Statistical Analysis of Chromatid Interference

Hongyu Zhao, Mary Sara McPeek and Terence P. Speed

Department of Statistics

University of California at Berkeley

Berkeley, CA 94720

February 1995

Running head: Chromatid Interference Analysis

Abstract

The non-random occurrence of crossovers along a single strand during meiosis can be caused by either chromatid interference or crossover interference or both. While crossover interference has been consistently observed in almost all organisms since the time of the first linkage studies, chromatid interference has not been as thoroughly discussed in the literature, and the evidence provided for it is inconsistent. In this paper, with virtually no restrictions on the nature of crossover interference, we describe the constraints which follow from the assumption of no chromatid interference for single spore data. These constraints are necessary consequences of the assumption of no chromatid interference, but their satisfaction is not sufficient to guarantee no chromatid interference. Models can be constructed in which chromatid interference clearly exists, but is not detectable with single spore data. We then extend our analysis to cover tetrad data, which permits more powerful tests of no chromatid interference. We note that the traditional test of no chromatid interference based on tetrad data does not make full use of the information provided by the data, and we offer a statistical procedure for testing the no chromatid interference constraints which does make full use of the data. The procedure is then applied to data from several organisms. Although no strong evidence of chromatid interference is found, we do observe an excess of two strand double recombinations, i.e. negative chromatid interference.

INTRODUCTION

During meiosis in diploid cells, each chromosome is paired with its homologue. Then each member of a given homologous pair duplicates, producing two identical sister chromatids, so that each synapsed paired structure now consists of a bundle of four homologous chromatids. Usually one or more crossovers occur among the four chromatids. A crossover is a precise breakage-and-reunion event occurring between two non-sister chromatids.

In genetic experiments of the kind considered here, there are two types of data: single spore data, in which the products of a single meiosis are recovered separately, and tetrad data, in which all four meiotic products are recovered together. In either case, the occurrence of crossovers cannot be detected directly, but must be inferred from observed recombination events. In the case of single spore data, for each pair of marker loci, a given meiotic product may be scored as recombinant or non-recombinant. In the case of tetrad data, for each pair of marker loci, there are three possible observed outcomes: parental ditype, non-parental ditype, and tetratype. In order to make inferences about crossovers from these types of data, a model is needed which connects the process of crossing-over to the observed single spore or tetrad recombination outcomes.

The two aspects of crossing-over which are relevant to the observed recombination outcome are first, the distribution of crossover events along the bundle of four chromatids and second, the selection of pairs of nonsister chromatids to be involved in each crossover. We say that there is *crossover interference* if the crossover points are not randomly distributed along the chromosome, and that there is *chromatid interference* if it is not true that any pair of non-sister chromatids is equally likely to be involved in any crossover, independent of which pairs were involved in other crossovers. Chromatid interference is said to be *positive* if the non-sister chromatid pair involved in one crossover is less likely to be involved in another crossover, and *negative* if this chance is increased, see BAILEY (1961 *p*.16). Crossover interference was observed shortly after the rediscovery of Mendel's work, and is seen in almost all organisms. On the other hand, the evidence of chromatid interference is sparse in the literature. It has been reported in *Neurospora crassa* by LINDEGREN and LINDEGREN (1942), in *Saccharomyces cerevisiae* by HAWTHORNE and MORTIMER (1960), in *Aspergillus nidulans* by STRICKLAND (1958) and in a few other organisms. The hypothesis of no chromatid interference implies certain constraints, discussed below, and in the above studies, only a subset of the constraints are tested to assess the existence of chromatid interference. Furthermore, in these studies, markers are considered in groups of three, even when there are more than three markers in the experiment, and no chromatid interference (NCI) is tested for each group. As a result, the actual p-value for these multiple tests should be higher than desired. In this paper, we propose a procedure to test all the constraints simultaneously for any number of loci, and we discuss the choice of the critical value. Some of the data in the literature are reanalyzed and the results are compared with the previous results.

CHARACTERIZATION OF NO CHROMATID INTERFERENCE

Single spore data: In single spore recombination data, if there are $n+1$ markers, $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_{n+1}$, involved in an experiment, then the pattern of each observation can be recorded as $i = (i_1 i_2 \ldots i_n)$, where $i_j = 1$ when $\mathcal{A}_j$ and $\mathcal{A}_{j+1}$ have recombined, and $i_j = 0$ otherwise. It is easy to see that there are 2^n distinct recombination patterns for single spore data involving $n+1$ markers. In the following, we will use p_i to denote the probability of recombination pattern $i = (i_1 i_2 \ldots i_n)$.

Tetrad data: For some organisms, such as *Saccharomyces cerevisiae* (baker's yeast) and *Neurospora crassa* (red bread mold), all four products of a single meiosis can be recovered together in an unordered tetrad. There are many advantages in using these organisms for genetic analysis. In particular, since they allow examination of the distribution of crossovers among all four strands, they are well-suited for investigating the possibility of chromatid interference. As described above, there are three possible observed tetrad patterns when two marker loci are considered. We let p_0, p_1 and p_2 denote the probabilities of observing parental ditype, tetratype and non-parental ditype respectively, between a given pair of markers. For three markers $\mathcal{A}_1$, $\mathcal{A}_2$ and $\mathcal{A}_3$, there are three possible tetrad types between $\mathcal{A}_1$ and $\mathcal{A}_2$, and three between $\mathcal{A}_2$ and $\mathcal{A}_3$, so it seems that there should be $3 \times 3 = 9$ different tetrad types among three markers. In fact, there are 12 distinguishable tetrad types, as we have depicted using ordered pair notation in Table 1, with (P,P) denoting parental ditype between $\mathcal{A}_1$ and $\mathcal{A}_2$ and $\mathcal{A}_2$ and $\mathcal{A}_3$, respectively, and similarly for (P,T), etc. There are seen to be four distinguishable tetrad configurations corresponding to tetratypes between $\mathcal{A}_1$ and $\mathcal{A}_2$ and $\mathcal{A}_2$ and $\mathcal{A}_3$, labeled $(\mathrm{T,T})_1, \ldots, (\mathrm{T,T})_4$. These correspond to configurations in which we have parental ditype, one of two distinguishable tetratypes, or nonparental ditype, respectively, between $\mathcal{A}_1$ and $\mathcal{A}_3$, the two tetratypes being distinguishable through the parental origin of the one (and only one) meiotic product which has a parental pattern of alleles at the three markers. A simple inductive argument similar in nature to the reasoning just used leads to the conclusion that there are 4^{k-1} distinguishable tetrad types among n+1 markers $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_{n+1}$, having tetratypes between exactly k of the n pairs $\mathcal{A}_i, \mathcal{A}_{i+1}$, i=1, ..., n, and either a parental or a nonparental ditype specified for each of the remaining n-k pairs. This observation can be used to prove that there are $2^n + (3^n - 1)2^{n-2}$ distinguishable tetrad types for unordered tetrad data involving $n + 1$ markers, a fact which is stated without proof in SHULT and LINDEGREN (1959). A proof is provided in the APPENDIX. To keep our notation simple, we write $p_i, i = (i_1 i_2 \ldots i_n)$ for the joint probability of having tetrad type i_j between $\mathcal{A}_j$ and $\mathcal{A}_{j+1}$, j=$1, 2, \ldots, n$, where $i_j = 0, 1, 2$ denotes parental ditype, tetratype, and nonparental ditype, respectively, and if $i_j = 1$ (corresponding to tetratype) for exactly $k > 1$ of these intervals, we will use $p_i(h)$, h=$(h_1, \ldots, h_{k-1})$ to denote the probability of the tetrad type uniquely specified by $(i_1, \ldots, i_n)$ and $(h_1, \ldots, h_{k-1})$ where each h_j is 1, 2, 3 or 4, $j = 1, \ldots, k - 1$.

Characterization of NCI for single spore data: Assume we have $n+1$ ordered loci $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_{n+1}$, and let I_j denote the interval between $\mathcal{A}_j$ and $\mathcal{A}_{j+1}$. As before, we let p_i be the probability of recombination pattern $i = (i_1 i_2 \ldots i_n)$ among the loci $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_{n+1}$, where $i_j = 1$ or 0, depending on whether or not there is recombination in I_j. We now define q_i, $i = (i_1 i_2 \ldots i_n)$, $i_j = 0$ or 1 for all $1 \leq j \leq n$, to be the chance of zero crossovers in each of the intervals I_j for which $i_j = 0$ and at least one crossover in each of the intervals I_j for which $i_j = 1$. SPEED, MCPEEK and EVANS (1992) show that under the assumption of NCI, the following relationship holds between $\mathbf{p} = (p_i)$ and $\mathbf{q} = (q_i)$:

$$p_i = \sum_{h \geq i} \frac{1}{2^{h \cdot 1}} q_h,$$

and inverting,

$$q_i = 2^{i \cdot 1} \times \sum_{h \geq i} (-1)^{(h-i) \cdot 1} p_h,$$

where $h \geq i$ means that $h_j \geq i_j$ for all j, and $h \cdot i = \sum_{j=1}^{n} h_j i_j$. As a result, they prove that a necessary and sufficient condition for the existence of a point process model for crossovers which could give rise to a particular set of recombination probabilities $\mathbf{p}$ under NCI is

$$\text{for all} \quad i, \quad 0 \leq \sum_{h \geq i} (-1)^{(h-i) \cdot 1} p_h.$$

The relationship between $\mathbf{p}$ and $\mathbf{q}$ and the constraints on $\mathbf{p}$ under NCI can be reexpressed in tensor (Kronecker) product notation as follows: write $\mathbf{p}$ as a column vector in lexicographical order (last component moving fastest), and likewise $\mathbf{q}$. For instance, in the three-marker case, write $\mathbf{p} = (p_{00}, p_{01}, p_{10}, p_{11})$. Now we define a class of matrices which plays an important role, namely

$$\mathbf{R_n} = \begin{pmatrix} 1 & .5 \\ 0 & .5 \end{pmatrix} \otimes \ldots \otimes \begin{pmatrix} 1 & .5 \\ 0 & .5 \end{pmatrix} \qquad \text{(n terms)},$$

where $\otimes$ is the standard tensor product (see e.g. BELLMAN (1970)). For example, in the

2×2 case, where

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \text{ and } B = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix},$$

then $A \otimes B$ is defined to be the 4×4 matrix

$$A \otimes B = \begin{pmatrix} a_{11}B & a_{12}B \\ a_{21}B & a_{22}B \end{pmatrix}.$$

Then, the relation between $\mathbf{p}$ and $\mathbf{q}$ can be written as $\mathbf{p} = \mathbf{R_n q}$ with inverse $\mathbf{q} = \mathbf{R_n^{-1} p}$, and the NCI constraint on $\mathbf{p}$ can be written as $\mathbf{R_n^{-1} p} \geq \mathbf{0}$.

Characterization of NCI for tetrad data: We now proceed to derive a similar relationship and set of constraints for tetrad data under NCI. First we give several definitions to simplify the following discussion. Given the joint recombination or tetrad probabilities $\mathbf{p}=(p_i)$ for a set of loci, we say that a crossover point process is *compatible* with $\mathbf{p}$ if, under the assumption of NCI, it gives rise to these joint probabilities. Two underlying crossover point processes are said to be *equivalent* with respect to a given set of loci if they are compatible with the same joint recombination or tetrad probabilities $\mathbf{p}$.

In his 1935 paper, MATHER proved that if $k \geq 1$ crossovers occur between a pair of markers on the four-strand bundle, then the conditional probabilities $p_0^{(k)}, p_1^{(k)}$ and $p_2^{(k)}$ of observing a tetrad with parental ditype, tetratype and non-parental ditype, respectively, are given by:

$$\begin{aligned} p_0^{(k)} &= \tfrac{1}{3}(\tfrac{1}{2} + (-\tfrac{1}{2})^k) \\ p_1^{(k)} &= \tfrac{2}{3}(1 - (-\tfrac{1}{2})^k) \\ p_2^{(k)} &= \tfrac{1}{3}(\tfrac{1}{2} + (-\tfrac{1}{2})^k). \end{aligned}$$

If $\mathbf{r} = (r_k)$ is the distribution of the number of crossovers between the two markers, where r_k is the chance of exactly k crossovers occurring between the markers, then

$$
\begin{aligned}
p_0 &= r_0 &&+ \textstyle\sum_{k\geq 2} \frac{1}{3}(\frac{1}{2} + (-\frac{1}{2})^k) r_k \\
p_1 &= r_1 &&+ \textstyle\sum_{k\geq 2} \frac{2}{3}(1 - (-\frac{1}{2})^k) r_k \qquad (*) \\
p_2 &= &&\textstyle\sum_{k\geq 2} \frac{1}{3}(\frac{1}{2} + (-\frac{1}{2})^k) r_k \quad .
\end{aligned}
$$

The possible underlying crossover processes can be grouped together into classes according to which joint tetrad probabilities $\mathbf{p}$ they are compatible with on a given set of loci, i.e. all the processes in a given class are equivalent in the sense defined above. In the two-locus case, any crossover distribution has an equivalent process in which at most two crossovers can occur between the markers. To see this, given any crossover process, suppose (r_k) is the distribution of the number of crossovers between the two markers. Let

$$
\begin{aligned}
q_0 &= r_0 \\
q_1 &= r_1 + \textstyle\sum_{k\geq 2}(\frac{1}{3} - \frac{4}{3}(-\frac{1}{2})^k) r_k \\
q_2 &= 4\textstyle\sum_{k\geq 2} \frac{1}{3}(\frac{1}{2} + (-\frac{1}{2})^k) r_k \\
q_l &= 0 \qquad l \geq 3.
\end{aligned}
$$

It is easy to verify that q_0, q_1, q_2 are all non-negative, and then we can define a new underlying crossover process in which k crossovers occur between the markers with probability q_k, and conditional on the number of crossovers, their distribution in the interval could be taken to be, say, uniform. By substituting $\mathbf{q}$ for $\mathbf{r}$ in (*), we find that this new process gives rise to the same p_0, p_1 and p_2 as does the original process, i.e. they are equivalent.

The relation between $\mathbf{p} = (p_i)$ and $\mathbf{q} = (q_k)$ is

$$
\begin{pmatrix} p_0 \\ p_1 \\ p_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & \frac{1}{4} \\ 0 & 1 & \frac{1}{2} \\ 0 & 0 & \frac{1}{4} \end{pmatrix} \begin{pmatrix} q_0 \\ q_1 \\ q_2 \end{pmatrix},
$$

with inverse

$$
\begin{pmatrix} q_0 \\ q_1 \\ q_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -2 \\ 0 & 0 & 4 \end{pmatrix} \begin{pmatrix} p_0 \\ p_1 \\ p_2 \end{pmatrix}.
$$

We argue that a necessary and sufficient condition for there to be some underlying crossover process compatible with a given set of tetrad probabilities is that

$$\begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -2 \\ 0 & 0 & 4 \end{pmatrix} \begin{pmatrix} p_0 \\ p_1 \\ p_2 \end{pmatrix} \geq \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

To see necessity, suppose there were such a process. Define $\mathbf{q}$ in relation to this process as above ($\mathbf{q}$ in terms of $\mathbf{r}$). Then $\mathbf{p}$ and $\mathbf{q}$ must satisfy the relation given above, and since the q's are nonnegative, the necessity of the constraint is established. To see sufficiency, define $\mathbf{q}$ in terms of $\mathbf{p}$ by the inverse relation given above. Then the constraint implies that the q's are nonnegative, and the q's sum to one since the p's do, so this is sufficient for $\mathbf{q}$ to define a crossover process in which the chance of k crossovers occurring between the two markers is q_k, and the crossovers are, say, uniformly distributed in the interval. This crossover process is compatible with $\mathbf{p}$, and so sufficiency of the constraint is established.

For tetrad data involving three markers, as discussed before, there are 12 distinguishable types including (i_1, i_2), $i_1, i_2 = 0, 1, 2$, and three additional subcells of (1,1). Under the assumption of NCI, the four subcells in (1,1) have equal probabilities, i.e., $p_{11}(1) = p_{11}(2) = p_{11}(3) = p_{11}(4) = \frac{1}{4}p_{11}$. It can be proved that for tetrad data involving three markers, for any underlying crossover process on these two intervals, there always exists a process equivalent to it with at most two crossover events in each interval. We let $\mathbf{q} = (q_{i_1 i_2})$ denote the joint distribution of the number of crossover events in the latter process. Let $\mathbf{p}$ and $\mathbf{q}$ be the vectors of $(p_{i_1 i_2})$ and $(q_{i_1 i_2})$ in lexicographic order, and write

$$\mathbf{T_n} = \begin{pmatrix} 1 & 0 & \frac{1}{4} \\ 0 & 1 & \frac{1}{2} \\ 0 & 0 & \frac{1}{4} \end{pmatrix} \otimes \ldots \otimes \begin{pmatrix} 1 & 0 & \frac{1}{4} \\ 0 & 1 & \frac{1}{2} \\ 0 & 0 & \frac{1}{4} \end{pmatrix} \qquad \text{(n terms)}.$$

Then under the assumption of NCI, the relation between $\mathbf{p}$ and $\mathbf{q}$ can be written as $\mathbf{p} = \mathbf{T_2 q}$ with inverse $\mathbf{q} = \mathbf{T_2^{-1} p}$. The characterization of NCI for tetrad data involving three loci follows naturally from the above results: for any distribution $\mathbf{p}$=$(p_{i_1 i_2})$ of tetrad data across three loci, there is an underlying crossover process satisfying NCI that is compatible with $\mathbf{p}$ if and only if $\mathbf{T_2^{-1}p} \geq \mathbf{0}$ and $p_{11}(1) = p_{11}(2) = p_{11}(3) = p_{11}(4) = \frac{1}{4}p_{11}$.

To generalize to tetrad data involving $n+1$ markers, we first divide all the tetrad data into 3^n categories labeled $i = (i_1 i_2 \ldots i_n)$, where i_j can be 0, 1, or 2, corresponding to parental ditype, tetratype or non-parental ditype, respectively, in the j-th interval. And if, for a given i, there are $k \geq 2$ tetratypes in the n intervals, then we can then subdivide these tetratype combinations further into 4^{k-1} distinguishable sub-categories. We can denote the probability of each subcell by $p_i(h) = p_{i_1 i_2 \ldots i_n}(h_1, h_2, \ldots, h_{k-1})$, where each h_j could be 1,2,3 or 4.

The properties and characterization of NCI for tetrad data involving $n+1$ markers can be summarized as follows, see the APPENDIX for details: For any underlying crossover process, there is a crossover process with at most two events between each consecutive pair of markers which is equivalent to it with respect to the given set of markers. Under the assumption of NCI, the relation between the joint distribution of the number of crossovers $\mathbf{q} = (q_{i_1 i_2 \ldots i_n})$ and the tetrad probabilities $\mathbf{p} = (p_{i_1 i_2 \ldots i_n})$ can be expressed as $\mathbf{p} = \mathbf{T_n q}$, and for those $i = (i_1 i_2 \ldots i_n)$ with $i_r = 1$ for more than one r in $\{1,2, \ldots, n\}$, i.e. more than one tetratype, all the subcells defined by the strands involved have equal probability. Finally, for a given set of joint tetrad probabilities, $\mathbf{p} = (p_{i_1 i_2 \ldots i_n})$, there is an underlying crossover process satisfying NCI compatible with $\mathbf{p}$ if and only if $\mathbf{T_n^{-1}p} \geq \mathbf{0}$, and all the subcell probabilities $p_i(j)$ are equal in any cell i with more than one $i_h = 1$.

STATISTICAL TESTING PROCEDURE

The preceding results characterize NCI in terms of both single spore and tetrad recombination probabilities. It is thus possible to test for chromatid interference not only from tetrad data, on which most of the genetic analysis of chromatid interference is based, but also from single spore data, although it may be more difficult to detect in that case.

Note that the constraints on $\mathbf{p}$ described above are sufficient to ensure the existence of a crossover process satisfying NCI which is compatible with $\mathbf{p}$, but for any $\mathbf{p}$ there exist crossover processes compatible with $\mathbf{p}$ which do not satisfy NCI. However, violation of the constraints would certainly imply chromatid interference, in the absence of other extenuating experimental circumstances such as differential viability of gametes.

In previous analyses of chromatid interference based on tetrad data, loci have been analyzed three at a time. In these studies, only those tetrads having a tetratype in both intervals are typically considered, and in those cases, the four different recombination patterns are combined into three groups: those in which the same two strands show recombination in both intervals, known as 2-strand double recombinants; those in which two strands recombine in one interval and the other two recombine in the other interval, known as 4-strand double recombinants; and the rest which are called 3-strand double recombinants. Since each nonsister pair has an equal chance of recombining in the second interval, the ratio of 2:3:4 strand types should be 1:2:1 under NCI. The χ^2 test is then used to examine whether the ratio expected under the assumption of NCI is observed. However, the analysis just described does not make full use of the data. In order to do this, one must consider the full complement of markers at once and test the linear inequalities as well as the equiprobability constraints characterizing NCI.

Using the constraints on the joint recombination and tetrad probabilities that characterize NCI, we now develop some procedures to test whether these constraints are violated for recombination data or tetrad data from multilocus crosses. The likelihood ratio test is a natural one to use and should have reasonable properties. The corresponding test statistic is defined by:

$$T = -2 \log \frac{\prod_i \hat{p}_i^{x_i}}{\prod_j \dot{p}_j^{x_j}}$$

where x_i is the observed frequency of spores or tetrads with recombination or tetrad type $i = (i_1 i_2 \ldots i_n)$, $\dot{\mathbf{p}} = (\dot{p}_i)$ is the maximum likelihood estimate (MLE) of $\mathbf{p}$ under the NCI constraints, and $\hat{\mathbf{p}} = (\hat{p}_i)$ is the unconstrained MLE of $\mathbf{p}$.

It is straightforward to prove that the unconstrained MLE

$$\hat{p}_j = \frac{x_j}{\sum_i x_i}$$

and we turn to the problem of maximizing $\prod {p_i}^{x_i}$ subject to the no chromatid interference constraints on $\mathbf{p}$, i.e. of computing the constrained MLE $\dot{p}$ of $\mathbf{p}$.

For single spore data, all of the constraints are inequalities except the requirement that the p's must sum to one. For tetrad data, the constraints consist of both linear inequalities and equalities. Notice that all the equality constraints are equiprobability constraints on the subcells of a cell $i = (i_1 i_2 \ldots i_n)$ for which $i_r = 1$ for more than one r. So if we know p_i for such a cell, the probability of each subcell is then determined by the constraint. Thus for tetrad data, we can first maximize the likelihood as if there were only 3^n different recombination patterns subject to the linear constraints. Then the equality constraint allows us to assign equal probability to each subcell in the appropriate cells. For example, consider tetrad data with three markers. We first maximize the likelihood subject to $\mathbf{T_2^{-1} p} \geq \mathbf{0}$. Suppose the likelihood so constrained is maximized when $p_{11} = \dot{p}_{11}$. Then $\dot{p}_{11}$ together with $\dot{p}_{11}(1) = \dot{p}_{11}(2) = \dot{p}_{11}(3) = \dot{p}_{11}(4) = \frac{1}{4}\dot{p}_{11}$ give the maximum likelihood estimate of $\mathbf{p}$ under the full set of no chromatid interference constraints. Therefore for both recombination data and tetrad data, we will maximize the likelihood subject to a set of linear inequalities and the condition $\sum p_i = 1$.

The constrained MLE of p: This optimization problem can be viewed as a special case of the geometric programming problem (Duffin, Peterson and Zener (1967)).

MAZUMDAR and JEFFERSON (1983) use this technique to find the maximum likelihood estimates for multinomial probabilities. The difficulty of the problem increases as the number of constraints increases. When there are many constraints, one approach to reducing the degree of difficulty of this optimization problem is to replace all the linear inequality constraints by a single such constraint. PHILLIPS and BEIGHTLER (1973) give a summary of the so-called *surrogate geometric programming* problem. Under certain conditions, the solution to this simplified optimization problem is the same as the original one. COOKE (1983) applied this technique to maximum likelihood estimation for multinomial probabilities $\mathbf{p} = (p_i)$ when the (p_i) satisfy a set of linear inequality constraints. Unfortunately, his algorithm does not always lead to the correct answer; sometimes, the solution (p_i) from his algorithm does not sum to one. One way to deal with this problem is to replace all the inequality constraints by a single linear inequality constraint, and use this inequality constraint together with the equality constraint $\sum p_i = 1$ as the constraints for a surrogate program.

The geometric programming approach makes no use of the implicit structure of the joint probabilities $\mathbf{p} = (p_i)$. If we think of $\mathbf{p}$ as being derived from an underlying unobservable crossover process, we are led to another algorithm. This algorithm may be viewed as an instance of the EM algorithm, DEMPSTER, LAIRD and RUBIN (1977). Recall that all the constraints are derived from the requirement that the underlying joint crossover probabilities should be non-negative. We may think of unobservable crossover frequencies $\mathbf{y} = (y_i)$ as constituting the complete data, with parameters the joint crossover probabilities $\mathbf{q} = (q_i)$. The incomplete data are then the observed recombination or tetrad frequencies $\mathbf{x} = (x_i)$, with parameters $\mathbf{p}$. Writing $\mathbf{p} = \mathbf{Cq}$, we have $\mathbf{C} = \mathbf{R_n}$ for recombination data and $\mathbf{C} = \mathbf{T_n}$ for tetrad data.

Using the above notation, the algorithm can be described as:

1. Start with some initial estimate $\mathbf{q^0}$ of $\mathbf{q}$.

2. E-step: Compute the expectation of $\mathbf{y}$ conditional on $\mathbf{q}$ and $\mathbf{x}$,

$$y_i^{k+1} = \sum_{j=1}^{N} \{(c_{ji} q_i^k)/(\sum_{s=1}^{N} c_{js} q_s^k)\} x^j.$$

3. M-step: Estimate $\mathbf{q}^{k+1}$ from the expected $\mathbf{y}^{k+1}$ obtained in the previous E-step,

$$q_i^{k+1} = y_i^{k+1}/n.$$

   where n is the number of observations.

4. Repeat until $\mathbf{q}^k$ converges.

For the data we have analyzed, this algorithm converges very quickly. Since the function is convex in $\mathbf{p}$, the likelihood in $\mathbf{p}$ is increased after each iteration, and $\mathbf{p}^k = \mathbf{C}\mathbf{q}^k$ must converge to the global solution.

Estimation of the p-value: In order to have a level-α likelihood ratio test, the critical value c_α for the test statistic T must be chosen as:

$$c_\alpha = max_{\mathbf{p} \in \Theta_0} c_{\alpha,\mathbf{p}} \qquad \text{where} \quad P_{\mathbf{p}}(T > c_{\alpha,\mathbf{p}}) = \alpha,$$

and Θ_0 is the set of all $\mathbf{p}$ satisfying the NCI constraints. To determine c_α, we must search through the whole restricted parameter space Θ_0. Note that apart from the condition that $\mathbf{p}$ be a set of multinomial probabilities and so $\sum p_i = 1$, the other relevant constraints are inequality constraints which can be written in the following form $\mathcal{K}(A) = \{p \in P^n : A'p \geq 0\}$, where $\mathcal{K}(A)$ is *polyhedral cone* in R^k. For the definition and some properties of polyhedral cones, see ROBERTSON, WRIGHT and DYKSTRA (1988 *p*.109). If the observation $\mathbf{x}$ comes from a multivariate normal distribution, PERLMAN (1969) and RAUBERTAS, LEE and NORDHEIM (1986) discuss in detail how to test the hypothesis that the means of this multivariate normal distribution lie in a polyhedral cone. They showed that if the likelihood ratio test is used, the critical value is largest when the multivariate normal distribution has mean $\mathbf{0}$. ROBERTSON, WRIGHT and DYKSTRA (1988 *Chap*.5) discuss the properties of the likelihood ratio test statistic when the hypothesis puts order restrictions on the multinomial probabilities. They proved that asymptotically, $c_{\alpha,\mathbf{p}}$ is largest when all p_i are the same.

Under NCI, the joint probabilities (p_i) lie in the intersection of a hyperplane, $\sum p_i = 1$, and a polyhedral cone. For example, for tetrad data involving two markers, the constraints are $p_i \geq 0$, and

$$\begin{array}{ccccccc} p_0 & + & p_1 & + & p_2 & = & 1 \\ p_0 & & & - & p_2 & \geq & 0 \\ & & p_1 & - & 2p_2 & \geq & 0 \\ & & & & 4p_2 & \geq & 0 \end{array}$$

The last inequality constraint $4p_2 \geq 0$ can be deleted since it must be true for any multinomial distribution. Changing the second and third inequalities into equalities, there is a unique solution $(p_0, p_1, p_2) = (0.25, 0.50, 0.25)$ to the three linear equations. This occurs when the underlying crossover probabilities are $(q_0, q_1, q_2) = (0, 0, 1)$.

For single spore data and tetrad data, the very last inequality is always redundant since it has the form $p_h \geq 0$, where h is the last index in the lexicographical order. Removing this constraint and setting the left-hand side of each remaining inequality equal to 0 gives a system of linear equations with the same number of equations as unknowns. It can be shown that $\mathbf{p^0} = \mathbf{Cq^0}$ is the unique solution to these equations, where $q_i^0 = 0$ for $i \neq h$, $q_h^0 = 1$, $\mathbf{C} = \mathbf{R_n}$ for recombination data, and $\mathbf{C} = \mathbf{T_n}$ for tetrad data.

Using arguments similar to those of ROBERTSON, WRIGHT and DYKSTRA (1988 *Chap.*5), it is not hard to prove that asymptotically, $c_{\alpha,\mathbf{p}}$ is maximized over all $\mathbf{p}$ in the constraint region when $\mathbf{p} = \mathbf{p^0}$. The exact distribution of the test statistic T is not tractable even in the multivariate normal case, so the distribution must be approximated. We may use a Monte Carlo method to approximate the distribution of T in the following way: If the dataset under study is of size m, then simulate $\mathbf{B}$ independent samples, or versions of the dataset, $\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_B}$, each of size m, from probability distribution $\mathbf{p}$. For each sample, calculate the likelihood ratio test statistic $T_i, i = 1, \ldots, B$. When B is large, the empirical distribution of the T_i is a good approximation to the real distribution of T. The $1 - \alpha$ percentile for T can be approximated from the empirical distribution T_i and used as the critical value $c_{\alpha,\mathbf{p}}$. From the asymptotic result that $c_{\alpha,\mathbf{p^0}}$ is the largest of all the $c_{\alpha,\mathbf{p}}$ for $\mathbf{p}$ satisfying the constraints, the following testing procedure should yield a level α test:

1. Calculate the likelihood ratio test statistic T for the observed data (x_i);
2. Use the Monte Carlo method to approximate $c_{\alpha,\mathbf{p^0}}$ as explained above;
3. Reject the hypothesis of no chromatid interference if $T \geq c_{\alpha,\mathbf{p^0}}$.

Although using $c_{\alpha,\mathbf{p^0}}$ as the critical value leads to a level α test asymptotically, simulations suggest it is not a good choice even when the sample size is m=1250, which is quite usual for the data analyzed here. For example, in the case of tetrad data involving three markers, when $\mathbf{q^0} = (0, \ldots 0, 1)$, the estimated $c_{0.05,\mathbf{p^0}}$ is 0.04 from 2000 simulations, and the estimated $c_{0.05,\mathbf{p}}$ for certain other $\mathbf{p}$ based on 2000 simulations can be as large as 0.06. Since the asymptotic result is not valid even for a sample of size 1250, and a numerical search for the biggest $c_{\alpha,\mathbf{p}}$ would be very time consuming, we still need a reliable, fast testing procedure that asymptotically is of level α and performs well for medium sample sizes. When the sample size is large, the maximum likelihood estimate $\dot{\mathbf{p}}$ should be close to the true parameter $\mathbf{p}$, so instead of searching for the biggest $c_{\alpha,\mathbf{p}}$ over all possible $\mathbf{p}$, we might use $c_{\alpha,\dot{\mathbf{p}}}$ as the critical value for the test. GEYER (1991) proposed the same method for estimating the critical value for the likelihood ratio test statistic in the context of isotonic convex logistic regression.

We have conducted some simulations to assess the performance of this procedure. For each simulation, 1000 independent samples of recombination data, each with sample size m=1000, are generated from the probability distribution $\mathbf{p}$. For each sample, we estimate $\mathbf{p}$ by the maximum likelihood estimate $\dot{\mathbf{p}}_\mathbf{i}$ and calculate the likelihood ratio test statistic T_i. Then 1000 independent samples of recombination data, each with size m=1000 are simulated from $\dot{\mathbf{p}}_\mathbf{i}$ to estimate the critical value $c_{\alpha,\dot{\mathbf{p}}_\mathbf{i}}$. If the test procedure proposed above works well, we expect $T_i > c_{\alpha,\dot{\mathbf{p}}_\mathbf{i}}$ with probability α , since the test is set to be of level α. Three different $\mathbf{p}$'s for tetrad data are used as the *true* parameter, with 3, 4 and 5 markers respectively. The components of $\mathbf{p}$ are determined by the following rule: $p_{i_1 i_2 \ldots i_n} = 0$ if $i_1 < 2, i_2 < 2, \ldots, i_n < 2$ and $p_{i_1 i_2 \ldots i_n} = c$ otherwise, i.e., probabilities for all cells $i_1 i_2 \ldots i_n$ with at least one $i_k = 2$ are equal, and are 0 if none of i_k is 2. Out of 1000 samples, $T_i > c_{\alpha,\dot{\mathbf{p}}_\mathbf{i}}$ occurs 51 times when there are 3 markers in tetrad data, 47 times when there are 4 markers and 46 times when there are 5 markers. These results show that in some respect the bootstrap testing procedure has the expected type-I error rate, and since the critical value is estimated from the distribution of likelihood ratio statistic T at $\dot{\mathbf{p}}$, this procedure should have more power than the test where a global maximal critical value is used.

TESTING NCI IN DIFFERENT ORGANISMS

We use the bootstrap testing procedure to test whether chromatid interference exists in a number of previously analyzed datasets. The organisms *Neurospora crassa*, *Saccharomyces cerevisiae*, *Aspergillus nidulans*, and *Drosophila* are treated in the next four subsections.

Neurospora crassa: In their experiments, Lindegren and Lindegren (1942) use four intervals across the centromere on the sex chromosome of *Neurospora crassa.* Intervals I and II are on the left of the centromere and III and IV on the right. These regions are symmetrically placed across the centromere. By examining the 2:3:4 strand ratio in different pairs of adjacent intervals, i.e. the relative proportions of parental ditypes, tetratypes and nonparental ditypes across the combined interval, in tetrads exhibiting tetratype in each of the separate intervals, they conclude that there are locally specific patterns of chromatid interference. Among those tetrads having tetratype in both interval I and IV, the 2:3:4

strand ratio is 15:7:6, showing an excess of 2-strand double recombinants. There is also an excess of 2-strand double recombinants for those tetrads having tetratype in both intervals II and III, where the 2:3:4 strand ratio is 10:2:1. However, there is no significant deviation from the expected 1:2:1 ratio for those tetrads with tetratype in any other pair of intervals. If we use the bootstrap testing procedure to analyze the whole dataset together (sample size 1577), the calculated p-value is less than 0.01.

STRICKLAND (1961) contains the results from four experiments each of which uses the same four markers (with the same phase) in linkage group V of *Neurospora crassa*. He concludes that all four crosses show chromatid interference, and that the interference pattern varies from experiment to experiment. There is a significant deficiency of 3-strand double recombinants in the first experiment, and for the other three experiments, an excess of 2-strand double recombinants is observed for adjacent pairs of intervals, while there is no evidence of chromatid interference in the nonadjacent pair of intervals. When the data from these four experiments are analyzed by the bootstrap testing procedure, the p-values are 0.01, 0.67, 0.55 and 0.73 with sample sizes of 1802, 1968, 2239, and 2810 respectively, so the conclusion based on our proposed testing procedure is that only the data from the first experiment are statistically significant, rejecting the hypothesis of NCI. Indeed, in the first experiment, among 17 tetrads with tetratype in two particular intervals, only 1 of them is of 3-strand type, which shows a substantial deficiency.

In his study, PERKINS (1962) uses 6 markers within the right arm of linkage group I and observes an excess of 2-strand over 4-strand double recombinants. He takes each pair of intervals along the chromosome, counts 2-strand and 4-strand double recombinants, and then adds the numbers of 2-strand double recombinants together and the numbers of 4-strand double recombinants together to see if there is a statistically significant deviation from the expected 1:1 ratio. He estimates the p-value to be 0.05. The p-value estimated from our approach is 0.01 (sample size 1262), showing statistically significant evidence against the hypothesis of NCI. There appears to be an excess of 2-strand over 4-strand double recombinants.

Our method of testing chromatid interference in *Neurospora crassa* for data from different sources does not always give significant evidence against NCI. Among all the datasets, there does seem to be a consistent interference pattern, namely, an excess of 2-strand double recombinants over 4-strand double recombinants, suggesting some degree of *negative* chromatid interference in *Neurospora crassa*. It also seems that this interference is both locus specific and experiment specific.

Saccharomyces cerevisiae: HAWTHORNE and MORTIMER (1960) analyze chromatid interference in *Saccharomyces cerevisiae*. Three datasets in their paper give observed tetrad frequencies for markers in linkage group III, VII and IX. They observe that the frequency of 2-strand and 4-strand double recombinants is greater than what is expected on the basis of a 1:2:1 ratio of 2-, 3- and 4-strand double recombinants. Our analysis of their three datasets gives p-value of 0.57, 0.01 and 0.83 with sample sizes of 278, 93, and 213, respectively. In the second experiment, in which four markers were used, the 2:3:4 strand ratio was 6:4:10 for tetrads showing tetratype in the first and third interval, and was 1:0:2 for tetrads with tetratypes in interval II and III. This shows a deficiency of 3-strand double recombinants and an excess of 4-strand over 2-strand double recombinants, but this pattern is not observed in the other two datasets, where the calculated p-values are not significant.

Aspergillus nidulans: STRICKLAND (1958) analyzes chromatid interference using three different crosses in *Aspergillus nidulans*. He concludes that there is no disagreement between the data from his experiment and the expected 1:2:1 ratio for 2-, 3- and 4-strand double recombinants, although he observes an excess of 2-strand double recombinants among the adjacent intervals in the second cross. With some caution, he then uses WHITEHOUSE's formula to count the undetected double crossovers within intervals and shows some evidence of an excess of 2-strand double recombinants over what would be expected under NCI. Using our method, the estimated p-values for the data from his experiments are 0.33, 0.10 and 0.56 with sample sizes 392, 264 and 575, respectively. So if there is any chromatid interference in *Aspergillus nidulans*, it does not seem to be strongly exhibited in these experiments.

Drosophila: In addition to tetrad data, we have examined two large well-known *Drosophila* (single spore) datasets collected by MORGAN, BRIDGES and SCHULTZ (1935) and WEINSTEIN (1935). We have found that the linear inequality constraints which characterize NCI for single spore data are so rarely violated in these datasets that we did not see any need to perform a formal test; the data did not demonstrate any incompatibility with NCI. Since we put no restriction on the crossover process along the bundle of chromatids, evidence against NCI from single spore data is essentially based on testing if the recombination probability is larger than the non-recombination probability within intervals and if the chance of recombination increases as the interval is enlarged. Thus, it is rather difficult to gather evidence against NCI from single spore data based on our characterization, though WRIGHT (1947) observed that in the house mouse the factors wv_2 and sh_2 showed recombination probability with *sex* of 0.56 and 0.57, respectively.

DISCUSSION

If chromatid interference exists, then current genetic mapping methods, which assume no chromatid interference, may not lead to correct genetic maps. Thus, it is important to make the best possible use of the data on chromatid interference in order to assess its true extent. Not that chromatid interference is not always observable; all that we can detect is chromatid interference which leads to violation of the NCI constraints on multilocus recombination or tetrad probabilities.

A number of datasets from the literature were reanalyzed by our bootstrap testing procedure, and some did show significant deviation from NCI. An excess of 2-strand double recombinants over 4-strand double recombinants, i.e. *negative* chromatid interference, seems to be a common feature of these experiments. There are some other patterns which might account for the deviation from NCI, such as a deficiency of 3-strand double recombinants, but overall, there is no consistent pattern among all these experiments. Our analysis suggests that when there is chromatid interference, it varies from experiment to experiment and from organism to organism. This observed effect may be due to something other than varying chromatid interference, such as differential viability of zygotes.

Constraints for tetrad data can be separated into two sets, one consisting of linear inequality constraints, and the other of equality constraints. Some of the equality constraints have been used extensively in the literature to examine chromatid interference in different organisms. In the inequality constraints on tetrad probabilities $\mathbf{p} = (p_{i_1 i_2 \ldots i_n})$, only those $p_{i_1 i_2 \ldots i_n}$ with $i_r = 2$ for at least one r, i.e. with at least one interval showing non-parental ditype, actually decrease the left-hand side of the inequality. Thus, these particular p_i are important for detecting deviation from NCI. If the markers are close to each other, the observed frequency of any recombination pattern $i_1 i_2 \ldots i_n$ with $i_r = 2$ for at least one r will be rather small. This is the case for all the tetrad datasets analyzed above. As a result, these linear inequalities are not often violated because of the limits of the given sample sizes. Thus, the equality constraints actually play the central role in our analysis. The inequality constraints will be important for the analysis if we analyze a set of well separated markers, or some organism such as *Schizosaccharomyces pombe* which has a large number of crossovers during meiosis. If the testing procedure results in a small p-value, we can search for the pattern of chromatid interference, but a large p-value does not necessarily imply that there is no chromatid interference. When there are many markers in the experiment and the sample size is not large, it is hard to observe a significant deviation from the expectation under NCI. This is because the number of possible patterns increases exponentially with the number of markers, and hence the expected count for each pattern is very small, so it becomes difficult to distinguish chromatid interference from no chromatid interference. Grouping the data with respect to some intervals may be necessary to give the test a reasonable amount of power.

ACKNOWLEDGEMENTS

This work was supported by National Science Foundation grant DMS-9113527.

The authors would like to thank Uri Liberman for his helpful comments.

APPENDIX

Proposition 1: For tetrad data involving $n+1$ markers, there are $2^n + (3^n - 1)2^{n-2}$ distinguishable tetrad patterns.

Proof: There are 3 possible tetrad types between each pair of consecutive markers. Among the overall patterns for which the type within every interval is either parental or non-parental ditype, there are 2^n different types. For recombination pattern $i_1 i_2 \ldots i_n$ with k intervals having tetratype, there are 4^{k-1} different patterns among the tetratype intervals; see discussion in text. There are $\begin{pmatrix} n \\ i \end{pmatrix}$ ways to choose which k intervals will have the tetratypes, and there are 2^{n-k} ways to assign parental and non-parental ditypes to the other n-k intervals. The number of subcells with at least one tetratype is thus

$$\sum_{k=1}^{n} \begin{pmatrix} n \\ k \end{pmatrix} 2^{n-k} 4^{k-1} = 2^{n-2}(3^n - 1).$$

We conclude that the total number of distinguishable types is

$$2^n + (3^n - 1)2^{n-2}.$$

Proposition 2: Under the assumption of NCI, the relation between the joint recombination probabilities $\mathbf{p}$=($p_{i_1 i_2 \ldots i_n}$) and the underlying crossover distribution $\mathbf{q}$=($q_{i_1 i_2 \ldots i_n}$) is $\mathbf{p} = \mathbf{R_n q}$.

Proof: Given $\mathbf{q}$=($q_{j_1 j_2 \ldots j_n}$), let $r_0^0 = 1$, $r_0^1 = 0$, $r_1^0 = 1/2$ and $r_1^1 = 1/2$. Then

$$p_{i_1 i_2 \ldots i_n} = \sum_{j_1} \sum_{j_2} \ldots \sum_{j_n} r_{j_1}^{i_1} r_{j_2}^{i_2} \ldots r_{j_n}^{i_n} q_{j_1 j_2 \ldots j_n}$$

and so

$$\mathbf{p} = (\frac{1}{2} \begin{pmatrix} 2 & 1 \\ 0 & 1 \end{pmatrix}) \otimes \ldots \otimes (\frac{1}{2} \begin{pmatrix} 2 & 1 \\ 0 & 1 \end{pmatrix}) = \mathbf{R_n q}.$$

The following proposition uses notation introduced in the second section of the paper.

Proposition 3: Assume NCI. Then for every combination $\mathbf{i} = (i_1, \ldots, i_m)$ where $i_j = 0$, 1 or 2, $j = 1, \ldots, m$, with exactly k j's such that $i_j = 1$, the 4^{k-1} probabilities $p_i(\mathbf{h})$, $\mathbf{h} = (h_1, \ldots, h_{k-1})$, with h_j=1, 2, 3 or 4, are all equal.

Proof: The proof requires the introduction of a fair amount of additional notation and for this reason its details will be omitted. It uses a counting argument involving an inductive step, similar to that sketched prior to the definition of $p_i(\mathbf{h})$. The NCI assumption enters at the end in the form of an equiprobability assumption concerning the strands involved in crossovers.

Proposition 4: In the case of tetrad data, for any underlying crossover process, there is a crossover process with at most two events between each consecutive pair of markers, inducing the same tetrad probabilities.

Proof: For $k \geq 1$, let $c_k^1 = \frac{1}{3} - \frac{4}{3}(-\frac{1}{2})^k$, $c_k^2 = \frac{4}{3}(\frac{1}{2} + (-\frac{1}{2})^k)$, $c_0^1 = c_0^2 = 0$, and let $c_0^0 = 1$,

$c_k^0 = 0$ for $k \geq 1$. Given any $\mathbf{q} = (q_{i_1 i_2 \ldots i_n})$, the desired $\mathbf{q}^* = (q^*_{i_1 i_2 \ldots i_n})$ can be constructed as follows:

$$q^*_{i_1 i_2 \ldots i_n} = \sum_{j_1 \geq 0} \sum_{j_2 \geq 0} \cdots \sum_{j_n \geq 0} c_{j_1}^{i_1} c_{j_2}^{i_2} \cdots c_{j_n}^{i_n} q_{j_1 j_2 \ldots j_n}$$

for $i_r = 0$, 1 or 2, and $q^*_{i_1 i_2 \ldots i_n} = 0$ if for some r, $i_r \geq 3$. It is easy to see that $\mathbf{q}^*$ and $\mathbf{q}$ give rise to the same tetrad probabilities.

Proposition 5: In the tetrad case, under the assumption of NCI, the relation between the crossover distribution $\mathbf{q} = (q_{i_1 i_2 \ldots i_n})$ and the tetrad probabilities $\mathbf{p} = (p_{i_1 i_2 \ldots i_n})$ is $\mathbf{p} = \mathbf{T_n q}$, and for those $i_1 i_2 \ldots i_n$ with $i_r = 1$ for more than one r in {1,2, …,n}, i.e. more than one tetratype, all the subcells defined by the strands involved have equal probability.

Proof: The relation $\mathbf{p} = \mathbf{T_n q}$ can be proved using an argument similar to that used in the proof of Proposition 2, and equiprobability for the subcells of the relevant cells follows from Proposition 3.

Proposition 6: For a given set of $\mathbf{p} = (p_{i_1 i_2 \ldots i_n})$ of joint tetrad probabilities, there is an underlying crossover process satisfying NCI compatible with $\mathbf{p}$, if and only if $\mathbf{T_n^{-1} p} \geq \mathbf{0}$, and all the subcell probabilities $p_{i_1 i_2 \ldots i_n}(j_1, j_2, \ldots, j_{k-1})$ in a cell $i_1 i_2 \ldots i_n$ with $i_r = 1$ for more than one r, are equal.

Proof: Necessity follows from non-negativity of $\mathbf{q}$ and Proposition 5. If $\mathbf{T_n^{-1} p} \geq \mathbf{0}$, we put $\mathbf{q} = \mathbf{T_n^{-1} p}$, and it is straightforward to construct a crossover point process with distribution $\mathbf{q}$ which is compatible with $\mathbf{p}$.

LITERATURE CITED

Bailey, N. T. J., 1961 *Introduction to the Mathematical Theory of Genetic Linkage.* Oxford University Press, London.

Bellman, R. E., 1970 *Introduction to Matrix Analysis. 2nd ed.* McGraw-Hill, New York.

Cooke, W. P., 1983 Surrogate geometric programming estimates of restricted multinomial proportions. Communications in Statistics **12:** 291-305.

DEMPSTER, A. P., N. M. LAIRD, and D. B. RUBIN, 1977 Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, Series B **39:** 1-22.

DUFFIN, R. J., E. L. PETERSON, and C. ZENER, 1967 *Geometric Programming.* John Wiley and Sons, Inc., New York.

GEYER, C. J., 1991 Constrained maximum likelihood exemplified by isotonic convex logistic regression. Journal of the American Statistical Association **86:** 717-724.

HAWTHORNE, D. C., and R. K. MORTIMER, 1960 Chromosome mapping in Saccharomyces: centromere-linked genes. Genetics **45:** 1085-1110.

LINDEGREN, C. C., and G. LINDEGREN, 1942 Locally specific patterns of chromatid and chromosome interference in Neurospora. Genetics **27:** 1-24.

MATHER, K., 1935 Reduction and equational separation of the chromosomes in bivalents and multivalents. J. Genet. **30:** 53-78.

MAZUMDAR, M., and T. R. JEFFERSON, 1983 Maximum likelihood estimates for multinomial probabilities via geometric programming. Biometrika **70:** 257-261.

MORGAN, T. H., C. B. BRIDGES, and J. SCHULTZ, 1935 Report of investigations on the constitution of the germinal material in relation to heredity. Carnegie Institute of Washington **34:** 284-291.

PERKINS, D. D., 1962 Crossing-over and interference in a multiply marked chromosome arm of Neurospora. Genetics **47:** 1253-1274.

PERLMAN, M. D., 1969 One sided problems in multivariate analysis. Annals of Mathematical Statistics **40:** 549-567.

PHILLIPS, D. T., and C. S. BEIGHTLER, 1973 Geometric programming: a technical state-of-the-art survey. AIIE Transactions **5:** 97-112.

RAUBERTAS, R. F., C. C. LEE, and E. V. NORDHEIM, 1986 Hypothesis tests for normal means constrained by linear inequalities. Communications in Statistics. **15:** 2809-2833.

ROBERTSON, T., F. T. WRIGHT, and R. L. DYKSTRA, 1988 *Order Restricted Statistical*

Inference. Wiley, New York.

SHULT, E. E., and C. C. LINDEGREN, 1959 A survey of genetical methodology from Mendelism to tetrad analysis. Can. J. Genet. Cytol. **1:** 189-201.

SPEED, T. P., M. S. MCPEEK, and S. N. EVANS, 1992 Robustness of the no-interference model for ordering genetic markers. Proc. Natl. Acad. Sci. U.S.A. **89:** 3103-3106.

STRICKLAND, W. N., 1958 An analysis of interference in *Aspergillus nidulans.* Proc. Roy. Soc. Series B **149:** 82-101.

STRICKLAND, W. N., 1961 Tetrad analysis of short chromosome regions of *Neurospora crassa.* Genetics **46:** 1125-1141.

WEINSTEIN, A., 1935 The theory of multiple strand crossing over. Genetics **21:** 155-199.

WHITEHOUSE, H. L. K., 1973 *Towards an Understanding of the Mechanism of Heredity.* St. Martin's Press, New York.

WRIGHT, M. E., 1947 Two sex linkages in the house mouse. Heredity **1:** 349-354.

Table 1

<table>
<tr><td>(P,P)</td><td colspan="2">(P,T)</td><td>(P,N)</td></tr>
<tr><td rowspan="2">(T,P)</td><td>$(T,T)_1$</td><td>$(T,T)_2$</td><td rowspan="2">(T,N)</td></tr>
<tr><td>$(T,T)_3$</td><td>$(T,T)_4$</td></tr>
<tr><td>(N,P)</td><td colspan="2">(N,T)</td><td>(N,N)</td></tr>
</table>

Legend: Table 1 shows the different possible observed tetrad types among three markers. P, N and T stand for parental ditype, non-parental ditype and tetratype, respectively.