the dominant model, MAX is identical to Pearson's test. In other words, MAX captures all the information in Pearson's test. In addition, when the true GRRs are ordered from the recessive model to the dominant models, MAX is more powerful than Pearson's test due to its restriction. On the other hand, when the true GRRs are not ordered (i.e. a large family of genetic models is considered), Pearson's test may be more powerful than MAX. Hence, Pearson's test should be more robust than MAX when the family of genetic models include the overdominant (underdominant) model. The above results are summarized in Table 1. The comparison in Table 1 also includes MAX3.

NUMERICAL RESULTS

Table 2 reports the results from a simulation study to examine how often the data-driven scores (x_0^*, x_1^*, x_2^*) are ordered when the true GRRs are ordered. We chose $\theta_0 = \theta/(2\pi) \in [0, 1]$ with increments of 0.001 for a genetic model and $\lambda = 1.0$ for the alternative hypothesis. The sample sizes were $r = s = 150$ with a minor allele frequency of 0.3. The frequency that the scores (x_0^*, x_1^*, x_2^*) are ordered, denoted by *freq*, was estimated using 1,000 replicates. Results in Table 2 show that the frequencies with the ordered scores (x_0^*, x_1^*, x_2^*) are usually higher when the true GRRs are ordered (order = yes). Figure 2 shows the box plots of the frequencies of ordered data-driven scores when the true GRRs are ordered or not. We noticed that when the true GRRs are ordered, the frequency of ordered data-driven scores ranges from 48.2% (in Table 2) to 99.7% with a mean and median frequencies of 79.2% and 77.0%, and a standard deviation of 15.4% (data not shown).

Table 3 reports some of the simulated powers of the three test statistics given $\theta_0 \in [0, 1]$ or $\theta \in [0, 2\pi]$. The genotype counts (r_0, r_1, r_2) in cases and (s_0, s_1, s_2) in controls were simulated from the multinomial distributions $Mul(r; p_0, p_1, p_2)$ and $Mul(s; q_0, q_1, q_2)$, respectively, where $p_i = \text{pr}(G_i)f_i/k$ and $q_i = \text{pr}(G_i)(1 - f_i)/(1 - k)$ for $i = 0, 1, 2$, and k is the disease prevalence. Genotype frequencies $q_i = \text{pr}(G_i)$ were calculated under Hardy-Weinberg equilibrium for a given minor allele frequency. Under H_0 , $f_0 = f_1 = f_2 = k$ and $p_i = q_i = \text{pr}(G_i)$ for all $i = 0, 1, 2$. Under H_1 , (λ_1, λ_2) were obtained given (λ, θ) . Then, $f_0 = k/\{\text{pr}(G_0) + \lambda_1\text{pr}(G_1) + \lambda_2\text{pr}(G_2)\}$, $f_1 = \lambda_1 f_0$ and $f_2 = \lambda_2 f_0$ were calculated. Finally, the probabilities p_i and q_i were calculated. In all simulations, k and the minor allele frequency were fixed. We chose $k = 0.1$ and minor allele frequencies $p = 0.1, 0.3$ and 0.5 . We first simulated data under H_0 so that the critical values for the three test statistics were determined. The sample sizes are $r = s = 150$. The critical values for the three statistics (Pearson's test, the trend test and MAX) were determined using 100,000 replicates. Then we simulated data under H_1 and estimated the power using 10,000 replicates and the previously simulated critical values. The sample sizes for the power comparison were also $r = s = 150$. All test statistics have good control of Type I errors using this Monte-Carlo simulation approach (results not shown).

In Table 3, $\lambda = 1$. The largest power among the three tests is highlighted (in some cases, the second largest power is also highlighted when the difference of power is less than 0.1%). We see that, in the top panel in Table 3 (for the ordered GRRs), either the CATT or MAX is most powerful, and Pearson's test is usually less powerful, in particular that MAX is always more powerful than Pearson's test when $\theta_0 \in [0.125, 0.250]$ corresponding to the ordered GRRs with $\theta \in [\pi/4, \pi/2]$. (Note that when $\theta_0 \in [0.625, 0.750]$ or $\theta \in [5\pi/4, 3\pi/2]$, the genetic model is also ordered.) The CATT is often most powerful for θ_0 in the middle around the additive model ($\theta_0 = 0.176$ with $(\lambda_1, \lambda_2) = (1.448, 1.894)$), while MAX is often most

powerful for the models around the recessive model ($\theta_0 = 0.250$ with $(\lambda_1, \lambda_2) = (1, 2)$) and the dominant model ($\theta_0 = 0.125$ with $(\lambda_1, \lambda_2) = (1.707, 1.707)$).

On the other hand, when the GRRs are not ordered (the bottom panel), i.e. $\theta_0 \notin [0.125, 0.250]$, we only present some results for $\theta_0 \in [0.275, 0.400]$. When the GRRs are not ordered, Pearson's test is always most powerful under the models of consideration, while the CATT loses substantial power when the ordered scores are still used. For minor allele frequency $p = 0.3$ and $\lambda = 0.8$, the maximum power of the three test statistics was chosen over all values of $\theta_0 \in [0, 1]$ and plotted in Figure 3, where D ($\theta_0 = 0.125$ with $(\lambda_1, \lambda_2) = (1.566, 1.566)$) and 0.625 with $(\lambda_1, \lambda_2) = (0.434, 0.434)$), A ($\theta_0 = 0.176$ with $(\lambda_1, \lambda_2) = (1.359, 1.715)$) and 0.676 with $(\lambda_1, \lambda_2) = (0.641, 0.285)$) and R ($\theta_0 = 0.25$ with $(\lambda_1, \lambda_2) = (0, 1.8)$ and 0.75 with $(\lambda_1, \lambda_2) = (0, 0.2)$) stand for the dominant, additive and recessive models, respectively. Figure 3 shows that within the range from the recessive model to the dominant model ($\theta_0 \in [0.125, 0.250]$) or in the range from the dominant model to the recessive model ($\theta_0 \in [0.625, 0.750]$), either the CATT or MAX is most powerful. The CATT is most powerful when the genetic model is close to the additive to the dominant models while MAX is most powerful when the genetic model is close to the recessive. Pearson's test is often most powerful outside of the two ordered ranges. The results also show that the CATT and MAX may be also most powerful for some θ_0 outside of the two ordered ranges.

DISCUSSION

We provided comparisons among the three commonly used test statistics for case-control genetic association studies: Pearson's Chi-square test, the Cochran-Armitage trend test and MAX. We showed that Pearson's test is just a trend test with unrestricted data-driven scores, compared to the trend test using prespecified scores. In addition, we showed that MAX captures all information contained in Pearson's test when the data-driven scores used by Pearson's test are ordered or restricted. These comparisons provide us insights into the power performance among the three statistics. It explains the trade-off in power and robustness between Pearson's test and the trend test, and also explains why MAX is always more powerful than Pearson's test when the underlying genetic model is restricted to the family of genetic models between the recessive model and the dominant model. Similar conclusions can be obtained when MAX3 (Freidlin *et al.* 2002; Gonzalez *et al.* 2008) is used.

In practice, if we know a genetic model, we should choose the trend test corresponding to that model. Also, if we know the scientifically plausible model is restricted to the genetic models between the recessive model and the dominant model, MAX should be used. On the other hand, if the genetic model outside of the above range is likely, e.g. the overdominant model, we should consider Pearson's test. Recently, WTCCC (2007) proposed a new test statistic for genome-wide association studies, which takes the minimum of the p-values for the trend test optimal for the additive model and Pearson's test. This robust test, referred to as MIN2 in Joo *et al.* (2008), utilizes the strength of both the trend test and Pearson's test to resolve the trade-off between them. The asymptotic null distribution for MIN2 is reported in Joo *et al.* (2008) together with the power comparison between MIN2 and MAX3.

Acknowledgments

The authors would like to thank Nancy Geller and two referees for their helpful comments.

Appendix

Proof of Result 1

The proof is given for any $2 \times K$ table ($K \geq 3$). Denote the first row of a $2 \times K$ table by $(n_{11}, \dots, n_{1K})$ with the total $n_{1\cdot}$ and the second row by $(n_{21}, \dots, n_{2K})$ with the total $n_{2\cdot}$. The column totals are denoted by $n_{\cdot 1}, \dots, n_{\cdot K}$. The grand total is $n_{\cdot\cdot} = n_{1\cdot} + n_{2\cdot}$. Denote $\omega_i = n_{i\cdot}/n_{\cdot\cdot}$, $\hat{p}_{kj} = n_{ij}/n_{i\cdot}$, and $\bar{p}_j = n_{\cdot j}/n_{\cdot\cdot} = \sum_{i=1}^2 \omega_i \hat{p}_{ij}$.

Pearson's test for association, denoted by $T_{\chi^2_{K-1}}$ can be written as

$$T_{\chi^2_{K-1}} = \sum_{i=1}^2 \sum_{j=1}^K \frac{(n_{ij} - n_{\cdot j} n_{i\cdot} / n_{\cdot\cdot})^2}{n_{\cdot j} n_{i\cdot} / n_{\cdot\cdot}}.$$

We first show that $T_{\chi^2_{K-1}}$ can also be written as

$$T_{\chi^2_{K-1}} = n_{\cdot\cdot} \omega_1 \omega_2 \sum_j \frac{(\hat{p}_{1j} - \hat{p}_{2j})^2}{\bar{p}_j} = n_{\cdot\cdot} \omega_1 \sum_j \frac{\hat{p}_{1j}(\hat{p}_{1j} - \hat{p}_{2j})}{\bar{p}_j}.$$

Note that

$$\begin{aligned} T_{\chi^2_{K-1}} &= n_{\cdot\cdot} \sum_i \sum_j \frac{(n_{i\cdot}/n_{\cdot\cdot})(n_{ij}/n_{i\cdot} - n_{\cdot j}/n_{\cdot\cdot})^2}{n_{\cdot j}/n_{\cdot\cdot}} = n_{\cdot\cdot} \sum_i \sum_j \frac{\omega_i (\hat{p}_{ij} - \bar{p}_j)^2}{\bar{p}_j} \\ &= n_{\cdot\cdot} \sum_j \frac{\omega_1 (\hat{p}_{1j} - \omega_1 \hat{p}_{1j} - \omega_2 \hat{p}_{2j})^2}{\bar{p}_j} + n_{\cdot\cdot} \sum_j \frac{\omega_2 (\hat{p}_{2j} - \omega_1 \hat{p}_{1j} - \omega_2 \hat{p}_{2j})^2}{\bar{p}_j} \\ &= n_{\cdot\cdot} \omega_1 \omega_2 \sum_j \frac{(\hat{p}_{1j} - \hat{p}_{2j})^2}{\bar{p}_j}. \end{aligned}$$

From $\omega_1 \hat{p}_{1j} / \bar{p}_j + \omega_2 \hat{p}_{2j} / \bar{p}_j$ it follows

$$\omega_1 \sum_j \frac{\hat{p}_{1j}(\hat{p}_{1j} - \hat{p}_{2j})}{\bar{p}_j} + \omega_2 \sum_j \frac{\hat{p}_{2j}(\hat{p}_{1j} - \hat{p}_{2j})}{\bar{p}_j} = \sum_j (\hat{p}_{1j} - \hat{p}_{2j}) = 0. \quad (5)$$

Therefore, we find

$$\omega_1 \sum_j \frac{\hat{p}_{1j}(\hat{p}_{1j} - \hat{p}_{2j})}{\bar{p}_j} = \omega_2 \sum_j \frac{(\hat{p}_{1j} - \hat{p}_{2j})^2}{\bar{p}_j} - \hat{p}_{1j}(\hat{p}_{1j} - \hat{p}_{2j}),$$

where the right hand side is $-\omega_2 \sum_j \{\hat{p}_{2j}(\hat{p}_{1j} - \hat{p}_{2j}) / \bar{p}_j\}$. Thus, the claim follows from $\sum_j \{\hat{p}_{1j}(\hat{p}_{1j} - \hat{p}_{2j})\} / \bar{p}_j = -\omega_2 \sum_j \{\hat{p}_{1j}(\hat{p}_{2j})^2\} / \bar{p}_j$. The CATT in (1) is

$$T_{\text{CATT}}(x_0, x_1, x_2) = \frac{n'' \left\{ \sum_j x_j (n_{1j}/n_{1\cdot}) - \sum_j x_j (n_{\cdot j}/n'') \right\}^2}{(n_{2\cdot}/n_{1\cdot}) \left[\sum_j x_j^2 (n_{\cdot j}/n'') - \left\{ \sum_j x_j (n_{\cdot j}/n'') \right\}^2 \right]}$$

where $x_0 = 0$, $x_1 = x$, and $x_2 = 1$. For the $2 \times K$ table, the data-driven scores are

$x_1^* = (n_{1j}/n_{1\cdot})(n_{1\cdot}/n'')(n_{\cdot j}/n'') = \hat{p}_{1j}\omega_1/\bar{p}_j$. Note that $\sum_j x_j^* (n_{1j}/n_{1\cdot}) = \omega_1 \sum_j (\hat{p}_{1j}^2/\bar{p}_j)$ and $\sum_j x_j^* (n_{\cdot j}/n'') = \omega_1$. Further, $\hat{p}_{1j} - \bar{p}_j = \omega_2 (\hat{p}_{1j} - \hat{p}_{2j})$. Thus, $\sum_j x_j^* (n_{1j}/n_{1\cdot}) - \sum_j x_j^* (n_{\cdot j}/n'') = \omega_1 \sum_j \hat{p}_{1j} (\hat{p}_{1j} - \bar{p}_j) / \bar{p}_j$. Hence the numerator of $T_{\text{CATT}}(x_0, x_1, x_2)$ using x^* can be written as

$$n'' \omega_1^2 \omega_2^2 \left\{ \sum_j \frac{\hat{p}_{1j} (\hat{p}_{1j} - \hat{p}_{2j})}{\bar{p}_j} \right\}^2. \quad (6)$$

Also, since $\sum_j (x_j^*)^2 (n_{\cdot j}/n'') = \omega_1^2 \sum_j (\hat{p}_{1j}^2/\bar{p}_j)$, the denominator of $T_{\text{CATT}}(x_0, x_1, x_2)$ using x^* can be written as

$$\omega_1 \omega_2 \left(\sum_j \frac{\hat{p}_{1j}^2}{\bar{p}_j} - 1 \right) = \omega_1 \omega_2 \sum_j \frac{\hat{p}_{1j} (\hat{p}_{1j} - \bar{p}_j)}{\bar{p}_j} = \omega_1 \omega_2^2 \sum_j \frac{\hat{p}_{1j} (\hat{p}_{1j} - \hat{p}_{2j})}{\bar{p}_j}. \quad (7)$$

Using (6) and (7), $T_{\text{CATT}}(x_0, x_1, x_2) = n'' \omega_1 \sum_j \hat{p}_{1j} (\hat{p}_{1j} - \hat{p}_{2j}) / \bar{p}_j = T_{\chi^2_{k-1}}$.

Proof of Result 2

Kimeldorf et al. (1992) considered MAX of one-sided asymptotically normally distributed trend test over $x \in [0, 1]$. From Kimeldorf et al. (1992), when $r_0/r \leq s_0/s$ and $r_2/r \geq s_2/s$, cases are stochastically greater than controls and, for any $x \in [0, 1]$, the one-sided trend test is non-negative. On the other hand, when $r_0/r \geq s_0/s$ and $r_2/r \leq s_2/s$, cases are stochastically smaller than controls. Then the trend test is non-positive for any $x \in [0, 1]$. In this case, we can switch the two alleles. Since we consider two-sided trend tests, we do not need to consider this case. If cases are neither stochastically greater nor smaller than the controls, they are incomparable. Then there exists $x \in [0, 1]$ such that the trend test equals 0. Based on the results of Kimeldorf et al. (1992), MAX (two-sided) is reached using the scores r_0/n_0 , r_1/n_1 , and r_2/n are used in the trend test when the scores are ordered (increasing or decreasing); otherwise the pool adjacent violators algorithm (PAVA) can be applied to find MAX. The PAVA would pool the two adjacent columns into a single column. Thus, the same score will be given to the two pooled columns. For a 2×3 table, this is equivalent to using scores (0, 0, 1) if the first two columns are pooled or (0, 1, 1) if the last two columns are pooled. Note that these two scores are also “extreme” scores, corresponding to the two boundaries of the genetic models.

References

- Balding D. A tutorial on statistical methods for population association studies. *Nature Review Genetics*. 2006; 7:781–791.
- Conneely KN, Boehnke M. So many correlated tests, so little time! Rapid adjustment of P values for multiple correlated tests. *American Journal of Human Genetics*. 2007; 81:1158–1168. [PubMed: 17966093]
- Davies RB. Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*. 1977; 64:247–254.
- Davies RB. Hypothesis-testing when a nuisance parameter is present only under the alternative. *Biometrika*. 1987; 74:33–43.
- 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]
- Haberman SJ. Log-linear models for frequency tables with ordered classifications. *Biometrics*. 1974; 30:589–600.
- Gonzalez JR, Carrasco JL, Dudbridge F, Armengol L, Estivill X, Moreno V. Maximizing association statistics over genetic models. *Genetic Epidemiology*. 2008; 32:246–254. [PubMed: 18228557]
- Goodman LA. Simple models for the analysis of association in cross classifications having ordered categories. *Journal of the American Statistical Association*. 1979; 74:537–552.
- Guedj M, Nuel G, Prum B. A note on allelic tests in case-control association studies. *Annals of Human Genetics*. 2008; 72:407–409. [PubMed: 18355390]
- Joo J, Kwak M, Ahn K, Zheng G. A robust genome-wide scan statistic of the Wellcome Trust Case-Control Consortium. *Biometrics*. 2008 to appear.
- Kimeldorf G, Sampson AR, Wright LR. Min and max scorings for two-sample ordinal data. *Journal of the American Statistical Association*. 1992; 87:241–247.
- Knapp M. On the asymptotic equivalence of allelic and trend statistic under Hardy-Weinberg equilibrium. *Annals of Human Genetics*. 2008; 72:589–589. [PubMed: 18503608]
- Lette G, Lange C, Hirschhorn JN. Genetic model testing and statistical power in population-based association studies of quantitative traits. *Genetic Epidemiology*. 2007; 31:358–362. [PubMed: 17352422]
- Li Q, Zheng G, Li Z, Yu K. Efficient approximation of P-value of the maximum of correlated tests, with applications to genome-wide association studies. *Annals of Human Genetics*. 2008a; 72:397–406. [PubMed: 18318785]
- Li Q, Yu K, Li Z, Zheng G. MAX-rank: a simple and robust genome-wide scan for case-control association studies. *Human Genetics*. 2008b; 123:617–623. [PubMed: 18491142]
- Robertson, T.; Wright, FT.; Dykstra, RL. *Order Restricted Statistical Inference*. John Wiley & Sons; 1988.
- SAS JMP Genomics 3.2. N.C.: SAS Institute Inc. Raleigh; 2008.
- Sladek R, Rocheleau G, Rung J, Dina C, Shen L, Serre D, Boutin P, Vincent D, Belisle A, Hadjadj S, Balkau B, Heude B, et al. A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature*. 2007; 445:881–885. [PubMed: 17293876]
- Sasieni PD. From genotype to genes: doubling the sample size. *Biometrics*. 1997; 53:1253–1261. [PubMed: 9423247]
- The Wellcome Trust Case Control Consortium (WTCCC). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007; 447:661–678. [PubMed: 17554300]
- Wang K, Sheffield VC. A constrained-likelihood approach to marker-trait association studies. *American Journal of Human Genetics*. 2005; 77:768–780. [PubMed: 16252237]
- Zheng G. Use of max and min scores for trend tests for association when the genetic model is unknown. *Statistics in Medicine*. 2003; 22:2657–2666. [PubMed: 12898550]
- Zheng G, Chen Z. Comparison of maximum statistics for hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrics*. 2005; 61:254–258. [PubMed: 15737101]

Zheng, G.; Freidlin, B.; Gastwirth, JL. IMS Lecture Notes - Monograph Series. 2006. Comparison of robust tests for genetic association using case-control studies; p. 320-336.(2nd special issue in honor of E.L. Lehmann)

Figure 1.
Reparameterization of the GRRs.

Figure 2.
Box plots of the frequencies that the data-driven scores are ordered given the true GRRs are ordered or not (ordered = Yes or No).

Figure 3.

Plot the maximum power of Pearson's test (χ^2), the trend test (CATT) and MAX over all genetic models (θ_0). The models is restricted if θ_0 is between 0.125 and 0.250 and between 0.625 and 0.750.

Table 1

Relations among Pearson's test, the trend test, MAX3 and MAX in terms of the score x

Tests	Expressions
Trend test	$T_{\text{CATT}}(0, x, 1)$ with fixed x in $[0,1]$
Pearson's test	Trend test with random x in $(-\infty, \infty)$
MAX	Pearson's (trend) test with random x in $[0,1]$
MAX3	Maximum of the three trend tests with x fixed at $0, 1/2$, and 1

Table 2

Simulated frequencies (freq %) of ordered data-driven scores given the true GRRs are ordered or not: $\theta_0 = \theta/(2\pi)$ and $\theta \in [0, 2\pi]$. The GRRs are calculated based on $\lambda_1 = 1 + \cos \theta$ and $\lambda_2 = 1 + \sin \theta$.

θ_0	θ	λ_1	λ_2	freq	ordered
0	0	2.000	1.000	5.1%	no
0.125	$\pi/4$	1.707	1.707	48.2%	yes
0.200	$2\pi/5$	1.309	1.951	77.6%	yes
0.500	π	0.000	1.000	0.0%	no
0.625	$5\pi/4$	0.293	0.293	50.4%	yes
0.700	$7\pi/5$	0.691	0.049	98.0%	yes
0.800	$8\pi/5$	1.309	0.049	22.3%	no

Table 3

Empirical powers (%) for the three test statistics when GRRs are ordered or not ordered: minor allele frequencies $p = 0.1, 0.3, \text{ and } 0.5$, prevalence 0.1 , sample sizes $r = s = 150$, the alternative is specified by $\lambda = 1$ and $k = 0.1$. The results are based on 10,000 replicates. The genetic model is specified by $\theta_0 = \theta/2\pi$. The recessive, additive, and dominant models correspond to $\theta_0 = 0.25, 0.176, \text{ and } 0.125$. The GRRs are ordered if $\theta_0 \in [0.125, 0.250]$; otherwise the GRRs are not ordered. The highlighted powers are the most or nearly most powerful among the tests under comparison

θ_0	GRRs (λ_1, λ_2)	$p = 0.1$			$p = 0.3$			$p = 0.5$		
		CATT	MAX	$T_{\chi^2_2}$	CATT	MAX	$T_{\chi^2_2}$	CATT	MAX	$T_{\chi^2_2}$
0.125	(1.71,1.71)	45.4	45.7	44.9	61.7	63.2	61.4	36.4	45.4	42.4
0.150	(1.59,1.81)	38.6	38.4	36.9	60.2	58.1	54.6	46.8	47.2	43.0
0.176	(1.45,1.89)	29.6	29.4	27.8	57.3	52.6	47.9	56.9	52.5	47.1
0.200	(1.31,1.95)	21.5	21.6	19.9	52.3	48.4	42.8	64.1	60.3	55.2
0.250	(0.00,2.00)	7.4	9.5	9.6	36.0	50.3	48.4	74.8	82.3	80.8
0.275	(0.84,1.99)	6.5	10.6	14.0	26.7	58.9	64.1	77.9	90.7	91.4
0.300	(0.69,1.95)	12.3	19.0	26.0	15.6	67.3	80.9	80.5	97.0	98.1
0.325	(0.55,1.89)	25.5	35.1	44.0	8.0	79.9	94.6	80.4	99.2	99.8
0.350	(0.41,1.81)	42.4	53.3	59.6	4.6	91.3	99.5	78.6	99.2	100.
0.400	(0.19,1.59)	65.4	69.5	70.1	22.2	99.9	99.9	65.1	100.	100.