Statistical methods in genetic relatedness and pedigree analysis

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

Solutions for exercise set VII

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) The prior odds is $P(H_1)/P(H_2) = p/q$. The posterior odds is

$$\text{prior odds} \cdot \text{LR} = (p/q) \cdot 100.$$

b) Using Bayes theorem we find

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

c) The prior odds is 1, the posterior odds is 100 and

$$P(H_1 \mid \text{data}) = \frac{\text{LR}}{\text{LR} + 1} = \frac{100}{101} \approx 0.99.$$

The posterior odds of 100 means that H_1 is 100 times more probable than H_2. The posterior probability that AF is the father is 0.99.

Exercise VII-2   (Exclusion probability for one marker)

a) We find

$$EP = P(AF = A/A \mid H_2) = 0.1^2 = 0.01$$

b) We can go

```r
# Define pedigrees
claim = nuclearPed(fa = "AF",  child = "CH")
true = list(singleton("AF"), singleton("CH"))

# Add marker
claim = claim |> addMarker(CH = "a/a", afreq = c(A = 0.1, a = 0.9))

# Compute EP for the alleged father
exclusionPower(claim, true, ids = "AF")

## Potential mismatches: 1 (1)
## Expected mismatches: 0.01
## P(at least 1 mismatch): 0.01
```

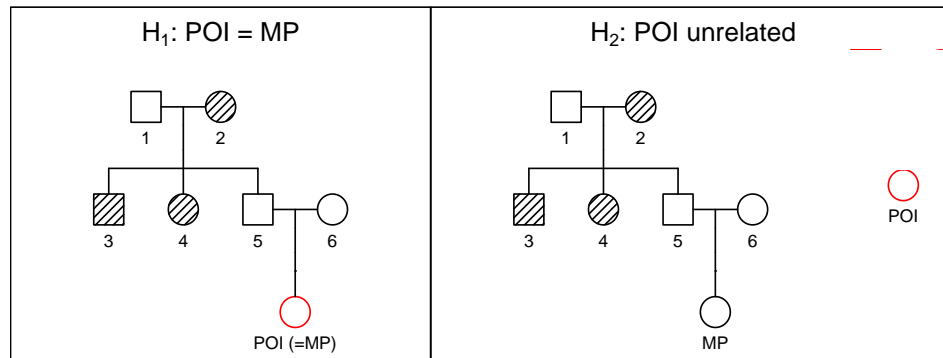**Exercise VII-3   (Exclusion probability for two markers)**

We find

$$P(\text{exclusion}) = 1 - (1 - 1/10) \cdot (1 - 1/2) = 0.55.$$

Exercise VII-4   (Missing person identification)

a) (Answer omitted.)

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

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

```
## Potential mismatches: 8 (D3S1358, D7S820, CSF1PO, PENTA_D, VWA, TPOX, D19S433, D2S1338)
## Expected mismatches: 1.679
## P(at least 1 mismatch): 0.863
```

The overall probability of exclusion is EP $= 0.86$, which is normally considered to be sufficient.

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

```
## Mean LR: 46998.29
## Mean log10(LR): 3.223
## Estimated power:
##   P(LR >= 10000) = 0.282
```

The output shows that only 28% of the simulations give an LR exceeding 10000. The answer depends to some extent on the seed, but with 1000 simulations the numbers will not change much.
If you wonder why the mean LR is much higher than $10^{\text{mean } \log(\text{LR})}$, this is a simple consequence of the fact that the function $log(x)$ is concave. (Google *Jensen's inequality* for more details.)
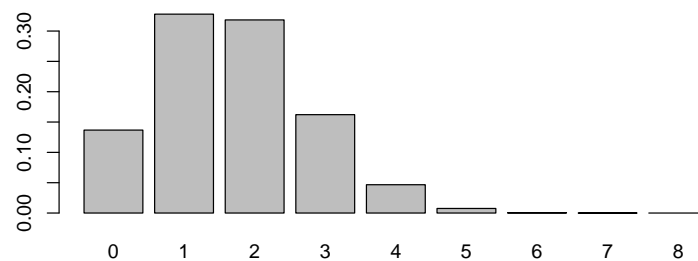
e) `powerPlot(ep, ip)`
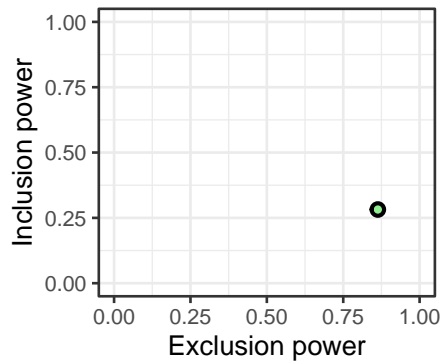
f) The commands give LR $= 0$, which excludes POI1 as the missing person (adding the usual caveats).

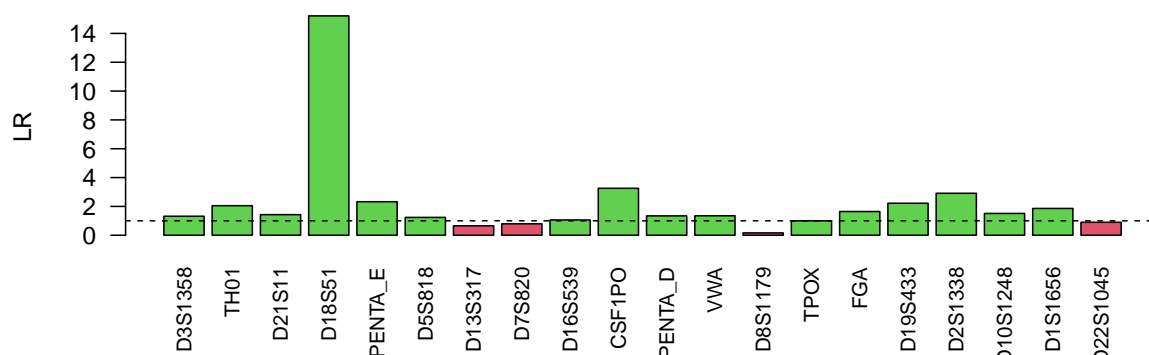g) There are 5 markers in favor ($LR > 1$, green bars) and 15 against (LR < 1, red bars).
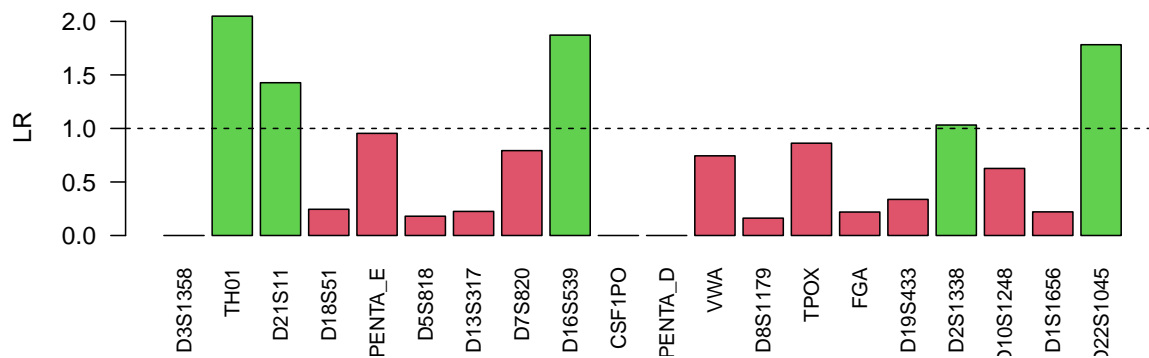
h) 3 exclusions. This is within the normal range for this reference data, as indicated by this plot:

```
barplot(ep$distribMismatch)
```


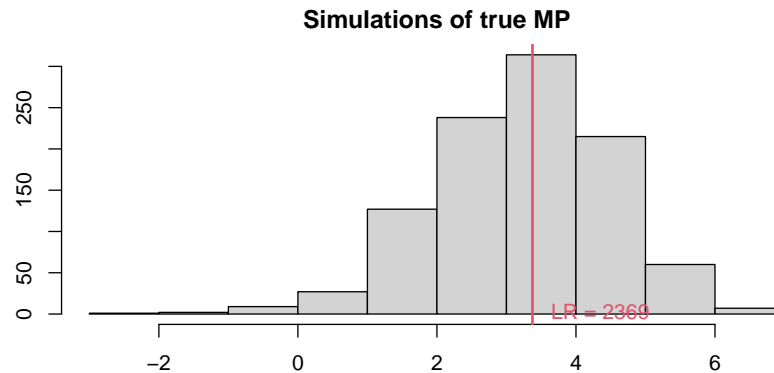i) The suggested code gives LR $= 2369.091$ for POI2, with the marker-wise LRs as shown below.

The marker D18S51 gives the largest LR:

```
test2$LRperMarker["D18S51"]
```

```
##  D18S51
## 15.2168
```

j) The LR for POI2 is quite as expected for the true POI, as indicated by the distribution plot:

Simulations of true MP

We conclude that POI2 is quite likely to be the missing person, but that some further evidence is needed to declare a positive match.

Exercise VII-5   (DVI analysis)

a) (Answer omitted.)

b) The plot was produced with the following command:

```
refs = typedMembers(am)
plot(am, marker = 1, hatched = refs, labs = c(refs, missing),
     col = list(red = missing, blue = refs)
```

c) The total *a priori* number of solutions is:

```
ncomb(5,5,3,3)
```

```
## [1] 52564
```

The numbers in the function call are, respectively:

- the number of female victims
- the number of female missing persons
- the number of male victims
- the number of male missing persons

d) MP6 has inbreeding coefficient $f = 1/16$, and her parents are first cousins. Code:

```
inbreeding(am, "MP6")
verbalise(am, parents(am, "MP6"))
```

e) The pairwise LR matrix is found as follows:

```
prw = pairwiseLR(pm, am, missing)
prw$LRmatrix
```

```
##              MP1          MP2         MP3         MP4         MP5         MP6
## V1 479971259            0 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
## V2         0 67760107189 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
```

```
## V3          0           0 6.409841e+14 0.000000e+00 0.000000e+00 0.000000e+00
## V4          0           0 0.000000e+00 1.803600e+12 1.803600e+12 0.000000e+00
## V5          0           0 0.000000e+00 1.030067e+11 1.030067e+11 0.000000e+00
## V6          0           0 0.000000e+00 0.000000e+00 0.000000e+00 8.817392e+12
## V7          0           0 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
## V8          0           0 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
##          MP7         MP8
## V1          0   0.0000000
## V2          0   0.5512209
## V3          0   0.0000000
## V4          0   0.0000000
## V5          0   0.0000000
## V6          0   0.0000000
## V7   16946051 295.8389523
## V8          0   0.2684890
```

f) The joint solutions are found as follows:

```
res = jointDVI(pm, am, missing)
head(res)
```

```
##    V1  V2  V3  V4  V5  V6  V7  V8    loglik          LR   posterior
## 1 MP1 MP2 MP3 MP4 MP5 MP6 MP7   * -737.0038 1.374125e+90 6.904732e-01
## 2 MP1 MP2 MP3 MP4 MP5 MP6 MP7 MP8 -737.8061 6.159953e+89 3.095266e-01
## 3 MP1 MP2 MP3 MP4 MP5 MP6   *   * -752.3418 2.997856e+83 1.506369e-07
## 4 MP1 MP2 MP3 MP4 MP5 MP6   * MP8 -753.3430 1.101543e+83 5.535057e-08
## 5 MP1 MP2 MP3   * MP5 MP6 MP7 MP8 -773.8441 1.375599e+74 6.912137e-17
## 6 MP1 MP2 MP3   * MP5 MP6 MP7   * -774.8047 5.263521e+73 2.644825e-17
```

Comment: The top two joint solutions, which are much more likely than the rest, agree that $V_i = MP_i$ for $i = 1, ..., 7$, while disagreing whether V_8 is MP_8 or has no match. The uncertainty about V_8 cannot be resolved with the given data. While the pairwise LRs could not separate the sisters V_4 and V_5, the joint solution uniquely identifies them directly.