Statistical methods in genetic relatedness and pedigree analysis

NORBIS course, Oslo, June 2022
Magnus Dehli Vigeland and Thore Egeland

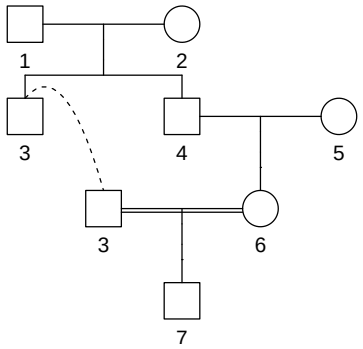
Exercise set IV. Realised relatedness

For these exercises you need the **ibdsim2** package in addition to the core ped suite packages.

```
library(pedsuite)
library(ibdsim2)
```

Exercise IV-1 (Realised inbreeding)

In this exercise we study the realised inbreeding in children of uncle–niece parents:



a) Use the following code to simulate 500 genomes of children with uncle–niece parents.

```
x = avuncularPed("uncle", "niece") |>
  addSon(parents = c(3,6))
plot(x)
sims = ibdsim(x, N = 500, ids = 7, seed = 123)
```

b) Extract the first simulation and plot the haplotypes for chromosome 1:

```
s = sims[[1]]
haploDraw(x, s, chrom = 1, cols = 2:7, pos = 2, height = 5, margin = c(2,1,1,1))
```

Experiment with `pos`, `height` and `margin` to make the plot look nice. Why do you need six colours?

c) How many autozygous segments can you see on chromosome 1? Check your eyesight by listing the coordinates of all autozygous segments:

```
findPattern(s, pattern = list(autozygous = 7))
```

d) The command below produces an informative plot of the distribution of autozygous segments. How many segments does a typical child of uncle–niece have?

```
plotSegmentDistribution(sims)
```

e) (Optional) Rerun the simulations after making individual 4 female (so that 3 becomes a *maternal* uncle of 6). Plot the distributions together and comment on the result.
Hint: Try `plotSegmentDistribution(M = sims, F = sims2)`, where `sims2` are the new simulations.

Exercise IV-2 (Realised IBD between siblings)

In this exercise we look at the distribution of realised IBD coefficients between full siblings.

a) Use the following code to simulate 500 pairs of full siblings:

```
x = nuclearPed(2)
sims = ibdsim(x, N = 500, ids = 3:4, seed = 123)
```

b) Compute the realised IBD coefficients of each simulation:

```
k = realisedKappa(sims)
```

Study the output. What is the sample mean of each coefficients? Round to 3 decimals.

c) The `perSimulation` slot of `k` holds the realised coefficients of each simulation. Plot them in the triangle with the command below, and comment on the result:

```
showInTriangle(k$perSimulation)
```

Exercise IV-3 (DNA sharing between siblings)

This is a continuation of the previous exercise, where we investigate this frequently heard claim:

Siblings share 50% of their DNA.

a) Explain that the fraction of (autosomal) DNA that is shared IBD, equals twice the realised kinship coefficient. (Hint: Use that $\varphi^R = \frac{1}{4}\kappa_1^R + \frac{1}{2}\kappa_2^R$.)

b) Use the simulations from Exercise IV-2 to estimate the mean and standard deviation of the DNA sharing between siblings.
Hint: Let `phi = realisedKinship(sims)` and work with `2 * phi$perSimulation`.

c) Formulate a more precise version of the original claim.

Exercise IV-4 (Fruit fly siblings)

The fruit fly *Drosophila melanogaster* has two autosomal chromosomes of lengths approx. 107 and 110 cM.

a) Make a simplified map of the fruit fly genome as follows:

```
mapFF = list(uniformMap(cM = 107, chrom = 1),
             uniformMap(cM = 110, chrom = 2))
```

b) Simulate 500 pairs of fruit fly siblings, and plot their realised κ coefficients in the IBD triangle. Compare with the plot for humans (Exercise IV-2) and comment on the result.
Hint: Reuse the code from Exercise IV-2, but add `map = mapFF` in the `ibdsim()` command.

Exercise IV-5 (Zero IBD)

Suppose that you are a 10th generation, male-line descendant of Napoleon Bonaparte.

a) What is the probability that none of your (autosomal) DNA originates from Napoleon? Use the following code to give a realistic estimate:

```
x = linearPed(10)
sims = ibdsim(x, N = 500, ids = c(1, 21))
zeroIBD(sims)
```

b) Repeat the analysis for a female line of 10 generations. Why is the result different?