The Canadian Journal of Statistics
Vol. 31, No. ?, 2003, Pages ???-???
La revue canadienne de statistique

The universal validity of the possible triangle constraint for Affected-Sib-Pairs

Zeny Z. Feng, Jiahua Chen and Mary E. Thompson

Key words and phrases: ASP method; linkage analysis; possible triangle constraint; identical by descent (IBD).
MSC 2000: Primary 92D10; secondary 92D30.

Abstract: In Affected-Sib-Pair analysis, genetic marker data are collected from families with at least two sibs affected by a disease under investigation. At any locus not linked to the disease gene, a sib pair shares 0, 1 and 2 alleles identical by descent (IBD) with probabilities of 0.25, 0.5 and 0.25 respectively. With linkage, the IBD value increases stochastically. Louis, Payami & Thomson (1987) and Holmans(1993) were the first ones who discovered that the IBD distribution satisfies the "possible triangle constraint" in some situations. Consequently, more powerful statistical procedures can be designed in detecting linkage. It is of statistical and genetical importance to investigate whether the possible triangle constraint remains true under general genetic models. In this paper, we introduce a new technique to prove the possible triangle constraint. The proof is particularly simple for the single disease locus case. The general case is proved by linking IBD distributions between marker loci through a transition probability matrix. Thus, our proof is less dependent on genetic concepts.

Title in French: we can supply this

Résumé : In Affected-Sib-Pair analysis, genetic marker data are collected from families with at least two sibs affected by a disease under investigation. At any locus not linked to the disease gene, a sib pair shares 0, 1 and 2 alleles identical by descent (IBD) with probabilities of 0.25, 0.5 and 0.25 respectively. With linkage, the IBD value increases stochastically. Louis, Payami & Thomson (1987) and Holmans(1993) were the first ones who discovered that the IBD distribution satisfies the "possible triangle constraint" in some situations. Consequently, more powerful statistical procedures can be designed in detecting linkage. It is of statistical and genetical importance to investigate whether the possible triangle constraint remains true under general genetic models. In this paper, we introduce a new technique to prove the possible triangle constraint. The proof is particularly simple for the single disease locus case. The general case is proved by linking IBD distributions between marker loci through a transition probability matrix. Thus, our proof is less dependent on genetic concepts.

1. INTRODUCTION

The sib-pair method is a commonly used non-parametric statistical method for linkage analysis introduced by Penrose (1935). Examples include the study of complex traits such as type I diabetes (MIM 222100), type II diabetes (MIM 125853), alcoholism (MIM 103780), bipolar disorder (MIM 125480), schizophrenia (MIM 181500) (Leal & Ott 2000) and the study of inflammatory bowel

disease (Mirea, Bull, Silverberg & Siminovitch 2000, Chen, Kalbfleisch & Romero-Hidalgo 2000, Darlington & Paterson 2000 and Roslin, Loredo-Osti, Greenwood & Morgan 2000). Based on genetic theory, a pair of sibs who are concordant at a given genetic trait would be expected to be concordant at closely linked traits. Consequently, if both sibs are affected by an inheritable disease, the traits linked to the disease-susceptibility genes are more likely to be concordant. By analyzing frequencies of pairs of sibs being like or unlike for two traits, it is possible to detect deviations from their normal values and therefore discover whether or not two traits are linked.

The sib-pair method can be applied by collecting genetic information from families of only two generations. In comparison, most other methods need information from families of three or more generations. It is much easier to collect genetic information from families of two generations. Thus, the sib-pair method is a very useful tool in linkage analysis. Out of many possible kinds of sib pairs, the affected sib pair contains most information for the purpose of linkage analysis between a disease-susceptibility gene and the marker under consideration (Suarez, Rice & Reich 1978 and Liang, Huang & Beaty 2000). It is therefore cost efficient to collect genetic information only from families with at least two affected sibs. The disease gene is believed to lie near the markers of unusual similarity among affected sib pairs (Ott 1999, Chapter 13).

At each locus, an individual possesses two copies of the same gene. One gene can have several different types called alleles. An offspring inherits one allele from each of two parents. If the same allele from one parent is inherited by both sibs, they share this allele identically by descent. The number of alleles shared identical by descent (IBD) by an affected-sib-pair can be 0, 1 or 2. If there is no linkage between a disease gene and the marker under investigation, the affected sib pairs sampled are representative of the general population. Hence, the distribution of the IBD random variable, denoted by (π_0, π_1, π_2), is $(1/4, 1/2, 1/4)$. The affected-sib-pair method detects deviations from this null distribution.

Interestingly, under a number of simple genetic models, the IBD distribution is found to satisfy the "possible triangle constraint" defined by two inequalities: $2\pi_0 \leq \pi_1$ and $\pi_1 \leq 1/2$. This discovery implies that only deviations toward the possible triangle region are suggestive of linkage between the disease locus and the marker under consideration. Taking this into consideration effectively reduces the type I error and allows us to improve the power of the statistical significance test. Thus, it is of statistical and genetical importance to investigate whether the possible triangle constraint remains true in general.

A number of geneticists have worked on this problem recently. Under a single disease locus model and equal male and female recombination rate assumption, Suarez *et al.* (1978) established a relationship between (π_0, π_1, π_2) and population prevalence, genetic variances. Using their result, Louis *et al.* (1987) found that (π_0, π_1, π_2) satisfies the two inequalities. Holmans (1993) generalized this result to the multiple unlinked disease loci model, for the IBD distribution at or linked to any one of the disease loci, and named these two inequalities as the "possible triangle constraint". Farrall (1997) extended the result in another direction to the two disease loci model with sex-specific recombination rates. This result is applicable to the IBD distribution at each of the two disease loci, and is derived using a result of Cordell, Todd, Bennett, Kawaguchi & Farrall (1995). Greenwood & Bull (1999) demonstrated that when sibs live in different environments, the possible triangle constraint can be violated. In view of these important results, it is even more pressing to investigate to what extent the possible triangle constraint remains valid.

In this paper, by using a new technique, we solve this problem under only very general genetic assumptions. We show that under Hardy-Weinberg equilibrium and linkage equilibrium (for two or more disease loci model) and the assumption of no crossover interference, the possible triangle constraint remains valid, regardless of differential male-female recombination fractions, the disease model, or the additivity of the gene effects. It is worth mentioning that our proof avoids the introduction of many genetic concepts. At the same time we note that, at a single disease locus, the two inequalities which form the possible triangle are alternative forms of the additive and dominance variance components of the genetic variance.

Our proof proceeds in a progressive way. The result is proved to be true from simple cases to more complex cases. For the ease of presentation, we organize the paper as follows. In Section 2, we give a new proof for the case when there is only one disease locus. The possible triangle is expressed by two inequalities. The proof of the first inequality in Section 2 is found to be generally applicable. Hence, in Section 3, we give the general proof of the first inequality of the possible triangle constraint. The proof of the second inequality is done case by case. Thus, in Sections 4 and 5, we prove the second inequality of the possible triangle constraint when there are exactly two linked disease loci and when there are multiple linked disease loci, respectively. The proof of the most general case is presented in Section 6.

2. THE POSSIBLE TRIANGLE CONSTRAINT UNDER SINGLE DISEASE LOCUS MODEL

We divide this case into two situations. The first situation is where the marker and the disease locus coincide. We then consider the situation where they are merely linked.

2.1 The marker and the disease locus coincide

Let X_1, X_2, X_3, X_4 be four alleles at the disease locus of two parents randomly sampled from the general population. To fix the notation, X_1, X_2 will be the mother's maternal and paternal alleles and X_3, X_4 will be the father's maternal and paternal alleles, respectively. Under Hardy-Weinberg equilibrium, which we assume, the X_i's are independent and identically distributed random variables.

Let A be the event that a sib is affected by the disease. Let a_1a_2 be the genotype at the disease locus of an individual. We define

$$g(a_1a_2) = P(A|a_1a_2), \tag{1}$$

which is the disease penetrance of individuals with genotype (a_1a_2). Clearly, the population disease prevalence is $E[g(X_1X_2)]$.

Let IBD_M denote the number of alleles shared identical by descent at the marker locus under investigation. Let BA be the event that both sibs are affected. For a randomly selected sib pair from the population, and $j = 0, 1, 2$, we define

$$\pi_j = P(\mathrm{IBD}_M = j|\mathrm{BA})$$

and

$$P_j = P(\mathrm{BA}, \mathrm{IBD}_M = j).$$

By definition, we have

$$P_j = \pi_j P(\mathrm{BA}).$$

The possible triangle constraint at the locus is defined as

$$\pi_1 \geq 2\pi_0, \quad \pi_1 \leq \frac{1}{2}. \tag{2}$$

It is easily seen that (**??**) is equivalent to

$$P_0 + P_2 \geq P_1, \quad P_1 \geq 2P_0. \tag{3}$$

We will say that the vector (π_0, π_1, π_2) satisfies the possible triangle constraint if its elements satisfy the two inequalities, and use a similar convention for (P_0, P_1, P_2). Thus, to prove that (P_0, P_1, P_2) satisfies the two inequalities in (**??**) is equivalent to showing that (π_0, π_1, π_2) satisfies the possible triangle constraint. In this paper, we refer to $P_0 + P_2 \geq P_1$ as the first inequality and $P_1 \geq 2P_0$ as the second inequality. Throughout this paper, our objective is to show that these two inequalities remain under quite general situations.

Recall that under the single disease locus assumption, the disease status is solely determined by the genotype of the individual at X. Hence,

$$\begin{aligned} P_0 &= \frac{1}{16}E\{g(X_1X_3)g(X_2X_4) + g(X_1X_4)g(X_2X_3) \\ &\quad +g(X_2X_3)g(X_1X_4) + g(X_2X_4)g(X_1X_3)\} \\ &= \frac{1}{4}[E\{g(X_1X_3)\}]^2 \\ &= \frac{1}{4}K^2, \end{aligned}$$

where $K = E\{g(X_1X_3)\}$ is the population prevalence. Note that expression $g(X_1X_3)$ implicitly indicates that the sib 1 inherits grand-maternal alleles from both parents while $g(X_2X_4)$ indicates that sib 2 inherits grand-paternal alleles. Therefore, they share 0 alleles IBD. Similarly, we calculate that

$$\begin{aligned} P_1 &= \frac{1}{16}E\{g(X_1X_3)g(X_1X_4) + g(X_1X_4)g(X_1X_3) + g(X_2X_3)g(X_2X_4) \\ &\quad +g(X_2X_4)g(X_2X_3) + g(X_1X_3)g(X_2X_3) + g(X_1X_4)g(X_2X_4) \\ &\quad +g(X_2X_3)g(X_1X_3) + g(X_2X_4)g(X_1X_4)\} \\ &= \frac{1}{2}E\{g(X_1X_3)g(X_1X_4)\} \end{aligned}$$

and

$$\begin{aligned} P_2 &= \frac{1}{16}E\{g^2(X_1X_3) + g^2(X_1X_4) + g^2(X_2X_3) + g^2(X_2X_4)\} \\ &= \frac{1}{4}E\{g^2(X_1X_3)\}. \end{aligned}$$

Let us further define $g(X_1) = E\{g(X_1X_3)|X_1\}$. Hence, $(P_0, P_1, P_2) = \frac{1}{4}([E\{g(X_1X_3)\}]^2, 2E\{g^2(X_1)\}, E\{g^2(X_1X_3)\})$. Recall that X_i, $i = 1, \ldots, 4$ are independent and identically distributed. Therefore, we have

$$\begin{aligned} 4(P_0 + P_2 - P_1) &= E\{g^2(X_1X_3)\} + [E\{g(X_1X_3)\}]^2 - 2E\{g(X_1X_3)g(X_1X_4)\} \\ &= E[\{g(X_1X_3) - g(X_1) - g(X_3) + E\{g(X_1)\}\}^2] \geq 0. \end{aligned} \tag{4}$$

Similarly,

$$\begin{aligned} 2(P_1 - 2P_0) &= E\{g(X_1X_3)g(X_1X_4)\} - [E\{g(X_1X_3)\}]^2 \\ &= E\{g^2(X_1)\} - [E\{g(X_1)\}]^2 \\ &= \mathrm{var}\{g(X_1)\} \geq 0. \end{aligned} \tag{5}$$

Note that $V_A = \mathrm{var}\{g(X_1)\}$ and $V_D = E[\{g(X_1X_3) - g(X_1) - g(X_3) + E\{g(X_1)\}\}^2]$ are respectively the additive and dominance variances defined in genetics. See Kempthorne (1954). The projection $g(X_1)$ of $g(X_1X_3)$ onto X_1 is referred to as the additive effect in the genetics literature. The possible triangle constraint is the consequence of (**??**) and (**??**). When the dominance variance is zero, the first inequality becomes equality, which implies $\pi_1 = 1/2$.

2.2 The marker and the disease locus are linked

Next, we consider the case when a disease is caused by a single locus and the marker under investigation is linked to the disease locus. Let θ_m and θ_f be the recombination fractions of the

maternal and paternal meioses. Put $\bar{\theta}_m = 1 - \theta_m$ and $\bar{\theta}_f = 1 - \theta_f$. Throughout the paper, we use notation M to represent the marker under investigation. Let X represent the disease locus.

Under the current disease model, the penetrance of an individual depends solely on its genotype at X. The probability of both members of a sib pair being affected by the disease is influenced by the IBD value at the marker due to the linkage between M and X. Our derivation of the possible triangle constraint starts with working out the transition probabilities between the IBD values at these two loci.

Let IBD_M and IBD_X be the number of alleles identical by descent of the sib pair at M and X respectively. Employing the same technique as in Haseman & Elston (1972), for $i, j = 0, 1, 2$, the transition probability

$$\mathbb{P}_{ij} = P(\mathrm{IBD}_M = j | \, \mathrm{IBD}_X = i)$$

is found to be the $(i,j)^{\mathrm{th}}$ entry of

$$\mathbb{P} = \begin{bmatrix} \varphi_m\varphi_f & \varphi_m\bar{\varphi}_f + \varphi_f\bar{\varphi}_m & \bar{\varphi}_m\bar{\varphi}_f \\ \frac{1}{2}(\varphi_m\bar{\varphi}_f + \varphi_f\bar{\varphi}_m) & \varphi_m\varphi_f + \bar{\varphi}_m\bar{\varphi}_f & \frac{1}{2}(\varphi_m\bar{\varphi}_f + \varphi_f\bar{\varphi}_m) \\ \bar{\varphi}_m\bar{\varphi}_f & \varphi_m\bar{\varphi}_f + \varphi_f\bar{\varphi}_m & \varphi_m\varphi_f \end{bmatrix}, \tag{6}$$

where $\varphi_m = \theta_m^2 + \bar{\theta}_m^2$, $\varphi_f = \theta_f^2 + \bar{\theta}_f^2$, $\bar{\varphi}_m = 1 - \varphi_m$, $\bar{\varphi}_f = 1 - \varphi_f$.

It turns out that the transition probability matrix has an interesting mathematical property. It is so useful that we state it as a theorem.

THEOREM 1. *Suppose (p_0, p_1, p_2) is a non-negative vector which satisfies the possible triangle constraint as described by (??). Let $\mathbb{P}$ be a transition probability matrix defined by (??), and*

$$(P_0, P_1, P_2) = (p_0, p_1, p_2)\mathbb{P}.$$

Then, (P_0, P_1, P_2) also satisfies the possible triangle constraint as described by (??).

We leave the proof of this theorem to the Appendix. For $j = 0, 1, 2$, let

$$p_j = P(\mathrm{BA}, \mathrm{IBD}_X = j), \quad P_j = P(\mathrm{BA}, \mathrm{IBD}_M = j).$$

It is seen that

$$\begin{aligned} P(\mathrm{BA}, \mathrm{IBD}_M = j) &= \sum_{i=0}^{2} P(\mathrm{BA}, \mathrm{IBD}_M = j, \mathrm{IBD}_X = i) \\ &= \sum_{i=0}^{2} \sum_{G_i^*:\mathrm{IBD}_X = i} P(\mathrm{BA}, \mathrm{IBD}_M = j, G_i^*). \end{aligned}$$

Here, $\{G_i^* : \mathrm{IBD}_X = i\}$ represents all possible genotypes of the sib pair at X such that the IBD_X value equals i. The penetrance is determined by genotypes at X only. Thus,

$$\begin{aligned} P(\mathrm{BA}, \mathrm{IBD}_M = j) &= \sum_{i=0}^{2} \sum_{G_i^*:\mathrm{IBD}_X = i} P(\mathrm{BA}|G_i^*) P(G_i^*, \mathrm{IBD}_M = j) \\ &= \sum_{i=0}^{2} \sum_{G_i^*:\mathrm{IBD}_X = i} P(\mathrm{BA}|G_i^*) P(G_i^*) P(\mathrm{IBD}_M = j|G_i^*) \\ &= \sum_{i=0}^{2} \sum_{G_i^*:\mathrm{IBD}_X = i} P(\mathrm{BA}, G_i^*) P(\mathrm{IBD}_M = j|G_i^*). \end{aligned}$$

Note that $P(\text{IBD}_M = j|G_i^*)$ does not depend on the particular genotype, but on the IBD_X value only. Thus, $P(\text{IBD}_M = j|G_i^*) = \mathbb{P}_{ij}$ and hence

$$\begin{aligned} P(\text{BA}, \text{IBD}_M = j) &= \sum_{i=0}^{2} \sum_{G_i^*:\text{IBD}_X=i} P(\text{BA}, G_i^*)\mathbb{P}_{ij} \\ &= \sum_{i=0}^{2} P(\text{BA}, \text{IBD}_X = i)\mathbb{P}_{ij}. \end{aligned}$$

In matrix notation, we have

$$(P_0, P_1, P_2) = (p_0, p_1, p_2)\mathbb{P}. \tag{7}$$

Since (p_0, p_1, p_2) satisfies the possible triangle constraint, the conclusion at the marker hence follows by Theorem 1.

3. PROOF OF THE FIRST INEQUALITY

It turns out that the technique used to establish (**??**), is applicable to general cases. Hence, we prove the validity of the first inequality in (**??**), $P_2 + P_0 - P_1 \geq 0$, of the possible triangle constraint for general cases in this section. Suppose a disease trait is controlled by n disease loci and the n disease loci are spread out over several chromosomes. Let G_n denote the genotypes of the two parents of the n disease loci. Let (M_1M_2, M_3M_4) denote the genotypes of the two parents at the marker M. Here, we let M_1 and M_2 be the mother's maternal and paternal alleles and M_3 and M_4 be the father's maternal and paternal alleles, respectively. We use notation G_M as shorthand for the parents' genotypes (M_1M_2, M_3M_4). We further define

$$f(M_1M_3; G_M, G_n) = P(A|M_1M_3 \text{ at } M; G_M, G_n). \tag{8}$$

Similarly, we define $f(M_1M_4; G_M, G_n)$, $f(M_2M_3; G_M, G_n)$, and $f(M_2M_4; G_M, G_n)$. With these definitions, we have

$$\begin{aligned} &P(\text{BA}, \text{IBD}_M = 0|G_M, G_n) \\ &= \frac{1}{4}\{f(M_1M_3; G_M, G_n)f(M_2M_4; G_M, G_n) + f(M_2M_3; G_M, G_n)f(M_1M_4; G_M, G_n) \\ &\quad + f(M_1M_4; G_M, G_n)f(M_2M_3; G_M, G_n) + f(M_2M_4; G_M, G_n)f(M_1M_3; G_M, G_n)\} \\ &= \frac{1}{2}\{f(M_1M_3; G_M, G_n)f(M_2M_4; G_M, G_n) + f(M_2M_3; G_M, G_n)f(M_1M_4; G_M, G_n)\}. \end{aligned}$$

Therefore, we have

$$\begin{aligned} P_0 &= P(BA, \text{IBD}_M = 0) \\ &= E\{P(BA, \text{IBD}_M = 0|G_M, G_n)\} \\ &= \frac{1}{2}[E\{f(M_1M_3; G_M, G_n)f(M_2M_4; G_M, G_n)\} \\ &\quad + E\{f(M_2M_3; G_M, G_n)f(M_1M_4; G_M, G_n)\}], \end{aligned}$$

where E is the expectation over the distribution of (G_M, G_n). Similarly, we have expressions for P_1 as:

$$\begin{aligned} P_1 &= \frac{1}{2}[E\{f(M_1M_3; G_M, G_n)f(M_1M_4; G_M, G_n)\} + E\{f(M_1M_3; G_M, G_n)f(M_2M_3; G_M, G_n)\} \\ &\quad + E\{f(M_2M_4; G_M, G_n)f(M_2M_3; G_M, G_n)\} + E\{f(M_2M_4; G_M, G_n)f(M_1M_4; G_M, G_n)\}], \end{aligned}$$

and for P_2 as:

$$\begin{aligned} P_2 &= \frac{1}{2}[E\{f^2(M_1M_3;G_M,G_n)\} + E\{f^2(M_1M_4;G_M,G_n)\} \\ &\quad + E\{f^2(M_2M_3;G_M,G_n)\} + E\{f^2(M_2M_4;G_M,G_n)\}]. \end{aligned}$$

It is easy to verify that

$$\begin{aligned} 4(P_2 + P_0 - P_1) &= E[\{f(M_1M_3;G_M,G_n) - f(M_1M_4;G_M,G_n) \\ &\quad + f(M_2M_4;G_M,G_n) - f(M_2M_3;G_M,G_n)\}^2] \\ &\geq 0, \end{aligned}$$

which is the first inequality of the possible constraint.

4. PROOF OF THE SECOND INEQUALITY IN TWO LINKED DISEASE LOCI MODEL

Let X and Y be the two disease loci. Here, we consider three cases separately. The first case occurs when the marker locus M is at X. The second case occurs when the marker locus M is flanked by X and Y. The third case occurs when the marker locus M is outside of and linked to X and Y.

We use notation (X_1X_2, X_3X_4) and (Y_1Y_2, Y_3Y_4) for the genotypes of the two parents at X and Y, where X_1Y_1 and X_2Y_2 denote the mother's maternal and paternal haplotypes and X_3Y_3 and X_4Y_4 denote the father's maternal and paternal haplotypes respectively.

Hence, our notation also provides the phase information between X and Y. We also use G_X and G_Y as shorthand for the parents' genotypes (X_1X_2, X_3X_4) and (Y_1Y_2, Y_3Y_4). Under the Hardy-Weinberg equilibrium and linkage equilibrium, which we assume, the X_i's and Y_i's are two independent sets of independent identically distributed random variables.

4.1. The marker is one of the two disease loci.

Assume that the marker locus M is at the disease locus X. As in Section 2.2, we denote the recombination fractions between X and Y as θ_m and θ_f for mother and father respectively. We continue to use $P_j = P(\text{BA}, \text{IBD}_M = j) = P(\text{BA}, \text{IBD}_X = j)$.

Under the current model, the penetrance of an individual depends only on its genotypes at X and Y. Let a_1a_2 and b_1b_2 be genotypes at X and Y of a sib. Define

$$g(a_1a_2;b_1b_2) = P\{\text{A}|\text{genotypes of a sib being } a_1a_2;b_1b_2\}.$$

With this definition, $g(X_1X_3;Y_1Y_3)$ will be the probability of a sib being affected given that he/she inherits X_1, X_3, Y_1 and Y_3 from his/her parents. Note that the phase information is not relevant in this definition. As for the single locus model, we prove the validity of the second constraint by working out an explicit expression for each of the P_j's. Given the parental haplotypes at X and Y, the genotype of a child at X has 4 possibilities given by

$$\Omega_X = \{(X_1X_3), (X_1X_4), (X_2X_3), (X_2X_4)\}.$$

Suppose the sib 1 has genotype (X_1X_3) at X; then, the possible genotype of sib 2 can be classified into three groups according to the value of IBD at X:

IBD	0	1	2
Genotype	X_2X_4	X_1X_4, X_2X_3	X_1X_3

.

Due to the possibility of crossover between X and Y, the genotype of the first sib at Y can be any one of the allele pairs in

$$\Omega_Y = \{(Y_1Y_3), (Y_1Y_4), (Y_2Y_3), (Y_2Y_4)\}.$$

The sib pair may have any one of the 16 combined genotypes at Y in $\Omega_Y \times \Omega_Y$. The distribution, however, is not uniform due to linkage. In particular, given the first sib having genotype (X_1X_3) and parental phase information, the genotype distribution of the first sib at Y is given by

Y_1Y_3	Y_2Y_3	Y_1Y_4	Y_2Y_4
$\bar{\theta}_m\bar{\theta}_f$	$\theta_m\bar{\theta}_f$	$\bar{\theta}_m\theta_f$	$\theta_m\theta_f$

. (9)

Given $\text{IBD}_X = 0$ and sib 1 having genotype (X_1X_3), the second sib has genotype (X_2X_4) at X, and consequently, the genotype distribution of sib 2 at Y is

Y_1Y_3	Y_2Y_3	Y_1Y_4	Y_2Y_4
$\theta_m\theta_f$	$\bar{\theta}_m\theta_f$	$\theta_m\bar{\theta}_f$	$\bar{\theta}_m\bar{\theta}_f$

.

Let us further define

$$f(X_1X_3; G_X, G_Y) = P(\text{A}|\text{the sib has } X_1X_3 \text{ at } X, G_X, G_Y). \tag{10}$$

That is, $f(X_1X_3; G_XG_Y)$ is the conditional probability that a sib is affected, given parental genotypes G_X and G_Y and that the sib inherits maternal alleles X_1 and X_3 from the parents. By (**??**), it is seen that

$$\begin{aligned} f(X_1X_3; G_X, G_Y) &= \bar{\theta}_m\bar{\theta}_f g(X_1X_3; Y_1Y_3) + \theta_m\bar{\theta}_f g(X_1X_3; Y_2Y_3) \\ &\quad +\bar{\theta}_m\theta_f g(X_1X_3; Y_1Y_4) + \theta_m\theta_f g(X_1X_3; Y_2Y_4). \end{aligned} \tag{11}$$

In the above expression, the phase information is important. When the X and Y entries in g have the same subindex, the coefficient will have a factor $\bar{\theta}_m$ or $\bar{\theta}_f$, which indicates no recombination occurred. Otherwise, the coefficient will have a factor θ_m or θ_f. With this convention, we define and compute $f(X_1X_4; G_X, G_Y)$, $f(X_2X_3; G_X, G_Y)$ and $f(X_2X_4; G_X, G_Y)$ similarly. We notice that when θ is small, an individual who takes maternal allele X_3 at X is more likely to take allele Y_3 at Y. Similarly, an individual who takes allele X_4 at X is more likely to take allele Y_4 at Y. Thus, due to the phase information

$$E\{f(X_1X_3; G_X, G_Y)\} \neq E\{f(X_1X_4; G_X, G_Y)\}.$$

As in the last section, we use the same expression for P_0, P_1 and P_2 except that we specified (G_X, G_Y) in function f instead of the general notation of (G_M, G_n). Hence, for example, we have

$$P_0 = \frac{1}{2}[E\{f(X_1X_3; G_X, G_Y)f(X_2X_4; G_X, G_Y)\} + E\{f(X_1X_4; G_X, G_Y)f(X_2X_3; G_X, G_Y)\}$$

To prove the second constraint, we need to introduce notations

$$\begin{aligned} g(;) &= E[g(X_1X_3; Y_1Y_3)], \\ g(X_1;) &= E[g(X_1X_3; Y_1Y_3)|X_1], \\ g(; Y_1) &= E[g(X_1X_3; Y_1Y_3)|Y_1], \\ g(; Y_1Y_3) &= E[g(X_1X_3; Y_1Y_3)|Y_1Y_3] \end{aligned}$$

and similarly

$$\begin{aligned} g(X_1;Y_1) &= E[g(X_1X_3;Y_1Y_3)|X_1,Y_1)], \\ g(X_1;Y_1Y_3) &= E[g(X_1X_3;Y_1Y_3)|X_1,Y_1Y_3)], \end{aligned}$$

and so on. Note that the above definitions are not dependent on the phase information. With this notation, we have

$$\begin{aligned} &E[f(X_1X_3;G_X,G_Y)f(X_2X_4;G_X,G_Y)] \\ &= E[E\{f(X_1X_3;G_X,G_Y)|G_Y\}E\{f(X_2X_4;G_XG_Y)|G_Y\}] \\ &= E[\{\bar\theta_m\bar\theta_f g(;Y_1Y_3)+\theta_m\bar\theta_f g(;Y_2Y_3)+\bar\theta_m\theta_f g(;Y_1Y_4)+\theta_m\theta_f g(;Y_2Y_4)\} \\ &\quad \times\{\bar\theta_m\bar\theta_f g(;Y_2Y_4)+\theta_m\bar\theta_f g(;Y_1Y_4)+\bar\theta_m\theta_f g(;Y_2Y_3)+\theta_m\theta_f g(;Y_1Y_3)\}] \\ &= (\theta_m^2+\bar\theta_m^2)(\theta_f^2+\bar\theta_f^2)E\{g^2(;)\}+2\{\theta_f\bar\theta_f(\theta_m^2+\bar\theta_m^2)+\theta_m\bar\theta_m(\theta_f^2+\bar\theta_f^2)\}E\{g^2(;Y_1)\} \\ &\quad +4\theta_m\theta_f\bar\theta_m\bar\theta_f E\{g^2(;Y_1Y_3)\}. \end{aligned}$$

Note, we use notation $E\{g^2(;)\}$ in the above expression even though $E\{g^2(;)\}=g^2(;)$. We may compute $E[f(X_2X_3;G_X,G_Y)f(X_1X_4;G_X,G_Y)]$ similarly and find it to be the same. Thus,

$$\begin{aligned} P_0 &= (\theta_m^2+\bar\theta_m^2)(\theta_f^2+\bar\theta_f^2)E\{g^2(;)\} \\ &\quad +2\{\theta_f\bar\theta_f(\theta_m^2+\bar\theta_m^2)+\theta_m\bar\theta_m(\theta_f^2+\bar\theta_f^2)\}E\{g^2(;Y_1)\} \\ &\quad +4\theta_m\theta_f\bar\theta_m\bar\theta_f E\{g^2(;Y_1Y_3)\} \\ &= \varphi_m\varphi_f E\{g^2(;)\}+\{\varphi_m\bar\varphi_f+\varphi_f\bar\varphi_m\}E\{g^2(;Y_1)\}+\bar\varphi_m\bar\varphi_f E\{g^2(;Y_1Y_3)\} \\ &= \mathbb{P}_{00}E\{g^2(;)\}+\mathbb{P}_{01}E\{g^2(;Y_1)\}+\mathbb{P}_{02}E\{g^2(;Y_1Y_3)\}. \end{aligned}$$

Using exactly the same technique, it is straightforward to find

$$P_1 = 2\mathbb{P}_{10}E\{g^2(X_1;)\}+2\mathbb{P}_{11}E\{g^2(X_1;Y_1)\}+2\mathbb{P}_{12}E\{g^2(X_1;Y_1Y_3)\}.$$

From the well known fact that $\mathrm{var}(X)\geq\mathrm{var}(E(X|Y))$ for any two random variables, we obtain $E\{g^2(X_1;)\}\geq E\{g^2(;)\}$, $E\{g^2(X_1;Y_1)\}\geq E\{g^2(;Y_1)\}$ and $E\{g^2(X_1;Y_1Y_3)\}\geq E\{g^2(;Y_1Y_3)\}$. Hence,

$$P_1 \geq 2\mathbb{P}_{10}E\{g^2(;)\}+2\mathbb{P}_{11}E\{g^2(;Y_1)\}+2\mathbb{P}_{12}E\{g^2(;Y_1Y_3)\}.$$

Therefore,

$$\begin{aligned} P_1-2P_0 &\geq 2(\mathbb{P}_{10}-\mathbb{P}_{00})E\{g^2(;)\}+2(\mathbb{P}_{11}-\mathbb{P}_{01})E\{g^2(;Y_1)\} \\ &\quad +2(\mathbb{P}_{12}-\mathbb{P}_{02})E\{g^2(;Y_1Y_3)\} \\ &= (\mathbb{P}_{01}-2\mathbb{P}_{00})E\{g^2(;)\}+2(\mathbb{P}_{11}-2\mathbb{P}_{10})E\{g^2(;Y_1)\} \\ &\quad +(\mathbb{P}_{21}-2\mathbb{P}_{20})E\{g^2(;Y_1Y_3)\} \\ &= (E\{g^2(;)\},2E\{g^2(;Y_1)\},E\{g^2(;Y_1Y_3)\})\mathbb{P}\psi_1 \end{aligned} \tag{12}$$

with $\psi_1=(-2,1,0)$. The first equality is the result of the special structure of $\mathbb{P}$.

Notice that $(E\{g^2(;)\},2E\{g^2(;Y_1)\},E\{g^2(;Y_1Y_3)\})$ has the same structure as $([E\{g(X_1X_3)\}]^2,2E\{g^2(X_1)\},E\{g^2(X_1X_3)\})$ in Section 2.1, hence it satisfies the possible triangle constraint (**??**). By Theorem 1, $(E\{g^2(;)\},2E\{g^2(;Y_1)\},E\{g^2(;Y_1Y_3)\})\mathbb{P}$ also satisfies possible triangle constraint. Therefore, P_1-2P_0 in (**??**) is non-negative. This proves the second inequality of the possible triangle constraint.

Together with the result we have for the first constraint in Section 3, we proved that the possible triangle constraint remains at the marker locus that is at one of the two linked disease loci. This result has been obtained by Farrall (1997) as he expressed the joint IBD probabilities at the two

linked disease loci in terms of the genetic variance components. As the variance components could not be negative, he noted that possible triangle constraint remains at each of the two loci. Our proof in this genetic model employs a different technique, which also allows us to extend our proof to the case of more complicated situations, e.g., when the marker under investigation is flanked by two linked disease loci, or in a more general situation, when the marker is linked to one of multiple loci. In the following subsections and sections, we will prove that the second constraint remains fulfilled in the rest of these other situations. With the result we have in Section 3, the proof of the possible triangle constraint is then complete.

4.2. The marker is flanked by two disease loci.

In this subsection, we consider the case when the marker M is flanked by two linked disease loci X and Y. Let θ_{1m} and θ_{1f} be the recombination fractions between M and X and θ_{2m} and θ_{2f} be the recombination fractions between M and Y. We similarly define $\bar{\theta}_{1m}$, $\bar{\theta}_{1f}$, $\bar{\theta}_{2m}$ and $\bar{\theta}_{2f}$. We assume that there is no crossover interference, so that the recombination between X and M is independent of recombination between M and Y.

Since the penetrance of an individual depends on genotypes at X and Y only, we retain the notation $g(X_1X_2;Y_1Y_2)$, $g(;)$, $g(X_1;)$ $g(;Y_1)$, G_X, G_Y, Ω_X and Ω_Y. We also retain the notation G_M for the marker M and define

$$f(M_1M_3;G_M,G_X,G_Y) = P(\mathrm{A}|M_1M_3 \text{ at } M;G_M,G_X,G_Y).$$

Even though the probability of a sib being affected does not depend on his/her marker allele types, the marker information induces a distribution on which parts of the parental genotypes G_X and G_Y are transmitted. That is, the phase information is contained in the definition implicitly. Mathematically, $f(M_1M_3;G_M,G_X,G_Y)$ can be expressed in terms of g's as:

$$\begin{aligned} & f(M_1M_3;G_M,G_X,G_Y) \\ = \;& \bar{\theta}_{1m}\bar{\theta}_{1f}\bar{\theta}_{2m}\bar{\theta}_{2f}\; g(X_1X_3;Y_1Y_3) + \theta_{1m}\bar{\theta}_{1f}\bar{\theta}_{2m}\bar{\theta}_{2f}\; g(X_2X_3;Y_1Y_3) \\ & + \cdots + \theta_{1m}\theta_{1f}\theta_{2m}\theta_{2f}\; g(X_2X_4;Y_2Y_4) \end{aligned}$$

which contains 16 terms. The coefficient of each term in the above expression depends on whether X and M or Y and M have the same subscripts. If a pair of X and M have the same subscript, then its coefficient will have a factor of $\bar{\theta}$'s representing that no recombination occurred. Otherwise, the coefficient will have a factor of θ's.

We define $f(M_1M_4;G_M,G_X,G_Y)$, $f(M_2M_3;G_M,G_X,G_Y)$ and $f(M_2M_4;G_M,G_X,G_Y)$ similarly. We use the same technique to prove the second constraint as in Section 4.1. Let $\mathbb{P}^{(1)}$ and $\mathbb{P}^{(2)}$ denote the transition probability matrices of IBD values from M to X and from M to Y respectively. By a routine calculation, it is found that

$$P_0 = (\mathbb{P}^{(1)}_{00},\mathbb{P}^{(1)}_{01},\mathbb{P}^{(1)}_{02})\;\mathcal{Q}\;(\mathbb{P}^{(2)}_{00},\mathbb{P}^{(2)}_{01},\mathbb{P}^{(2)}_{02})^\tau \tag{13}$$

and

$$P_1 = (\mathbb{P}^{(1)}_{10},\mathbb{P}^{(1)}_{11},\mathbb{P}^{(1)}_{12})\;\mathcal{Q}\;(\mathbb{P}^{(2)}_{10},\mathbb{P}^{(2)}_{11},\mathbb{P}^{(2)}_{12})^\tau, \tag{14}$$

where

$$\mathcal{Q} \;=\; \begin{pmatrix} E\{g^2(;)\} & E\{g^2(;Y_1)\} & E\{g^2(;Y_1Y_3)\} \\ E\{g^2(X_1;)\} & E\{g^2(X_1;Y_1)\} & E\{g^2(X_1;Y_1Y_3)\} \\ E\{g^2(X_1X_3;)\} & E\{g^2(X_1X_3;Y_1)\} & E\{g^2(X_1X_3;Y_1Y_3)\} \end{pmatrix}. \tag{15}$$

Let $\mathcal{Q}_{\cdot j}$ be the jth column of $\mathcal{Q}$. Let $\mathcal{E}_{0j} = (\mathbb{P}^{(1)}_{00},\mathbb{P}^{(1)}_{01},\mathbb{P}^{(1)}_{02})\mathcal{Q}_{\cdot j}$, and $\mathcal{E}_{1j} = (\mathbb{P}^{(1)}_{10},\mathbb{P}^{(1)}_{11},\mathbb{P}^{(1)}_{12})\mathcal{Q}_{\cdot j}$, for $j = 0,1,2$. Note that, each column and each row of the matrix $\mathcal{Q}$ has the same structure as $(\,[E\{g(X_1X_3)\}]^2, E\{g^2(X_1)\}, E\{g^2(X_1,X_3)\})$ in Section 2.1. Hence, for example, $(E\{g^2(;)\},$

$2E\{g^2(X_1;)\}, E\{g^2(X_1X_3;)\})$ satisfies the possible triangle constraint. With these properties and applying Theorem 1, for example, we have

$$\mathcal{E}_{10} - 2\mathcal{E}_{00} = (E\{g^2(;)\}, 2E\{g^2(X_1;)\}, E\{g^2(X_1X_3;)\})\mathbb{P}^{(1)}\psi_1 \geq 0.$$

Consequently, defining $\mathcal{E}_0 = (\mathcal{E}_{00}, \mathcal{E}_{01}, \mathcal{E}_{02})$ and $\mathcal{E}_1 = (\mathcal{E}_{10}, \mathcal{E}_{11}, \mathcal{E}_{12})$, we get,

$$\begin{aligned} P_1 - 2P_0 &= \mathcal{E}_1 \begin{pmatrix} \mathbb{P}^{(2)}_{10} \\ \mathbb{P}^{(2)}_{11} \\ \mathbb{P}^{(2)}_{12} \end{pmatrix} - 2\mathcal{E}_0 \begin{pmatrix} \mathbb{P}^{(2)}_{00} \\ \mathbb{P}^{(2)}_{01} \\ \mathbb{P}^{(2)}_{02} \end{pmatrix} \\ &\geq 2\mathcal{E}_0 \begin{pmatrix} \mathbb{P}^{(2)}_{10} - \mathbb{P}^{(2)}_{00} \\ \mathbb{P}^{(2)}_{11} - \mathbb{P}^{(2)}_{01} \\ \mathbb{P}^{(2)}_{12} - \mathbb{P}^{(2)}_{02} \end{pmatrix} \\ &= \mathcal{E}_{00}(\mathbb{P}^{(2)}_{01} - 2\mathbb{P}^{(2)}_{00}) + 2\mathcal{E}_{01}(\mathbb{P}^{(2)}_{11} - 2\mathbb{P}^{(2)}_{10}) + \mathcal{E}_{02}(\mathbb{P}^{(2)}_{21} - 2\mathbb{P}^{(2)}_{20}) \\ &= (\mathcal{E}_{00}, 2\mathcal{E}_{01}, \mathcal{E}_{02})\mathbb{P}^{(2)}\psi_1 \end{aligned}$$

with the last equality obtained in the same way as in (**??**). It turns out that we have

$$(\mathcal{E}_{00}, 2\mathcal{E}_{01}, \mathcal{E}_{02}) = (\mathbb{P}^{(1)}_{00}, \mathbb{P}^{(1)}_{01}, \mathbb{P}^{(1)}_{02})\ \mathcal{Q}^*$$

and hence

$$P_1 - 2P_0 \geq (\mathcal{E}_{00}, 2\mathcal{E}_{01}, \mathcal{E}_{02})\mathbb{P}^{(2)}\psi_1 = (\mathbb{P}^{(1)}_{00}, \mathbb{P}^{(1)}_{01}, \mathbb{P}^{(1)}_{02})\ \mathcal{Q}^*\ \mathbb{P}^{(2)}\psi_1$$

with

$$\mathcal{Q}^* = \begin{pmatrix} E\{g^2(;)\} & 2E\{g^2(;Y_1)\} & E\{g^2(;Y_1Y_3)\} \\ E\{g^2(X_1;)\} & 2E\{g^2(X_1;Y_1)\} & E\{g^2(X_1;Y_1Y_3)\} \\ E\{g^2(X_1X_3;)\} & 2E\{g^2(X_1X_3;Y_1)\} & E\{g^2(X_1X_3;Y_1Y_3)\} \end{pmatrix}.$$

It is obvious now that each row vector of $\mathcal{Q}^*$ satisfies the possible triangle constraint. By Theorem 1, $\mathcal{Q}^*\ \mathbb{P}^{(2)}\psi_1 \geq 0$ component-wise. Hence, we have

$$P_1 - 2P_0 \geq 0.$$

4.3. The marker is outside of the two linked disease loci.

In this subsection, we consider the case when the marker M is outside of and linked to the two linked disease loci X and Y. Without loss of generality, the marker and the two disease loci are in the order of M-X-Y. Let $\mathbb{P}$ be the transition probability matrix between M and X. Let (p_0, p_1, p_2) be the IBD distribution of affected sib pair at X, i.e. $p_j = P(\text{BA}, \text{IBD}_X = j)$, for $j = 0, 1, 2$. By Section 4.1, (p_0, p_1, p_2) satisfies the possible triangle constraint. Let $P_j = P(\text{BA}, \text{IBD}_M = j)$, for $j = 0, 1, 2$. Then $(P_0, P_1, P_2) = (p_0, p_1, p_2)\mathbb{P}$. By Theorem 1, (P_0, P_1, P_2) satisfies the possible triangle constraint. That is, the IBD distribution of an affected sib pair at M satisfies the possible triangle constraint. The assumption that p_j is well defined would be satisfied under the assumption of linkage equilibrium for X and Y.

5. SECOND INEQUALITY UNDER MULTIPLE LINKED DISEASE LOCI MODEL

In this section, we consider the case when all disease loci are on the same chromosome but the number of disease genes may exceed 2. We consider two situations in this case. In the first situation, the marker M is flanked by disease loci. In the second situation, the marker M is at a disease locus and flanked by other disease loci.

5.1. The marker of interest is flanked by disease loci.

First of all, we consider a simple case that there are three disease loci underlying the disease trait. We denote the disease loci by X, Y and Z. Without loss of generality, we may assume that the order of the disease loci and the marker is X-M-Y-Z. Let G_Z represent parental genotypes at Z of a sib. We define

$$g(X_1X_3; Y_1Y_3; G_Z) = P(A|X_1X_3; Y_1Y_3; G_Z),$$

which is the conditional probability of a sib being affected given the genotype X_1X_3 and Y_1Y_3 at X and Y and the parental genotypes at Z. We further define

$$f(M_1M_3; G_M, G_X, G_Y, G_Z) = P(A|M_1M_3 \text{ at } M; G_M, G_X, G_Y, G_Z),$$

which is the conditional probability of a sib being affected given the genotype M_1M_3 at M and parental genotypes at X, Y and Z. With exactly the same routine as in Section 4.2, we find that

$$P_0 = (\mathbb{P}_{00}^{(1)}, \mathbb{P}_{01}^{(1)}, \mathbb{P}_{02}^{(1)}) \; \mathcal{Q} \; (\mathbb{P}_{00}^{(2)}, \mathbb{P}_{01}^{(2)}, \mathbb{P}_{02}^{(2)})$$

and

$$P_1 = (\mathbb{P}_{10}^{(1)}, \mathbb{P}_{11}^{(1)}, \mathbb{P}_{12}^{(1)}) \; \mathcal{Q} \; (\mathbb{P}_{10}^{(2)}, \mathbb{P}_{11}^{(2)}, \mathbb{P}_{12}^{(2)})$$

where $\mathcal{Q}$ has the same form as (**??**) except that the functions g in the components of $\mathcal{Q}$ are given by:

$$\begin{aligned} g(;) &= E[g(X_1X_3; Y_1Y_3; G_Z)], \\ g(X_1;) &= E[g(X_1X_3; Y_1Y_3; G_Z)|X_1], \\ g(; Y_1) &= E[g(X_1X_3; Y_1Y_3; G_Z)|Y_1] \end{aligned}$$

and so on. The proof of $P_1 - 2P_0 \geq 0$ is the same as that in Section 4.2.

If there are more than one disease loci beyond Y, or there are disease loci ahead of X, we may use Z to denote all these disease loci and use G_Z to summarize parental genotypes at these disease loci. The proof of the possible triangle constraint remains the same.

5.2. The marker and one of the disease loci coincide.

If the marker under investigation is a disease locus and is flanked by other disease loci, the possible triangle constraint remains true by the following purely mathematical proof. Symbolically, the current case is X-$(M = Y)$-Z which is the limiting case of X-M-Y-Z when $M \to Y$. The equal sign between M and Y indicates that M and Y are identical. Since the possible triangle constraint is shown to be true for the case of X-M-Y-Z regardless of the distance between M and Y, then, as $M \to Y$, the possible triangle constraint at Y must be true at the limit due to continuity of the recombination functions at 0.

6. GENERAL MULTI-LOCUS DISEASE MODEL

In this section, we prove the validity of the possible triangle constraint in general multi-locus disease models. Suppose a disease trait is controlled by n disease loci and is also affected by environmental factors. The n disease loci may spread out over several chromosomes. Suppose l of them are linked to the marker under investigation. Let G_l be the genotypes of a sib at these l linked loci, and G_e be his/her genotypes of the other $n - l$ disease loci. If the sib pair is concordant with respect to environmental factors, (usually, this may be a reasonable assumption as if siblings grow up in a similar environment in general), conceptually, we can then divide the affected-sib-pair population into subpopulations such that sib pairs in each subpopulation have the same value of G_e. Thus, the

possible triangle constraint is true for the IBD distribution defined on each subpopulation. The IBD distribution of the whole population is simply a mixture of the subpopulation IBD distributions. Consequently, the possible triangle constraint remains true.

7. DISCUSSION

In this paper, the IBD distribution of an affected sib pair at any marker is shown to satisfy the possible triangle constraint in a general disease model under the assumptions of Hardy-Weinberg equilibrium, linkage equilibrium (where applicable), and no crossover interference. By focusing on deviations in the direction of the possible triangle region only, we can effectively reduce the type I error of the statistical significance test and improve the power of the test. For example, when the level of significance is set at 5%, the critical value for the likelihood ratio test under the possible triangle constraint is 3.417, while it would be 5.991 when the constraint is not taken into consideration. Consequently, the sample size requirement will be reduced substantially when the possible triangle constraint is known to hold.

Our proof is largely statistical and relies relatively little on genetic concepts. We defined the gene specific penetrance function in various situations. Under the single disease locus model, the additive variance in genetics is found to be the variance of the conditional expectation of the penetrance function. The dominance variance is found to be the variance of the residual penetrance. With this notion, the derivation is particularly simple and straightforward. We linked the IBD distributions at the marker and at the disease locus by a simple linear transformation (see Theorem 1). This property allowed us to build our proofs from simple models to more complex models.

Greenwood & Bull (1999) studied the affected sib pairs method under various genetic models with environmental co-variates. They noted that the possible triangle constraint remains valid at the presence of the gene-environment interactions, provided that the environmental modification of the genetic effect does not change the direction of the genetic effect. When an exposure changes the direction of the genetic effect, and the two sibs have different exposures, the IBD distribution of this affected sib pair may fall outside the possible triangle. Thus, the possible triangle constraint may not be applicable in this situation. In Section 6, we showed that if each sib pair in the sample is concordant with respect to environmental factors, the possible triangle constraint remains valid.

APPENDIX: PROOF OF THEOREM 1

Let $\xi_1 = (1, 2, 1)$ and $\xi_2 = (-1, 0, 1)$. It is seen that $\xi_1\mathbb{P} = \xi_1$ and $\xi_2\mathbb{P} = (\varphi_m + \varphi_f - 1)\xi_2$. Here, ξ_1 and ξ_2 are eigenvectors of $\mathbb{P}$ from the left.

We now decompose the vector (p_0, p_1, p_2) into a linear combination of the eigenvectors of $\mathbb{P}$ plus a residual:

$$(p_0, p_1, p_2) = \frac{p_1}{2}\xi_1 + (\frac{p_1}{2} - p_0)\xi_2 + (p_2 + p_0 - p_1)(0, 0, 1).$$

Thus, if $(P_0, P_1, P_2) = (p_0, p_1, p_2)\mathbb{P}$, then

$$(P_0, P_1, P_2) = \frac{p_1}{2}\xi_1 + (\frac{p_1}{2} - p_0)(\varphi_m + \varphi_f - 1)\xi_2 + (p_2 + p_0 - p_1)(\mathbb{P}_{20}, \mathbb{P}_{21}, \mathbb{P}_{22}).$$

Further, let $\psi_1 = (-2, 1, 0)^\tau$ and $\psi_2 = (1, -1, 1)^\tau$. It is obvious that $\xi_1\psi_1 = 0$, $\xi_2\psi_1 = 2$ and $\xi_1\psi_2 = \xi_2\psi_2 = 0$. Then, we have

$$\begin{aligned} P_2 + P_0 - P_1 &= (P_0, P_1, P_2)\psi_2 \\ &= (p_2 + p_0 - p_1)(1 - 2\varphi_m)(1 - 2\varphi_f) \\ &\geq 0 \end{aligned}$$

because $\varphi_m, \varphi_f \geq 0.5$ and (p_0, p_1, p_2) are assumed to satisfy the possible triangle constraint (**??**). This proves the first inequality for the possible triangle constraint. For the second inequality, we have

$$\begin{aligned} P_1 - 2P_0 &= (P_0, P_1, P_2)\psi_1 \\ &= (\varphi_m + \varphi_f - 1)(p_1 - 2p_0) + (p_2 + p_0 - p_1)\{\bar{\varphi}_m(2\varphi_f - 1) + \bar{\varphi}_f(2\varphi_m - 1)\} \\ &\geq 0 \end{aligned}$$

This proves the second inequality and hence Theorem 1.

ACKNOWLEDGMENTS

This work was partially supported by Natural Science and Engineering Research Council of Canada (NSERC).

REFERENCES

J. Chen, J. D. Kalbfleisch & S. Romero-Hidalgo (2000). Genetic data analysis of affected sib pairs. *Canadian Journal of Statistics*, 30, 145–151.

H. J. Cordell, J. A. Todd, S. T. Bennett, Y. Kawaguchi & M. Farrall (1995). Two-locus maximum lod score analysis of a multifactorial trait: Joint consideration of IDDM2 and IDDM4 with IDDM1 in type 1 diabetes. *American Journal of Human Genetics*, 57, 920–934.

G. A. Darlington & A. D. Paterson (2000). Genetic analysis of chromosome 6 in inflammatory bowel disease. *Canadian Journal of Statistics*, 30, 152–157.

M. Farrall (1997). Affected sib pair linkage tests for multiple linked susceptibility genes. *Genetic Epidemiology,* 14, 103–115.

C. M. T. Greenwood & S. B. Bull (1999). Analysis of affected sib pairs, with covariates- with and without constraints. *American Journal of Human Genetics*, 64, 871–885.

J. K. Haseman & R. C. (1972). The inverstigation of linkage between a quantitative trait and a marker locus. *Behavior Genetics*. 3, 3-19.

P. Holmans (1993). Asymptotic properties of affected-sib-pair linkage analysis. *American Journal of Human Genetics*, 52, 362–374.

O. Kempthorne (1954). *An Introduction to Genetic Statistics*. John Wiley and Sons, New York.

S. M. Leal & J. Ott (2000). Effects of stratification in the analysis of affected-sib-pair data. *American Journal of Human Genetics*, 66, 567–575.

K. Liang, C. Huang & T. H. Beaty (2000). A unified sampling approach for multipoint analysis of qualitative and quantitative traits in sib pairs. *American Journal of Human Genetics*, 66, 1631–1641.

E. J. Louis, H. Payami & G. Thomson (1987). The affected sib pair method. V: Testing the assumptions. *American Journal of Human Genetics*, 51, 75–92.

L. Mirea, S. B. Bull, M. S. Silverberg & K. A. Siminovitch (2000). The genetic analysis of a complex disease. *Canadian Journal of Statistics*, 30, 138–144.

J. Ott (1999). *Analysis of Human Genetics Linkage*. 3rd edn. Johns Hopkins University Press, Baltimore MD.

L. S. Penrose (1935). The detection of autosomal linkage in data which consist of pairs of brothers and sisters of unspecified parentage. *Annals of Eugenics*, 6, 133–138.

N. M. Roslin, J. C. Loredo-Osti, C. M. T. Greenwood & K. Morgan (2000). Genetic analysis of the role of the HLA region in inflammatory bowel disease. *Canadian Journal of Statistics*, 30, 158–166.

B. K. Suarez, J. Rice & T. Reich (1978). The generalized sib pair IBD distribution: its use in the detection of linkage. *Annals of Human Genetics*, 42, 87–94.

Received ???
Accepted ???

Zeny Z. FENG: zzqfeng@uwaterloo.ca
Jiahua CHEN: jhchen@uwaterloo.ca

Mary E. THOMPSON: methomps@uwaterloo.ca
Department of Statistics and Actuarial Science, University of Waterloo
Waterloo, Ontario, Canada and N2L 3G1