Computational Statistics

Exercise sheet 1

Nelson Lima and Luca Amendola, ITP, Heidelberg

26/04/2017

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

Problem 1

This first problem will get you familiar with the R software. The idea is to reproduce the survey of $N = 120$ people on the months of their birthdays that Luca describes in the first pages of his lecture notes.

Hence, considering the available months in one calendar year, start by drawing 120 uniformly distributed answers. Then, compute with your own code the average and the standard deviation of your distribution. Get the same results using the built-in functions of R. Lastly, draw the survey's histogram with bars equal to the number of months p_{j} with n_{j} people, as in Fig. 1.1 of the lecture notes.

Now, expand the survey by repeating it 50, 200 and 1000 times. Get the same results as before, but now also plot the histogram for the frequency of months (normalized to unity). What does this behave with the number of experiments?

Problem 2 - The Birthday "Paradox"

You are supposed to verify numerically the so-called "birthday paradox", which we describe below. In the first place, you should verify if the theoretical prediction as a number of N people in the room is a good approximation to the exact result that can be obtained by combinatorics. This should be done with a plot (up to N = 50). For what value of N does the probability approach 50%?

As a bonus, you can try something more complex. Perform n (this should be a sufficiently large number) simulations with 1 to 50 people in a room. Compute the probability of matching birthdays for the aggregate of simulations as a number of people in the room, and see if you recover the previous results!

Describing the birthday "Paradox"

Let us estimate the probability that in N random people there are at least two with the same birthday. A person B has the same birthday of person A only once in 365. Then $P(coinc., N = 2) = 1/365$ and the probability of non-coinc. is $P(non - coinc., N = 2) = 1 - 1/365 = 364/365$. Let's add a third person. His/her birthday will not coincide with the other two 363 times over 365. The joint probability that the 3 birthdays does *not* coincide is then

$$P(non-coinc., N=3) = \frac{364}{365}\frac{363}{365} \tag{1}$$

It is clear then that for N persons we have

$$P(non-coinc., N) = \frac{365}{365}\frac{364}{365}\frac{363}{365}...\frac{365-N+1}{365} \tag{2}$$

We can now use

$$e^{-x} \approx 1-x \tag{3}$$

to write

$$\frac{365-N+1}{365} = 1 - \frac{N-1}{365} \approx e^{-(N-1)/365}$$

and therefore

$$P(non-coinc, N) = e^{-1/365}e^{-2/365}e^{-3/365}...e^{-(N-1)/365} = e^{-\frac{N(N-1)}{2}\frac{1}{365}} \tag{4}$$

Finally, the probability of having at least one concidence must be the complement to unity to this, i.e.

$$P(coinc, N) = 1 - e^{-\frac{N(N-1)}{2}\frac{1}{365}} \approx 1 - e^{-\frac{N^2}{730}} \tag{5}$$

.