ISFG 2024 Pre-congress Workshop

Kinship Statistics and Pedigree Analysis: Advanced

Magnus Dehli Vigeland and Thore Egeland

Exercise set III. Pedigree reconstruction

Before you start, load the **pedsuite** packages needed in the exercises.

```
library(pedsuite)
library(pedbuildr)  # not a core package; must be loaded separately
```

Exercise III-1   (Pairwise estimates)

This exercise uses the built-in `trioData` dataset from the **pedbuildr** package. The dataset is a matrix containing genotypes for three male individuals.

a) Print `trioData` and study the output. How many markers are there? What kind of markers? How are the alleles labelled? What is the genotype of the 1st person for the 3rd marker?

b) Use the following code to create the three individuals and attach the data.

```
x = singletons(1:3) |> setMarkers(alleleMatrix = trioData, locusAttr = "snp12")
x
```

c) Use `ibdEstimate()` to estimate the pairwise relatedness between the individuals, and plot the corresponding points in the IBD triangle.

```
kap = ibdEstimate(x)
showInTriangle(kap, labels = T)
```

How do you interpret the relationship between each pair? How confident are you?

d) Use the pairwise estimates to deduce how the three individuals are related.
   Hint: Start by connecting the closest pairs in the diagram below.

| 1 | | 2 |

| 3 |
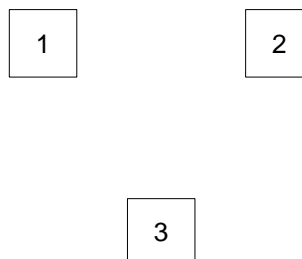
Exercise III-2 (Pedigree reconstruction)

We will now use the full **pedbuildr** machinery to reconstruct the `trioData` pedigree.

a) With `x` as before, run a reconstruction with default options and plot the most likely pedigrees:

```
r = reconstruct(x)
plot(r, top = 6)
```

Is the top pedigree the same as you found in the previous exercise? How much more likely is it compared with the second best?

b) Do the reconstruction again, but this time allowing up to 3 connecting individuals. Limit the search space by specifying that 2 is the father of 3. (Hint: Use options `extra`, `knownPO` and `age`.)

Do you get the same winner as before? Comment on the result.

c) [Bonus question] How are 1 and 3 related in the *second* most likely pedigree in part b)?

Exercise III-3 (A question about fathers)

The following is based on an actual case from Australia. Genotypes are available from two sisters and their children. The first sister has one child, the other has two children. The question we must answer is: *Do any of the children have the same father?*
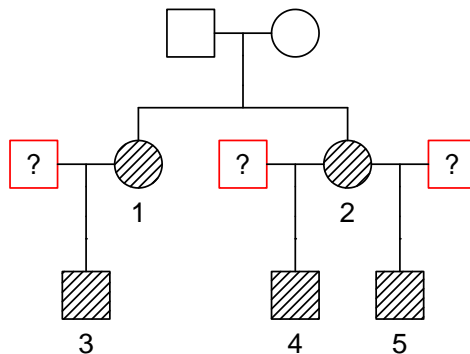
The data is given in the files `reconstruct-fathers.ped` and `reconstruct-fathers.freq` among the datasets for this course, which you can download with the following code:

```
download.file("https://magnusdv.github.io/pedinr/datasets/data.zip", dest = "data.zip")
unzip("data.zip")
```

a) Load the data with the commands below, and use `summary()` to inspect the data. Check that the labels and sexes match the figure. How many markers are used?

```
x = readPed("data/reconstruct-fathers.ped") |>
  setFreqDatabase("data/reconstruct-fathers.freq")
```

b) Use **pedbuildr** to perform a pedigree reconstruction on the data set. *Hint*: Use undisputed parts of the family to restrict the search space. For example, consider the parameters `knownPO` and `noChildren`.

c) Plot the six most likely pedigrees and study the paternity constellations. What is your conclusion?