Computational Statistics

Exercise sheet 2

To be handed in for marking until tutorial of 3/05/2017

Nelson Lima and Luca Amendola, ITP, Heidelberg

3/05/2017

www.thphys.uni-heidelberg.de/~amendola/teaching.html

Problem 1 - Bayes Theorem [5 points]

Imagine you meet a guy who claims to know that the casino in Frankfurt has two sorts of slot machines: one that pays out 10% of the time, and one that pays out 20% of the time (note these numbers may not be very realistic). The two types of machines are colored blue and red, respectively. The only problem is, the guy is so drunk he can not quite remember which color corresponds to which kind of machine.

Next day you go to the casino to find out more. You see a red and a blue machine side by side. You toss a coin to decide which machine to try first; based on this you then put the coin into the red machine. It does not pay out. How should you update your estimate of the probability that this is the machine you are interested in? What if it had paid out - what would your new estimate be then?

Problem 2 - Bayes Theorem [5 points]

You go to see the doctor about an ingrowing toenail. The doctor selects you at random to have a blood test for swine flu, which for the purposes of this exercise we will say is currently suspected to affect 1 in 10,000 people in Germany. The test is 99% accurate, in the sense that the probability of a false positive is 1%. The probability of a false negative is zero. You test positive. What is the new probability that you have swine flu? Now imagine that you went to a friend‘s wedding in Brazil recently, and (for the purposes of this exercise) it is know that 1 in 200 people who visited Brazil recently come back with swine flu. Given the same test result as above, what should your revised estimate be for the probability you have the disease?

Problem 3 - Average distance [10 points]

In this exercise, you are supposed to find the average distance between two uniformly distributed points within a rectangle of height 60 cm and width 50 cm. Design your own code to compute this as efficiently as possible, and compare your result to the exact analytical answer given by **[typo in the analytical formula has been corrected]**

$$\overline{d}_{1,2} = \frac{1}{15}\left[a\lambda^2 + \frac{b}{\lambda^2} + d\left(3 - \lambda^2 - \frac{1}{\lambda^2}\right) + \frac{5}{2}\left(\frac{b}{\lambda}\log\frac{a+d}{b} + a\lambda\log\frac{b+d}{a}\right)\right]. \tag{1}$$

This result was derived for a rectangle of sizes $a \geq b$, with $\lambda = a/b$ and $d = \sqrt{a^2 + b^2}$.

Now, fixing one of the sides of the rectangle to 1, plot $\overline{d}_{1,2}/d$ for λ between 0.01 and 100, using a logarithmic scale on the x-axis. What can you say about this plot?

For this exercise, you are supposed to email the tutor your own (working!!) R code and the results obtained from running it, such as the plot that is asked and a screenshot of results on the console (for example).