BM5163 Bayesian Inference in Bioengineering

Exam 1

Instructions

1. This exam is open notes where you can use any hand-written material. Photocopies/prints/books/electronic devices are not permitted to be used.
2. You are expected to work on these problems on your own. **Any reasonable signs of 'copying/plagiarism' will attract penalties.**

Questions

1. A hospital is studying a progressive neuroinflammatory disorder that develops through an intermediate biological state. Individuals first develop a latent inflammatory state (I^+), which can (in some) develop into full-blown disease (D^+). A person does not develop this disease without going through the latent state. The sequence can be denoted as $(I^-, D^-) \to (I^+, D^-) \to (I^+, D^+)$. A blood biomarker test detects active disease (I^+, D^+) but may occasionally give positive results in response to inflammation even in the absence of full disease, that is, in case of (I^+, D^-). In the population 10% individuals develop the latent inflammatory state, and among these, 40% develop the disease. However, individuals without inflammation never develop the disease. The biomarker test detects the full disease with a sensitivity of 0.90. However, it also produces false positives at a rate of 0.30 when inflammation is present without the disease, and at a rate of 0.02 when inflammation is absent. A randomly selected individual from the population undergoes three independent tests, of which two are positive, and one is negative (order unknown). Assume conditional independence of test outcomes, and answer the following

   There are three states in the system (I^-, D^-), (I^+, D^-), and (I^+, D^+). Let's call them S, L, and P, respectively.

   (a) Compute the prior probability of disease for this person. **(5)**

   Given that 10% of individuals develop the inflammatory state, and among these, 40% develop the disease, we have:

   $$P(I^+) = 0.10, \quad P(D^+ \mid I^+) = 0.40$$

   Since disease occurs only through the inflammatory state,

   $$P(D^+) = P(I^+) \cdot P(D^+ \mid I^+)$$

   $$P(D^+) = 0.10 \times 0.40 = 0.04$$

   Note: here we have $P(I^-) \cdot P(D^+ \mid I^-) = 0$

   (b) Compute the posterior probability of disease for this person. **(10)**

   As mentioned earlier, we have

   $$P = (I^+, D^+), \quad L = (I^+, D^-), \quad S = (I^-, D^-)$$

   These give

   $$P(P) = 0.10 \times 0.40 = 0.04, \quad P(L) = 0.10 \times 0.60 = 0.06, \quad P(S) = 0.90$$

   From this we get

   $$P(+ \mid P) = 0.90, \quad P(+ \mid L) = 0.30, \quad P(+ \mid S) = 0.02$$

   Now we calculate the likelihood of two positive and one negative test

   $$P(2+, 1- \mid S) = \binom{3}{2} p^2(1-p) = 3p^2(1-p) = 3(0.02)^2(0.98) \approx 0.001176,$$
   $$P(2+, 1- \mid P) = 3(0.9)^2(0.1) = 0.243,$$
   $$P(2+, 1- \mid L) = 3(0.3)^2(0.7) = 0.189$$

   Therefore,

   $$P(\text{data}) = 0.243 \times 0.04 + 0.189 \times 0.06 + 0.001176 \times 0.90 == 0.0221184$$

Using Bayes theorem, we get

$$P(P \mid \text{data}) = \frac{0.243 \times 0.04}{0.0221184} = \frac{0.00972}{0.0221184} \approx 0.439$$

(c) Compute the posterior probability that the person is in the inflammatory state but does not yet have disease. **(10)**
Using Bayes' theorem

$$P(L \mid \text{data}) = \frac{P(\text{data} \mid L)\, P(L)}{P(\text{data})}$$

Substituting the known values gives

$$P(\text{data} \mid L) = 0.189, \quad P(L) = 0.06, \quad P(\text{data}) = 0.0221184$$

This gives

$$P(L \mid \text{data}) = \frac{0.189 \times 0.06}{0.0221184} = \frac{0.01134}{0.0221184} \approx 0.513$$

(d) Redo parts (b) and (c), if it is surely known that the person has inflammation. **(15)**
Since we know that the person has an infection, the updated prior probabilities are

$$P(P \mid I^+) = 0.40, \quad P(L \mid I^+) = 0.60$$

Likelihood of two positives and one negative is

$$P(2+, 1- \mid P) = 0.243, \quad P(2+, 1- \mid L) = 0.189$$

Total probability of the data given I^+

$$P(\text{data} \mid I^+) = 0.243 \times 0.40 + 0.189 \times 0.60 = 0.0972 + 0.1134 = 0.2106$$

Posterior probability of disease

$$P(P \mid \text{data}, I^+) = \frac{0.243 \times 0.40}{0.2106} = \frac{0.0972}{0.2106} \approx 0.462$$

Posterior probability of inflammation without disease

$$P(L \mid \text{data}, I^+) = \frac{0.189 \times 0.60}{0.2106} = \frac{0.1134}{0.2106} \approx 0.538$$

2. Suppose one student misses this exam for some genuine reason. Once all the exams are over at the end of the course, can you design a systematic approach to award the marks to the student for the missed exam? You can assume that only one student has missed one exam. All other exams were attended by all the students. Give all the details of your calculation. **(30)**

   Note: This is not a question to assess your creative writing or literary capabilities. So avoid giving a verbose answer.

   Let X denote the score of the student in the missed exam. Let the PDF of scores of all other students in this exam be $p(X)$, which acts as a prior for us. This distribution also reflects the overall difficulty and grading pattern of the exam. In the absence of any other information, we can use this to estimate the missing score, which will turn out to be the class average $\Rightarrow$ the student will be rewarded with marks equivalent to the class average.
   Since we have information about the student's performance in other exams, we can do better. Let the student's scores in other exams be $Y_1, Y_2, \ldots, Y_n$. From historical data, model the conditional distribution $p(Y_1, Y_2, \ldots, Y_n \mid X)$.
   Now we are set to use Bayes theorem to get

$$p(X \mid Y_1, \ldots, Y_n) \propto p(Y_1, \ldots, Y_n \mid X)\, p(X)$$

   Therefore, the awarded score can be chosen as the estimate of X, that is

$$\hat{X} = \mathbb{E}[X \mid Y_1, \ldots, Y_n]$$