Statistics for the Doctoral School in Biomolecular Sciences
Academic year 2014-2015

Exercises after Lecture 2

Bayes' formula

EXERCISE 2.1 A diagnostic test has sensibility ($\mathrm{P}(T^+|I)$) of 90% and specificity ($\mathrm{P}(T^-|H)$) of 95%.
[Legend: I = ill, H = healthy, T^+ = positive test, T^- = negative test]

(a) If the prevalence of the disease in the population is 1/200, compute the probability that a randomly chosen individual is ill, knowing that the test has been positive.

(b) Which would be the prevalence of the disease in the population if the probability that a test on a randomly chosen individual is positive were 20%?

EXERCISE 2.2 A diagnostic test has sensibility ($\mathrm{P}(T^+|I)$) of 90% and specificity ($\mathrm{P}(T^-|H)$) of 70%. Half of the tests in a mass screening yield, positive result. Which is the probability that an individual is ill if the test is positive?

EXERCISE 2.3 In a famous trial, Sally Clark was convicted of murdering her two children, who had both suddenly died when they were few months old. At the trial, a physician testified that the probability of a child without apparent ailments dying of SIDS (Sudden Infant Death Syndrome) is about 1 in $8,500$. On this basis, it was estimated that the probability that both children would have died of SIDS (the only alternative considered to them having being murdered) was extremely low.

- How can the probability that both children would have died of SIDS can be computed on the basis of the information presented? *[As far as I understand, this can be done only using an extra assumption, as was actually done at the trial]*. Is the assumption reasonable?
- It has been argued that the probability really relevant for the trial was the probability that Sally Clark was guilty, given that her two children had died and that the only alternatives are SIDS or murder. State then the problem in a Bayesian fashion, writing down the resulting formula and the information needed. It may of interest knowing that the probability that a UK baby is murdered has been estimated to $1.1 \cdot 10^{-5}$.

EXERCISE 2.4 On the basis of the experience, I deem that the probability p that an orange taken from a chest is rotten follows a distribution[1] with density $f(x) = x^{-0.5}(1-x)^{8.5}/B$ where $B(a,b) = \int_0^1 x^{-0.5}(1-x)^{8.5}\,dx$ is a constant such that f is a density, i.e. $\int_0^1 f(x)\,dx$.

I choose at random 5 oranges in the chest, and I find 1 rotten. Following the Bayesian framework, what would be now the probability distribution of p? [*It is enough writing down the ingredients of the new formula without performing any computation but the obvious ones. If one has the time and access to a pc, it may be interesting completing the computations and drawing the density functions before and after the experiment.*]

[1]this is named a Beta distribution