GHEP kinship exercise 2022, advanced level

Thore Egeland 2022-03-25

General instructions

This is a multiple-choice test containing 20 questions. For each question you are asked to choose the correct alternative. There may be issues related to e.g., rounding, and so your answer may differ slightly from the correct one. If your answer does not agree exactly with any alternative, you should choose the closest option.

Throughout we make the following simplifying assumptions:

- independent autosomal markers
- no silent alleles, no drop-out, no drop-in
- no mutations
- no deviations from Hardy Weinberg Equilibrium

In some cases, solutions using pen and paper are intended. However, you are free to use any software throughout the exercise.

Part A

This part deals with coefficients quantifying relatedness and inbreeding. The distinction between alleles identical by descent (IBD) and alleles identical by state (IBS) is essential. IBD alleles originate from the same ancestral allele within a given pedigree, while IBS alleles only have the same appearance, but they need not come from the same ancestor. Unrelated individuals may share IBS alleles, but not IBD alleles.

The kinship coefficient φ between two individuals is the probability that a random allele in one individual is IBD to a random allele at the same locus in the other individual, for a random locus. As an example, note that $\varphi = 0.5$ for identical twins. The inbreeding coefficient f of an individual equals the kinship coefficient of the parents.

The relationship between a pair of individuals can be described in more detail by Z, the number of IBD alleles shared by two related, noninbred, individuals. We define the IBD coefficients

$$\kappa_0 = P(Z = 0),\ \kappa_1 = P(Z = 1), \text{and } \kappa_2 = P(Z = 2).$$

Several of the questions below can be answered using the freely available QuickPed , developed by Magnus D Vigeland, and described here.

1) Consider the pedigree in Figure 1 below. We can deduce that

    a. the allele 11 is IBD in children 1, 2 and 3

b. there are no IBD alleles
c. the allele 11 is IBD in children 1 and 3, but there is a different copy (IBS) in child 2
d. the allele 11 is IBD in children 1 and 2
e. the allele 16 in children 1 and 3 is IBD but not IBS

2) For any relationship between two noninbred individuals, it is true that

a. $\kappa_0 + \kappa_1 + \kappa_2 = 1$
b. $\kappa_0 + \kappa_1 + \kappa_2 > 1$
c. $\kappa_0 + \kappa_1 + \kappa_2 < 1$
d. $\kappa_0 + \kappa_1 + \kappa_2 = 0$
e. $\kappa_0 + \kappa_1 + \kappa_2 = 0.5$

3) For any relationship between two noninbred individuals, it is true that

a. $\varphi = \kappa_0 + \kappa_1 + \kappa_2$
b. $\varphi = \kappa_0 + \kappa_2$
c. $\varphi = \kappa_1 + \kappa_2$
d. $\varphi = 2\kappa_0$
e. $\varphi = 0.25\kappa_1 + 0.5\kappa_2$

4) Consider the pedigrees in Figure 2 below, which all have the same IBD coefficients between the hatched individuals. The IBD coefficients $(\kappa_0, \kappa_1, \kappa_2)$ are

a. (0.50, 0.00, 0.50)
b. (0.50, 0.50, 0.00)
c. (0.40, 0.00, 0.60)
d. (0.00, 0.50, 0.50)
e. (0.25, 0.50, 0.25)

5) In each pedigree in Figure 2, the kinship coefficient between the hatched individuals is

a. φ = 0.50
b. φ = 0.25
c. φ = 0.75
d. φ = 1.00
e. φ = 0.125

6) The child A in Figure 3 has inbreeding coefficient

a. f = 0.50
b. f = 0.25
c. f = 0.75
d. f = 1.00
e. f = 0.00

7) Consider the built-in Habsburg pedigree of QuickPed. The inbreeding coefficient of Charles II at the bottom of the pedigree is closest to a child of

   a. third cousins
   b. second cousins
   c. first cousins
   d. half sibs
   e. full sibs

8) Assume non-inbred individuals A and B are half siblings and also first cousins. This relationship is sometimes denoted "3/4 siblings". Their kinship coefficient is then

   a. φ = 0.1875
   b. φ = 0.2225
   c. φ = 0.1575
   d. φ = 0.2525
   e. φ = 0.0250

Part B

This section illustrates the 'Blind search' formula used to calculate LRs for pairwise relationships in several programs. The likelihood function for one marker can be written

$$L(\kappa_0, \kappa_2) = L(H) = \kappa_0 P(\, G \mid Z = 0\,) + (1 - \kappa_0 - \kappa_2)(\, G \mid Z = 1\,) + \kappa_2 P(\, G \mid Z = 2\,).$$

where $G := (g_1, g_2)$ are the genotypes, and κ_0 and κ_2 are the coefficients specified by the hypothesis *H* (note that $\kappa_1 = 1 - \kappa_0 - \kappa_2$). We exemplify for a paternity case, where we assume that the genotype of the alleged father (AF) is 1/1 and the child (CH) is 1/2. The frequencies are *p* and *q = 1-p* for the alleles 1 and 2. The hypothesis *H_1, AF is the biological father of CH*, corresponds to $\kappa_0 = \kappa_2 = 0$, which gives the likelihood

$$L(H_1) = 0 * P(\, G \mid Z = 0\,) + 1 * P(\, G \mid Z = 1\,) + 0 * P(\, G \mid Z = 2\,) = p^2 q.$$

Note that the term $P(\, G \mid Z = 1\,) = p^2 q$ coincides with the likelihood for a parent offspring relationship. The competing hypothesis *H_2, AF and CH are unrelated*, gives $\kappa_0 = 1$ and $\kappa_2 = 0$, and therefore

$$L(H_2) = 1 * P(G \mid Z = 0) + 0 * P(G \mid Z = 1) + 0 * P(G \mid Z = 2) = p^2 2pq.$$

The term $P(\, G \mid Z = 0\,) = p^2 2pq$ corresponds to the unrelated case since there is no IBD sharing. The last term is 0, i.e., $P(\, G \mid Z = 2\,) = 0$, since the marker data is not compatible with two IBD alleles. Anyway, this term also cancels since $\kappa_2 = 0$. The likelihood ratio in this case is

$$LR = \frac{L(H_1)}{L(H_2)} = \frac{p^2 q}{p^2 2pq} = \frac{1}{2p}.$$

In exercises 9-13 we consider the LR comparing the following hypotheses for two samples A and B:

H_1: *A and B are first cousins*

H_2: *A and B are unrelated*

9) Assume A and B both have genotype 14/16 for the marker D2S1358, with allele frequencies $p_{14} = 0.1124$ and p_{16} = 0.2392. The LR, is then

   a. 3.27
   b. 6.53
   c. 1.57
   d. 2.13
   e. 1.00

10) For any marker where A and B share no alleles, the LR is

   a. 1/8
   b. 1/4
   c. 3/4
   d. 1/16
   e. 0

11) What is the smallest possible LR with 22 independent markers?

   a. 1
   b. 0.10
   c. 0.0018
   d. 1/16
   e. 0.012

12) 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 individuals are homozygous for allele d?

   a. 1/16
   b. 0.170
   c. 10.2
   d. 1
   e. 1.107

13) Consider an autosomal marker with alleles a, b, c, and d, with frequencies p_a , p_b , p_c , *and* p_d. What is the LR if A and B are both homozygous for allele d?

   a. $0.75 + 0.25/p_d$
   b. $0.75 + p_d/0.5$
   c. $0.25/p_d$
   d. $1/p_d$
   e. *1.75*

Part C

This part addresses exemplifies blind search. We consider *n = 100* DNA profiles and for each pair of profiles we would like to check if some close family relationship is indicated. Implementation of this search typically uses the blind search formula introduced in the previous part. For some of the below questions you will need software. The input files are

- Norwegian_DB.fam A database of Norwegian allele frequencies for 35 STR markers, in Familias format. If you use Familias, you should remove mutation models, or set mutation rates to 0
- Norwegian_DB.txt The above database in plain text format. (In Familias you may be asked to scale frequencies to sum to 1 when importing. You should answer 'Yes'.)
- pm.txt 100 DNA profiles, labelled V1, …, V100
- ref.txt One DNA profile, labelled ref

14) One hundred profiles give rise to the following number of pairwise comparisons:

   a. 99
   b. 4950
   c. 9900
   d. 10000
   e. 100000

15) Assume $n = 100$ profiles from unrelated people are subjected to a blind search. Let H_1: *full sibs* and H_2: *unrelated*. Assume the false positive rate is *P(LR > 10000 | H_2) = 1/1000000*. If we assume that the comparisons are independent and H_2 is true, the probability that at least one pairwise comparison will give an LR exceeding 10000 is

   a. 1/1000000
   b. 1/100000
   c. 1/10000
   d. 0.0049
   e. 0.05

16) If we search for parent child-relationships, i.e., test H_1: *parent-child* against H_2: *unrelated* with an LR threshold of 10000 using the profiles in the file pm.txt, we find that

   a. there are no parent-child relationships
   b. there is a parent-child relation between V1 and V2, and between V1 and V3, but the analysis doesn't tell the direction of the relationships
   c. V1 is the parent of V2 and V1 is the parent of V3
   d. V11 is the parent of V2 and V10 is the parent of V3
   e. V19 is the parent of V2

17) Consider the following output:

Person 1	Person 2	Relationship	LR
V1	V2	Parent-Child	1.6412901e+015
V1	V2	Siblings	6.8562885e+013

The LR comparing H_1: *Parent-child* to H_2: *(Full) siblings* is

   a. 1.64e+15
   b. 23.9
   c. 6.86e+13
   d. 0.04
   e. 10000

18) (Continuation of the previous exercise.) Assume the priors $P(H_1) = P(H_2) = 0.5$. Then the posterior $P(H_1|data)$ is

   a. 0.96
   b. 0.99
   c. 0.59
   d. 0.04
   e. 0.50

19) The profile in ref.txt most likely comes from

   a. someone unrelated to all profiles in pm.txt
   b. the mother or daughter of V2
   c. a first cousin of V2
   d. a sibling of V3
   e. a half sibling of V3

20) Consider the hypotheses H1, H2 and H3 in Figure 4. The LRs comparing H3 to H2 and H3 to H1 are, respectively,

a. 626014 and 1.8e+15
b. 626014 and 3.6e-010
c. 1.8e+15 and 2.8e+09
d. 1 and 1.8e+15
e. 626014 and 0

Figures

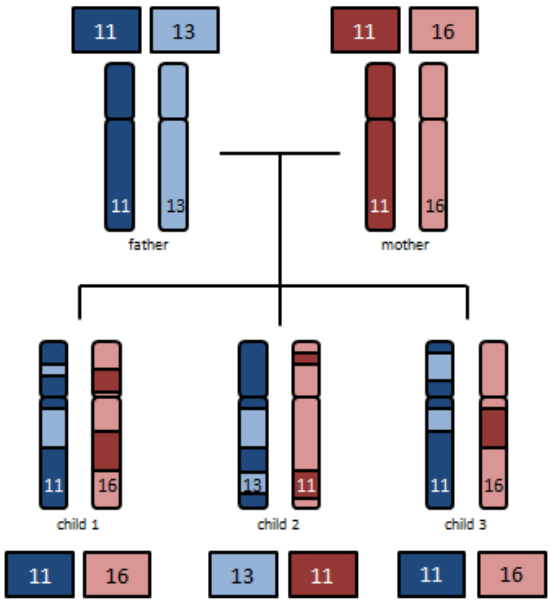

Figure 1 For Question 1. Reproduced from Kling et al., Mass Identifications, Academic Press, 2021.

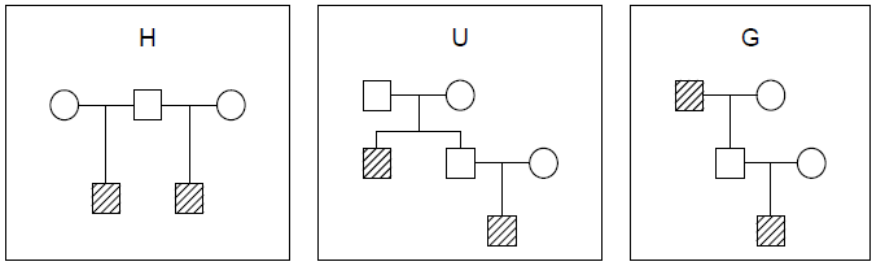

Figure 2 For Questions 4 and 5. Reproduced from Vigeland, Pedigree Analysis in R, Academic Press, 2021.

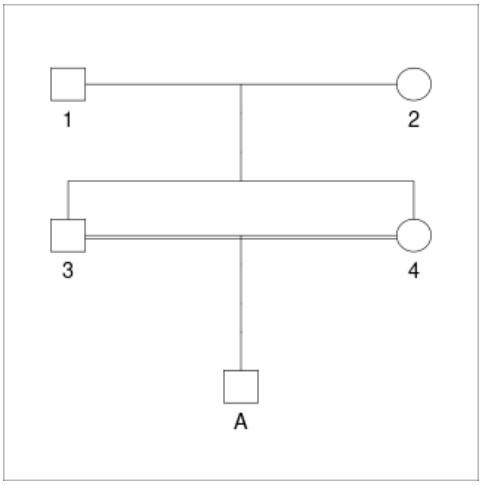

Figure 3 For Question 4.

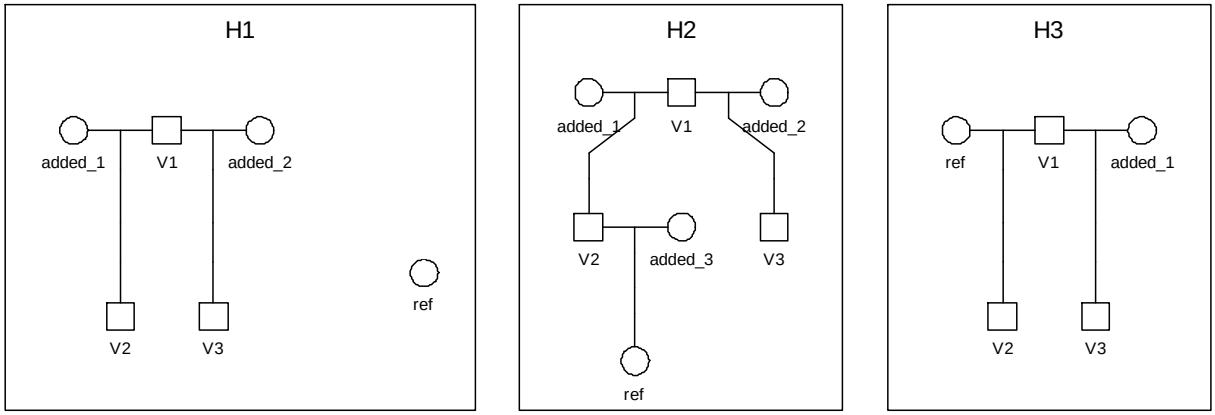

Figure 4 For Question 20.