CS114. Assignment 2. Continuous Random Variables.

Instructions

- Complete all the problems below and show your work.
- In addition to the LOs listed next to each problem, you will be graded on #MathTools, #CompTools, #Communication, and #Professionalism for the overall quality of your text, math, and code.
- You may upload one Python notebook containing all your work, or a PDF with your text **and** math and a Python notebook with your code separately. No handwritten submissions.

Problem 1. Bus Lines (#Probability)

Maria arrives at a bus stop at a uniformly random time. When Maria arrives at the bus stop, either of two independent bus lines (both of which can take her home) may come by. Company A's bus arrival times are exactly 10 minutes apart, whereas the time from one Company B bus to the next is distributed as $\text{Expo}(\frac{1}{10})$.

1. What is the probability that Company B's bus arrives first?

2. What is the PDF of Maria's waiting time for a bus? Be sure to state what the support of the distribution is.

3. Write a simulation to check that your answer in part (1) and your expression for the PDF in part (2) are correct.

Problem 2. Counting Votes (#Distributions)

In a small voting district, $\text{Poisson}(\lambda)$ people arrive to vote during a referendum. Each voter votes *for* the proposal with probability p and *against* the proposal with probability $(1-p)$. Assume everyone casts their vote independently.

1. If D is the difference between the number of votes *for* the proposal and the number of votes *against* the proposal, determine $E(D)$ and $\text{Var}(D)$.

2. Implement a simulation to check your answers.

Problem 3. Hereditary Heights (#Probability)

We model the heights of a family as $\text{Normal}(\mu, \sigma^2)$ random variables. These heights have the same distribution but they are not necessarily independent. Assume there is a mother, a father, and 4 children and that they are all biologically related.

1. This is an oversimplification but assume for this part that the heights are all independent. On average, how many of the 4 children are taller than both parents?

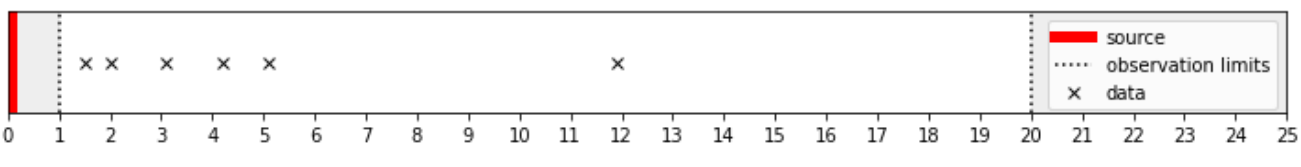
2. Let X_1 be the height of the mother, X_2 be the height of the father, and $Y_1, \ldots, Y_4$ be the heights of the children. Suppose that $(X_1, X_2, Y_1, \ldots, Y_4)$ is Multivariate Normal, with $\text{Normal}(\mu, \sigma^2)$ marginals and $\text{Corr}(X_1, Y_j) = \rho$ for $j \in \{1, 2, 3, 4\}$, with $\rho < 1$.

   On average, how many of the children are at least 1 centimeter taller than their mother?

Problem 4. Radioactive Decay (#ParameterEstimation)

Radioactive particles are emitted from a source and decay independently at a random distance X from the source. The experimental setup allows for decay events to be observed only if they occur at a distance between $x = 1$ cm and $x = 20$ cm from the source.

As data, use $\{x_i\} = \{1.5, 2.0, 3.1, 4.2, 5.1, 11.9\}$.



If $X \sim \text{Expo}(\lambda)$, what is λ?

1. Provide your best estimate for λ and motivate in what sense this is the "best" estimate.

2. Provide a 95% interval for the value of λ. This part can be a bit hand-wavy, meaning you don't have to compute a precise 95% interval but should provide some reasonable justification for how you arrived at the interval you got.

3. **Stretch goal.** This part is entirely optional. If you provide a **excellent, well-written solution** to this problem, you will get an **extra score of** ⑤ on either #MathTools or #CompTools depending on which is most appropriate for your work. This is an all-or-nothing problem. You will either get a ⑤ or you won't. You cannot score less than ⑤ on this problem — instead, you will get no extra grade if the attempt is incorrect or insufficient.

   Problem: Do Bayesian parameter estimation in the problem above. Provide and motivate for an appropriate prior over λ. Compute the posterior distribution over λ and determine a 95% credible interval for λ correct to at least 3 significant digits. Show all your work clearly. Make plots as needed to explain your work.