ISFG summer school - virtual edition 2021

Pedigree analysis in R

Magnus Dehli Vigeland and Thore Egeland

Solutions for exercise set V

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

Exercise V-1

a) The answers are given by the summary output:

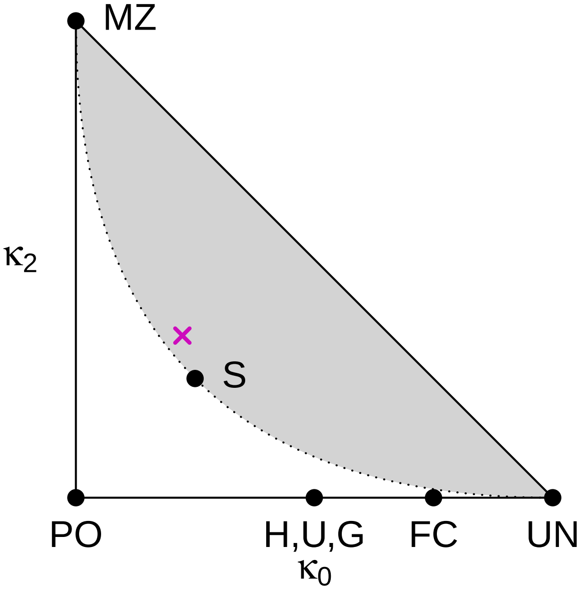
```
## List of 4 singletons.
## Labels: 1 (female), 2 (male), 3 (female), 4 (male).
## 500 attached markers.
```

b)

```
# Estimate kappa
k24 = ibdEstimate(w, ids = c(2, 4), verbose = F)
k24
```

```
##   id1 id2   N      k0      k1      k2
## 1   2   4 500 0.22335 0.43665 0.34001
```

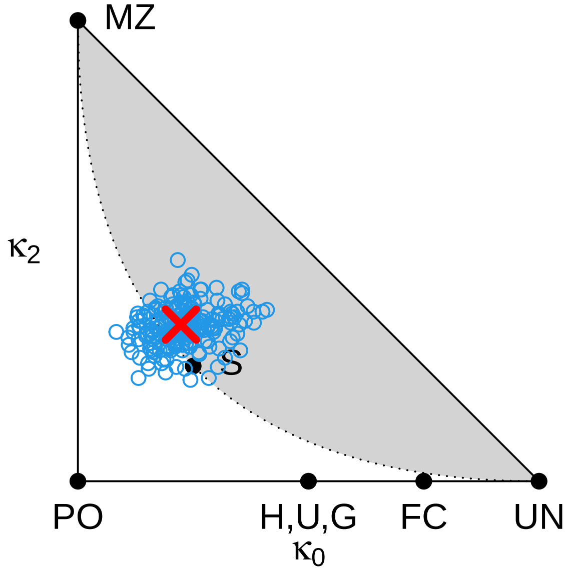
```
# Plot
showInTriangle(k24)
```



The estimate, `k24` = (0.22, 0.44, 0.34) is close to the S point, which suggests that they are full siblings. However, we don't know anything about the uncertainty at this point, so we cannot conclude anything with confidence.

c) R code:

```
bs = ibdBootstrap(w, ids = c(2, 4), param = "kappa", N = 200)
```



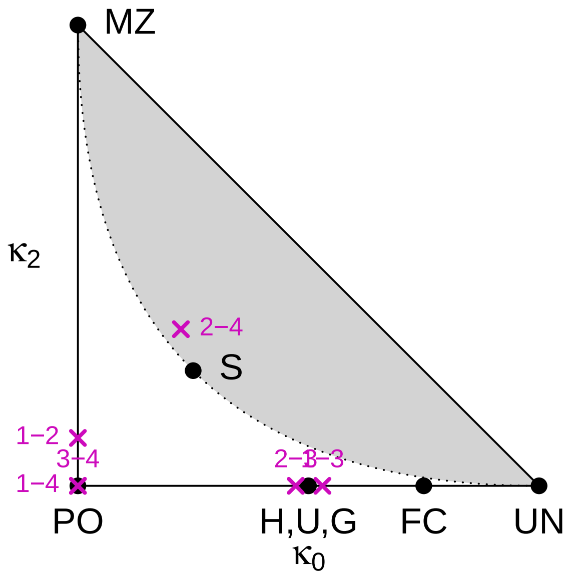
The bootstrap estimates are clustered fairly closely together – in particular they don't overlap with other common relationships. This strengthens our confidence in that 2 and 4 are full siblings.

d) We compute and plot kappa estimates for all pairs:

```
k = ibdEstimate(w, verbose = FALSE)
k
```

```
##   id1 id2   N      k0      k1      k2
## 1   1   2 500 0.00000 0.89603 0.10397
## 2   1   3 500 0.53032 0.46968 0.00000
## 3   1   4 500 0.00000 1.00000 0.00000
## 4   2   3 500 0.47243 0.52757 0.00000
## 5   2   4 500 0.22335 0.43665 0.34001
## 6   3   4 500 0.00000 1.00000 0.00000
```

```
showInTriangle(k, labels = T, pos = c(2,3,2,3,4,3))
```



We conclude that 1-2, 1-4 and 3-4 are most likely parent-child pairs, 2-4 full siblings, and 1-3 and 2-3 either half siblings, uncle-nephew (etc) or grandparent-grandchild. We are most confident about the parent-child relationships, which normally are the easiest to detect.

Exercise V-2

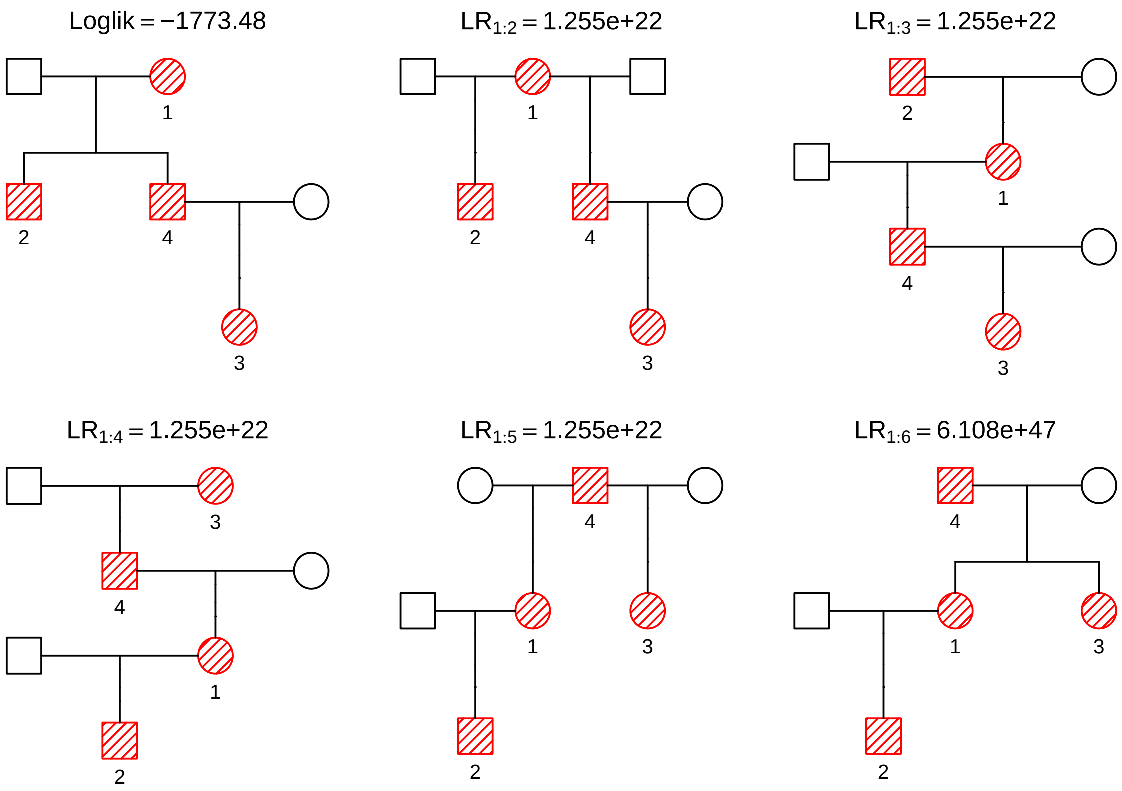
a) The first argument, `w` is the list of singletons, including the marker data. `inferPO = TRUE` instructs the program to first perform a pairwise estimation and detect pairs that are confidently parent-child or confidently *not* parent-child. `linearInb = FALSE` excludes pedigrees involving incest between individuals in a direct line, e.g., parent-child incest.

b) 6 pedigrees.

c) The most likely pedigree is the first one in the figure below. It is more than 10^{22} times more likely than the next (as seen in the plot titles).

```
# Reconstruct
r = reconstruct(w, inferPO = TRUE, linearInb = FALSE)

# Plot the most likely pedigrees
plot(r)
```



Loglik = −1773.48

$LR_{1:2} = 1.255e+22$

$LR_{1:3} = 1.255e+22$

$LR_{1:4} = 1.255e+22$

$LR_{1:5} = 1.255e+22$

$LR_{1:6} = 6.108e+47$

Exercise V-3

a) 24 markers, as reported by `summary()`:

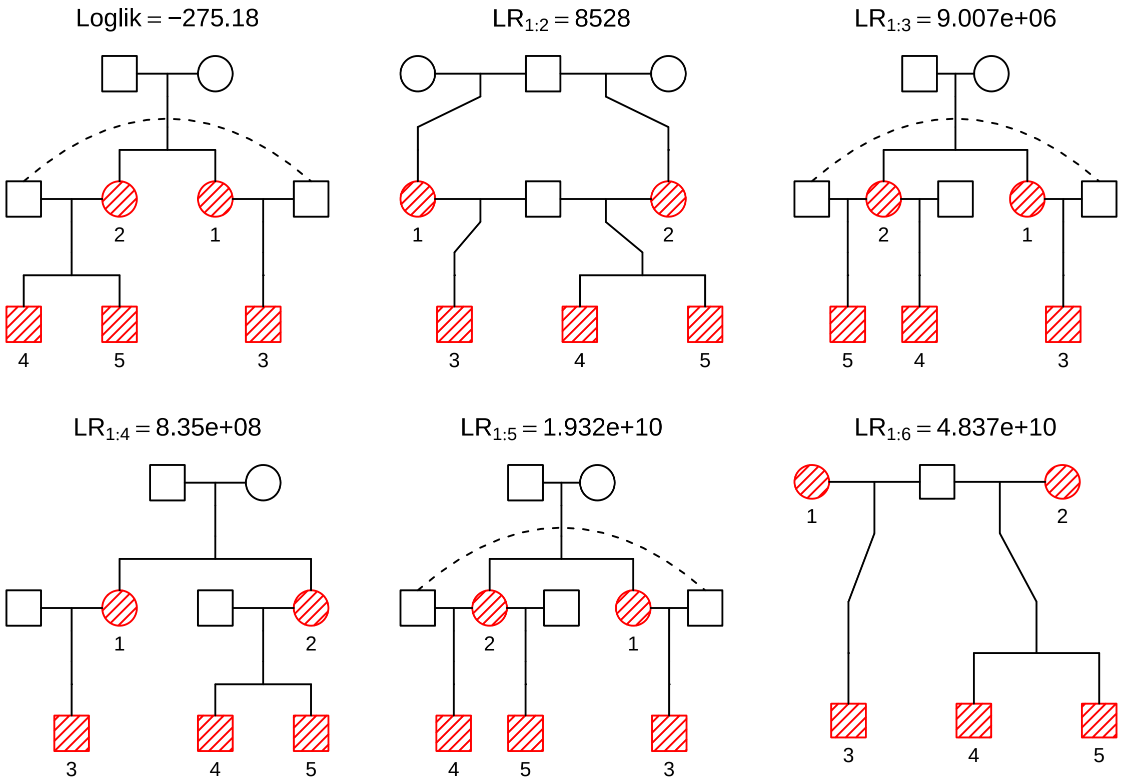
```
x = readPed("data/reconstruct-fathers.ped")
x = setFreqDatabase(x, "data/reconstruct-fathers.freq")
summary(x)
```

```
## List of 5 singletons.
## Labels: 1 (female), 2 (female), 3 (male), 4 (male), 5 (male).
## 24 attached markers.
```

b) Here is one way to do it:

```
# Reconstruct
r = reconstruct(x, noChildren = 3:5, knownPO = list(c(1,3), c(2,4), c(2,5)),
                linearInb = FALSE)

# Plot 6 most likely pedigrees
plot(r, top = 6)
```



All three children have the same father! Indeed, this is supported by both the two most likely pedigrees above.

Exercise V-4

e) The maximum likelihood estimate is the point $(k_0, k_2) = (0, 1)$, indicating a relationship where the individuals are genetically identical. (Note that the simplest such relationship, MZ twins, is impossible in this case, since the individuals have different sex.)

One lesson from this exercise is that relatedness estimation using a single marker is not very informative. The likelihood function will always be linear (i.e. flat!), and therefore maximising in a corner of the IBD triangle. Only when data from several markers are collected, the likelihood function in general becomes strictly concave, and the maximum may occur anywhere in the IBD triangle.