Homework 3

Due October 2 at 11 pm

Unless stated otherwise, justify any answers you give. You can work in groups, but each student must write their own solution based on their own understanding of the problem.

When uploading your homework to Gradescope you will have to select the relevant pages for each question. Please submit each problem on a separate page (i.e., 1a and 1b can be on the same page but 1 and 2 must be on different pages). We understand that this may be cumbersome but this is the best way for the grading team to grade your homework assignments and provide feedback in a timely manner. Failure to adhere to these guidelines may result in a loss of points. Note that it may take some time to select the pages for your submission. Please plan accordingly. We suggest uploading your assignment at least 30 minutes before the deadline so you will have ample time to select the correct pages for your submission. If you are using LaTeX, consider using the minted or listings packages for typesetting code.

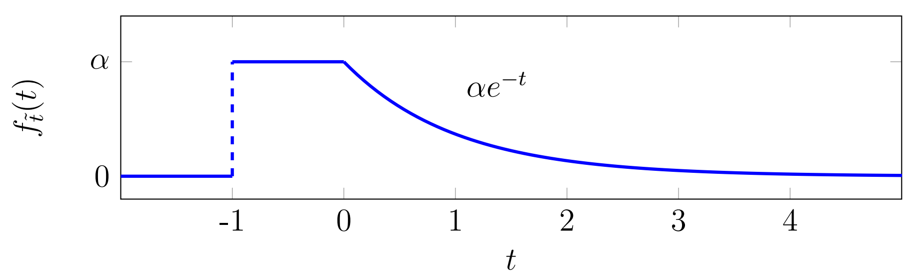
1. (Fish) A biologist is studying a rare species of fish. She captures four individuals and measures their weights, which are 5, 8, 5 and 6 kg.

   (a) What is the empirical conditional probability that a fish weighs more than 7 kg given that they weigh more than 6 kg?

   (b) Plot an estimate of the pdf of the fish weight using kernel density estimation with a rectangular kernel of width 2.

   (c) What is the conditional probability that a fish weighs more than 7 kg given that they weigh more than 6 kg according to your estimated pdf?

2. (Nuclear power plant) The random variable $\tilde{t}$ with the following pdf



   models the time at which there is a leak in a nuclear power plant. The pdf is constant during the time the station is built (between -1 and 0) and exponential with parameter 1 afterwards (from 0 to $+\infty$).

   (a) Compute the value of the constant α.

   (b) Compute the cdf of $\tilde{t}$ and plot it.

   (c) Compute the pdf of $\tilde{t}$ conditioned on $\tilde{t} < 0$.

3. (Measurements) You have access to the readings of a device that indicates whether a radioactive particle has decayed. However you do not get a continuous reading, you get a reading every second.

   (a) A reasonable model for the time the particle takes to decay is that it is a random variable with pdf

   $$f_{\tilde{t}}(t) := \begin{cases} \lambda \exp(-\lambda t), & \text{if } t \geq 0, \\ 0 & \text{otherwise,} \end{cases} \tag{1}$$

   where λ is a fixed constant. Taking into account that the measurement device rounds up the time and outputs an integer number of seconds (if the time is 0.1 it outputs 1, if it is 13.4 it outputs 14), compute the pmf of the reading from the device. What kind of random variable is this?

   (b) What is the pdf of the error between your reading and the true time of decay?

4. (Applying the cdf) The array in `samples.npy` contains $n := 1,000$ i.i.d. samples from a certain distribution.

   (a) Compute the empirical cdf of the data F_X and plot it.

   (b) If you apply the empirical cdf to each data point x_i , $1 \leq i \leq n$, to obtain a new data point $y_i := F_X(x_i)$, what are the new data equal to? Does your answer depend on the distribution of the data?