Lod Scores: What are they and how can I get them?

Lora's easy how-to for urban students on the go.

This topic is covered in books, in class, and on the web, but I find that people always need to hear it 2 or 3 different ways before really understanding what it is all about.

What is the Lod score?

Lod stands for "log of the odds ratio." In this case, the odds ratio is the likelihood that two markers are linked divided by the likelihood that they are not linked. Recall from genetics that the closer two markers are to each other, the lower the odds of a recombination (crossing over event) occurring between them in meiosis. In our lab, we calculate map distances by setting up crosses between flies carrying the genetic markers of interest and counting the recombinant animals yielded by the cross. In human genetics, matings are generally uncontrolled, and the number of offspring is much lower than the numbers seen in our fly lab, but it is still possible to use data from pedigrees to calculate map distances from recombination frequencies.

Things to remember

- A logarithm is the power to which 10 is raised to give a certain value. For example, the logarithm of 100 is 2, because 10 squared is 100. Logs do not have to be whole numbers; in Lod analysis, they usually are not. The idea, however, is still the same. The logarithm of any number greater than 1 is positive. The log of any number less than 1 is negative (10^-1 = 0.1, so the log of 0.1 is -1). The log of 1 is 0 (10^0 = 1). The log of 0 is negative infinity. (Note: You may remember from math class that logs can be calculated for any base number, but we are interested here only in "base-10" logarithms.) If you are unclear about the idea of logarithms, get a scientific calculator and try taking the log of a few numbers. Make a graph if you need to - pictures are always helpful!

- Recombination frequencies range from 0% to 50%. A recombination frequency of 50% between two markers means that a crossing-over event occurs in half of the meioses; the same "recombination frequency" would be observed from two markers on different chromosomes. Only when the RF is less than 50% can you be sure that two markers are on the same chromosome, much less close to each other!

- In statistics, the probability of several events all happening is found by multiplying together the probabilities of each one happening independently. Let's say we want to find the probability that I will get to work by 10:00. In order for that to happen, I have to get out of bed at the right time, I can't fall back to sleep in the shower, and my dog cannot meet too many other dogs in Riverside Park. The probability that I will get to work at 10am is equal to the probability of each of these events, multiplied together:

Probability(Work by 10) = Probability(Get up by 8:30) * Probability(No shower sleep) * Probability(Very few dogs in park)

Looking at this, it is a wonder I get to work at all!

- It may be obvious, but I will put this out there: If the probability of something happening is p, the probability of its NOT happening is 1-p.

Determining the Lod score

The chart you see in the breast cancer linkage paper is showing the Lod values for each family for various values of θ (theta, the recombination frequency). In a nutshell, you would choose different values of θ, calculate Z (the Lod score) for each one, and then see which value of θ gives the highest value for Z. This value of θ is your estimate of the map distance between the markers. Here is how it is done:

1. From the pedigree, determine the number of recombinant and nonrecombinant individuals. This is the problematic part of Lod score analysis. In this class you will be able to look at a pedigree chart and determine the number of recombinants; in "real life," you would have to genotype many individuals and construct your own pedigree chart. Inherent in this process are the problems of incomplete penetrance and mild forms of disease, which we discussed in class.
2. Choose a value for the recombination frequency. This can be any number between 0 and 0.50, and you will be doing several, so just pick a value to start.
3. Determine the likelihood that the two genes are linked. The likelihood is the overall product of the frequencies of recombination and the frequencies of nonrecombination. The probability of getting a recombinant individual (this is θ, the recombination frequency) is raised to a power equal to the number of recombinant individuals. If the number of recombinant individuals is *k*, then the probability of getting *k* recombinants is θ^k. Likewise, the probability of getting *n-k* nonrecombinants (*n* is the total number of people examined) is $(1-\theta)^{n-k}$. There is an example of this calculation later, but this is the general form.
4. Determine the likelihood that two genes are not linked. This is simple compared to step 3! Remember that the recombination frequency between two unlinked markers is always 0.50. The probability of getting n individuals with any genotype is just $(0.50)^n$. This is always true, regardless of the number of recombinant individuals (remember, in this step we are assuming that the two genes are NOT linked!)
5. Calculate the Lod score. The general formula for the Lod score is below:

$$Z(\theta) = \log \frac{(\theta^k)((1-\theta)^{n-k})}{(0.5)^n}$$

6. Repeat steps 2-5 with other values of θ. You can do this by hand for small pedigrees, but for larger ones, you will want to use a spreadsheet program like Excel. There are computer programs that calculate Lod scores for you directly from pedigree data as well.
7. Examine your chart of θ versus Lod score. The θ that gives you the highest Lod score is your estimation of the distance between the two markers, measured in cM.

Sample calculation of Lod score

Let's do the Lod score for Family A in Figure 11.4 of the Strachan and Read book (page 275; this was also handed out in the overhead packet in class).

1. In phase III, there are 6 individuals, 4 of which are affected by the disease. By looking at their grandmother (affected female in phase I) and their mother (affected female in phase II), we see that the allele associated with the disease is allele A1. Of the 4 affected individuals in phase III, only 1 of them does not have allele A1 (individual III6 is A2A3). Neither of the unaffected individuals in phase III has allele A1. We can therefore say that in phase III, there are 5 nonrecombinants and 1 recombinant individual. Note that you should count both affected and nonaffected individuals!

2. Let's start with a value of 0.10 (fairly close linkage).

3. For θ=0.10, *n*=6, *k*=1...the probability of getting *k* recombinants is θ^k; in this case, that would be (0.10)^(1). The probability of getting *n*-*k* nonrecombinants is $(1-\theta)^{n-k}$; in this case, *n*-*k*=6-1=5 and 1-θ=1-0.10=0.90...so the probability of getting 5 nonrecombinants is (0.90)^(5).

4. Easy! The likelihood that the genes are not linked is always $(0.50)^n$, so here it is (0.50)^(6).

5. The Lod score for this value of theta is:

```
Z(0.10) = log ( (0.10)^1) * ((0.90)^5) / ((0.50)^6) )
        = log ( 0.10 * 0.59049 / 0.015625 )
        = log ( 3.779136 )
        = 0.577392521
```

6. I calculated Z for this pedigree for several values of θ using an Excel spreadsheet. Results are shown below:

θ	Z
0.001	-1.195992585
0.10	0.577392521
0.15	0.629365862
0.20	0.622659905
0.25	0.579426300
0.30	0.508791429
0.35	0.414814802
0.40	0.298996217
0.45	0.161205935
0.50	0.000000000

7. The highest Lod score (Z) on this chart is 0.629365862 (in red above), which arose from a recombination frequency of 0.15. We could go back and get a more precise value for q by choosing values between 0.15 and 0.20, for example, but this is probably not necessary since you will start doing experiments in the lab to find your gene at this point.

Also note that the value is too low to indicate on its own whether two genes are linked (in general, you would like to have Z>3 to conclude linkage), but it is a start. You would want to sample more members of this family, or find more families with the same disease and test them to see if they carry the mutant allele. As more individuals are added, the Z score gets higher and becomes a better indicator of linkage and map distance. You can add Lod scores from different families as long as you are sure they are mutant at the same spot!

This website has information on recombination and meiosis if your genetics is a little shaky. It also calculates Lod scores. It may be a little mathy for some, but the genetics review is worthwhile and I found their MLE method section useful.

LEB