

Write your full name in the top right corner of each page handed in, and use separate pages for solutions of different problems. Good luck!

1. Linkage and association are two important concepts in this course. Define the concepts and discuss differences and similarities. (2p)

2. A genetic epidemiologist who has identified multiple cases of a rare cancer syndrome in a single large pedigree assumes that the disease is caused by a single gene, that the inheritance pattern is autosomal dominant, and that the penetrance is 80% for disease allele carriers. All the individuals in the pedigree were genotyped for a set of 400 evenly spaced polymorphic microsatellite markers. Simulation under the assumption of a tightly linked marker and a rare disease allele ($p=0.0001$) has shown that the maximum possible two-point parametric LOD-score in this family is 4.0, but no LOD-score peak above 1.5 was found in the genomewide scan. Discuss possible reasons why this study was not more successful. (2p)

3. Assume Hardy-Weinberg equilibrium at a disease locus with two possible alleles A (disease related) and a (normal).
 - (a) Which genotypes are possible at the disease locus? (1p)

 - (b) If the relative frequency of the A-allele in the population is p , which are the probabilities of the genotypes at the disease locus? (1p)

 - (c) Assume further that the penetrance increases linearly from 0% to 100% as the number of A-alleles increases from 0 to 2. What is, under this genetic model, the probability of being affected? (1p)

- (d) Assume that you are affected by this disease. For which values of p is it more likely that you are a homozygous than a heterozygous disease allele carrier? (1p)

4. Eight of the ten members of the pedigree shown in Figure 1 have been genotyped at two marker loci.

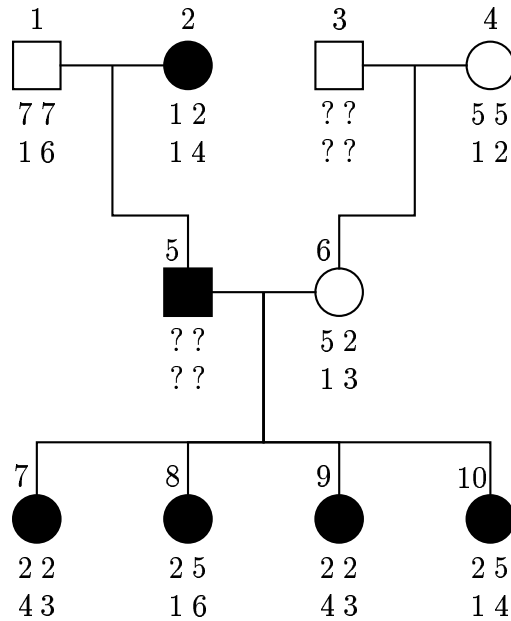


Figure 1: A three-generation pedigree with incomplete genotype data.

- (a) Some of the eight missing alleles can be inferred with 100% certainty. Where possible, fill in the missing alleles. (2p)
- (b) Find the two-marker haplotypes for the couple in the second generation. (1p)
- (c) Count the number of recombinant and nonrecombinant gametes in generation three and use that information to estimate the recombination fraction θ between the two markers. Note that you

are supposed to count recombinants and nonrecombinants with respect to the marker loci - the disease locus has not been used yet in this problem. Hint: all the eight meioses are informative. (2p)

- (d) What is the expected number of meioses (out of the eight from the second to the third generation) that shows no recombination under the null hypothesis of no linkage between the two marker loci?(1p)
- (e) Assume that only one of the eight chromosomes in the top generation has a disease allele (A) and that this allele explains all the occurrences of the disease in the pedigree. Will the maximum LOD-scores be the same for the two markers? A more detailed explanation (than yes or no) is required for 2p. (2p)

5. A single SNP-marker with alleles labelled 1 and 2 was genotyped in a trio consisting of parents and one affected child. Both parents turned out to be heterozygous whereas the child had a (1,1)-genotype.

- (a) Calculate the TDT-statistic for this single trio. (1p)
- (b) What is the minimum number of trios with exactly this marker pattern (heterozygous parents and (1,1)-homozygous child) that will lead to a significant association (at the 1%-level) between the 1-allele and the disease? Hint: You are allowed to assume that the TDT-statistic follows a chi-squared distribution with 1 degree of freedom under the null hypothesis of no association even though the approximation is poor for small samples. (2p)
- (c) In total, 40 trios were genotyped and 70 out of the 80 meioses turned out to be informative. Explain why some meioses are uninformative. (1p)
- (d) 50 meioses showed transmission of the 1-allele but not of the 2-allele whereas the opposite pattern was observed for the remaining 20 informative meioses. Calculate the TDT-statistic and comment

on the result. (1p)

6. Back in the mid 1990s, an Icelandic research group identified a large number of sister-pairs affected by breast cancer by merging data from two registries - the population registry and the cancer registry. At that time, the BRCA1-gene had been identified but not the BRCA2-gene. All the identified cases were screened for mutations in the BRCA1-gene and all sister-pairs with mutations in BRCA1 were excluded from the study. A genomewide linkage analysis pin-pointed a wide region on chromosome 13q where the allele sharing seemed to be higher than expected by chance.

(a) Explain the difference between IBD-sharing and IBS-sharing of alleles at a marker locus. (1p)

(b) Construct an example where the the IBD-count is observable even though none of the parents have been genotyped. Note: the only data available is the marker genotypes of the two sisters. (1p)

(c) Back to the Icelandic study. The IBD-count at the marker D13SX was observable in 50 sister pairs. 22 pairs shared 2 alleles IBD, 20 pairs shared 1 allele IBD and the remaining 8 pairs shared 0 alleles IBD. Calculate the NPL-statistic and comment on the result.(2p)

7. A and B are two microsatellite loci which are linked with the recombination frequency θ .

(a) Consider a case when two parents have the following genotypes $A_1A_2 B_1B_2$ and $A_3A_3 B_3B_3$. From earlier generations it can be deduced that the phase of the double heterozygote is A_1B_1/A_2B_2 . What is the probability of an offspring with the genotype $A_1A_3 B_1B_3$? (1p)

(b) Consider a case when two parents have the following genotypes and phases A_1B_1/A_2B_2 and A_3B_3/A_4B_4 . What is the probability of an offspring with the genotype $A_1A_3 B_1B_4$? (1p)

- (c) How many informative gametes can be scored in a) and b) respectively? (1p)
8. Q_1 and Q_2 are two alleles at a QTL. The allele frequencies are p and q respectively. Consider a sib-pair (full sibs). What is the probability that they are homozygous for the same allele if they are
- (a) (i) IBD = 0 and (ii) IBD = 2 at the QTL? (2p)
- (b) What is the probability that they are homozygous for different alleles if they are (i) IBD = 0 and (ii) IBD = 2 at the QTL? (2p)
9. In a QTL-analysis of blood-pressure using sib-pairs the following result was found for a marker locus. Siblings with IBD=0: 2.1, 4.2, 5.8, and 7.1. Siblings with IBD=1: 2.2, 2.3, 3.5, 3.8, 4.0, 4.1 and 4.4. Siblings with IBD=2: 2.1, 2.2 and 2.4. The figures are the square of the difference between the siblings.
- (a) Draw a rough diagram that illustrates the results above. (1p)
- (b) Why are approximately 50% of the observations IBD=1? (1p)
- (c) Does the marker appear to be linked to a QTL? Motivate your answer. (1p)
- (d) Give two reasons why the 'base-line' appears to be clearly above 0. (1p)
- (e) Why is the shape of the data in the diagram a triangle rather than a cluster along a line? (1p)
10. A woman married a man and together they had a son. She then divorced that man and married his father and had another son with him. Draw a pedigree over these individuals. What is the coefficient of relatedness and coefficient of identity ($P(\text{IBD}=2)$) for these two 'brothers'.

(True story.) (3p)

11. In a population the following gamete frequencies were observed with respect to two loci. $P_{11} = P(A_1B_1) = 0.8$. $P_{12} = 0.05$, $P_{22} = 0.15$.

(a) Calculate the allele frequencies. (1p)

(b) Calculate the level of linkage disequilibrium (D). (1p)

(c) Ten generations later D was found to be 0.02. Estimate the recombination frequency. (1p)