Homework 3

Due October 3 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. (Half life) The half life of a radioactive substance is a way to quantify how rapidly the substance decays. Given a fixed quantity of the substance, the half time is the time that it takes for it to be reduced to half (i.e. half of the radioactive particles have decayed). It is not immediately apparent why the time should be the same for any quantity. Here we'll show that it is (probabilistically) if the particles decay following an exponential distribution.

   (a) Let $\tilde{t}$ be a random variable with a pdf of the form

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

   where λ is a fixed constant. We define the half life $t_{1/2}$ as the number that satisfies $\mathrm{P}(\tilde{t} > t_{1/2}) = 1/2$. Compute $t_{1/2}$ in terms of λ. Then explain intuitively why this is a reasonable definition for the half life.

   (b) Compute t such that $\mathrm{P}(t_{1/2} < \tilde{t} < t) = 1/4$, and express it in terms of only $t_{1/2}$. Explain why the result is consistent with the intuitive meaning of half life.

   (c) Compute $\mathrm{P}(\tilde{t} > k t_{1/2})$ for any integer k. Again, explain why the result is consistent with the intuitive meaning of half life.

2. (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

   $$f_{\tilde{t}}(t) := \begin{cases} \lambda \exp(-\lambda t), & \text{if } t \geq 0, \\ 0 & \text{otherwise,} \end{cases} \tag{2}$$

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?

3. (Triangular pdf) We are interested in fitting a model with a parametric pdf equal to

$$f_w(x) = \begin{cases} \frac{2x}{w^2}, & \text{for } 0 \leq x \leq w, \\ 0, & \text{otherwise,} \end{cases} \tag{3}$$

where the parameter w is nonnegative.

(a) The observed values are 1.25, 0.4, 1.5, 1, 1.2. What are the possible values of the parameter w?

(b) Compute the likelihood function corresponding to these data and sketch it.

(c) What is the maximum likelihood estimate of w?

(d) If we observe 100 independent samples that are generated according to the parametric model with a fixed value of w, do you think that there is any chance that the ML estimate of w is correct? Justify your answer intuitively.

(e) Let us assume $w = 2$. Generate a sample from a random variable following the model using a uniform sample from the interval $[0, 1]$ equal to 0.64.

4. (Applying the cdf) The array in `samples.npy` contains 1,000 i.i.d. samples from an exponential distribution with parameter $\lambda := 1$. Let F denote the cdf of this distribution.

(a) Apply F to the data (i.e. for each data point x compute $F(x)$) and plot the corresponding histogram. What shape does it have?

(b) Let $\tilde{a}$ be a random variable with an invertible cdf $F_{\tilde{a}}$. What is the distribution of $F_{\tilde{a}}(\tilde{a})$? Justify your answer mathematically.