Error Detection in Genetic Linkage Data for Human Pedigrees Using Likelihood Ratio Methods

Margaret Gelder Ehm [1]
Rice University

Marek Kimmel [2]
Rice University

Robert W. Cottingham [3]
Baylor College of Medicine

Journal of Biological Systems Volume 3, Number 1

[1]Research supported by the U. S. National Library of Medicine Medical Informatics Research Training Program and the W.M. Keck Foundation.

[2]Research supported by NSF Grant 9203436.

[3]Research supported by a grant from the National Center for Human Genome Research.

1 Introduction

In Gregor Mendel's early genetic experiments, he hypothesized that each allele pair within a gene segregates independently during gamete (egg or sperm) formation. This assumption proved correct for some experiments he performed, but has been shown to be incorrect if the genes are located close together on the same chromosome. Such genes are said to be linked. The analysis of the mode of inheritance of these genes is called genetic linkage analysis. *Genetic maps* describing the positions of genes and their intervening distances can be constructed using linked genes. While the process of locating genes on chromosomes using linkage analysis is well defined, the methods used are affected by errors in the data.

Gathering linkage data requires typing families for many *markers*, or portions of DNA that distinguish individuals from one another. The typing procedures are complicated and the volume of markers is large enough that errors occur. As genetic maps become more dense, typing errors can obscure the information making gene locations harder to define. Therefore, only in recent years have these errors began to seriously affect the estimation of gene location. We present a method for detecting errors in pedigree data and discuss its properties. The method is designed to identify individuals and loci likely to contain errors. While the method requires further research to fully determine its properties, it shows promise for detecting errors in human genetic linkage data.

2 Genetic Linkage Analysis

In order to describe genetic linkage analysis, we must introduce common genetic terms. We will (1) define the basic terms used to describe genes, (2) characterize the phenomena of interest, (3) delineate the biological mechanisms at work, and (4) define the mathematical model used to characterize these mechanisms.

2.1 Background

The diploid human genome contains 23 pairs of chromosomes. The elements of each pair are similar to each other (*homologous*) but not identical. They are called *homologs* or *homologous chromosomes* and one originates from the father and one from the mother. Each of the two homologs contains the same linear sequence of genes but each with one of the possible forms (*alleles*) of a gene. Each of the homologous chromosomes contains two identical *chromatids* or replicas produced by chromosome replication [?]. The position of a gene on a chromosome is known as the *gene locus.* A *genotype* is the pair of alleles at a locus [18]. *Homozygotes* are individuals with identical alleles at a locus and *heterozygotes* are those having differing alleles at a locus. A *haplotype* is the set of alleles on one homolog and represents the genetic information received from one parent.

The expression of a genotype is called the *phenotype.* The relationship between the genotype and phenotype is expressed by the *penetrance function* which is defined as the conditional probability of observing a particular phenotype given a specific genotype. To study genotypes and phenotypes, *pedigrees*, family trees showing the inheritance patterns of genotypes and phenotypes [?], are typed and studied for patterns at loci of interest.

The genotype is fully specified when the alleles are known and can be associated with specific homologs. This association is called *phase.* For example, an individual with the genotype $\begin{bmatrix} A\,a \\ b\,B \end{bmatrix}$ may have alleles $\begin{bmatrix} A \\ B \end{bmatrix}$ or $\begin{bmatrix} A \\ b \end{bmatrix}$ on one of the homologs depending on the phase. (This notation represents two loci each with two possible alleles, A or a and B or b.) If the phase is known, the genotype is written $\begin{bmatrix} A | a \\ B | b \end{bmatrix}$ indicating one parent passed the haplotype $\begin{bmatrix} A \\ B \end{bmatrix}$ and the other parent passed $\begin{bmatrix} a \\ b \end{bmatrix}$. Figure 1 shows two third-generation offspring with the two possible phases [18]. The circles and squares represent females and males respectively.

During meiosis homologous pairs of chromosomes align and undergo *recombination,* the exchange of homologous regions of DNA. The exchanges are the result of *crossovers* and consequently the genotype of the child reflects a phase change and exhibits *nonparental haplotypes.* Each child in a pedigree whose genotypes are composed entirely of their parents' haplotypes, is said to be *nonrecombinant.* Figure 1 shows one nonrecombinant male and one recombinant female offspring in the third generation. Figure 2 illustrates two chromatids of two homologous chromosomes crossing over. At the conclusion of meiosis, each chromatid (ultimately the haplotype) is then passed to either an egg or a sperm cell to be passed along to offspring. In figure 2 the result is nonrecombinant haplotypes for loci A and B and recombinant haplotypes 2 and 3 and nonrecombinant haplotypes 1 and 4 for loci B and C. By analyzing phenotypes, their corresponding genotypes, and recombination events, we can use genetic linkage analysis to study the occurrence of crossovers between genes or markers on chromosomes.

While crossovers between homologous chromosomes are common events, their frequency is not uniform throughout the chromosome. The number of observed crossovers may be affected by many factors, most important of which is the distance between the markers. Crossovers rarely occur between genes close to one another on the chromosome and consequently the two loci rarely exhibit a recombination event. In this sense, the loci are *linked.* Crossovers occur randomly between genes further apart on a chromosome. These loci exhibit what is called *free recombination* and recombine with probability approaching one half.

Linkage is the phenomenon whereby the number of recombinant haplotypes passed to offspring is less than would be expected under free recombination. The degree of linkage is measured by the recombination fraction, θ, the proportion of recombinant haplotypes among all haplotypes produced by a parent. This fraction ranges from 0, loci completely linked, to 0.5, no evidence of linkage. The rate of crossovers or recombinations can serve as a measure of the distance between two genes.

Often genotypic, phenotypic, or phase information may be unknown. Even when genotypic information is available, it may be impossible to determine if the individual is a recombinant for a pair of loci. For example, consider a parent with the genotype $\begin{bmatrix} B\,b \\ a\,a \end{bmatrix}$. If the child exhibits the haplotype $\begin{bmatrix} B \\ a \end{bmatrix}$, it is impossible to determine if a recombination event has occurred. Recombination events will be apparent if a parent is doubly heterozygous $\begin{bmatrix} A\,a \\ B\,b \end{bmatrix}$. Then will he/she be *informative for linkage* at the pair of loci.

Many laboratories are working on constructing genetic maps for experimental animals and humans. One such concerted effort is at Centre d'Etude du Polymorphisme Humain (CEPH) in France. Most of their work is based on a set of 65 reference families, which we will refer to as CEPH families [18]. We use data obtained from CEPH families in our analyses.

2.2 Model

2.2.1 Two-Point Analysis

The goal in two-point linkage analysis is to determine if two markers are linked. This is accomplished by estimating θ. We will introduce the model often used in genetic linkage analysis which requires several preliminary definitions. Let $\mathbf{X} = x_1, \ldots, x_m$ be the vector of phenotypes for individuals in a pedigree of size m. Let $\mathbf{G} = g_1, \ldots, g_m$ be the corresponding vector of genotypes (unobservable in general). These are actually joint genotypes and phenotypes which represent alleles observed at two loci. The recombination frequency between the two loci is denoted by θ. The model includes several assumptions.

1. Each genotype has a known population frequency, $\mathcal{P}(g)$.
2. Recombinant haplotypes are passed with probability θ.
3. Given a specific genotype, g, phenotype x is exhibited with probability $\mathcal{P}(x \mid g)$.

We measure the aptness of the model by calculating the likelihood function of the data:

$$\mathcal{P}(\mathbf{X} = \{x_1, \ldots, x_m\}) = L(\mathbf{X}; \theta) = \sum_g \mathcal{P}(\mathbf{X}, g)$$

where the sum is taken over all possible sets of genotypes assigned to individuals. This model allows us to estimate θ using maximum likelihood estimation. The estimate is given by:

$$\hat{\theta} = \left\{ \theta : \max_\theta L(\mathbf{X}; \theta) \right\}.$$

While maximum likelihood estimates enjoy many desirable properties, calculation of the likelihood function is sometimes difficult. Elston and Stewart [?] propose a method that represents the pedigree likelihood as the sum:

$$L(\mathbf{X}; \theta) = \sum_{g_1} \mathcal{P}(x_1 \mid g_1)\mathcal{P}(g_1 \mid \cdot) \sum_{g_2} \mathcal{P}(x_2 \mid g_2)\mathcal{P}(g_2 \mid \cdot) \ldots \sum_{g_m} \mathcal{P}(x_m \mid g_m)\mathcal{P}(g_m \mid \cdot).$$

$\mathcal{P}(g_i \mid \cdot)$ is the probability of the i^{th} person's genotype given θ and the genotypes of his/her parents, if the latter are included in the pedigree. If not, then $\mathcal{P}(g_i \mid \cdot)$ is the genotype frequency of g_i. As noted above $\mathcal{P}(x_i \mid g_i)$, the penetrance function, is the conditional probability that person i expresses phenotype x_i given that he/she has genotype g_i.

Incomplete penetrance is an elegant way to represent the fact that the expression of genes is affected by their environment. Unfortunately, when genotypes are incompletely penetrant, calculating the pedigree likelihood becomes more complex and time consuming.

2.2.2 Multipoint Analysis

We can extend the notion of the pedigree likelihood from two loci to $k \geq 2$ loci by letting $\theta = (\theta_{1,2}, \ldots, \theta_{k-1,k})$ be a vector of recombination fractions. Analyzing the recombination fractions of several intervals simultaneously is called multipoint analysis. The maximum likelihood estimate for $\theta = (\theta_{1,2}, \ldots, \theta_{k-1,k})$, is:

$$\hat{\theta} = \left\{ \theta : \max_{\theta_{1,2},\ldots,\theta_{k-1,k}} L(\mathbf{X};\theta) \right\}.$$

Lathrop *et al* [?] indicate that multipoint analysis uses pedigree information more efficiently and is particularly useful in situations where individuals are uninformative for linkage as is the case with humans. They investigate the relative efficiency of a three-loci versus a two-loci analysis where the phenotype information is available for all individuals at three loci: A, B, and C. For loci A and C, they assume codominant alleles with equal genotype frequencies and a large number of alleles. Locus B is either a codominant system similar to A and C or a rare disease trait. The large-sample variances of $\hat{\theta}$ and the efficiency, or inverse of the ratio of the corresponding variances, are calculated. Although gains in efficiency will depend on the situation, they ranged from 1.12 for the estimates of θ in phase-unknown families to 10.0 in phase-known families with reduced penetrance.

3 Detection of Errors in Linkage Data

Although the model for genetic linkage has proven to be very useful, errors can obscure the information obtained. Errors are usually classified as *pedigree* errors (pertaining to the links of the pedigree) or *typing* errors (all other errors). Examples of errors discussed include the following:

1. faults in historical documents,
2. nonpaternity,
3. unidentified adoption,
4. misreading of electrophoresis gels or serological reactions,
5. misrecording of results [12].

Although both types of errors have been present in pedigree analysis since the beginning, only since the advent of high resolution maps have laboratory typing errors become a major obstacle. Often individuals containing typing errors that are inconsistent with the model of inheritance ($L(\mathbf{X};\theta) = 0$) can be detected. If the situation proves difficult, Ott [19] has developed a test that can assist in identifying the persons and loci causing the inconsistencies.

The errors that are the subjects of this study are typing errors consistent with the model. These errors are often undetectable by simple inspection. Errors may be observed as double recombinants or negate true recombinants. When the loci of interest are close together and recombinations between them rarely occur, spurious recombinations will greatly affect the estimate of the recombination fraction. Errors that appear as double recombinants over a short distance can be detected with a sufficiently sensitive two-point or multipoint method. In contrast, errors that negate true recombination events are especially difficult to detect with a two-point method, but may be detectable given the additional information available in a multipoint method [2]. Since we will analyze linked genes (θ small) errors that negate true recombination events will occur even less frequently than the recombination events. We propose using multipoint methods analyzing as many loci as is computationally feasible.

Little work has been done on the detection of errors in genetic linkage data. As genetic linkage maps become more dense, errors in pedigree data will become an increasingly important factor. Unfortunately detection of errors is a difficult problem. Although much research has been devoted to improving the likelihood evaluation algorithms [4, ?, 20], multipoint analysis for large pedigrees with many alleles requires time and memory. Whether laboratories have to type each individual several times or use a mathematical algorithm to detect errors, errors will be an important factor in the development of high-resolution maps.

4 Likelihood Ratio Method

In this section we propose a method developed to detect errors in human pedigree data. While the method was specifically designed to detect typing errors, it is general enough to highlight any statistically unlikely typing. Consequently, errors detected may be typing or pedigree errors. The method was motivated by the likelihood ratio test.

4.1 Description

When pedigree data is obtained by typing individuals, the observed genotype is equal to the true genotype unless a typing error has occurred. Following [8, 19], we will represent error in pedigree data as incompletely penetrant genotypes. Let ϵ^* be the true error rate, ϵ_{ij} the assumed error rate for individual i at locus j, a the number of alleles, and h the number of possible combinations of two alleles. Note $h = \frac{a(a+1)}{2}$. The error function will have a uniform distribution and is given below:

$$\mathcal{P}(x_i|g_i) = \begin{cases} 1 - \epsilon_{ij} + \frac{\epsilon_{ij}}{h} & \text{if } x_i = g_i \\ \frac{\epsilon_{ij}}{h} & \text{if } x_i \neq g_i. \end{cases}$$

The pedigree likelihood is calculated using the error function defined above. Introducing the penetrance function in this way has the effect of adding an additional parameter, ϵ_{ij}, to the model.

Consider conducting a series of tests for one individual at one locus,

$$H_0 : \epsilon_{ij} = 0$$

versus

$$H_1 : \epsilon_{ij} > 0.$$

The likelihood function including error function of individual i at locus j is denoted $L_{ij}(\mathbf{X}; \theta, \epsilon_{ij})$. We will use an approximation of the likelihood ratio test statistic given by:

$$\tilde{\lambda}_{ij}(\mathbf{X}) = -2 \ln \left\{ \frac{L_{ij}(\mathbf{X}; \hat{\theta}_0, 0)}{L_{ij}(\mathbf{X}; \hat{\theta}_0, \epsilon_0)} \right\}$$

where, generally $\hat{\theta}_{\epsilon_{ij}}$ is defined in the following way:

$$\hat{\theta}_{\epsilon_{ij}} = \left\{ \theta : \sup_{\theta} L(\mathbf{X}; \theta, \epsilon_{ij}) = L(\mathbf{X}; \hat{\theta}_{\epsilon_{ij}}, \epsilon_{ij}) \right\}$$

and $\hat{\theta}_0$ corresponds to $\hat{\theta}_{\epsilon_{ij}}$ when $\epsilon_{ij} = 0$.

While using the above method increases the number of likelihood evaluations, each evaluation takes a little time. Note that the denominator involves evaluating the likelihood when individual i and locus j show incomplete penetrance, but all other individuals and loci have complete penetrances. This modification has a big effect on the pedigree likelihood calculation. Even though $\tilde{\lambda}_{ij}(\mathbf{X})$ must be calculated for all individuals and loci, its calculation involves completely penetrant genotypes for all individuals and loci except the individual at the locus of interest. Therefore the calculation time is closer to what would be expected for completely penetrant genotypes.

In order to use the above index, we must assess its properties. Under regularity conditions, the exact likelihood ratio

$$\lambda_{ij}(\mathbf{X}) = -2 \ln \left\{ \frac{\sup_{\theta} L_{ij}(\mathbf{X}; \theta, \epsilon_{ij} = 0)}{\sup_{\theta, \epsilon_{ij}} L_{ij}(\mathbf{X}; \theta, \epsilon_{ij})} \right\} = -2 \ln \left\{ \frac{L_{ij}(\mathbf{X}; \hat{\theta}_0, \epsilon_{ij} = 0)}{L_{ij}(\mathbf{X}; \hat{\theta}_{\hat{\epsilon}_{ij}}, \hat{\epsilon}_{ij})} \right\}$$

is asymptotically χ^2_1 under H_0. $\tilde{\lambda}_{ij}(\mathbf{X})$ involves approximating $L_{ij}(\mathbf{X}; \hat{\theta}_{\hat{\epsilon}_{ij}}, \hat{\epsilon}_{ij})$ with $L_{ij}(\mathbf{X}; \hat{\theta}_0, \epsilon_0)$. Figure 3 is a contour plot of the likelihood function $L_{ij}(\mathbf{X}; \theta, \epsilon)$ in the case where we are testing H_0 for one individual at one locus. If computation time was no obstacle, we would numerically find $\hat{\theta}_{\hat{\epsilon}_{ij}}$ in order to calculate the likelihood ratio statistic. The approximation we have proposed includes calculation of $\hat{\theta}_0$ and then uses $\hat{\theta}_0$ instead of $\hat{\theta}_{\hat{\epsilon}_{ij}}$. Therefore, how well the approximation works depends on the difference between the heights of the likelihood function at $\hat{\theta}_0$ and at $\hat{\theta}_{\hat{\epsilon}_{ij}}$.

Since the distribution of the test statistic is not known and appears to be difficult to calculate, we will investigate alternative ways to interpret the significance of the values of $\tilde{\lambda}_{ij}(\mathbf{X})$.

4.2 Computational Procedure

The error detection procedure consists of the following steps.

1. Select $3-4$ loci to test for typing errors.
2. Find the most likely order for the loci and verify that θ is small.
3. For each i and each j, calculate $\tilde{\lambda}_{ij}(\mathbf{X})$.
4. Identify individuals and loci with high values of $\tilde{\lambda}_{ij}(\mathbf{X})$ as typings likely to contain errors.

The likelihood evaluations used in this method were calculated using a slight modification of FASTLINK 2.1 [4, 20]. When selecting loci to be tested for error, among other considerations, we should be aware of the pedigree likelihood calculation time which is directly related to the number of loci and number of alleles at each locus. Choosing loci that are close together ($\hat{\theta}$ small) ensures that the probability of a true crossover is small making spurious crossovers easier to detect.

If our test statistic were indeed the likelihood ratio test statistic with regularity conditions satisfied then we would compare its value to the upper tail of a chi-square distribution with one degree of freedom. Since this is not the case, we use Monte Carlo simulation to determine the empirical cumulative distribution function for the test statistic under H_0 for a specific pedigree structure. If $\tilde{\lambda}_{ij}(\mathbf{X})$ falls into the tail of the simulated distribution, we identify individual i at locus j as likely to contain an error. In this way, testing $H_0 : \epsilon_{ij} = 0$ by calculating $\tilde{\lambda}_{ij}(\mathbf{X})$ for each individual and locus allows us to pinpoint the individuals and loci that may contain the errors.

4.3 Applications

We examined the results of our method on pedigrees of several sizes. Figure 4 shows three loci from a three-generation CEPH pedigree with 13 individuals. In order to test our method, we introduce an error in the pedigree and study how the method identifies it. The error was introduced at individual nine, locus two. We chose three loci, two of which had four alleles and one of which had two alleles to test for error. Using FASTLINK 2.1, we determined the most likely order for the loci and found $\hat{\theta} = (0.142, 0.001)$ assuming that order. We set $\epsilon_0 = 0.01$.

Table 1 gives the values of $\tilde{\lambda}_{ij}(\mathbf{X})$ for each of the 13 individuals and three loci. In order to test the hypothesis:

$$H_0 : \epsilon_{ij} = 0$$

versus

$$H_1 : \epsilon_{ij} > 0$$

we used Monte Carlo simulation to determine an empirical cumulative distribution function for the test statistic under H_0. The pedigree simulation procedure SLINK [17, 22] was used to generate pedigrees with the same familial structure, allele frequencies, and recombination fractions as the pedigree of interest. The values for $\tilde{\lambda}_{ij}(\mathbf{X})$ are compared to the empirical distribution function under the null hypothesis. Values of $\tilde{\lambda}_{ij}(\mathbf{X})$ that fall in the upper tail of this empirical distribution function correspond to values of the test statistic that are unlikely under the null hypothesis for individual, i, at locus, j.

To illustrate what the distribution functions look like, figure 5 gives the empirical distribution function of $\tilde{\lambda}_{4,1}(\mathbf{X})$ (individual 4 and locus 1). The observed value of $\tilde{\lambda}_{4,1}(\mathbf{X})$ (-0.011) is marked on the graph. Individuals and loci whose observed approximate likelihood ratio statistic is larger than a high enough percentile are labeled as likely to contain typing errors. Note that individual 4, locus 1 does not belong to that category. Table 2 gives the p-values of the approximate likelihood ratio statistic for each individual at each locus based on the empirical distribution function. Individual 9 is likely to contain an error at locus 2. Individuals 9 and 12 are likely to contain errors at locus 3. We conclude the method picked up the error we introduced in individual 9 at locus 2.

5 Discussion

While we have researched several aspects of this index, the approximate likelihood ratio test statistic, there are still many things to consider. Perhaps the foremost consideration is computation time. It is often the limiting factor in applications of genetic linkage analysis. Fortunately, substantial gains have been made with the implementation of FASTLINK 2.1 [4, 20] program, but the calculation of the true likelihood test statistic is still prohibitively expensive. Another aspect of our statistic to investigate is its behavior for terminal loci, founding individuals, and in the presence of untyped individuals. Its power will vary for these situations where the amount of information is reduced. Finally, we must address how our index compares with other indices.

The index we consider was designed in order to be computationally efficient. However, it will still be unusable in situations were linkage analysis is computationally infeasible. Since linkage analysis involves maximization of the pedigree likelihood, time efficient calculation of the likelihood function is a necessity. After the maximum likelihood estimates of θ have been determined, our method involves mn more likelihood evaluations. m is the number of individuals in the pedigree and n is the number of loci. Each of these likelihood evaluations is slightly more complicated than the evaluations used to estimate θ, in the sense that for each evaluation, one individual has incompletely penetrant phenotypes at one locus. While this complicates the pedigree likelihood, the computation time is not increased substantially. Therefore, the index should be computable in all cases where the maximum likelihood estimates for θ can be determined.

References

[1] K.H. Buetow. Influence of aberrant observations on high-resolution linkage analysis outcomes. *American Journal of Human Genetics*, 49:985–994, 1991.

[2] R.W. Cottingham, R.M. Idury, and A.A. Schaffer. Faster sequential genetic linkage computations. *American Journal of Human Genetics*, 53:252–263, 1993.

[3] R.C. Elston and J. Stewart. A general model for the genetic analysis of pedigree data. *Human Heredity*, 21:523–542, 1971.

[4] A.J.F. Griffiths, J.H. Miller, D.T. Suzuki, R.C. Lewontin, and W.M. Gelbart. *An Introduction to Genetic Analysis*. W.H. Freeman and Company, New York, fifth edition, 1993.

[5] B.J.B. Keats, S.L. Sherman, and J. Ott. Report of the committee on linkage and gene order. *Cytogenetics and Cell Genetics*, 55:387–394, 1990.

[6] G.M. Lathrop, A.B. Hooper, J.W. Huntsman, and R.H. Ward. Evaluating pedigree data. I. The estimation of pedigree error in the presence of marker mistyping. *American Journal of Human Genetics*, 25:241–262, 1983.

[7] G.M. Lathrop, J.M. Lalouel, C. Julier, and J. Ott. Multilocus linkage analysis in humans. *American Journal of Human Genetics*, 37:482–498, 1985.

[8] J. Ott. Computer-simulation methods in human linkage analysis. *Proceedings of the National Academy of Sciences*, 86:4175–4178, 1989.

[9] J. Ott. *Analysis of human genetic linkage*. Johns Hopkins University, Baltimore, revised edition, 1991.

[10] J. Ott. Detecting marker inconsistencies in human gene mapping. *Human Heredity*, 43:25–30, 1993.

[11] A.A. Schaffer, S.K. Gupta, K. Shriram, and R.W. Cottingham. Avoiding recomputation in genetic linkage analysis. *Human Heredity*, 44:225–237, 1994.

[12] D.E. Weeks, J. Ott, and G.M. Lathrop. SLINK: A general simulation program for linkage analysis. *American Journal of Human Genetics*, 47:A204, 1990. Abstract.

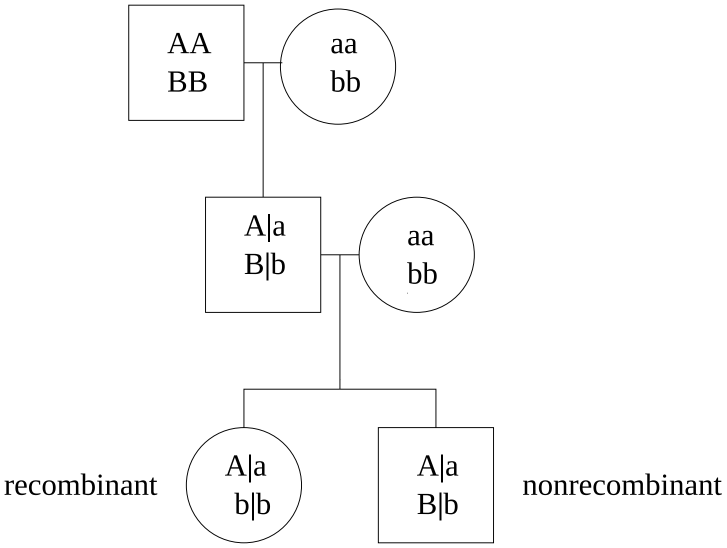

Figure 1: Example family with one recombinant and one nonrecombinant offspring.

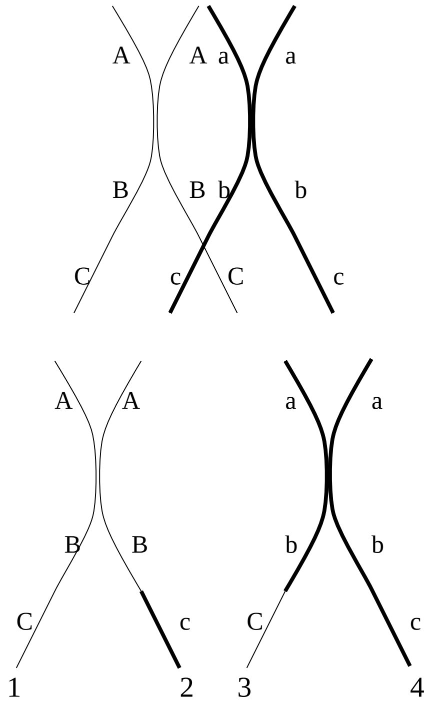

Figure 2: Example of 2 homologous chromosomes crossing over during meiosis.

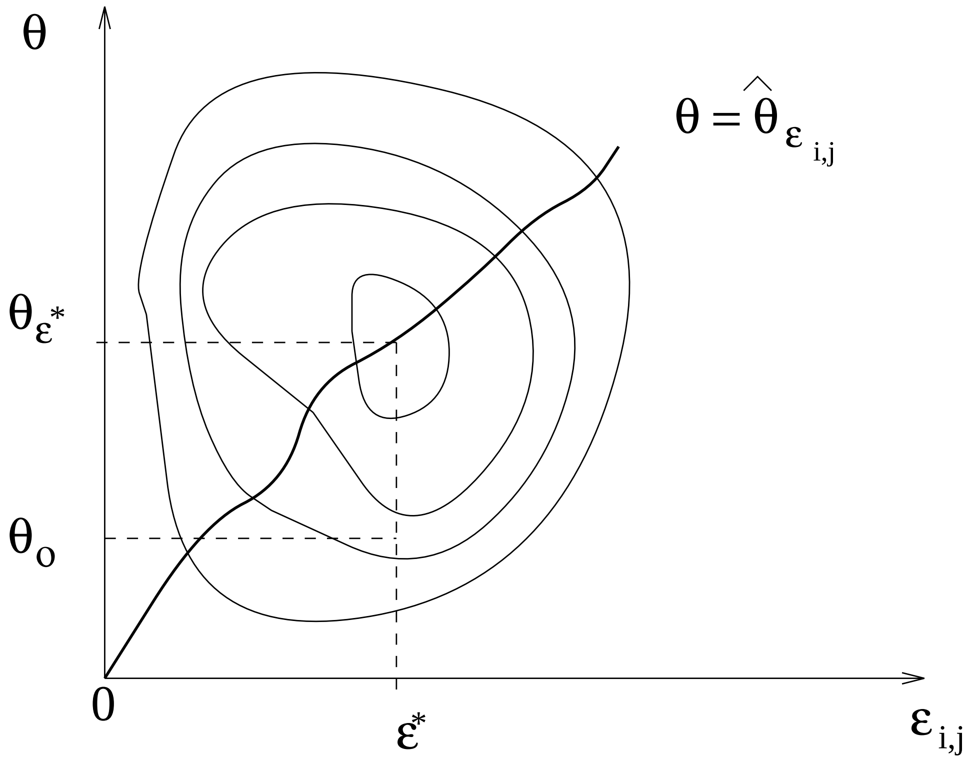

Figure 3: Contour Plot of $L_{ij}(\mathbf{X}; \theta, \epsilon)$.

Loci	Individuals (i)						
(j)	1	2	3	4	5	6	7
1	-.013	-.012	-.011	-.011	-.011	$\boxed{.003}$	-.011
2	-.018	-.018	-.018	-.018	-.018	-.018	-.018
3	-.018	-.018	-.018	-.018	-.018	-.018	-.018
	8	9	10	11	12	13	
1	-.011	$\boxed{.003}$	-.008	-.008	-.008	-.008	
2	-.018	$\boxed{.052}$	-.014	-.014	-.011	-.011	
3	-.018	$\boxed{.010}$	-.009	-.013	-.009	-.012	

Table 1: Values of $\tilde{\lambda}_{ij}(\mathbf{X})$ for CEPH pedigree 13292.

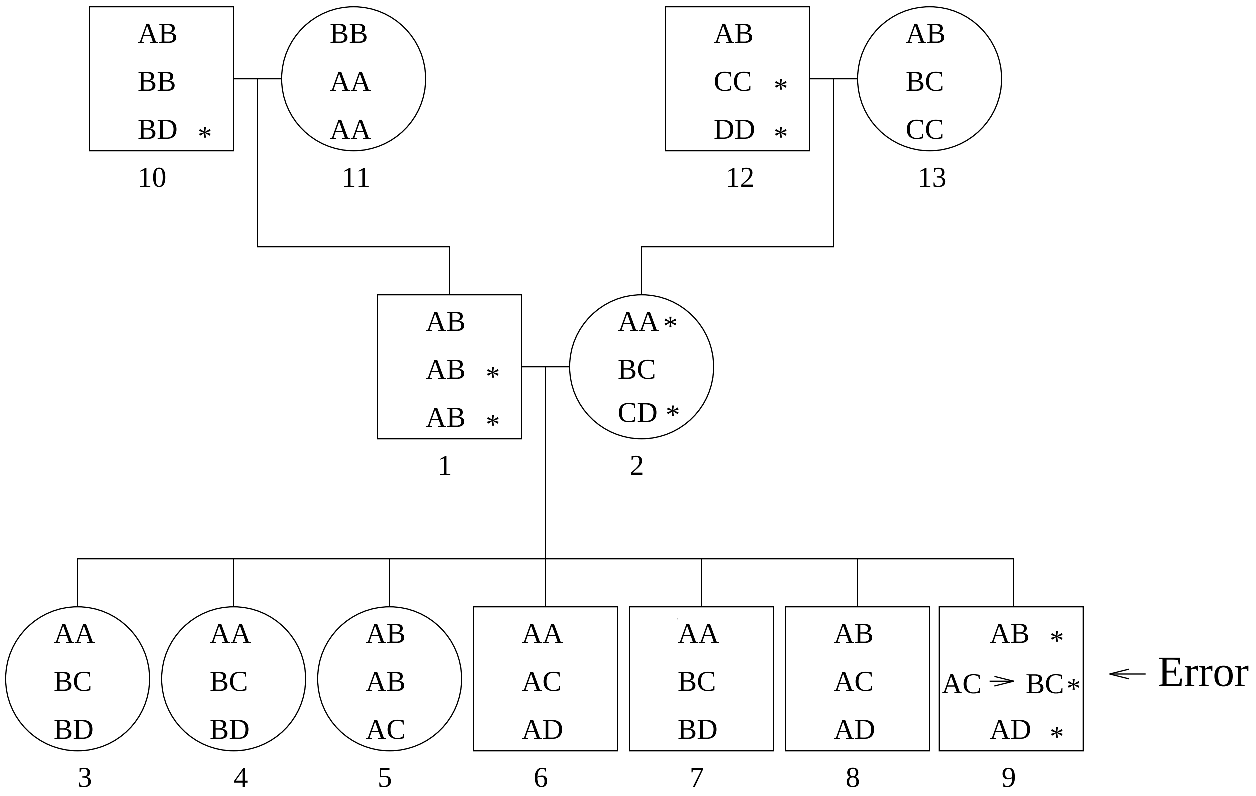

Figure 4: CEPH pedigree 13292 with typings at three loci.

Loci	Individuals (i)						
(j)	1	2	3	4	5	6	7
1	0.25	0.03	0.30	0.32	0.26	0.06	0.28
2	0.03	0.06	0.30	0.30	0.32	0.25	0.31
3	0.03	0.03	0.34	0.36	0.35	0.33	0.35
	8	9	10	11	12	13	
1	0.31	0.05	0.49	0.46	0.51	0.41	
2	0.32	$\boxed{0.00}$	0.38	0.75	0.22	0.22	
3	0.34	$\boxed{0.00}$	0.02	0.80	$\boxed{0.01}$	0.23	

Table 2: p-values for $\tilde{\lambda}_{ij}(\mathbf{X})$ for CEPH pedigree 13292.

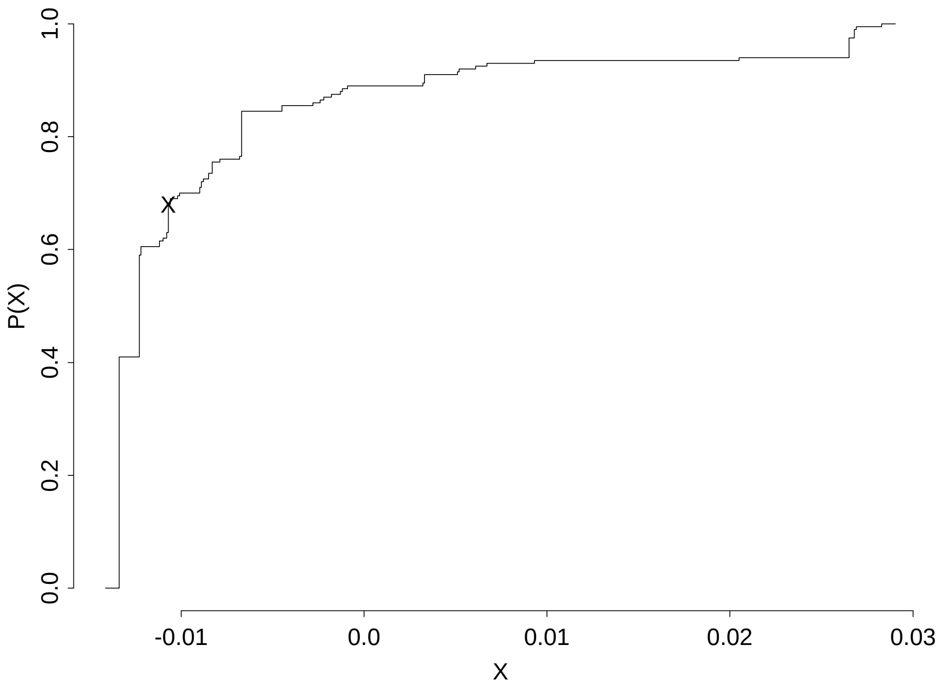

Figure 5: Empirical distribution function for $\tilde{\lambda}_{4,1}(\mathbf{X})$ under H_0 (CEPH pedigree 13292).