Statistical methods in genetic relatedness and pedigree analysis

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

Exercise set VII. Forensic exercises, part 2

Introduction

Before starting, load the core **pedsuite** packages and also **dvir**.

```
library(pedsuite)
library(dvir)
```

If you haven't downloaded the datasets (needed for the two last exercises) already, do so by running

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

Exercise VII-1 (Prior and posterior odds)

Assume $P(H_1) = p,\ P(H_2) = q = 1 - p$ and

$$\text{LR} = \frac{P(\text{data} \mid H_1)}{P(\text{data} \mid H_2)} = 100.$$

a) What is the prior and posterior odds?

b) Find $P(H_1 \mid \text{data})$.

c) Assume $p = 0.5$. Repeat a) and b). Intepret the answers.

Exercise VII-2 (Exclusion probability for one marker)

Consider the the standard hypotheses

- H_1: The alleged father (AF) is the biological father.
- H_2: The alleged father and the child are unrelated.

There is one SNP marker. The alleles are A and a with frequencies 0.1 and 0.9 respectively. Assume the genotype of the child is a/a.

a) Find the probability that a random man is excluded as the father using pen and paper.

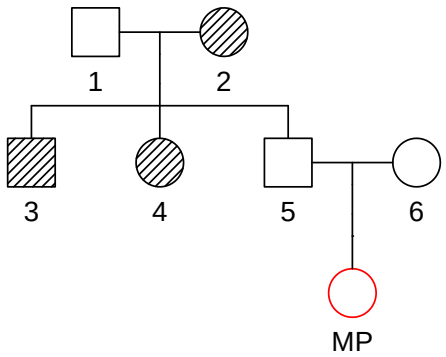
b) Confirm a) using `exclusionPower`.

Exercise VII-3 (Exclusion probability for two markers)

Assume the exclusion probabilities for two independent markers are EP1 = 1/10 and EP2 = 1/2. Find the combined exclusion probability.

Exercise VII-4 (Missing person identification)

In this exercise we will analyse a missing person case, matching two persons of interest against the reference family shown below. The genotype data is contained in the files `mpi-example.ped` and `mpi-example.freq`.



a) Load the data into R and inspect the dataset.

```
mpi = readPed("data/mpi-example.ped")
mpi = setFreqDatabase(mpi, "data/mpi-example.freq")
mpi
```

For convenience, extract the three components into separate variables.

```
ref = mpi$Reference
p1 = mpi$POI1
p2 = mpi$POI2
```

b) Make a missing person plot:

```
missingPersonPlot(ref, missing = "MP")
```

c) Find the exclusion power of the reference family, and interpret the output:

```
ep = missingPersonEP(ref, missing = "MP")
ep
```

d) Find the inclusion power of the reference family, and interpret the output:

```
ip = missingPersonIP(ref, missing = "MP", nsim = 1000, threshold = 10000, seed = 17)
ip
```

e) Plot the exclusion and inclusion powers together in a *power plot*.

```
powerPlot(ep, ip)
```

f) The following code computes the LR when matching POI1 against the reference. Run the commands and comment on the output.

```
test1 = missingPersonLR(ref, missing = "MP", poi = p1)
test1
```

g) Study the marker-wise LRs of the previous test, by running the commands below.

```
lr1 = test1$LRperMarker
cols = ifelse(lr1 > 1, 3, 2)
barplot(lr1, col = cols, ylab = "LR", las = 2, cex.names = 0.8)
abline(h = 1, lty = 2)
```

How many markers are in support of a match, and how many are against?

h) How many exclusions (markers with $LR = 0$) are there for POI1? Is this unexpectedly few/many for an unrelated individual compared with this reference family? *Hint*: Run `barplot(ep$distribMismatch)`.

i) Now we turn to POI2. First find the overall LR against the reference:

```
test2 = missingPersonLR(ref, missing = "MP", poi = p2)
```

Plot the LR for each marker like you did for POI1 previously. Which marker gives the largest LR?

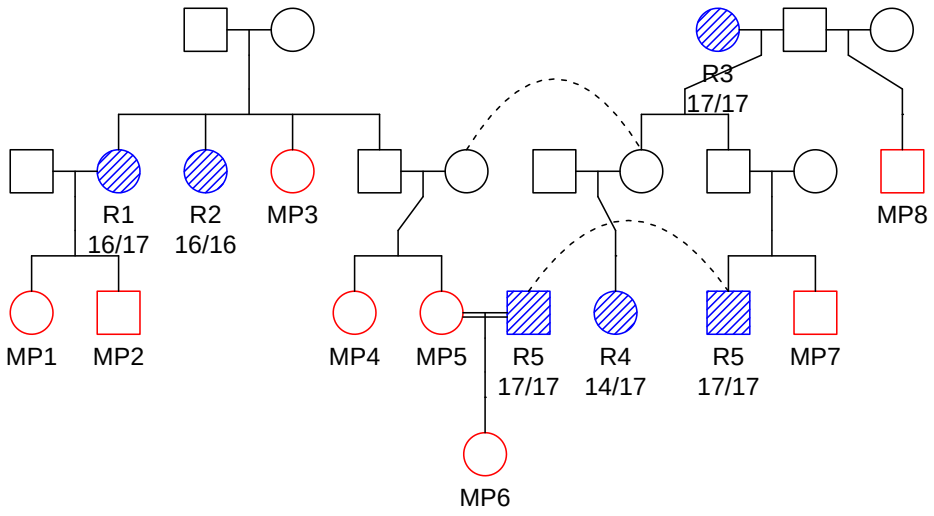
j) Recall that the inclusion power output `ip` contains the LR for 500 simulations of the true missing person. We can use this to compare the LR for POI2 with the LR distribution as follows:

```
hist(log10(ip$LRperSim), xlab = "log(LR)", main = "Simulations of true MP")
lr = test2$LRtotal
abline(v = log10(lr), col = 2, lwd = 2)
text(x = log10(lr), y = 0, label = paste(" LR =", round(lr)), col = 2, pos = 4)
```

Give an overall conclusion regarding POI2.

Exercise VII-5 (DVI analysis)

In this exercise we will analyse the DVI dataset `grave` included in the **dvir** package. The case involves 8 victim samples to be matched against a single family with multiple missing persons. Here is a plot of the reference pedigree, including genotypes for the first marker:



a) To save typing later on, extract the three components of the `grave` dataset:

```
pm = grave$pm            # Post mortem data (victim samples)
am = grave$am            # Ante mortem data (reference family)
missing = grave$missing # Names of the missing persons
```

Familiarise yourself with the dataset by inspecting each of the objects.

b) Try to reproduce the pedigree plot shown above. *Hint*: Here is a good start:

```
refs = typedMembers(am)
plot(am, labs = c(refs, missing), col = list(red = missing, blue = refs))
```

c) How many male/female victims are there, and how many male/female missing persons? Find the *a priori* total number of possible solutions. *Hint*: Check `?ncomb`.

d) Find the inbreeding coefficient of MP6. What is the relationship between her parents? *Hint*: For the last question you can use `verbalise()`.

e) Use `pairwiseLR()` to compute the matrix of pairwise likelihood ratios $\text{LR}_{i,j}$ comparing the assignment $V_i = M_j$ to the null hypothesis of no identification. *Hint*: Check `?pairwiseLR`.

f) Use `jointDVI()` to find the most likely joint solutions, and inspect the top five alternatives. Comment on your findings.