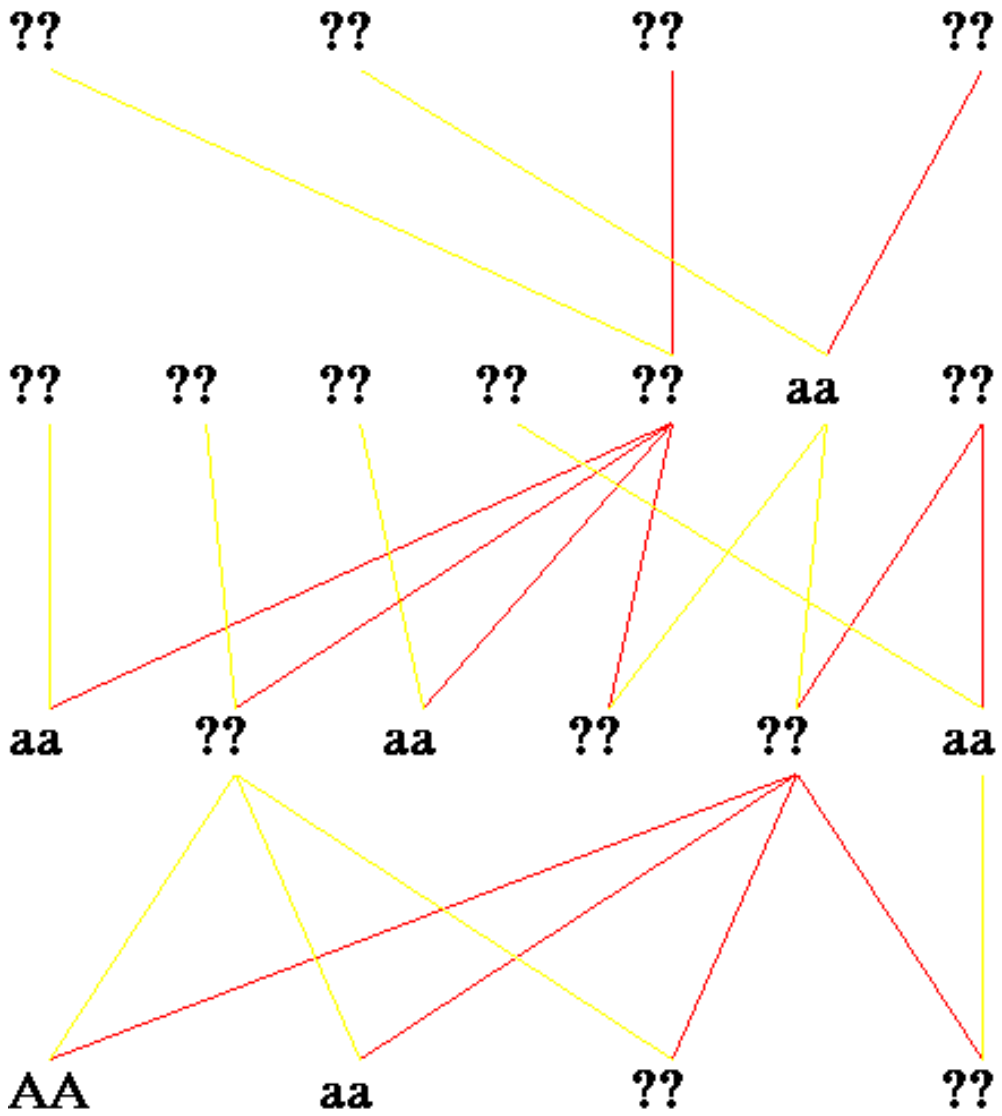
Exercise 1.2

Segregation analysis

The object of this exercise is to make a tool for large data sets available for you to use (free on your floppy disk, unsupported and not guaranteed), as well as to get some feel for segregation analysis, if you are not already expert.

You can construct your own pedigree to work on, or alternatively adopt the example given in this diagram:



For this pedigree, the file EXAMPLE.DAT is the input to the program GENEPROB.EXE which does the analysis. Results go to the screen and to the file GENEPROB.PED. This result file can be read by the Pedigree Viewer for easy browsing of the results. You will be shown how to do this.

EXAMPLE.DAT:

```
COMMENT: Example for segregation workshop
genefreq phenotypes(excl missing ones=9):
.5 3
phenotype then  g(f|u), u (aa Aa AA)  on columns ...
0   1  0  0
1   0  1  0
2   0  0  1
data format ...
 (3a5,i5)
data ...  (Note that id's don't have to be sequential as here)
    1    0    0    9
    2    0    0    9
    3    0    0    9
    4    0    0    9
    5    0    0    9
    6    0    0    9
    7    0    0    9
    8    0    0    9
    9    7    1    9
   10    8    2    0
   11    0    0    9
   12    9    3    0
   13    9    4    9
   14    9    5    0
   15    9   10    9
   16   11   10    9
   17   11    6    0
   18   16   13    2
   19   16   13    0
   20   16   13    9
   21   16   17    9
```

a. If necessary, copy files as instructed to your local hard or floppy disk.

b. Run GENEPROB from a DOS prompt and enter EXAMPLE.DAT as the input file. View the results in GENEPROB.PED using the Pedigree Viewer. Look at the declared genotypes (field 'Phen') and deduce the genotypes of as many of the ungenotyped animals as you can. View fields p(0) [probability of carrying zero A alleles], p(1) and p(2) to check your results.

c. Rerun after changing the prior estimate of gene frequency, and check that the changes in results are reasonable.

d. Rerun after making sensible changes to the penetrance values for the genotype/phenotype combinations (eg let there be some 'spillage' such that there is some finite probability that some genotypes are 'read' as belonging to the 'wrong' phenotypic class). You also can add extra phenotypes to represent, for example, dubious gel readings. Note that each of the three columns (aa Aa and AA genotypes) must add to unity no matter how many rows (phenotypic classes) they contain. Recall that each element in the array is the probability of observing phenotype *row* given knowledge that the genotype is (definitely) *col*.

Exercise 1-2 Segregation analysis and genotype probabilities

Segregation problem:

The spider syndrome in Suffolk sheep is a recessive lethal condition. Development at the ends of the long bones is impaired and lambs end up on the ground with legs played like a spider.

Spider.dat contains 167 sheep numbered sequentially, together with sire and dam number (unknown parents are denoted 0), tag number, and phenotype: '1' for normal and '2' for spider syndrome. There are no unknown phenotypes in this case as the trait is so easy to score.

Run a segregation analysis to calculate the probability of each genotype (++ +S SS) for each individual in the data set. (NB: you will need to construct a suitable header for your version of spider.dat.)

Use Pedigree Viewer to examine your results.

Why is the probability of being SS [p(2)] always either 0 or 1?

Sort left to right on probability of being Ss [p(1)] and 'Shade merit fields'. This will help to quickly identify likely carriers. Some animals are certain to be heterozygotes [p(1) = 1]. For each of these, deduce why this is so by inspecting the pedigree. Do this also for animals with values for p(1) which are high but less than unity. Can you always find good evidence to support this high probability of being a carrier?