Exercises using Familias 2[1]

Contents

---

[1] Thore Egeland, Oct 6, 2013
Norwegian University of Life Sciences
P.O Box 5003
N-1432 Aas Norway
Thore.Egeland@umb.no

The present document contains exercises for **Familias**. For these exercises we will be using Familias 2.0 (same as 1.97) which can be downloaded from http://familias.name/download.html. Some suggested solutions appear in this document . There are also some videos available. Input files used in the exercises are here. There is little overlap between the present exercises and the previous, appearing in the manual. In some cases, answers are given with ridiculously many digits. This is done since the exercises also serve to demonstrate that the program calculates correctly.[2]

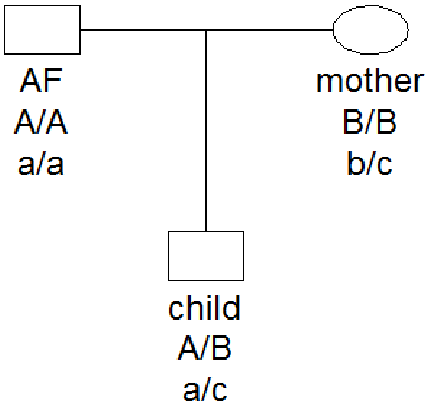
Exercises marked with a star (*) are more theoretical. Questions regarding theoretical calculations, can also be skipped if the focus is on Familias.

Exercise S1 Standard paternity case. Small example

The purpose of this exercise is to consider the simplest possible paternity case. Figure 1[3] shows a mother (undisputed), an alleged father (AF) and a child. We consider

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

Figure1. Standard paternity case.



a) Consider first only one one autosomal locus, called S1**,** with alleles A, B and C, see Figure 1. The allele frequencies are $p_A = p_B = 0.05$ and $p_C = 0.9$. Explain why the likelihood ratio is $LR = \frac{1}{p_A} = 20$. How do you interpret the *LR*?

b) Calculate the *LR* using **Familias**.

c) There is a second autosomal locus, called S2**,** with alleles a, b, c and d with allele frequencies 0.1, 0.1, 0.1 and 0.7, respectively. See Figure 1. Calculate the *LR* for this marker by hand. Calculate the likelihood ratio for the two first markers using **Familias**.

---

[2] More complete validation appears in “Validation of software for calculating the likelihood ratio for parentage and kinship”, Drábek, FSI: Genetics, Volume 3, Issue 2, Pages 112-118, (2008).

[3] All figures are made using the R-package paramlink available from http://cran.r-project.org/

d) It can be shown that the likelihood ratio for two first markers is $\left(1/p_A\right)*\left(1/p_a\right)$. Use this to verify the **Familias** answer.
e) Generate a report using the Report option. Save the report file. The report includes all input and all output. Check that the report file is correct. In particular check that the *LR* for markers S1 and S2 as well as the combined likelihood ratio is correct.
f) Save the **Familias** file. Exit **Familias**.
g) Start **Familias** and read the previously saved file.
h) We next consider *theta-correction* (θ*-correction*). For simplicity we will only use the first marker, S1. The θ- parameter is called kinship parameter in **Familias** and is set using the Options button in **Familias**. Set the kinship parameter to θ=0,02. Calculate the *LR* for the first marker S1. To get calculations for selected markers only, in this case S1, use the Systems included button. Check that your answer coincide with the following theoretical result

$$LR = \frac{1+3\theta}{2\theta+\left(1-\theta\right)p_A}$$

i) Discuss the assumptions underlying the calculations of this exercise.

Exercise S2. Paternity case. Mutation

We consider a motherless paternity case, see Figure 2, with one marker, VWA:

Allele	14	15	16	17	18	19	20	21
Freq	0,072	0,082	0,212	0,292	0,222	0,097	0,02	0,003

The case data and hypotheses are:

- Alleged father : 14, 15
- Child : 16, 17
- H_1: Alleged father is the true father.
- H_2: Alleged father and the child are unrelated.

a) Explain why *LR=0*. Confirm this answer using **Familias.**
b) Use the mutation model Equal probability with mutation rate *R*=0.007 for both males females and calculate *LR*.
c) Let

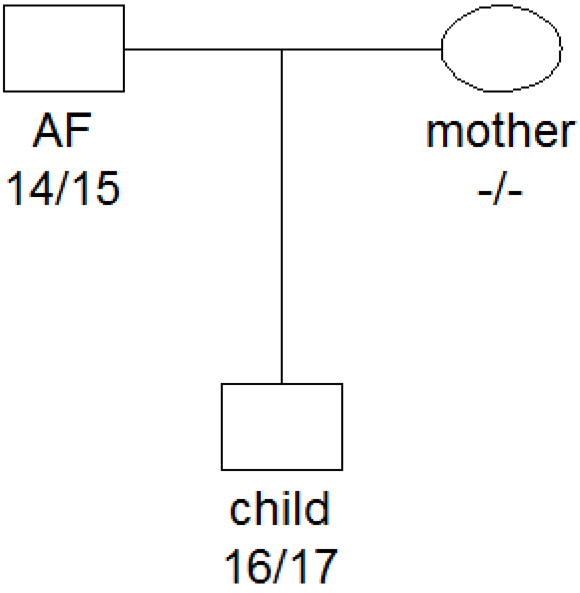
$$n = 8 \text{ (the no of alleles)},$$
$$m = R/(n-1) = 0.007/7 = 0.001,$$
$$p_{16} = 0.212 \text{ and } p_{17} = 0.292.$$

It can be shown (see Exercise S8) that

$$LR = \frac{m\left(p_{16} + p_{17}\right)}{2p_{16}p_{17}}.$$

Use this formula to confirm the **Familias** calculation.

Figure 2. Paternity case. Mutation



Exercise S3 Cap paper challenge ("bone")

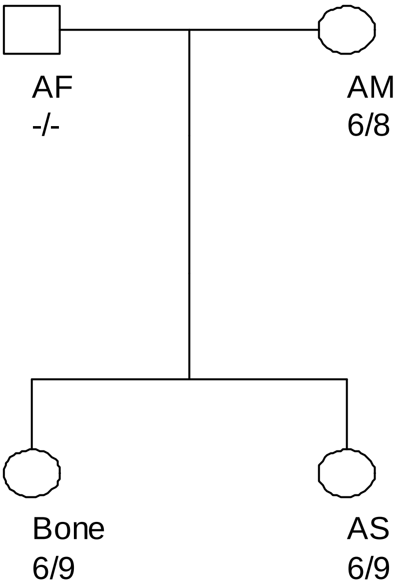
The College of American Pathologists (CAP) has several proficiency testing programs targeted to laboratories that perform DNA typing of STR loci. The below is a test from 2011: Hikers come across human skeletal remains in a forest. Evidence around the site provides a clue as to the identity of the individual. You are asked to test a bone to determine if the individual (bone) is related to an alleged mother (AM) and the mother's other daughter, the alleged full sister, (AS) see Figure 3. Consider the following two hypotheses:

- H_1: The individual providing the bone is the daughter of AM and sister of AS.
- H_2: The individual providing the bone is not related to the tested individuals (AM,AS) of Figure 3

a) Enter the data for the marker shown in Figure 3 manually into **Familias** and calculate the *LR*.
b) Read input from the file ExS3.txt. Calculate the *LR* based on all markers.
c) Use the output from **Familias** to find the *LRs* of the individual markers. Check that the answer for F13B corresponds to the one you found in problem a) above. Complete the below form:

LR	***LR(H1/H2)***
D3S1358	
D21S11	
D18S51	
D7S820	
D16S539	
CSF1PO	
F13B	
LPL	
TOTAL	

d) One of the markers, D7S820, gives a very large *LR*, namely 11189. What do you think is the reason for this large *LR*? What is the combined *LR* if marker D7S820 is removed?



Alle frequencies for F13B 14:0.122, 15:0.258, 17:0.197, rest:0.423.

Exercise S4 Cap paper challenge ("grandson-grandfather")

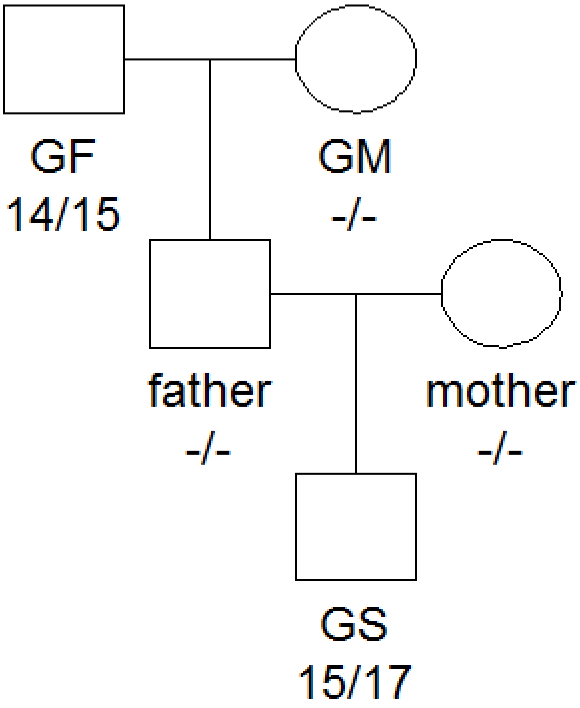
Two individuals, GF and GS, are submitted to the laboratory for testing. We consider the hypotheses

- H_1: GF is the grandfather of GS.
- H_2: The individuals GF and GS are unrelated.

Figure 4 shows the pedigree corresponding to H_1 for the first marker (D3S1358).

a) Enter the data manually into **Familias** and calculate the *LR* for the first marker shown in Figure 4. The allele frequencies are 14: 0,122; 15:0,258;17:0,197
b) Calculate the *LR* based on all markers. Read input from the file ExS4.txt.
c) *For the discussion:* Formulate a conclusion. In the CAP exercise it was stated that GF and GS share the same Y-haplotype and that the frequency of this haplotype is 0.0025. Can this information be used?

Figure 4. Grand father alternative



Exercise S5 Standard paternity case. Essen-Möller

We revisit Exercise S1. Consider the hypotheses

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

Rather than calculating the *LR* we will now calculate the Essen-Möller index *W* defined as the probability of H_1 conditional on the genotypic data.

Assume *apriori* that the hypotheses H_1 and H_2 are *apriori* equally likely. Then it can be shown that

$$W = P(H_1 \mid data) = \frac{LR}{LR+1}.$$

a) Recall that *LR=20* for the first marker. Calculate *W*.
b) Recall that *LR=200* for two markers. Calculate *W*.
c) In **Familias**, *W* is obtained pressing Probabilities. Use **Familias** to calculate *W* for the two above cases.
d) *For the discussion*: Do you prefer *LR* or *W*?

Exercise S6. Incest?

Consider the following hypotheses

- H_1: AF, the father of mother (undisputed), is the father also of her child.
- H_2: An unrelated man is the father of the child.

Figure 5 shows the pedigree corresponding to hypothesis H_1. The allele frequencies are $p_1 = p_2 = p_3 = 0.05$.

a) Use **Familias** to calculate the *LR*. Check that your result coincides with the theoretical $LR = 1/(2 * p_2)$. Does the incest influence the result *LR* in *this particular* case?
b) The defense claims that one should rather consider the following three hypotheses

   - H_1: AF, the father of mother, is the father also of her child.
   - H_2: An unrelated man is the father of child.
   - H_3: The brother of mother is the father of the child. See Figure 6.

   The *LR* can be calculated in several ways depending on the choice of the reference. Calculate

$$LR_{1,2} = \frac{P(data \mid H_1)}{P(data \mid H_2)} \text{ and } LR_{1,3} = \frac{P(data \mid H_1)}{P(data \mid H_3)}.$$

   Check that $LR_{1,3} = 2/(1+p_2)$.
c) When there are more than two hypotheses, as above, some prefer to rather calculate posterior probabilities for the hypotheses. Assume that each of the three hypotheses are equally likely and calculate the posterior probabilities

$$P(H_1 \mid data),\ P(H_2 \mid data) \text{ and } P(H_3 \mid data).$$

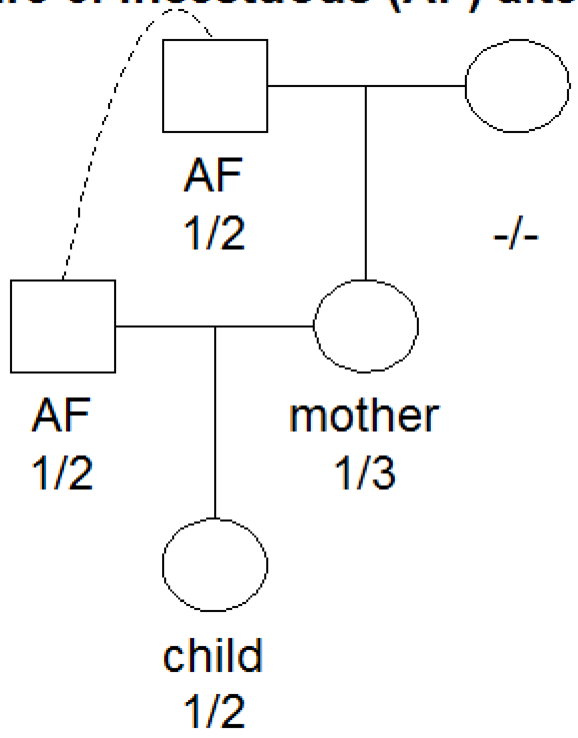
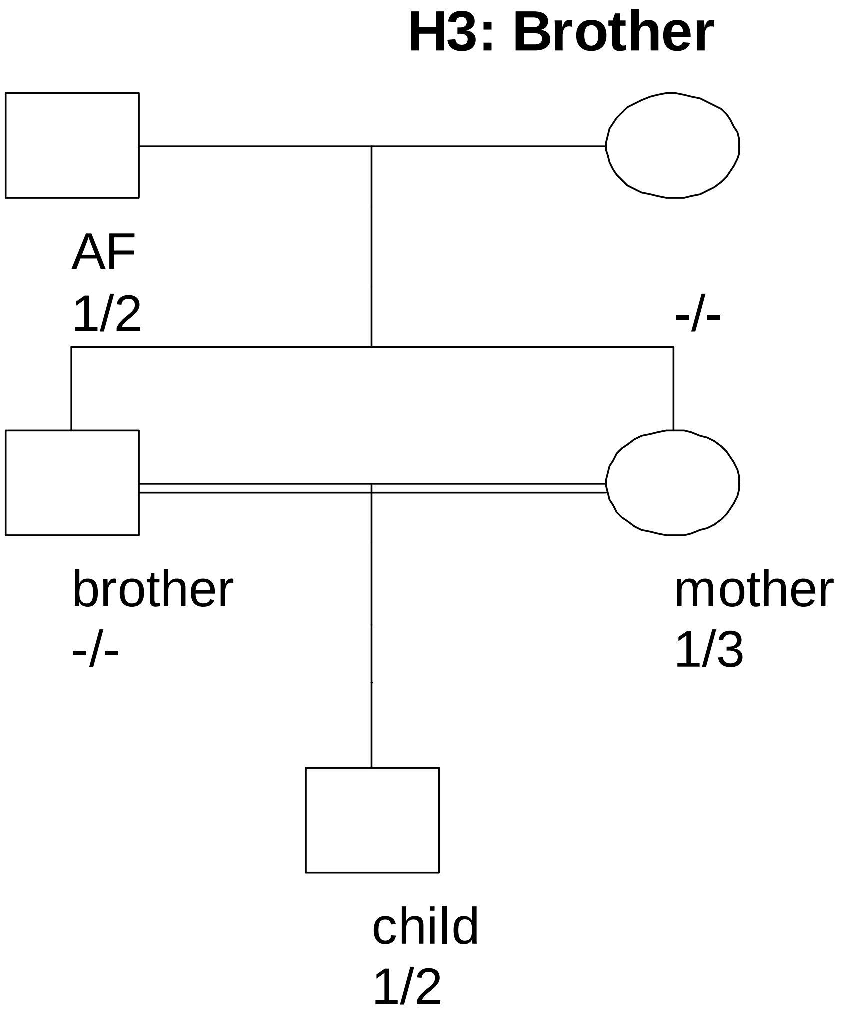


Figure 5. Incestuous (AF) alternative



Exercise S7*. Mutation

In this exercise, which extends on Exercise S2, we will try the different mutation models. Throughout we consider data from the system VWA:

Allele	14	15	16	17	18	19	20	21
Freq	0,072	0,082	0,212	0,292	0,222	0,097	0,02	0,003

The case data and hypotheses are:

- Alleged father : 14, 15
- child : 16, 17
- H_1: Alleged father is the true father.
- H_2: Alleged father and child unrelated.

There are four different mutation models in **Familias**. In this exercise, all four will be tried. We will use overall mutation rate R=0.005 and the same model for females and males. The answer for this exercise will be obtained using **Familias**; Exercise S8, on the other hand, is based on theoretical calculations.

a) Load the data ExS7.txt into **Familias**.
b) Use model four Equal probability. Calculate the *LR*.
   (Answer; *LR*=2.9E-03=0.0029. Comment: For this model mutations to all other alleles are equally likely).
c) Use model three Probability proportional to frequency. Calculate the LR.
   (Answer LR=6.3E-03=0.0063. Comment: For this model it is more likely to mutate to a common allele compared to a rare).
d) Use model two Probability decreasing with range (equal). For this model there are two parameters. The first is as before and should be set to R=0.005. Set the second parameter, Mutation range to r=0.5. Calculate the *LR*.
   (Answer: *LR*=4.7E-03=0.0047. Comment: For this model, mutation depends on the size of the mutation. With r=0.5, a two step mutation occurs with half the probability of a one-step mutation. A three step mutation occurs with half the probability of a two step mutation and so on.)
e) Use model one Probability decreasing with range (stable) with parameters as above.
   (Answer: *LR*=6.4E-03=0.0064.)
f) *Comment*: Model 1 (Probability decreasing with range (stable)) and model 3 (Probability proportional to frequency) and 4 are *stable* whereas model 2 and 4 are *unstable.* If a model is stable, introducing a new untyped person, say the father of the alleged father, does not change the *LR*. This is a reasonable property of a model as introducing irrelevant information should not change the result. For an unstable model, however, the *LR* will change slightly.

   Verify that for model 1 and model 3, that the *LR* does not change if a father of the alleged father is introduced, whereas a slight change occurs for the two other models.
g) *Comment*: There is an other subtle point of all mutation models: In this case only five alleles (14, 15, 16, 17 and ‘Rest allele’) are needed rather than the 8 alleles defined. A five allele model will lead to slightly changing *LR*s.

   Verify the above and comment.

Exercise S8* Mutation. Theoretical, requires some knowledge of probability

In this exercise, which serves to confirm the answers obtained in Exercise S7, we will fill in some mathematical details related to the mutation models. The data and hypotheses are as in Exercise S7.

a) Show that

$$LR = \frac{P(data \mid H_1)}{P(data \mid H_2)} = \frac{p_{16}(m_{14,17} + m_{15,17}) + p_{17}(m_{14,16} + m_{15,16})}{4 p_{16} p_{17}} \tag{1}$$

where p_{16} is the allele frequency for allele 16 and $m_{14,17}$ is the probability of a mutation from 14 to 17 and so on.

b) For the Equal probability model $m_{ij} = m = R/(n-1)$ where n is the number of alleles. Explain why the formula in Exercise S2 follows from Equation (1). Show that when n=8 and R=0.005, LR=0.0029, as in Exercise S7 b).

c) Consider next the proportional model. By definition,

$$\begin{aligned} m_{ij} &= kp_j \text{ for } i \neq j, \\ m_{ii} &= 1-k(1-p_i). \end{aligned} \tag{1.1}$$

Show that

$$LR = k = \frac{R}{\sum_{i=1}^{n} p_i(1-p_i)} = 0.0063$$

Comment: Exact calculations for the last two mutation models are more technical and are explained in Section A.1.4 of the manual.

Exercise S9. Paternity case with mutation

Load the file ExS9.txt. Consider the hypotheses

- H_1: Alleged father is the father
- H_2: Alleged father is not the father; an unrelated man is the father

a) Show that the $LR=0$.

b) There is one marker where the child and the alleged father do not share an allele. Find this marker.

c) Use model one Probability decreasing with range (stable) for females and males with mutation rate 0.001 and mutation range 0.5 for the marker PENTA_E and calculate *LR*.

d) Assume you are asked to consider the hypotheses H_3: Brother of alleged father is father. Calculate LR (H_1/H_3).

e) *For the discussion*: Is there a 'best' mutation model? Should a mutation model be used routinely for *all* markers?

Exercise S10. Sisters?

We would like to determine whether two girls (called sister1 and sister2 in the left part of Figure 7 below) are sisters (corresponding to hypothesis H_1) or if they are half sisters (corresponding to hypothesis H_2 shown on the right hand side of Figure 7). The alleles are given in Table 1 below. The allele frequencies are 0.1 for systems S1 and S2, and 0.05 for systems S3-S5.

Figure 7. H1 and H2 (right)

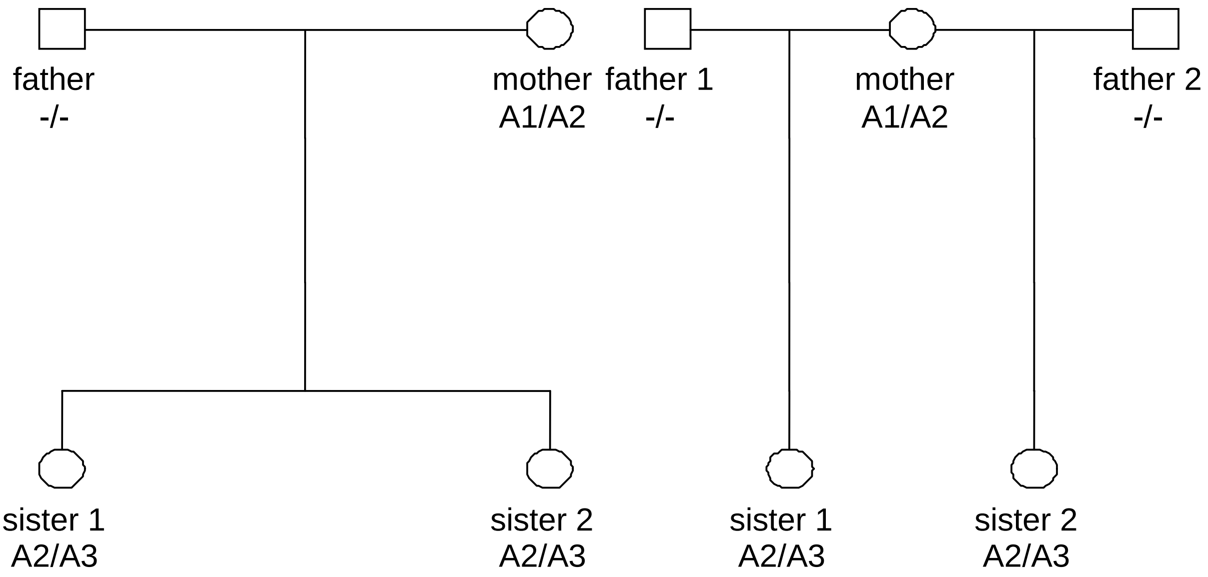


Table 1. Marker data for Exercise S10.

Person	S1	S2	S3	S4	S5
Mother	A_1A_2	A_1A_2	A_2A_3	A_2A_3	A_2A_3
Sister 1	A_2A_3	A_2A_3	A_3A_4	A_3A_4	A_3A_4
Sister 2	A_2A_3	A_2A_3	A_1A_3	A_1A_3	A_1A_3

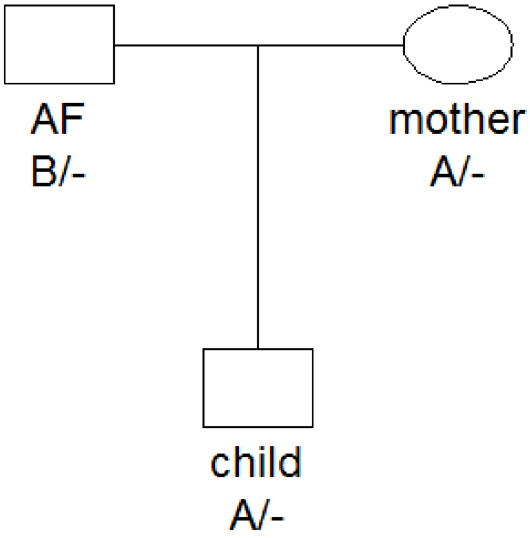
a) What is the *LR* comparing the full sisters-alternative compared to the half-sisters-alternative?

b) For the discussion: The *LR* in this case does not give rise to a clear conclusion. How many, which, further markers should be used?

Exercise S11 Silent allele?

See Figure 5. This is a paternity case where there is suspicion of a silent allele. Include a silent allele frequency of 0.05, and calculate the *LR* using **Familias**. The allele frequencies of A and

Figure 5. Silent allele?



B are both 0.1

Exercise S12* Theta-correction. Theoretical

Verify theoretically the formula in Exercise S1 h).

Exercise S13 Cap paper challenge ("grandson-grandfather continued")

This exercise expands on the previous Exercise S4 by introducing *theta correction*: Calculate the *LR* based on all markers. To save time you can read input from the file ExS4.txt and introduce a theta (kinship-parameter) of 0.02. (Hint: This is done in the pedigree window and in **Familias** 2.0 you should press the Options button.)

Exercise S14 Further input and output options

A practical way to start work with **Familias** is to begin by reading a **Familias** file containing the relevant database. Sometimes it is, however, of interest to read and write databases and case data and this will be the topic below.

a) Read input from the file ExS3.txt.
b) Write the data base from the 'General DNA data' window. Call the output file database.txt
c) Write the case data from the 'Case Related DNA Data'. Call the output file casedata.txt.
d) Open a new, blank, project in **Familias**.
e) Read database.txt
f) Read casedata.txt
g) Define the pedigrees, see Exercise S3, and calculate the *LR* once more.