

# Statistical methods in genetic relatedness and pedigree analysis

Oslo, January 2020

Magnus Dehli Vigeland and Thore Egeland

## Exercise set IV. Recombination and linkage

### Recap of theory

The *genetic map distance* between two loci on the same chromosome is defined as the expected number of crossovers between them per meiosis. The basic unit is 1 Morgan, corresponding to 1 expected crossover. A more common unit is 1 centiMorgan (cM) = 0.01 Morgan.

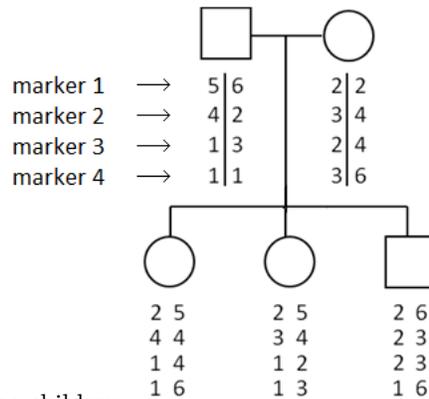
The *recombination rate*  $\theta$  between two markers is the long term fraction of gametes (meiotic products) whose alleles at the markers come from different parental strands. For instance, if the maternal strand has alleles  $A$  and  $B$  at the marker loci, and the paternal strand has the alleles  $a$  and  $b$ , then gametes with haplotypes  $Ab$  or  $aB$  are recombinants, while  $AB$  and  $ab$  are not.

For markers on the same chromosome, an odd number of crossovers between them result in an observable recombination. Several approximations exist for converting recombination rates to genetic distance, and *vice versa*. The simplest is Haldane's map function:

$$\theta = \frac{1}{2}(1 - e^{-2d}), \quad \text{and the inverse relation} \quad d = -\frac{1}{2} \ln(1 - 2\theta). \quad (1)$$

### Exercise IV-1

The pedigree below shows genotypes from 4 markers located close together on chromosome 1. The genotypes of the parents are phased, so that their haplotypes are known.



- Sort out the haplotypes of the children.
- Can you tell if there has been a recombination in this region in any of the meioses involved?

### Exercise IV-2

- Use Haldane's map function to calculate the recombination rate if the genetic distance is

- $d = 2$  Morgans,
- $d = 50$  cM,
- $d = 1$  cM.

- Use the inverse of Haldane's map function to calculate the genetic distance in cM if the recombination rate is

- $\theta = 0.49$ ,
- $\theta = 0.25$ ,
- $\theta = 0.05$ .

### Exercise IV-3

Three markers  $A, B, C$  are such that

$$\theta_{AB} = \theta_{BC} = 0.05.$$

Using Haldane's map function, compute the recombination rate  $\theta_{AC}$  between  $A$  and  $C$ .

### Exercise IV-4

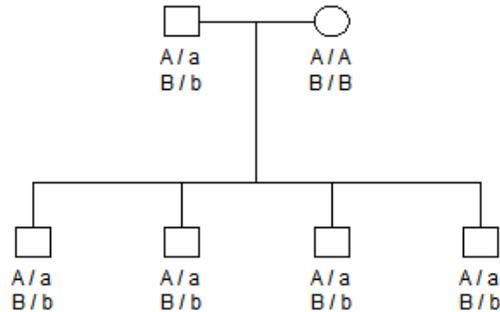
Experiments show that crossovers in practice never occur very close to each other. (Hence Haldane's assumption of no interference is not completely true.) Show that this implies that for very close markers, we have the simple relation

$$\theta = d.$$

(This is known as *Morgan's map function*).

### Exercise IV-5

The family below is genotyped with two autosomal SNPs, one with alleles  $A$  and  $a$ , and the other with alleles  $B$  and  $b$ . The minor allele frequencies are  $P(a) = 0.3$  and  $P(b) = 0.2$ .



a) Create the pedigree and marker objects in `pedtools`, for example with the following code:

```
x = nuclearPed(4)
m1 = marker(x, afreq = c(A = 0.7, a = 0.3))
m2 = marker(x, afreq = c(B = 0.8, b = 0.2))
genotype(m1, c(1, 3, 4, 5, 6)) = c('A', 'a')
genotype(m1, 2) = c('A', 'A')
genotype(m2, c(1, 3, 4, 5, 6)) = c('B', 'b')
genotype(m2, 2) = c('B', 'B')
```

```
plot(x, list(m1, m2))
```

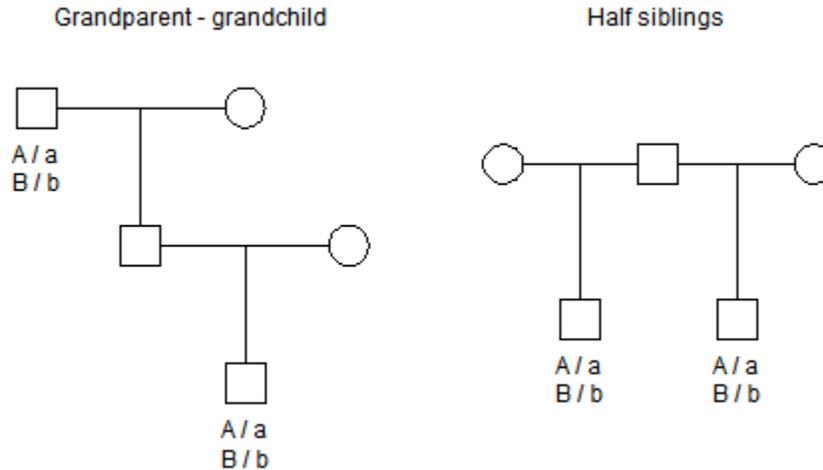
b) Use `pedprobr::likelihood()` to compute the pedigree likelihood, when the markers are:

- unlinked, i.e.  $\theta = 0.5$
- linked with recombination rate  $\theta = 0.25$
- completely linked, i.e.  $\theta = 0$

c) Try to explain why the likelihood is higher if the markers are linked, than if they are unlinked.

### Exercise IV-6

The relationship between grandparent-grandchild (G) is indistinguishable from that of half siblings (HS), when only using *unlinked* markers. In this exercise we will show that they can in fact be distinguished using *linked* markers.



- Create the two pedigrees, and the associated marker objects. Both markers are SNPs, with minor allele frequencies  $p_a = p_b = 0.01$ .
- Suppose first that the markers are unlinked, i.e., that  $\theta = 0.5$ . Show that the two pedigrees then have identical likelihoods, as predicted by theory.
- Suppose instead that the markers are linked, with recombination rate is  $\theta = 0.25$ . Compute the likelihood of each pedigree. Which is more likely?
- Compute the likelihood ratio

$$LR = \frac{P(\text{data} | G)}{P(\text{data} | HS)}$$

### Exercise IV-7 (Proof of Haldane's map function)

Haldane's model says that crossovers along a given chromosome occur as a Poisson process with rate 1. This implies that for two loci with genetic distance  $d$  Morgans, the number of crossovers between them follows the distribution given by:

$$P(\text{exactly } m \text{ crossovers}) = e^{-d} \cdot \frac{d^m}{m!}$$

- Explain that observable recombination between two markers is equivalent to an odd number of crossovers between them.
- If  $\theta$  is the recombination rate between the loci, show that  $\theta = e^{-d}(d + \frac{d^3}{3!} + \frac{d^5}{5!} + \dots)$
- Use this to arrive at Haldane's map function. Hint: Use the infinite series  $e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots$  to show that the expression inside the parenthesis in b) equals  $\frac{1}{2}(e^d - e^{-d})$ .