Exercises: using **familias**

Thore Egeland [1]

[1] Biostatistics
National University Hospital
0027 Oslo, Norway
Email: thore.egeland@basalmed.uio.no

Exercises: using **familias**

Introduction

The present document contains exercises for **familias** and some suggested solutions.
The program is obtainable as freeware from http://www.nr.no/familias. Documentation of the program, several example data files, including those for the exercises of this manual, and solutions to some exercises, are available from the mentioned site or from http://folk.uio.no/thoree/familias. And so it should not be necessary to perform all typing manually.

In some cases, answers are given with ridiculously many digits. This is done since the exercises may also serve to demonstrate that the program calculates correctly. Mutations, kinship and silent alleles are disregarded unless otherwise stated.

Exercise 1
We would like to investigate the relationship between a female (called Female), a man (called Man) and a boy (called Child). There is one locus, called S1, with alleles called A, B and C. The allele frequencies are all 0,05, 0,05 and 0,9. One possible relationship and genotypes of the individuals involved are shown in Figure 1.
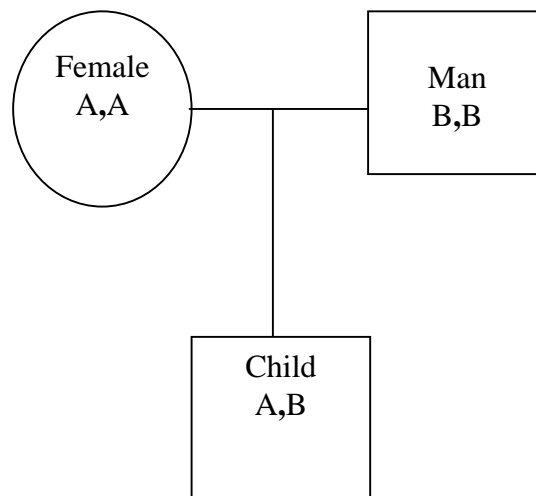

Figure 1: One possible pedigree, Exercise 1

It is believed that the real pedigree is the above suggestion, in which the female Female and the male Man are the parents of the child Child.

a) Enter the data. Remember to specify that the Child is a child!, i.e., this person is not old enough to have children of his own. Generate the possible pedigrees without including any extra persons. How many pedigrees does **familias** generate?
b) Assume a flat prior (default) prior. Describe this prior and show how it is calculated using familias. Determine the posterior probabilities for the different pedigrees. What is the posterior probability of the above suggestion?
c) Determine the paternity index (likelihood ratio) comparing the hypotheses

Exercises: using **familias**

- H_0: The alleged father is the real father (Female is the mother)
- H_1: The alleged father and the child are unrelated (Female is the mother)

Exercise 2

In this exercise you are asked to evaluate the likelihood of possible family relations between a female (called Female) and a male (called Man). Female is believed to be 20 years older than Man. The analysis should be based on 13 loci and assumes for simplicity that all the allele frequencies are 0.05. The DNA-data may be summarized as in Table 1. All allele frequencies are 0,05. For each system there should also be an Extra allele to secure that the frequencies sum to 1. The alternatives are

1. Man and Female are siblings,
2. Man and Female are half siblings,
3. Female is the aunt of Man,
4. Female is the mother of Man,
5. Man and Female are unrelated.

Table 1: The alleles of the female Female **and male** Man **are shown for 13 loci**

Locus	1	2	3	4	5	6	7	8	9	10	11	12	13
Female	ab	ac	bc	ab	ac	ab	ab	ab	ab	ac	ab	ac	aa
Man	ab	ab	ac	ab	bc	bc	ab	bc	bc	bb	ab	bb	ab

a) Is the age difference of any use?
   Read system data from the file AlleleSystemForEx2.txt..
   Read case data (Table) from the file CaseDataForEx2.txt.
b) Draw the five pedigrees (using pen and pencil) and give names to extra persons needed to define the pedigrees.
c) Assume a flat (default) prior. Calculate the posterior probabilities of the 5 alternatives using first all 13 loci and next only the first four loci.
d) What are the posterior probability ratios for the various alternatives when compared to Alternative 5 (the persons are unrelated)?
e) Repeat c) with inbreeding parameter M_1 equal to 0 and then 0.1. What are the probabilities now?

Exercise 3 (inspired by "Interpreting DNA evidence", Evett and Weir, 1998)

A woman, M, gives birth to a child, C. The father of M, AF, is suspected to be the father of C, but there may be other explanations, incestuous or not. Assume that M, C and AF have genotypes (A_1, A_3), (A_1, A_2) and (A_1, A_2), respectively. The allele frequencies are 0.05. Add an Extra allele, with frequency 0,85, so that the allele frequencies sum to 1. We consider five alternative pedigrees:
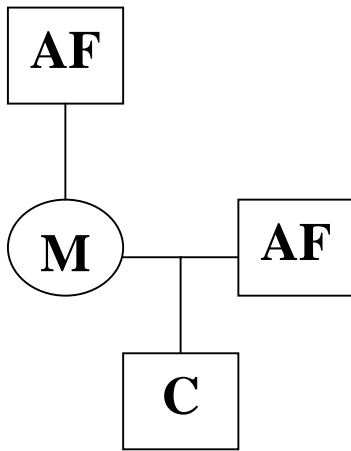
- **Alternative 1:** AF, the father of M, is the father of C.
- **Alternative 2:** The brother of M is the father of C.
- **Alternative 3:** An unrelated man is the father of C.
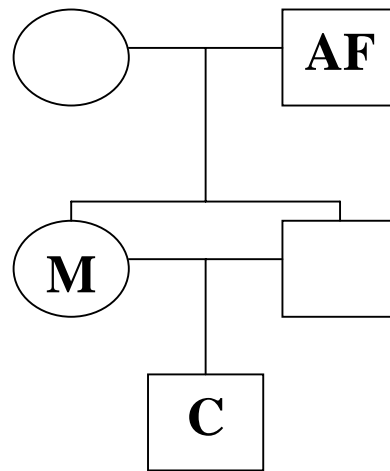
Exercises: using **familias**

- **Alternative 4:** A half-brother of M is the father of C.
- **Alternative 5:** A half-brother of C is the father of C.
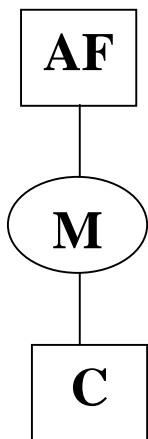
Exercises: using **familias**

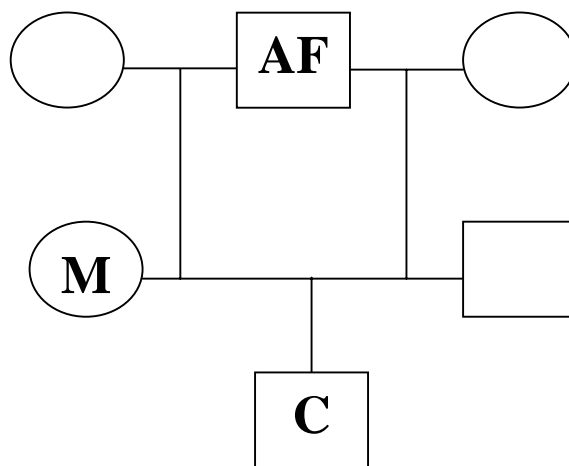**Alternative 1**

Alternative 2

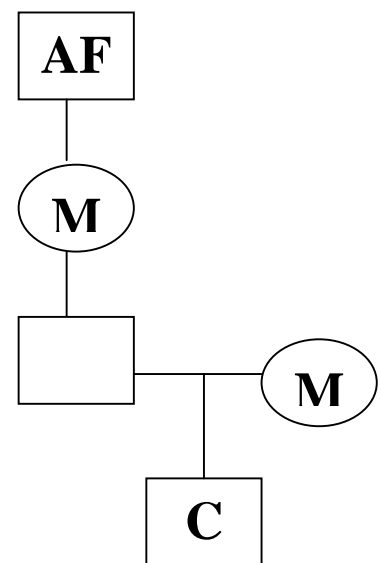**Alternative 3**   **Alternative 4**   **Alternative 5**

Figure 2 Five alternative pedigrees for Exercise 2

a) What are the posterior probabilities for the five different alternatives?

5

Exercises: using **familias**

b) What are the posterior probabilities for the five different alternatives if you alter the inbreeding parameter to zero?
c) Why is the probability in b) equal to 1 for the unrelated alternative?

Exercise 4

In this example there is an alleged father AF which is heterozygous (A, B), and a child CH which is homozygous (C, D). The allele frequencies are as listed in Table 2. The assumed pedigree is illustrated in Figure 3. The cross over the circle indicates that we have no DNA data for the child's mother. The letters in parenthesis indicate the alleles.
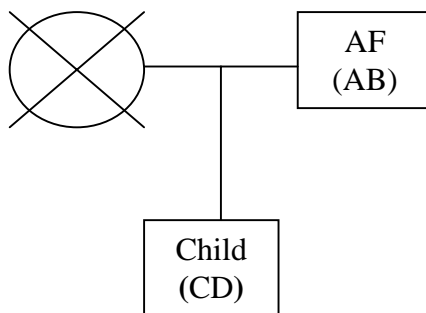

Figure 3: Assumed pedigree.

Table 2: Allele frequencies

A	0,1
B	0,2
C	0,3
D	0,4

a) What would the probability of the supposed pedigree be if the mutation rate were equal to zero?
b) Calculate the paternity index, that is, the likelihood ratio between the hypothesis naming AF as the father of CH and the hypothesis assuming an unknown man is the father of CH. Use mutation rate of 0.005 and and try all models of familias. Use same model for males and females. The parameter 'Mutation range', required for the two first models, should be set to 0,5.

Exercise 5

We would like to determine whether two girls (called sister1 and sister2 below) are sisters (corresponding to hypothesis H_1) or if they are half sisters (corresponding to hypothesis H_2). We know the alleles of their mother as well as the alleles of the sisters for the five allele systems S1-S5. The pedigrees are illustrated in Figure 4 and the DNA data are summarized in

Exercises: using **familias**

Table 3. The allele frequencies are 0.1 for systems S1 and S2, and 0.05 for systems S3-S5. We also assign an extra allele to each system so that the frequencies sum to 1.
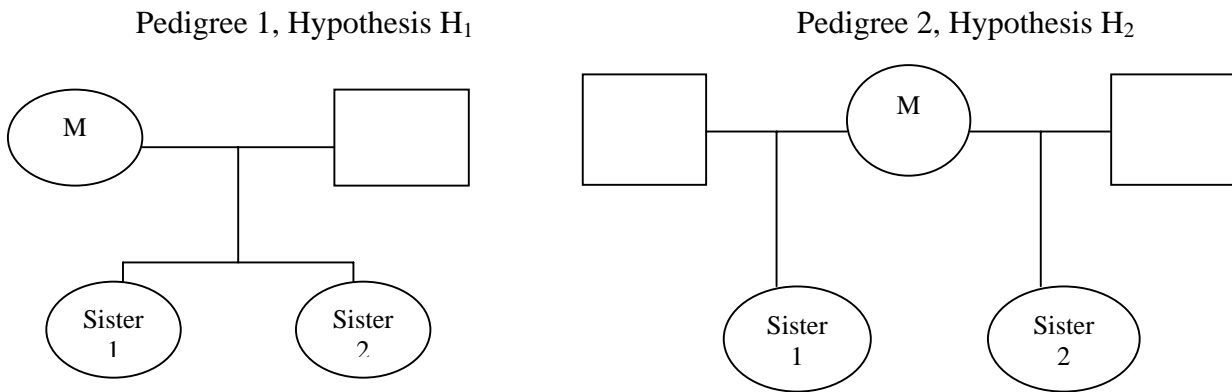
Pedigree 1, Hypothesis H_1 Pedigree 2, Hypothesis H_2

Figure 4: The pedigrees corresponding to the hypotheses of Exercise 5.

Table 3: DNA data

Person	S1	S2	S3	S4	S5
Mother (M)	A_1A_2	A_1A_2	B_2B_3	B_2B_3	B_2B_3
Sister1 (CH1)	A_2A_3	A_2A_3	B_3B_4	B_3B_4	B_3B_4
Sister2 (CH2)	A_2A_3	A_2A_3	B_1B_3	B_1B_3	B_1B_3

a) What does a flat prior mean in this case? Assume a flat prior. What are the posterior probabilities for the two hypotheses?
b) What is the posterior probability ratio for the full sisters-alternative compared to the half-sisters-alternative?

Exercise 6

This is another paternity case, where M is the mother of CH, and where it's claimed that AF is the father of CH. The persons are apparently homozygous A, A and B for M, CH and AF, respectively. The allele frequencies are 0.1. Include a silent allele frequency of 0.05, and an extra allele with frequency 0,75. Calculate the paternity index.

Exercise 7
We have a paternity case with an alleged father AF with genotype (A, B) and a child CH with genotype (C, D). The allele system is given in Table 4 where A=14,B=15, C=16 and D=17.

Exercises: using **familias**

Table 4: The allele system in Exercise 7.

Allele label	A	B	C	D	E	F	G	H
Repeat number	14	15	16	17	18	19	20	21
Count	44	49	127	175	133	58	12	2
Proportion	0.073	0.082	0.212	0.292	0.222	0.097	0.019	0.003

Assume mutation rate 0,005 and mutation range 0,5. Calculate the *transition matrix* and *paternity index* using the two first mutation models of familias
- Prob decreasing with range (stable)
- Prob decreasing with range (equal)

Exercise 9
This exercise deals with kinship and follows the approach suggested by Balding and Nichols (1995) and endorsed by NRCII. We will expand on Exercise 1. There is a man Man, a female Female and a child Child. The alleles in system S1 are respectively BB, AA and AB, and the allele frequencies are $p_1 = p_2 = 0.05$ for A and B (we consider only one locus).

a) Determine the paternity index (PI) comparing the hypotheses

- H_0: The alleged father is the real father (Female is the mother),
- H_1: The alleged father and the child are unrelated (Female is the mother).

Perform the calculations in familias with F_{ST} ('Kinship parameter') 0, 0.001, 0.01 and 0.1
b) Verify the results using the formula

$$PI = \frac{1 + 3F_{ST}}{2F_{ST} + (1 - F_{ST})p_B}$$

c) We will next derive the above formula using the Dirichlet distribution. The crucial formula, which is more general than presently needed, is the recursion

$$E(\tilde{p}_A^{r+1}\tilde{p}_B^{s}\tilde{p}_C^{t}\tilde{p}_D^{u}) = E(\tilde{p}_A^{r}\tilde{p}_B^{s}\tilde{p}_C^{t}\tilde{p}_D^{u}) \frac{rF_{ST} + p_A(1 - F_{ST})}{1 + (r + s + t + u - 1)F_{ST}} \quad (1)$$

where A,B,C and D are arbitrary indices and r,s,t and u non-negative integers.
We need to calculate two likelihoods, the first assuming H_0. The version without kinship, i.e., $p_1^2 p_2^2$ should now be replaced by

$$P(data \mid H_0) = E(\tilde{p}_1^2 \tilde{p}_2^2) \quad (2)$$

The likelihood under H_1 becomes

$$P(data \mid H_1) = E(\tilde{p}_1^2 \tilde{p}_2^3) \quad (3)$$

Exercises: using **familias**

To be able to use equation (1), we make the identifcations $p_1 = p_B$, $p_2 = p_A$, $r=2, s=2, t=u=0$.
Show that Eq. (3) becomes

$$P(data \mid H_1) = E(\tilde{p}_1^2 \tilde{p}_2^2) \; \frac{2F_{ST} + (1 - F_{ST})p_B}{1 + 3F_{ST}}$$

and verify the expression for PI.

Exercise 10.

This exercise discusses the Romanov case documented in Gill et al. (1994) and subsequent
papers. The Romanovs, i.e., Tsar Nicholay II, Tsarina and three of their five children as well
three servants and a doctor were executed in 1918. A grave was found in Ekaterinburg in
1991 and various DNA analyses were performed to verify that the Royal Family had been
found.

a) Use familias to show that
    - there are 4536 possible family relations between one male (called Tsar) and the four
    females (called Tsarina, Child1, Child2 and Child3),
    - there are 2020 possible family relations if incestuous alternatives are discarded,
    - there are 192 possible family relations if we assume Child1, Child2 and Child3 have no
    children of their one. How many are there if the incestuous alternatives are discarded?
b) Consider the last alternative, i.e., the one with 192 alternatives. What is the prior flat
    distribution on this set of pedigrees?
c)   Read the romanov, i.e., the file Ex10.txt data into familias. Verify the below table from
    Egeland et al. (2002) given below, using familias. Verify also that the accepted solution,
    i.e., Tsarina and Tsar are parents of the children comes out as the most likely in all cases.
    (Some of these calculations could take some time.)

Table 6 Results of Exercise 10

Alternative models	Number of pedigrees	Ratio of two most likely alternatives
I "paternity case"	2	2692
II "Child1-3 have no children"	192	63
III "All except incestuous"	2020	63
IV "All (inbreeding par 0.1)"	4536	20
V "All"	4536	1.2

Exercises: using **familias**

Exercise 11 (Database size. familias not used)

We will base calculations below on the following data:

ALLELE	No	Frequency
14	44	0,073
15	49	0,082
16	127	0,212
17	175	0,292
18	133	0,222
19	58	0,097
20	12	0,02
21	2	0,0033

a) A random person is typed. What are the probabilities of a homozygous {14,14} and heterozygous {16,17}?

b) A random sample of 1000 is typed. How many are expected to be {14,14} and {16,17}?

c) Assume the table above is based on a foreign database and we would like to make a database from our country assuming initially the same distribution applies. How many persons need to be typed before we can expect to observe the first {21,21}? How many persons need to be typed to be 95% certain to include a {21,21}?

d) Discuss assumptions underlying the above calculations.

Exercise 12 (requires some knowledge of probability)

In this exercise we will fill in some details related to the mutation models. Throughout we consider data from the system VWA:

Allele	A	B	C	D	E	F	G	H
Rep	14	15	16	17	18	19	20	21
No	44	49	127	175	133	58	12	2
Freq	.073	.082	.212	.292	.222	.097	.020	.003

The case data and hypotheses are:

- Alleged father : A,B
  Child   : C,D
- H_1: Alleged father true father
  H_2: Alleged father and Child unrelated

Assume mutation rate R=0.005.

a) Show that

$$LR = \frac{P(\text{data} \mid H_1)}{P(\text{data} \mid H_2)} = \frac{p_C(m_{AD} + m_{BD}) + p_D(m_{AC} + m_{BC})}{4 p_C p_D} \quad 10$$

Exercises: using **familias**

b) Consider the uniform mutation model . Explain why all mutation rates equal R/3
c) Calculate LR (answer 0.0049)
d) Consider next the proportional model. We introduce some notation

$$p_i, i = 1,...,n.$$

$$\sum_{i=1}^{n} p_i = 1$$

m_{ij} = probability of mutation i to j

Explain why

$$R = kp_j$$

$$k = R / \sum_{i=1}^{n} p_i(1-p_j)$$

e) Show that LR=k=0.0071 for the proportional model.
f) Consider lastly the "range" mutation model. Explain why the following is a reasonable model assumption

$$m_{ij} = k_i r^{|i-j|}$$

Show that

$$k_i = \frac{R(1-r)}{r(2 - r^{|i-1|} - r^{|n-i|})}$$

g) Assume r=0.5 and calculate LR (answer 0.0039).

Exercise 13

Consider one locus S1 and a Man, born 1940, with genotype A,B; a Female, born 1960, with genotype B,C; and a Child with genotype A,C. All allele frequencies are 0.05.

    a) Verify that there are 8 possible pedigrees and that the flat prior produced by familias is 0.125. (Hint: pedigree>probability and tick no 'Case specific DNA data')
    b) Are there any incestuous alternatives?
    c) What are the prior and posterior probabilities of the case where Man is the father of Female who is in turn mother of Child in the following cases given in the table below. (M_I, M_G and M_P denote respectively 'Generations parameter', 'Inbreeding parameter' and 'Promiscuity parameter' in familias terminology).

	$M_I=1,M_G=1,M_P=1$	$M_I=1,M_G=0.5,M_P=1$	$M_I=1,M_G=2,M_P=1$
prior probability			
posterior probability			

    d) Add a locus S2 with similar data and repeat item c)

Answers

Exercise 1

Answers:
   a) **familias** generates 12 different pedigrees (provided Child is entered as a child; otherwise, 22 pedigrees are generated).
   b) The probability of the suggestion in Figure 1 is 0,90497738.
   c) LR = 20.

How to do it:
   ▪ General DNA data: Click Add to enter an allele system. In the new window, enter allele system name S1 and the two alleles A and B, both with frequencies 0,05. Next give allele C with frequency 0,9. Press OK.
   ▪ Persons: Enter the persons Female, Man and Child and their gender. For Child also select Is Child.
   ▪ Case-Related DNA data: Double-click each person to enter his or her DNA data. In the new window, enter the appropriate allele system (use the pull-down menu) and the observed alleles for this person, then press Add and OK. Choosing the same allele twice enters homozygous loci. Repeat the procedure with the other persons.
   ▪ Pedigrees: Press Generate, and do not include any extra persons (just press OK). Press Probability to calculate probabilities for the alternative pedigrees. You can tick on 'Use case-specific DNA data' to choose to calculate the prior or posterior. The posterior corresponds to to the default option 'Yes'. To identify which pedigree is which, double-click them. Press Likelihood to obtain likelihoods instead of probabilities. Find the pedigree that indicates only Female as mother of Child, mark it, and press Scale. Now the likelihood ratio of interest is the number that comes behind the pedigree where in addition Man is indicated as the father of Child.

The analytical solution is derived in Section 2.6 of the manual and gives a paternity index of

$$PI = \frac{1}{p_B} = \frac{1}{0.05} = 20,$$

which is in concordance with the result obtained by familias.

Exercise 2

Answers:
a) The age difference is mostly useful when **familias** *generates* pedigrees; here, it only prevents users from entering relationships that contradict the given ages.
b) The posterior probabilities of the 5 alternatives are listed below . The figures in paranthesis are based only on 4 loci.

   1) Man and Female are siblings: 0.9986147 (0.88345)
   2) Man and Female are half siblings: 0.00069265 (0.01042)

Exercises: using **familias**

      3) Female is the aunt of Man: 0.00069265 (0.01042)
      4) Female is the mother of Man: 0 (0.09567)
      5) Man and Female are unrelated: 0 (0.000038)

c)

d) With M_I equal to 0 and 0,1 the probabilities do not change. This is because there are no incestuous alternatives.

How to do it:

- **General DNA data**: The allele systems are entered manually or more easily by reading the file AlleSystemForEx2.txt.
- **Persons**: Enter the two defined persons, and choose the woman's year-of-birth 20 years prior to the man's (use for example 1940 and 1960). (Actually this information is not used in the calculations of *this* exercise.)
- **Case-Related DNA data**: Enter the data given in Table 1 manually or reading the file CaseDataForEx2.txt.
- **Pedigrees**: The pedigrees must be entered manually. Click Add. Assuming persons Mother and Father have been defined, alternative 1 is given in Figure 2 below and the parent child-relations can be entered. Press Probability to calculate posterior probabilities for the various pedigrees. Do not alter the settings for the priors. To obtain posterior probability ratios, mark the alternative pedigree and press Scale. Press Scale again to return to the original probabilities. To calculate probabilities with altered inbreeding parameter, press Probability and change this parameter in the appearing window. To obtain results only for loci 1-4, press syst, select systems 1-4 and apply. Locate the results.
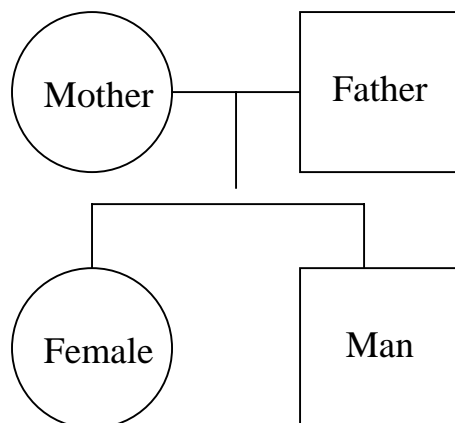

Figure 5: Entering the pedigree of alternative 1 in Exercise 2.

Exercise 3

Answers:
    a) The probabilities for the five different alternatives are listed below

        o **Alternative 1:** 0.4494382

Exercises: using **familias**

- o **Alternative 2:** 0.23595506
- o **Alternative 3:** 0.04494382
- o **Alternative 4:** 0.24719101
- o **Alternative 5:** 0.02247191

b) All the alternatives have posterior probabilities equal to zero except Alternative 3, which has probability 1.

c) Since Alternative 3 is the only non-incestuous alternative, it gets a posterior probability of 1 when all the others get zero prior probabilities.

How to do it:
Enter the allele system with mutation rates zero, and the persons without information of birth dates. The child's gender is arbitrary. Under Known relations, enter AF as the father of M, and M as the mother of C. Under Case-Related DNA data, enter the data given. Construct the alternative pedigrees manually, using Figure 2 as guidance (you could alternatively practice making such figures yourself. Note that it is unnecessary to enter the mothers of M and of the extra male as extra persons: Everybody has a mother, and entering one with a single child in the pedigree is thus he same as not entering any. The other pedigrees are entered similarly.

Exercise 4

Answers:
a) If the mutation rates were zero the supposed pedigree would not be possible, as the child could not have inherited any of his alleles from the alleged father.
b) The paternity indices should be 0,0067, 0,0039, 0,0071 and 0,0049.

How to do it:
Enter the allele system, the known persons (enter CH as a child) and their DNA data as usual. The pedigrees can be constructed either manually or automatically. The alternative pedigree involves no defined relations. Press Probability to obtain posterior probabilities, then Likelihood for the corresponding likelihoods. The likelihood ratio is then constructed by selecting the alternative (unrelated) pedigree and pressing Scale.

Exercise 5

Answers:
a) A flat prior implies that the alternatives are equally likely a priori and so the prior distribution gives 0.5 to both alternatives. The posterior probabilities are 0.791 for them being full sisters, and 0.209 for the half-sister alternative.
b) LR = 3.78.

How to do it:
Enter the allele systems or use the provided file. Enter the persons M, CH1 and CH2, and their DNA data as usual. Optionally, enter the known relations, that is, M as mother of both girls.

Exercises: using **familias**

Construct the alternative pedigree. Note that for the half-sister alternative, no extra males need to be entered. Press Probability to obtain posterior probabilities. Then select the pedigree where they are half-sisters and press Scale to obtain the posterior probability ratio.

Exercise 6

Answers:
The paternity index is 1.36.

How to do it:
Enter the allele system setting the silent allele frequency to 0.05. Enter the persons and their DNA data as usual. Construct the pedigrees manually. Use Probability → Likelihood → Scale to obtain the paternity index (likelihood ratio).

Let $p_A = p_B = 0.1$ and $p_S = 0.05$. According to http://dna-view.com/patform.htm, the answer equivalent is given by

$$PI = \frac{p_s(p_A + p_s)}{(p_A + p_s)^2(p_B + 2p_s) + p_s p_A(p_B + 2p_s)} = \frac{0.05 \cdot 0.15}{0.15^2 \cdot 0.2 + 0.05 \cdot 0.1 \cdot 0.2} = 1.36,$$

which is in exact concordance with the answer obtained by **familias**.

Exercise 7

The mutation matrices are written by familias to the home directory and given the names MutModelFemale.txt and MutModelMale.txt

Exercise 9
a and b)

Kinship	PI
0	20.0
0.001	19.3
0.01	14.8
0.1	5.3

c) Answer mostly given in text

How to do it:

Compared to previous exercises, the kinship parameter need to be provided in the pedigree window after probability has been entered.

Exercises: using **familias**

Exercise 10-12

Answers in problem text or not given.

Exercise 13

b) There is one, the Man has a child with his daughter.
c)

	$M_I=1,M_G=1,M_P=1$	$M_I=1,M_G=0.5,M_P=1$	$M_I=1,M_G=2,M_P=1$
prior probability	0,125	0,0833	0,1667
posterior probability	0,128	0,0845	0,171