

## Solutions

1. The lod score at recombination fraction  $\theta$ ,  $Z(\theta)$ , is defined as the logarithm (base 10) of the likelihood ratio  $\frac{L(\theta)}{L(0.5)}$ . Thus,

$$Z(0.2) = \log_{10} \left( \frac{L(0.2)}{L(0.5)} \right) = \log_{10}(0.1/0.01) = \log_{10}(10) = 1.$$

2. Assume that the genetic model is dominant with full penetrance and no phenocopies ( $f = (0, 1, 1)$ ). This model implies that all carriers are affected, but at least one of the parents must be a carrier and both are unaffected. This contradiction implies that the genetic model assumed above is impossible given the observed pedigree and the phenotypes of its members. The recessive model ( $f = (0, 0, 1)$ ) is possible since parents with one copy of the disease allele will be unaffected. On average, 50% of their children will be affected. The remaining two alternatives have reduced penetrance, meaning that parents may be carriers without developing the disease.
3. Under the assumption of a rare disease allele we "know" that the disease genotypes of the parents are (A a) and (a a), respectively, and that those of the two children are (a a) and (A a). Furthermore, the marker genotype of the father can be unambiguously inferred to be (3 4). The two meioses from the mother are impossible to score as recombinant or non-recombinant because she is homozygous at the disease locus, so let us concentrate on the two paternal meioses. To calculate the pedigree likelihood, we condition on the phase of the father and use the law of total probability.
  - Phase I: A and 3 on the same chromosome:
    - The first daughter (3) is recombinant since she has received a and 3 from the father.

- The second daughter (4) is also recombinant since she has received A and 4 from the father.
- Phase II: A and 4 on the same chromosome:
  - The first daughter (3) is non-recombinant since she has received a and 3 from the father.
  - The second daughter (4) is also non-recombinant since she has received A and 4 from the father.

$$L(\theta) = 0.5(\theta^2 + (1 - \theta)^2)$$

$$Z(0) = \log_{10} \left( \frac{L(0)}{L(0.5)} \right) = \log_{10} \left( \frac{0.5}{0.25} \right) = \log_{10}(2) = 0.3.$$

4. Let  $F_{11}$  denote paternal genotype (1 1),  $M_{12}$  maternal genotype (1 2),  $C1_{11}$  genotype of the first child (1 1), and finally  $C2_{12}$  genotype of the second child (1 2). Let, furthermore,  $G$  denote the vector of observed genotypes. Then

$$\begin{aligned}
 P(G) &= P(F_{11} \cap M_{12} \cap C1_{11} \cap C2_{12}) \\
 &= P(C1_{11} \cap C2_{12} \mid F_{11} \cap M_{12})P(F_{11} \cap M_{12}) \\
 &= P(C1_{11} \mid F_{11} \cap M_{12})P(C2_{12} \mid F_{11} \cap M_{12})P(F_{11})P(M_{12}) \\
 &= 0.5 \cdot 0.5 \cdot p^2 \cdot 2p(1 - p) \\
 &= 0.5p^3(1 - p) \\
 &= 0.0512.
 \end{aligned}$$

The definition of conditional probability was used in the second equality, independence between parental genotypes and conditional independence of the genotypes of the children in the third equality, and finally Mendelian segregation and HWE in the fourth equality. Note that children to these parents must have genotypes (1 1) or (1 2) and that these two genotypes are equally likely.

5. The genotypic value of a quantitative trait is, by definition, the average trait value given the genotype.
6. There are (at most)  $n(n+1)/2$  different genotypic values, corresponding to the number of different genotypes for a locus with  $n$  different alleles.

7. The genetic correlation of *any two relatives*,  $X_1$  and  $X_2$ , *sharing two alleles IBD at the QTL* is

$$\frac{C(X_1, X_2 | N = 2)}{V(X)} = \frac{V_A + V_D}{V_A + V_D} = 1.$$

8. An association between  $D$  and  $M_1$  exists if

$$P(DM_1) \neq P(D)P(M_1) = 0.00005.$$

9. To find the probability that an affected individual is homozygous for the disease allele  $A$ , we use Bayes' Theorem:

$$P(AA | \text{Affected}) = \frac{P(AA \cap \text{Affected})}{P(\text{Affected})} = \frac{P(\text{Affected} | AA)P(AA)}{P(\text{Affected})}.$$

The first probability in the numerator  $P(\text{Affected} | AA)$  is the penetrance parameter  $f_2 = 1$ . The second probability  $P(AA)$  is  $p^2 = 0.0001$  under the assumption of HWE. The probability in the denominator can be calculated using the law of total probability:

$$\begin{aligned} P(\text{Affected}) &= \sum_{i=0}^2 P(\text{Affected} | i \text{ disease alleles})P(i \text{ disease alleles}) \\ &= f_0 \cdot (1-p)^2 + f_1 \cdot 2p(1-p) + f_2 \cdot p^2 \\ &= 0.001 \cdot 0.99^2 + 0.001 \cdot 2 \cdot 0.01 \cdot 0.99 + 1 \cdot 0.01^2 \\ &= 0.0010999. \end{aligned}$$

Thus

$$P(AA | \text{Affected}) = \frac{1 \cdot 0.0001}{0.0010999} = 0.091.$$

10. The definition of the NPL-score for perfect marker information is:

$$NPL = \sqrt{\frac{2}{n}}(n_2 - n_0),$$

where  $n$  is the number of affected sib pairs, and  $n_2$  and  $n_0$  the number of pairs sharing 2 and 0 alleles IBD, respectively. Assuming  $n_2 = 0.45n$  and  $n_0 = 0.1n$ , we can expect an NPL-score above 4.1 if

$$NPL = \sqrt{\frac{2}{n}}(0.45n - 0.1n) \geq 4.1.$$

By solving this inequality for  $n$ , we see that  $n \geq 68.6$  affected sib pairs is required, so we recommend a sample of at least 69 ASP-families.

11. Note first that both parents have a 1-allele. Assume that the father passes on his 1-allele to the first daughter only and that the mother passes on her 1-allele to the second daughter only. The genotypes of the sibs will then be (1 3) and (1 2) and the number of alleles shared IBS and IBD 1 and 0, respectively. This is one of two possible solutions. The alternative is to switch the genotypes of the children.
12. Once again we condition on the phase of the affected parent (the father). If the disease allele A and the marker allele 1 are located on the same chromosome then three of the observed meioses are recombinant (those to 3, 4, and 7) and the remaining two are non-recombinant. The opposite is true under the second possible paternal phase. Using the law of total probability we find the pedigree likelihood

$$L(\theta) = 0.5(\theta^3(1 - \theta)^2 + \theta^2(1 - \theta)^3).$$

The lod score is defined as

$$Z(\theta) = \log_{10} \left( \frac{L(\theta)}{L(0.5)} \right)$$

Thus, the lod score at  $\theta = 0.1$  is

$$\begin{aligned} Z(0.1) &= \log_{10} \left( \frac{L(0.1)}{L(0.5)} \right) \\ &= \log_{10} \left( \frac{0.1^2 \cdot 0.9^2 (0.1 + 0.9)}{2 \cdot 0.5^5} \right) \\ &= \log_{10}(16 \cdot 0.1^2 \cdot 0.9^2) \\ &= -0.887. \end{aligned}$$

The likelihood function is an increasing function of  $\theta$  over the interval  $\theta \in [0, 0.5]$ . Thus, the maximum is reached at  $\theta = 0.5$  leading to a maximum lod score of

$$Z(0.5) = \log_{10} \left( \frac{L(0.5)}{L(0.5)} \right) = \log_{10}(1) = 0.$$

13. The three genotypic values are  $X_{11} = 1$ ,  $X_{12} = 3$ , and  $X_{22} = 4$ , where  $X_{ij}$  is the genotypic value for  $A_i A_j$ -individuals. Further, the given population frequency of the  $A_2$ -allele,  $p_2 = P(A_2) = 0.10$ , implies  $p_1 = P(A_1) = 0.90$  and hence, assuming Hardy-Weinberg equilibrium

and random mating, the corresponding genotype probabilities,  $P_{ij}$ , are given by

$$\begin{aligned} P_{11} &= P(A_1A_1) = p_1^2 = 0.81; \\ P_{12} &= P(A_1A_2) = 2p_1p_2 = 0.18; \\ P_{22} &= P(A_2A_2) = p_2^2 = 0.01. \end{aligned}$$

- (a) By definition,  $a = (X_{22} - X_{11})/2$  and  $(1+k)a = X_{12} - X_{11}$ . Hence  $a = 3/2$  and  $k = 1/3$ .
- (b) One possibility is to use

$$V_A = 2p_1p_2(a[1 + k(p_1 - p_2)])^2 = 0.6498.$$

Alternatively, first calculate the two additive allelic effects:

$$\begin{aligned} \alpha_1 &= X_{11}p_1 + X_{12}p_2 - \mu, \\ \alpha_2 &= X_{12}p_1 + X_{22}p_2 - \mu, \end{aligned}$$

where  $\mu = X_{11}P_{11} + X_{12}P_{12} + X_{22}P_{22}$  is the average trait value in the population. Then you will know the so called breeding values  $\Lambda_{ij} = \alpha_i + \alpha_j$  for each of the three genotypes. Finally, since  $V_A$  is by definition the variance of the breeding values in a population and the expected breeding value of a randomly drawn individual is zero,

$$V_A = \Lambda_{11}^2P_{11} + \Lambda_{12}^2P_{12} + \Lambda_{22}^2P_{22}.$$

- (c) From  $V(X) = E(X^2) - (E(X))^2$ , the (total) genetic variance is

$$V(X) = X_{11}^2P_{11} + X_{12}^2P_{12} + X_{22}^2P_{22} - \mu^2 = 0.6579.$$

Alternatively, if you first calculate the breeding values,  $\Lambda_{ij}$  (see (b) above), you might proceed by in addition calculating the dominance deviations,  $\delta_{ij} = X_{ij} - \mu - \Lambda_{ij}$ , for each of the three genotypes. In the next step you can then calculate the dominance genetic variance through  $V_D = \delta_{11}^2P_{11} + \delta_{12}^2P_{12} + \delta_{22}^2P_{22}$ , i.e. the variance of the dominance deviations. Finally,  $V(X) = V_A + V_D$ .

- (d)

$$H^2 = \frac{V(X)}{V(X) + V(e)} = \frac{0.6579}{1.6579} = 0.3968.$$

14. (a) First observe that two ('ordinary') cousins can never share two alleles IBD, i.e.  $P(N = 2) = 0$ . They have one allele each originating from their common grand-parents. The only possibility for these two alleles to be IBD is if the very same grand-parental allele is transmitted to both cousins. If the origin is the common grandfather this probability equals  $(1/2)^3$  and the same probability of course holds for the common grandmother. Hence,  $P(N = 1) = 2(1/2)^3 = 1/4$  and then, of course,  $P(N = 0) = 1 - P(N = 1) - P(N = 2) = 1 - 1/4 - 0 = 3/4$ . Therefore, the expected number of alleles shared IBD by two cousins is  $E(N) = 0 \cdot 0.75 + 1 \cdot 0.25 + 2 \cdot 0 = 0.25$ . Finally, the coefficient of kinship,  $\Theta = E(N)/4 = 1/16 = 0.0625$ .
- (b) By definition the coefficient of fraternity,  $\Delta$ , equals  $P(N = 2)$  which, as already has been observed in (a), equals 0 for two cousins.
- (c) The genetic covariance is

$$2\Theta V_A + \Delta V_D = 2/16 \cdot 1 + 0 \cdot V_D = 1/8 = 0.125.$$

15. (a) The TDT statistic is,

$$\text{TDT} = \frac{(b - c)^2}{b + c},$$

where  $b$  is the number of  $M_1M_2$ -parents transmitting an  $M_1$ -allele to the affected child and, similarly,  $c$  is the number of  $M_1M_2$ -parents who transmits an  $M_2$ -allele. Here the total number of marker heterozygous parents, i.e.  $b + c$ , equals 16 and hence the maximal possible value of the TDT statistic is  $16^2/16 = 16$  corresponding to the two cases where either  $b = 16, c = 0$  or  $b = 0, c = 16$ .

- (b) The proportion of  $M_1$ -alleles in the 17 affected children is

$$p = \frac{a + b}{2 \cdot 17} = \frac{13 + b}{34},$$

where  $a$  is the number of  $M_1M_1$ -parents (=13). Since under the null hypothesis an  $M_1M_2$ -parent is equally likely to transmit either

an  $M_1$  or an  $M_2$  allele, the expected value of  $b$  given  $b + c = 16$  is 8. Hence the expected proportion is

$$p = \frac{13 + 8}{34} = \frac{21}{34} = 0.6176.$$

(c) We have

$$p = \frac{a + b}{2 \cdot 17} = \frac{13 + b}{34} = \frac{1}{2},$$

implying  $b = 4$  (and hence also  $c = 12$ ). Therefore,

$$\text{TDT} = \frac{(b - c)^2}{b + c} = \frac{(4 - 12)^2}{16} = 4.$$