INSTITUT FÜR THEORETISCHE PHYSIK HEIDELBERG
PROF. DR. LUCA AMENDOLA
COMPUTATIONAL STATISTICS – SUMMER SEMESTER 2022 –

Tutors:

- Group 1: Ziyang Zheng
- Group 2: Nico Lorenz
- Group 3: Ziad Sakr
- Group 4: Adrian Schirra

Please, send (in pdf) or hand in the solution to this exercise sheet to your tutor, following their instructions, and carefully respecting the delivery date shown below. All exercise sheets will be graded. You can solve them individually or in pairs (with another student of the same tutorial group). In the latter case deliver please only one document with the solutions. Write the name of the file in the following format: X.Name1.Surname1 or X.Name1.Surname1-Name2.Surname2.pdf, with X being the number of the corresponding exercise sheet.

Exercise sheet 1

RECEIVED: Apr. 25 - DELIVERY: May. 2, 11:00 am

Exercise 1.1: Moments under (linear) transformations

The nth moment $\langle x^n \rangle$ of a distribution $p(x)\mathrm{d}x$ is defined as

$$\langle x^n \rangle = \int \mathrm{d}x\, p(x) x^n. \tag{1}$$

1. What is the transformation property of the moment $\langle x^n \rangle$ of a random variable x with the distribution $p(x)\mathrm{d}x$ under a linear transformation

$$x \to \alpha x + \beta? \tag{2}$$

(2pt)

2. What is the transformation of the first moment $\langle x \rangle$ for a nonlinear transformation of the form $x \to \alpha x^2$? (2pt)

3. Which would be the general desirable properties of a variable change for continuous distributions?(2pt)

4. For an arbitrary PDF $p(x)$, are the following two expressions always true ?

$$1 = \int_0^\infty \mathrm{d}x\, p(x) \tag{3}$$

$$p(x) \leq 1 \tag{4}$$

(2pt)

Exercise 1.2: Python Exercise

We now want to continue with a small Python exercise (you can use any other language), in which you will check on the properties of a PDF :

1. We want to start by the first example of the lecture where we compared the mean number birthdays of a month out of 120 people. Open a Python file and create a list of randomly generated birthdays. To do so import the python numpy module and use the command $numpy.random.uniform(0, 12, n_{data})$.
   Then we need to calculate (a) the histogram of this dataset. For that matter, create a list of bins by range(12) which will give you histogram bins for each month of the year. Now we need to take the histogram simply by $numpy.histogram(uniform, bins = range(12))[0]$ (2pt)

2. Now we want to run this experiment N_{exp} times. Wrap part (a) in a loop and run it N_{exp} times and compute the mean for each experiment. Plot a histogram of all the means. for $n_{data} = 120$ and $N_{exp} = 10, 50, 100, 1000, 100000$ (2pt)

Exercise 1.3: Python Exercise 2

You are supposed to verify numerically the so-called "birthday paradox", see below. To be more specific, you should verify the theoretical prediction (if it is a good approximation) as a function of the number N of people. The results should be shown with plots.

The birthday "paradox"

Let us estimate the probability that in N random people there are at least two with the same birthday. The 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 e$P(non - coinc., N = 2) = 1 - 1/365 = 364/365$. 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

$$P(non - coinc., N = 3) = \frac{364}{365}\frac{363}{365} \tag{5}$$

It is clear then that for N persons we have

$$P(non-coinc., N) = \frac{365}{365}\frac{364}{365}\frac{363}{365}...\frac{365-N+1}{365} \tag{6}$$

We can now use

$$e^{-x} \approx 1 - x \tag{7}$$

to write

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

and therefore

$$P(non-coinc, N) = e^{-1/365}e^{-2/365}e^{-3/365}...e^{-(N-1)/365} = e^{-\frac{N(N-1)}{2}\frac{1}{365}} \tag{8}$$

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

$$P(coinc, N) = 1 - e^{-\frac{N(N-1)}{2}\frac{1}{365}} \approx 1 - e^{-\frac{N^2}{730}} \tag{9}$$

For $N = 20$ one has, perhaps surprisingly (this is the "paradox") $P(N) = 0.5$ i.e. almost 50%.

- Plot the function $P(coinc, N)$ over N from $N = 0$ to $N = 100$. (1pt)
- Find the corresponding N for $P = 69$ % , 95% and 99%. (1pt)