Probability and Statistics Chapter 3 Questions

Sanjit K. Kaul

I. Textbook Problems

End of Chapter Problems 3.7.5, 3.7.6, 3.7.8, 3.7.9, 3.7.11, 3.7.13 from the textbook by Roy Yates and David Goodman (2nd Ed).

Question 1. Let X be a continuous random variable. For any $x \in \mathcal{R}$, define

$$x^+ = \begin{cases} x & x > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

1) Derive the CDF and PDF of the random variable $Y = (X - a)^+$, where $a > 0$, in terms of the CDF and PDF of X.
2) Derive $E[Y|X > a]$ and $E[Y|X \leq a]$.

Using the derived expressions, answer the following

(a) What is $E[Y]$ when X is an exponential random variable of rate 2 (same as mean $1/2$) and $a = 5$.
(b) What is $E[Y]$ when X is a continuous uniform random variable of rate with range space $S_X = (-5, 10)$ and $a = 5$.

Question 2. Assume that the time between two bus arrivals is well modeled by an exponential random variable with mean 0.5 hours. You arrive at the bus stop. You missed the previous bus by three minutes. What is the expected value of the time you wait for a bus to arrive? Suppose 15 minutes have passed since your arrival. What is the expected value of the additional time you will wait for a bus to arrive?

Now assume that the time to the next bus is a continuous uniform random variable with the range space $(0, 1)$ hours. You arrive at the bus stop. You missed the previous bus by three minutes. What is the expected value of the time you wait for a bus to arrive? What is the standard deviation of your wait time? Suppose 15 minutes have passed since your arrival. What is the expected value of the additional time you will wait for a bus to arrive? What is the standard deviation of this additional time?

Question 3. Consider the function

$$f(x) = \begin{cases} c_1 e^{-x^2} + c_2 e^{-x} & x > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Arrive at c_1 and c_2 that make the function a valid probability density function. Hint: Use known pdf(s)).

Question 4.

Suppose you send either -1 V (OFF) or 1 V (ON) over a channel to a device. You send 1 with probability 0.6. The channel flips the symbol you send with probability p. For example, a 1 sent becomes -1 at the device with probability p. Denote by A the event that the device gets a 1, and by B the event that a -1 is sent. Using the definition of independence answer the following.

1) Are the events A and B independent for all valid selections of p?

2) Is there a value of p for which the events are independent? If yes, what is it?

Let's now consider a different abstraction (instead of the symbol flipping one) for the channel. In this new abstraction, the channel adds a standard normal random variable to the sent symbol, which as before is -1 or 1. The resulting value is received by the device. The device uses a threshold based mechanism to map the received value to a -1 or 1. Specifically, if the received value is greater than a known threshold c, the device assumes that the symbol 1 was sent. Else, it assumes -1 was sent. As before you send 1 with probability 0.6. Answer the following questions. Table **??** may help with your calculations.

(aa) What is the distribution (pdf) of the values received at the device when the symbol sent by you is 1?

(bb) What is the distribution (pdf) of the values received at the device when the symbol sent by you is -1?

(cc) What is the distribution (pdf) of the values received at the device?

(dd) Recall the probability p in the symbol flipping abstraction. What is the value of p in this abstraction? Write it as a function of c. Use properties of the cdf of the standard normal.

(ee) The device makes an error when the symbol it assumes it has received is not the same as the one you sent. What is the probability of error? Write it as a function of c.

(ff) Say you can choose c from the set $\{-0.2, 0, 0.2\}$. Which c gives the smallest probability of error? What is this probability?

Question 5. A processing unit accepts jobs in batches. Specifically, all jobs that arrive during the same minute are grouped as a batch. Such a batch of jobs is scheduled to be processed at a certain time in the future.

The number of jobs in a batch is known to be a Poisson random variable of rate 1 job/minute. Once the processing of a batch starts, jobs in the batch are processed sequentially, in the order of their arrival. Let the *waiting time* of a job in a batch be defined as the time that elapses between the start of processing of the batch and the start of processing of the job. The first arrival, therefore, gets processed with zero waiting time, the second arrival waits for the first to get processed, and similarly, the k^{th} arrival waits for the first $k-1$ arrivals in the batch to get processed. Processing time of a job in any batch is an exponential of mean 0.5.

What is the expected value of the waiting time of a randomly selected job from a randomly selected batch?

Question 6. A Gaussian random variable with mean 0 and variance 10 and an Exponentially distributed random variable with mean 1 fall in love at first chat. As they chat about meeting in person, they decide to do the following. Each one of them draws a number from their distribution independently of the other. In case both the drawn numbers have the same sign (positive or negative), they will meet in person. Else, they will not meet. Derive the probability that they will meet?

Question 7. You create a coin that has the following properties. When the coin is tossed, the coin internally starts a timer that ends after a time interval that is exponentially distributed with rate 1 per microsecond. If the timer ends in 0.5 microseconds, then the coin toss gives heads. If the clock expires after 1.5 seconds, the coin toss gives tails. Otherwise, the coin keeps spinning (and no result is obtained). We will associate a value of 0.5 with heads, 1.5 with tails, and 1 with spinning. Answer the following questions. [Hints: Expectation of a sum of random variables is the sum of expectations of the random variables. The pdf of a rate λ exponential RV X is $f_X(x) = \lambda e^{-\lambda x}$, for $x \geq 0$, and 0 otherwise.]

Question 8. A number is drawn from a Gaussian distribution with mean μ and standard deviation σ. A priori your belief is therefore summarized by the PDF $f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/(2\sigma^2)}$, $x \in (\infty, \infty)$. You are told that the drawn number was larger than the mean. Express your revised belief given the information.

Suppose two such numbers are drawn independently of each other. In the absence of any information regarding the outcomes of the draw, calculate the probability that both the numbers were either larger than the mean or smaller than the mean?

Question 9. You have submitted your application to the Ministry of Procrastination (MoP). Historically, the time X MoP takes to respond is well described by an exponential distribution with mean 1 year. We have $f_X(x) = \mu e^{-\mu x}, x \geq 0$, with mean $1/\mu = 1$. Calculate the CCDF. Suppose that MoP has not responded to your application for 2 years. Write down this event in terms of the random variable X. Given this information, calculate the conditional CCDF of X and the corresponding PDF. Comment on how the PDF compares with $f_X(x)$.

Question 10. You start your day at a bus stop excited to reach the probability and stats lecture on time. A bus arrives a random B minutes later where B is exponentially distributed with mean $1/\mu = 10$ minutes. The bus takes 45 minutes to reach the lecture's venue. If the bus doesn't arrive for 10 minutes, you decide to take an Uber. The Uber arrives a random U minutes later where U is exponentially distributed with mean 5 minutes. The Uber takes 30 minutes to reach the lecture's venue. If the Uber doesn't arrive within 5 minutes, you decide to take an auto. The auto arrives A minutes later where A is exponentially distributed with mean 5 minutes. The auto, like Uber, takes 30 minutes to reach the lecture's venue. If the auto doesn't arrive within 5 minutes you decide not to attend the lecture and head home, which is a 5 minute walk from the bus stop. Answer the following questions.

1) Over any selection of 10 days, what is the probability that you take the Uber 2 days, the bus for 4 days, and the auto for 2 days. Choices are independent of each other. Don't simplify exponentials, if any.
2) Calculate the expected value of the time you spend waiting for transport.
3) Assuming you reach the lecture, calculate the expected value of the total time it takes from the start of day.
4) Calculate the expected value of the total time spent by you from the start of your day till you reach the lecture venue or are back home.

Question 11. The length of a radio advertisement is uniformly distributed over the continuum $(0, 1)$ minutes. What is the probability that an advertisement is exactly 0.5 minutes long? Calculate the probability that an advertisement that has already run for 0.5 minutes will not exceed 0.8 minutes in length.

Question 12. You add two independent Gaussian noise sources. One of them is zero mean and has unit standard deviation. The other has unit mean and a standard deviation of 4. What is the probability that the result of the addition is greater than 1? Explain your answer.

Calculate the probability that the addition results in a positive value. Leave your answer in terms of the CCDF of the standard Gaussian.

Question 13. You are given a weighted sum of PDF(s) $f_1(x)$ and $f_2(x)$. Specifically, we have

$$z(x) = w_1 f_1(x) + w_2 f_2(x), \; x \in (-\infty, \infty).$$

Obtain conditions that when satisfied by the scalar weights w_1 and w_2 ensure that $z(x)$ is also a PDF.

Question 14. Playing a stage in the game called FLUKE returns a number, which is distributed as a continuous uniform random variable, from the interval $(0, 1)$. This number is independent of the numbers returned on playing other stages. The game ends as soon as you win a stage. Else, it continues to the next stage. The winning criteria of each stage is different. In the first stage you win if the number returned lies in $(0, 0.5)$. You win stage 2 if the number returned on playing stage 2 lies in $(0, 0.25)$. More generally, you win stage $k \geq 1$, if the number returned from playing stage k lies in $(0, 0.5^k)$. Answer the following questions.

(a) (10 marks) Derive the probability that exactly n stages of the game are played.

(b) (15 marks) Derive the average number of stages that are played when it is known that the game lasted for at most 3 stages.

(c) (15 marks) Derive the probability that you will keep playing the game forever. In other words, derive the probability that you will never win a stage? [Hint: Approximate the value of the resulting expression by summing over some of the larger terms.]