Homework 7

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. (Spider on a wall) There's a spider living on a wall of your living room that has a painting behind which the spider likes to hide. Figure 1 shows a diagram of the wall; it is 10 feet high and 10 feet wide.
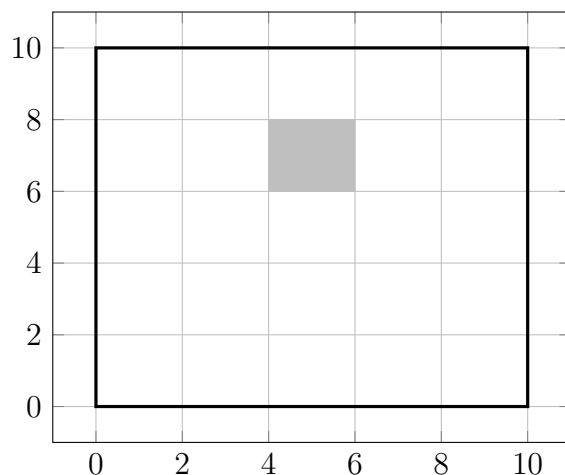

Figure 1: Wall and painting (in gray).

After observing the spider for a while you determine that (1) it spends twice the time behind the painting than on the rest of the wall, (2) it never crawls on the painting or leaves the wall, (3) if it is not behind the painting then it is equally likely to be anywhere on the wall. Since you cannot see it behind the painting, you assume that when it is there it is also equally likely to be at any spot.

(a) Model the position of the spider as a bivariate random variable and give its pdf.

(b) Compute the pdf of the height at which the spider is located and sketch it.

(c) Compute the conditional cdf of the height at which the spider is located, given that you can see it (i.e. it's not under the painting) and sketch it.

2. (Sonar) A scientist is trying to determine the depth of the sea at a certain location. She knows that it must be deeper than 5 km but that is all she knows. To capture this uncertainty she models the depth as uniformly distributed between 5 and 10 km. In order to measure the depth she uses sonar, taking 2 measurements, which we model as two random variables $\tilde{s}_1$ and $\tilde{s}_2$. If the depth is equal to x then each sonar measurement is uniformly distributed between $x - 0.25$ and $x + 0.25$. The two measurements are conditionally independent given the depth.

   (a) Compute and sketch the pdf of the first sonar measurement $\tilde{s}_1$.

   (b) Compute the conditional pdf of the depth conditioned on the measurements being equal to 7 km and 7.1 km.

   (c) Compute the joint pdf of the two sonar measurements $\tilde{s}_1$ and $\tilde{s}_2$. Are the two measurements independent? Justify your answer mathematically and explain it intuitively.

3. (Exotic fruit) A pomologist is studying a newly-discovered plant that produces edible fruit. They have available 10 specimens, which she measures to obtain the following data:

Fruit weight (kg)	1.5	2.3	0.8	1.2	2.0	1.2	0.7	2.7	2.3	0.6
Stem height (cm)	20	15	18	19	17	22	21	14	17	22
Stem radius (mm)	8	12	6	10	17	12	9	14	10	8
Number of leafs	110	94	152	123	78	60	111	83	85	90
Median leaf length (cm)	12	21	9	14	19	15	7	29	22	15

   The pomologist wants to use these data to model the conditional pdf of the fruit weight given the other four features.

   (a) Why is it difficult to estimate the conditional pdf using a simple nonparametric estimator where we apply kernel density estimation to a relevant subset of the data?

   (b) Model all the features as jointly Gaussian and use the model to compute the conditional pdf of the fruit weight for a plant that has a stem height of 15 cm, a stem radius of 20 mm, 120 leafs and median leaf length of 8 cm.

4. (Simulating a random vector) Explain how to simulate a random two-dimensional vector $\tilde{x}$ with a joint pdf $f_{\tilde{x}}(x)$ that is uniformly distributed in the shaded region. Assume that you have access to independent samples from a uniform distribution in $[0, 1]$. Explain your method, justifying why it works. Then implement it and submit the code along with a scatterplot of 1,000 samples.