

Exam Forensic Probability and Statistics

January 25 2024, 9.45 - 12.45

You can score 38 points. Grade = $\max\{(2 + \text{number of points}) \times 10/40, 1\}$.

Question 1. In December 1996 Sally Clark's firstborn son Christopher was found dead in his bed, aged 2 1/2 months. He died when Sally was alone in the house with him. In January 1998 her second child Harry also died, aged 2 months, in similar circumstances. Soon afterwards Sally Clark was accused of having murdered her two children by smothering.

Apart from the possibility of a double murder, the possibility of a double Sudden Infant Death Syndrome (SIDS) was considered, and the following hypotheses were formulated.

H_1 : Sally Clark murdered her two children;

H_2 : The two children died of SIDS.

(a) (2 points) Denoting the evidence of two dead children by E , explain why

$$LR_{H_1, H_2}(E) = 1.$$

Now let H_1^c be the hypothesis that Sally Clark did not murder her two children. We write μ for the prior probability of a double murder, and δ for the prior probability of a double SIDS. We assume that the circumstances of the two dead children were such that the only possible explanations are double murder or double SIDS.

(b) (3 points) Show that

$$LR_{H_1, H_1^c}(E) = \frac{1 - \mu}{\delta}.$$

(c) (3 points) The likelihood ratio in (b) is much larger than the one in (a). It therefore seems to be matter which pair of hypotheses is used in court. Comment on this.

Question 2. When forensic DNA profiles of traces are interpreted, sometimes errors are made in the following sense. Let C be an unknown criminal, who is the donor of a stain. We let G_C^{lab} be the DNA profile attributed to C by the forensic laboratory examining the stain. If $G_C^{lab} = g', G_C = g$ for $g' \neq g$, then an error has been made. We assume that $P(G_C^{lab} = g \mid G_C = g)$ is independent of the profile g , and that $P(G_C^{lab} = g) = P(G_C = g)$. Furthermore we assume that errors are made independently of whether $S = C$.

Now let $E = \{G_C^{lab} = g, G_S = g\}$ be the event that the profiles registered for suspect S and criminal C are both equal to g . Let $Z = 0$ be the indicator variable for an error: $Z = 0$ if $G_C^{lab} = G_C$ and $Z = 1$ otherwise.

(a) (4 points) Show that the likelihood ratio in favour of $Z = 0$ versus $Z = 1$ is equal to

$$1 + \frac{\delta}{p_g(1 - \delta)},$$

where $\delta = P(S = C)$ and p_g is the population frequency of g (you may assume that DNA profiles of different persons are independent).

(b) (3 points) Explain the limiting values of the likelihood ratio in (a) when $\delta \downarrow 0$ and when $\delta \uparrow 1$.

(c) (3 points) Derive the posterior probability $P(Z = 0 | E)$ and show that this probability tends to 1 if $p_g \rightarrow 0$. Can you explain this?

Question 3. (4 points) In the population-genetic approach to the formula for the θ -correction, we have looked at the probability that alleles in the population are identical by descent (IBD). When we follow the lineages of the alleles back in time, the “waiting times” for alleles to either coalesce or to arise in the population as a result of immigration/mutation, are exponentially distributed. Writing N for the population size, we assumed that the waiting time for an immigration/mutation event is exponentially distributed with parameter $\mu/(2N)$, whereas for any pair of lineages, the waiting time for coalescence is exponential with parameter $1/N$. For any two alleles, the probability of coalescence before mutation/immigration is then given by

$$\frac{1/N}{1/N + \mu/N}.$$

Consider four alleles. Compute the probability that they are all IBD.

Question 4. Consider the pedigree in Figure 1.

(a) (2 points) Identify the inbred individuals, if any.

(b) (3 points) Compute $\theta_{X,Y}$ for individuals $X = 5, Y = 6$ in this pedigree.

(c) (3 points) Consider a locus with allele frequencies $p_a = 0.2, p_b = 0.8$. Suppose that individual 5 has genotype (a, a) on this locus. Give the probability distribution for the genotype of individual 6 on this locus.

Question 5. We carry out a familial search to identify the father F of an unknown offender in a database $\mathcal{D}$ consisting of $N = 500,000$ persons.

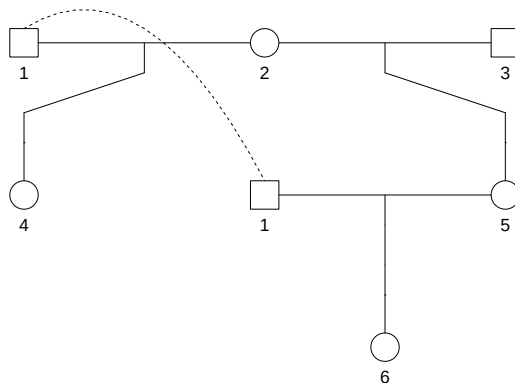


Figure 1: Pedigree for Question 4.

We estimate that $P(F \in \mathcal{D}) = 1/11$. Let E denote the event that for the individuals in $\mathcal{D}$ the likelihood ratios in favour of being F , versus being unrelated to F , are $r_1 = 1,000,000, r_2 = 500,000$ and $\sum_{i=3}^N r_i = 500,000$.

(a) (2 points) Calculate $P(F \in \mathcal{D} \mid E)$ and $P(F = 1 \mid E)$.

(b) (3 points) If it would now become known that $F \neq 2$, what would be the impact on the probabilities that $F \in \mathcal{D}$ and that $F = 1$?

Question 6. (3 points) Show that

$$\frac{P(H_1 \mid LR(E) \geq t)}{P(H_2 \mid LR(E) \geq t)} = E(LR \mid LR \geq t, H_2) \frac{P(H_1)}{P(H_2)}.$$