Statistical methods in genetic relatedness and pedigree analysis

Oslo January 2020

Magnus Dehli Vigeland and Thore Egeland

Exercise set VI. Forensic Exercises.

Complicating factors like mutation are ignored unless otherwise stated. Exercises marked with a * are more mathematical and can be skipped. [1]

Exercise 1

The purpose of this exercise is to illustrate the basic concepts of forensic genetics and paternity testing by means of a simple paternity case. We consider the following two hypotheses

H_1**:** The alleged father (AF) is the biological father.

H_2**:** The alleged father and the child are unrelated.

The alleged father and the child are genotyped. The mother is not disputed. The figure below illustrates the hypotheses for the two markers used in this exercise.

a) Consider initially only the first marker. The allele frequencies are p_A and p_B. Explain why the likelihood ratio is $LR_1 = 1/p_A$.

b) Use R to plot LR_1 as a function of $p_A \in (0.001, 0.2)$. Explain why the likelihood ratio increases as p_A approaches 0.

c) There is a second autosomal marker, called S2, with alleles including a, b, and c with allele frequencies p_a, p_b, and p_c. Calculate the likelihood ratio for this marker and also for both markers.

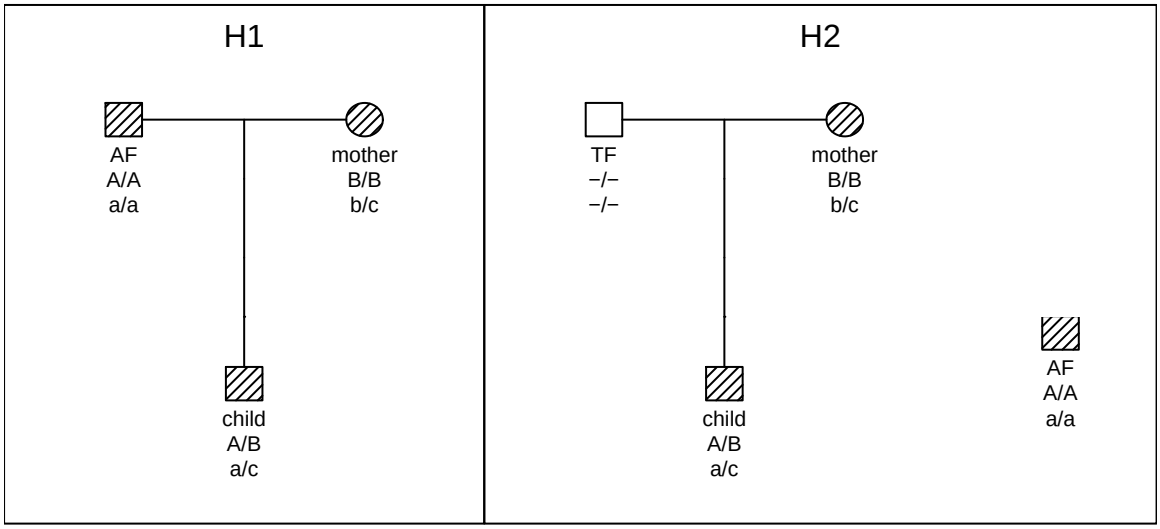
d) Assume $p_A = 0.05$ and $p_a = p_b = p_c = 0.1$ and find the likelihood ratio in this case. How do you interpret LR?

e) Try the following commands which can be used to check answers above:

```
library(forrel)
x = nuclearPed(1, father = "AF", mother = "mother",
               children = "child")
m1 = marker(x, alleles = c("A", "B", "C"), afreq = c(0.05, 0.05, 0.9),
            AF = "A", mother = "B", child = c("A", "B"))
m2 = marker(x, alleles = c("a", "b", "c", "d"), afreq = c(0.1, 0.1, 0.1, 0.7),
            AF = "a", mother = c("b", "c"), child = c("a", "c"))
H1 = setMarkers(x, list(m1, m2))
H2 = list(nuclearPed(1, father = "TF", mother = "mother",
               children = "child"), singleton("AF"))
```

[1] The online version of the exercises was revised 2020-01-11. Most importantly, the previous Exercise 9 is removed (numbering not changed) and some code in Exercise 4 has been simplified.

```
H2 = transferMarkers(H1, H2)
plotPedList(list(H1, H2), marker = 1:2, shaded = typedMembers,
            frametitles = c("H1", "H2"))
```




```
res = kinshipLR(list(H1, H2), ref = 2)
```

Exercise 2

We consider the same hypotheses as in the previous exercise. There is one marker with alleles denoted 1, 2, and 3. The alleged father is 1/2, the child is 1/3; the mother is untyped. The allele frequencies are p_1, p_2, and p_3.

a) Find the likelihood ratio.

b) Is it possible to have $LR < 1$ in a case where the father and son share an allele? Explain.

Exercise 3

Confirm the answers of the previous exercise when $p_1 = 0.2$, $p_2 = 0.3$ and $p_3 = 0.5$ by running:

```
library(forrel)
H1 = nuclearPed(1, father = "AF",  children = "CH")
p1 = 0.2; p2 = 0.3; p3 = 0.5
m1 = marker(H1, AF = c(1,2), CH = c(1,3), afreq = c(p1, p2, p3))
H1 = setMarkers(H1, m1)
H2 = list(singleton("AF"), singleton("CH"))
H2 = transferMarkers(H1, H2)
kinshipLR(list(H1 = H1, H2 = H2), ref = 2)
c("LR formula" = 1/(4*p1))
```

Exercise 4

This exercise demonstrates how you can continue projects in Windows `Familias`, freely available from https://familias.no/, in R. We consider a standard paternity case. The hypotheses are as in Exercise

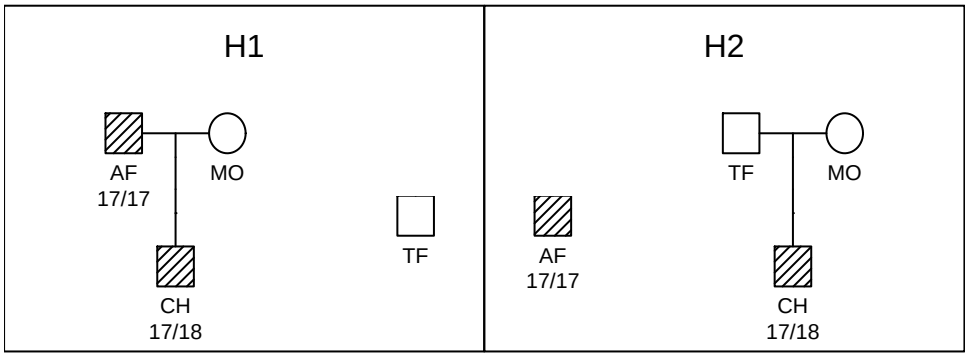
1, but now there are 21 markers. Once the `Familias` data, i.e., the `.fam` file has been converted to ped objects, we have access to all the functionality of R. The basic function, `forrel::readFam`, used for conversion from `Familias` to R is illustrated first:

a) Run the below commands

```
library(forrel)
dat = readFam("http://familias.name/norbisRelatedness/paternityCase.fam",
              verbose = FALSE)
plotPedList(dat)
```

The next step is to understand the structure of the converted data returned from the function `readFam`, i.e., `dat` above. The plot generated by `plotPedList` is helpful. In this case the `.fam` file includes pedigree data and the output is a list of ped objects.

b) Produce the below plot where the genotypes are for the first marker.



c) Show that LR comparing H1 to H2 is 0 by running

```
res = kinshipLR(dat, ref = 2)
```

d) Find the marker with $LR = 0$.

e) Calculate the LR once more, but now with the marker giving 0 likelihood ratio removed (a practice we advise strongly against) by running

```
prod(res$LRperMarker[-5])
```

Rather than removing incompatible markers, we introduce a mutation model. The possible mutation models "custom", "equal", "proportional", "random", "stepwise" and "onestep" are described in the documentation of `pedmut::mutationModel`. Different models can be used for females and males. Note that "custom" is completely general as you can define the mutation matrix. Below we use the `proportional` model. This model has nice mathematical properties like being stationary, but is biologically unsound.

f) Run the below commands

```
library(pedmut)
mutmod(dat, marker = "PENTA_E") = list("proportional", rate = 0.005)
kinshipLR(dat, ref = 2)$LRperMarker["PENTA_E", ]
```

g) Experiment with some different mutation models. Describe the mutation model defined below and interpret the output.

```
nA = nAlleles(dat,"PENTA_E")
als = alleles(dat[[1]][[1]],"PENTA_E")
m = diag(nA)
dimnames(m) = list(als, als)
m["10", "11"] = 0.005
m["10", "10"] =  1 - 0.005
mutmod(dat, marker = "PENTA_E") = list("custom", matrix = m)
kinshipLR(dat, ref = 2)
kinshipLR(dat, ref = 2)$LRperMarker["PENTA_E", ]
```

h) This exercise demonstrates a common strategy in forensic case work: mutations are introduced only when needed. Would you rather prefer modelling mutations for all markers before any calculations are done? Give reasons for your answer.

Exercise 5

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

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

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 6

Consider the the hypotheses of Exercise 1. 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. Find the probability that a random man is excluded as the father.

Exercise 7

Assume the exclusion probabilities for two independent markers are $EP_1 = 1/10$ and $EP_2 = 1/2$. Find the combined exclusion probability.

Exercise 8*

Consider the hypotheses

H_1**:** A and B are first cousins.

H_2**:** A and B are unrelated.

a) What is the LR if A and B share no alleles?

b) What is the smallest possible LR with 16 independent markers? Consider an autosomal marker with alleles a, b, c and d with allele frequencies 0.1, 0.1, 0.1 and 0.7, respectively. What is the LR if both individiduals are homozygous for allele d?

Exercise 9*

A child was conceived as a result of a rape.[2] The DNA-profiles of the defendant, mother, and child are available, see Table 1. The Likelihood Ratio $LR_{1,2}$ of this genetic evidence for the hypotheses

H_1 : The defendant is the father of the child,

H_2 : The defendant is unrelated to the father of the child,

is very high, thus providing strong evidence for paternity of the defendant.

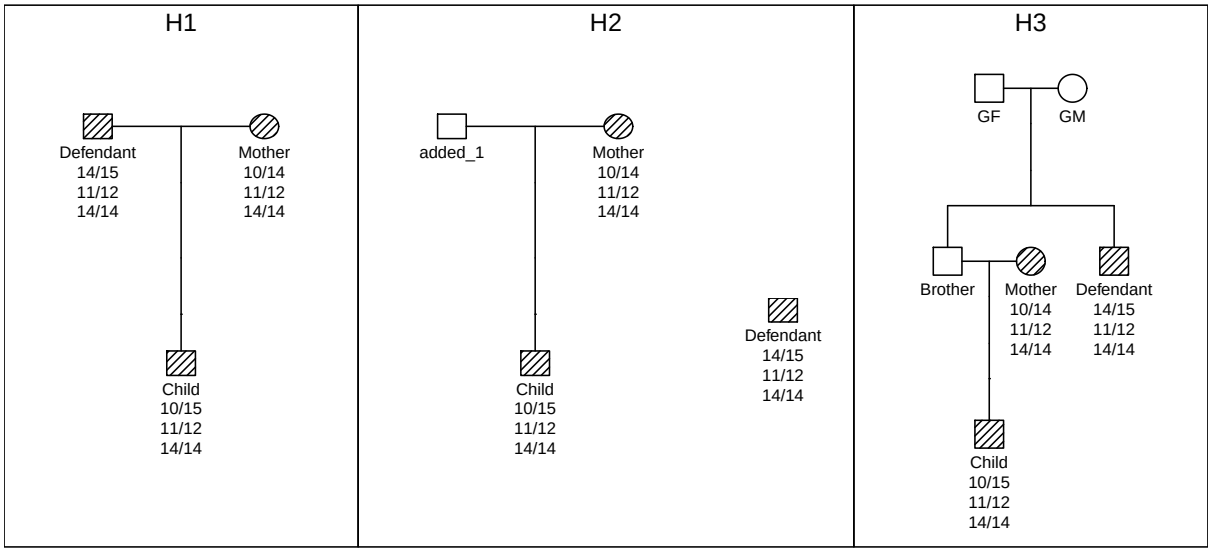
	marker	Mother	Child	Defendant	$LR_{1,2}$
1	CSF1PO	10/14	10/15	14/15	4.56
2	D2S1338	17/17	17/24	17/24	4.26
3	D3S1358	14/16	14/17	17/18	2.36
4	D5S818	11/13	12/13	11/12	2.83
5	D7S820	11/12	11/12	11/12	2.92
6	D8S1179	10/14	10/15	14/15	4.56
7	D13S317	8/13	12/13	12/12	3.24
8	D16S539	9/10	9/9	9/12	4.81
9	D18S51	13/14	14/18	13/18	5.45
10	D19S433	14/14	14/14	14/14	2.93
11	D21S11	29/29	29/30	30/33.2	2.15
12	FGA	22/24	24/24	22/24	3.63
13	TH01	9.3/9.3	9.3/9.3	7/9.3	1.64
14	TPOX	8/8	8/8	8/8	1.84
15	vWA	15/18	15/16	16/16	4.96
16	All				50218439.00

Table 1: Data for Exercise 9.

Now suppose that the defendant claims that he is innocent, but that he believes his brother is the actual father of the child. We formulate a third hypothesis

H_3 : The defendant's brother is the father of the child.

The figure below illustrates the hypotheses with genotypes for CSF1PO, D7S820, and D19S433.



Note that item i) below provides a *ped suite* solution to some of the questions.

[2]Exercise 9 in course version has been removed. This is the previous Exercise 10.

a) Give the algebraic formula for the Likelihood Ratio $LR_{1,2}$ for loci CSF1PO, D7S820, and D19S433.

b) Give the algebraic formula for the Likelihood Ratio $LR_{3,2}$ for the same loci.

c) Can you compute $LR_{3,2}$ numerically with the information above, or do you need a table of allele frequencies? Can you explain why?

d) What is the Likelihood Ratio for H_1 versus H_3 based on these three loci?

e) In the algebraic formula for $LR_{1,3}$ for marker CSF1PO, calculate its limits for $p_{15} \to 1$ and $p_{15} \to 0$, and explain the outcome.

f) Do the same for $LR_{1,3}$ on marker D19S433 when $p_{14} \to 1$ and $p_{14} \to 0$.

g) Discuss in the same way marker D7S820.

h) It can be shown that $LR_{3,2} = 500$. Can you calculate the probability that each hypothesis is true?

i) The code below can be used to check some results

```
library(forrel)
dat = readFam("http://familias.name/norbisRelatedness/Solutions2_18.fam")
H1 = dat[[1]][[4]]
H2 = list(dat[[2]][[4]], dat[[2]][[5]])
H3 = dat[[3]]
peds = list(H1, H2, H3)
# plotPedList(list(H1, H2, H3), shaded = typedMembers, marker = 1:3,
#     skip.empty.genotypes = T, cex = 1,frametitles =c("H1","H2","H3"))
kinshipLR(peds,ref = 2)$LRperMarker

p15 = as.double(afreq(H1,"CSF1PO")["15"])
c("LR.12 CSF1PO" = 1/(2*p15))
c("LR.32 CSF1PO" = (1+2*p15)/(4*p15) )

p11 = as.double(afreq(H1,"D7S820")["11"])
p12 = as.double(afreq(H1,"D7S820")["12"])
c("LR.12 D7S820" = 1/(p11 + p12))
c("LR.32 D7S820" = (1 + p11 + p12)/(2*(p11 + p12)) )

p14 = as.double(afreq(H1,"D19S433")["14"])
c("LR.12 D19S433" = 1/p14)
c("LR.32 D19S433" = (1 + p14)/(2*p14) )
```