This week’s programs, autolink and sim_ibd_link both evaluate multilocus ibd probabilities. autolink, a generalization of kin, uses recursion to compute two-locus inbreeding coefficients— the probability that the two haplotypes of an individual are ibd at both of two linked loci. It does this for a specified number of specified individuals ("probands"), and at a set of (hard-wired) recombination fractions from 0 to 0.5. The program sim_ibd_link is as before, but now we use it to look at patterns of ibd over linked loci -- the example file specifies one unlinked "trait" locus, and 5 linked "markers", with recombination fractions 0.18, 0.18, 0.1, 0.1. Since we are interested in dependence among the loci we must score the IBD patterns over the loci. To avoid having too much output, we just score yes/no (1/0) for some ibd event over windows of loci of a specified size (3 in the example). The ibd events are ibd/not for the two haplotypes of an individual (and one can do this simultaneously for several individuals).

The example files for this week are found (in the usual location) at:

http://depts.washington.edu/statgen/Data/class550.shtml

I have saved the example file for autolink as lab4auto.dat and the example for sim_ibd_link as lab4sim.dat.

Practicing Using Autolink:

Look at the file lab4auto.dat. It consists of the jvped information, followed by the number 2 and two identification numbers. This tells the program that it will calculate two-locus inbreeding coefficients for the two different individuals whose identification numbers are listed. To run autolink, type:

% autolink lab4auto.dat

Take a look at the output. Notice that when \( r = 0 \) (i.e. for two loci that are right on top of each other), the probability that the individual’s two haplotypes are ibd at both the loci is the same as the individual’s inbreeding coefficient (you can check this using kin if you want). Also, the value when \( r = 0.5 \) (i.e. for two unlinked loci) is the square of the value when \( r = 0 \). Think about why these patterns should hold.
Practicing Using sim_ibd_link:

Look at lab4sim.dat. In many ways, it looks like the file we used for this program last time. There are a few differences, though. After the pedigree data is the number 10000 (the number of simulations), followed by the number 3. The 3 indicates that we will be looking at loci three at a time. Next comes information telling the program that there are 5 marker loci and the recombination frequencies between them are 0.18, 0.18, 0.10, and 0.10. The next line says that the trait locus is unlinked to the other loci \( r = 0.5 \) and it is on the left of the markers (position 0). Finally, the program is to look at the 2 individuals—the individuals numbered 531 and 431. To run the program, type:

```
%sim_ibd_link lab4sim.dat
```

One very simple but useful thing included in the output is a list of the loci and the recombination fractions between them. Remember that locus 0 is the trait locus. The line looks something like:

```
0 0.5 1  .18 2  .18 3  .10 4  .10 5
```

This is just a list of our loci and the recombination frequencies between them. This is exactly the information we gave the program in lab4sim.dat, but it’s nice to have it to refer to when looking at the results. The program gives data on windows of three loci. The first such window is loci 0,1, and 2.

The output consists of a list of possible ibd patterns for the three loci and four columns of relative frequencies. Remember that we are looking at 3 loci at a time. The ibd pattern 0 1 0, for example, is the event that the individual’s two alleles are not ibd at the first locus, are ibd at the second, and aren’t ibd at the third. The four columns each represent a way of taking three of the loci. The first column refers to the window including loci 0, 1, and 2. The second column refers to loci 1, 2, and 3.

When I look at output from an unfamiliar program, I like to do a couple quick checks to see that the output makes sense. Here are some of the things I thought about:

1. The second and fourth columns of relative frequencies each has the property that the markers considered are evenly spaced. This should mean that, for these columns, there should be symmetry between patterns 001 & 100, 011 & 110.
2. Is the output from this program consistent with the output of autolink? Look at the first column. Here, the recombination fraction between the first two loci is 0.5. For the simulations that sim_ibd_link generated, the relative frequency of 531 being ibd at both of the first two loci is \( \Pr(110) + \Pr(111) = 0.123 \) (for my simulations). According to autolink, the probability should be 0.011963. That looks good to me. You could do the same sort of analysis to check the results of your simulations against autolink’s results for intervals with recombination fraction 0.18 or 0.1.
3. One more thing: Note that, in the absence of interference, one interval with \( r = 0.18 \) is the same as two adjacent intervals with \( r = 0.1 \). In column 4, \( \Pr(101) + \Pr(111) = \Pr(\text{both ends of an interval with } r = 0.18 \text{ are } \text{ibd}) \). Also, in column 1, \( \Pr(011) + \Pr(111) = \Pr(\text{both ends of an interval with } r = 0.18 \text{ are } \text{ibd}) \). My simulations give pretty good agreement between these.

Your Assignment:

Create input files for your pedigrees similar to the example files you’ve been using. Use these files to run autolink and sim_ibd_link on your pedigrees. Since this week’s examples looked at ibd probabilities within individuals, you will want to choose your inbred individual(s) as the ones to include in your programs. You may space your markers however you like, but please look at windows of three loci (sim_ibd_link is not behaving well for other window sizes—this may or may not be fixed in the next few days). Turn in the output from these programs.