Solutions: Exercises not requiring computer

Monterrey Nov 11-13, 2017

Thore Egeland

Exercise 1.

a) This is not a SNP marker as there are more than two different alleles. It is autosomal since males and females have two alleles.

b) 4, namely: 113, 114, 121, 130.

c) 4 and 5 are 113/114.

d) Individual 3 must be 121/130 since alleles these cannot come from the mother.

e) Individual 9 can be 113/113, 113/114 or 114/114 with probabilities $\frac{1}{4}$, $\frac{1}{2}$, $\frac{1}{4}$.

f) $LR = \frac{\frac{1}{2}}{2p_{13}p_{14}} = \frac{1}{4p_{13}p_{14}}$.

g) $LR = \frac{\frac{1}{2}}{1} = \frac{1}{2}$.

Exercise 2.

a) B

b) A/B

c) A/B. She inherited A from her mother.

d) Not possible. Individual 10 must be A or B.

e) 9 and 11 are maternal half sibs. 5 is the uncle of 11. 8 and 10 are unrelated.

Exercise 3.

a) We can write

$$LR = \frac{P(\text{child} \mid \text{mother,father})}{P(\text{child} \mid \text{mother})}$$

Consider first the numerator. The only possible genotype for the child, given the mother and AF, is A/B, and therefore the probability is 1. For the denominator, the father must have passed on the A allele, and therefore the probability is p_A. Hence $LR = 1/p_A = 20$. The standard interpretation is "The data is 20 times more likely assuming AF to be the father compared to the alternative that some unknown man is the father".

b) For the second locus $LR = 1/p_a$ and so the combined likelihood ratio is $(1/p_A) \cdot (1/p_a) = 20 \cdot 10 = 200$.

c) We find

$$P(H_1 \mid \text{data}) = \frac{200}{200+1} \approx 0.995$$
$$P(H_2 \mid \text{data}) = 1 - P(H_1 \mid \text{data}) \approx 0.005.$$

d) Discussed in class.

Exercise 4. Exercises a)-e) were discussed in class, see `http://familias.name/mty/Day1-Part2-TE-kinship-mty.pdf`.

f) First cousins *and* great grandparent – great (!) grandchild. From the above lecture, $\kappa_0 = \frac{3}{4}$ and $\kappa_1 = \frac{1}{4}$ for first cousins. A great grandparent – great grandchild never share two alleles IBD. They share 1 with probability $2\left(\frac{1}{2}\right)^3 = \frac{1}{4}$. Therefore, the relevant IBD probabilities are identical to first cousins.

g) See the presentation mentioned above. If there is no allele sharing or both alleles are shared, $SI > PI$. If precisely one allele is shared $PI > SI$. As an example, consider the case when both are homozygous for the allele a, $0 < p_a < 1$. Then

$$SI - PI = \frac{(1-p_a)^2}{4p_a^2} > 0.$$

Exercise 5. There is no consensus.

Exercise 6.

a) $LR = \kappa_0 = 3/4$

b) Smallest possible: $(3/4)^{16} \approx 0.01$. If both are d/d, then

$$LR = \frac{3}{4} + \frac{1}{4}\frac{1}{0.7} = 31/28 = 1.10714.$$

Exercise 7. Assuming linkage equilibrium (and for simplicity HWE), the probabilities of the mentioned combinations are both $const. \cdot p_{a_1}p_{a_2}p_{b_1}p_{b_2}$ for the same constant $const.$.

⋆**Exercise 8** (A fictitious paternity case)**.**

a) CSF1PO: The mother has transmitted allele 10, so the father must transmit allele 15. This happens with probability 1/2 under H_1 and with probability p_{15} under H_2. Thus, LR is $1/(2p_{15})$.

D7S820: Under H_1, the child's genotype has probability 1/2; under H_2 it has probability $\frac{1}{2}(p_{11} + p_{12})$), so $LR_{1,2} = 1/(p_{11} + p_{12})$.

D19S433: Under H_1 the child's genotype has probability 1; under H_2 it has probability p_{14}, so $LR_{1,2} = 1/p_{14}$.

b) CSF1PO: Now, the brother of the defendant must transmit allele 15. The allele he transmits is equal to allele 15 with probability 1/4, to allele 14 with probability 1/4, and is randomly drawn from the population with probability 1/2. Thus, he transmits allele 15 with probability $1/4+p_{15}/2$. Therefore

$$LR_{3,2} = \frac{P(\text{child} \mid H_3)}{P(\text{child} \mid H_2)} = \frac{\frac{1}{2}(\frac{1}{4} + \frac{1}{2}p_{15})}{\frac{1}{2}p_{15}} = \frac{1 + 2p_{15}}{4p_{15}}.$$

The answer can be checked using the file `Solutions2_18.fam`.

D7S820: The brother must transmit 11 or 12. This is done with probability $1/2 + (p_{11} + p_{12})/2$. The mother's allele is the other one with probability 1/2, so probability of child's genotype under H_3 is $1/4 + (p_{11} + p_{12})/4$. Under H_2 we had probability $\frac{1}{2}(p_{11} + p_{12})$ so

$$LR_{3,2} = \frac{1 + p_{11} + p_{12}}{2(p_{11} + p_{12})}.$$

D19S433: The brother transmits allele 14 to the child with probability $\frac{1}{2}(1 + p_{14})$ so $LR_{3,2} = (1 + p_{14})/(2p_{14})$.

c) Yes: CSF1PO: Since $1/(2p_{15}) = 4.56$ we know p_{15} and this yields $LR_{3,2} = 2.78$.

D7S820: Since $1/(p_{11} + p_{12}) = 2.92$, we know $p_{11} + p_{12}$ and this yields $LR_{3,2} = 1.96$.

D19S433: Since $1/p_{14} = 2.93$, we know p_{14} and this yields $LR_{3,2} = 1.97$.

It may seem surprising that this is possible, but in the $LR_{1,2}$ only the matching allele(s) play a role. If there are two matching alleles such as 11 and 12 for D7S820, they may be viewed as a "11 or 12" allele.

d) Dividing yields $2/(1+2p_{15})$ for CSF1PO, $2/(1+p_{11}+p_{12})$ for D7S820 and $2/(1+p_{14})$ for D19S433.

e) The limit $p_{15} \to 1$ yields $LR_{1,3} = 2/3$ (support for the brother being the father), limit $p_{15} \to 0$ yields $LR_{1,3} = 1$. If allele 15 is very common, we would expect the brother to have more alleles 15 than the defendant and so have better chances to pass an allele 15 to an offspring. If, 15 is very rare, the evidence is neutral.

f) Now $LR_{1,3}$ is always greater than or equal to one, with equality if $p_{14} = 1$, in which case the brother must have two alleles 14 as well and we cannot distinguish them anymore.

g) Same kind of analysis. If $p_{11} + p_{12}$ is close to 1, $LR_{1,3} \approx 1$. If $p_{11} + p_{12}$ is close to 0, $LR_{1,3} \approx 2$ and there is evidence against the defendant.

h) No, since we do not have prior probabilities nor do we know if even more scenarios are possible (e.g., father of defendant is father of child?).

Exercise 9. Consider four individuals labelled 1, 2, 3, 4, with genotypes g_1, g_2, g_3, g_4 obtained on 15 autosomal loci. A sibling test yields $SI(g_1, g_2) = 100$ and also $SI(g_3, g_4) = 100$.

a) No. We would need priors.

b) From slide 8 in Block-II-Part1-KS-ISFG2017.pdf we see that, whenever one allele is shared $((aa), (aX))$ or $((ab), (aX))$, then HSI=SI+1/4 so HSI=1.5

c) No. See Block-II-Part1-KS-ISFG2017.pdf

d)

$$SI(g_1, g_2) = \frac{1}{4} + \frac{1}{2}\frac{1}{4p(1-p)} + \frac{1}{4}\frac{1}{2p(1-p)},$$
$$p = \frac{5}{6} \text{ or } \frac{1}{6}.$$

Exercise 10. Discussed in class.

Exercise 11.

a) Consider the point $(\approx -5, \approx 0.6)$ on the NGM curve of page 4. Then

$$P(\log(LR) > 4 \mid H_D) = P(LR > 10000 \mid H_D) \approx 10^{-5} < 1/10000$$

in agreement with theory. Other points are checked similarly.

b) This was explained in class and is an inevitable effect for low thresholds in paternity cases: LR is the same (0) as long as there is one inconsistency, and if there are no inconsistencies then the LR is almost always large (in this case at least ≈ 1000 for the NGM kit and 10 for the SGMPlus kit). There do not occur LR's smaller than a certain minimum.

c) An example of the first claim is that $P(LR \geq 3.5 \mid PI) \approx 1$, $P(LR \geq 3.5 \mid SI) \approx 0.67$. An example of the second claim claim is that $P(LR \geq 7.5 \mid PI) \approx 0.08$, $P(LR \geq 7.5 \mid SI) \approx 0.1$. If two individuals are heterozygous for a pair of rare alleles, SI could get very large and larger than PI.

d) $x \approx 10,000$. Differences between SI, PI and also NGM and SGMplus are small.

e) TPR is higher for PI for low thresholds. For instance 10^4 gives ≈ 0.92 for PI and ≈ 0.59 for SI. It's the other way around for high thresholds.

Exercise 12. The calculations appear as the first example in Slooten and Egeland, International Journal of Legal Medicine 128 (2014): 415-425.

Exercise 13.

a) This is seen by reading of $TPR = 0.60$.

b)

$$FPR = 10^{-5},\ 10^{-3.5}, \text{ average} = \frac{1}{2}(10^{-5} + 10^{-3.5}) = 0.00016$$

c) This is seen from the curve.

d) $FPR = 10^{-4} < 0.00016$.