EECS 126 Probability and Random Processes
Thomas Courtade

UC Berkeley, Spring 2021
February 18, 2021

Midterm 1

Last Name	First Name	SID

***Rules*.**

- **Unless otherwise stated, all your answers need to be justified and your work must be shown. Answers without sufficient justification will get no credit.**
- You have 70 minutes to complete the exam and 10 minutes exclusively for submitting your exam to Gradescope. (DSP students with $X\%$ time accomodation should spend $70 \cdot X\%$ time on the exam and 10 minutes to submit).
- Collaboration with others is strictly prohibited.
- You should not discuss the exam with anyone (this includes your roommate, your parents, social media, reddit, etc.) until 24 hours after the exam concludes (Feb 19th, 2:10pm).
- You may reference your notes, the textbook, and any material that can be found through the course website. You may use Google to search up general knowledge or use calculators. However, searching up a question is not allowed.
- For any clarifications you have, please create a private Piazza post. We will have a Google Doc that shows our official clarifications.

Problem	points earned	out of
Honor Code		5
Problem 1		10
Problem 2		11
Problem 3		26
Problem 4		24
Problem 5		24
Total		100

Honor Code [5 points]

Please copy the following word for word, and sign afterwards.

By my honor, I confirm that

1. *this work is my own original work;*
2. *I have not and will not discuss this exam with anyone during the exam and for 24 hours after the exam;*
3. *I have not and will not Google/search for any of these exam problems.*

1 Smoking [7 + 3 points]

A recent study of smokers reported that 20% of smokers reported a form of lung disease, 10% had a form of heart disease, and one-third of smokers that had either lung or heart disease had both.

(a) What is the probability that a smoker has both lung and heart disease?

(b) Let q denote the answer to part (a). Conditioned on a smoker having heart disease, what is the probability they have lung disease in terms of q?

(a) Let L be the event of a smoker having lung disease and H be the event of a smoker having heart disease. We are given the information $\Pr[L \cap H | L \cup H] = 1/3$. Expanding using the definition of conditional probability and noting $\Pr[(L \cap H) \cap (L \cup H)] = \Pr[L \cap H]$:

$$1/3 = \Pr[L \cap H | L \cup H] = \frac{\Pr[(L \cap H) \cap (L \cup H)]}{\Pr[L \cup H]} = \frac{\Pr[L \cap H]}{\Pr[L] + \Pr[H] - \Pr[L \cap H]}$$

Rearranging and solving, for $\Pr[L \cap H]$ we find that $\Pr[L \cap H] = 3/40$.

(b) By definition of conditional probability, we have $\Pr[L|H] = q/\Pr[H] = 10q$.

2 Middle School [3 + 8 points]

Assume a middle school is composed of 40% 6th graders, 40% 7th graders and 20% 8th graders. The average height of students in these grades are 4ft, 4.5ft, and 5ft, respectively. The variance of heights within each grade are 1 ft, 1/2 ft and 1/2 ft, respectively.

Suppose you pick a student at random. Let X denote their grade, and Y denote their height.

(a) What is $\mathbb{E}[Y]$?

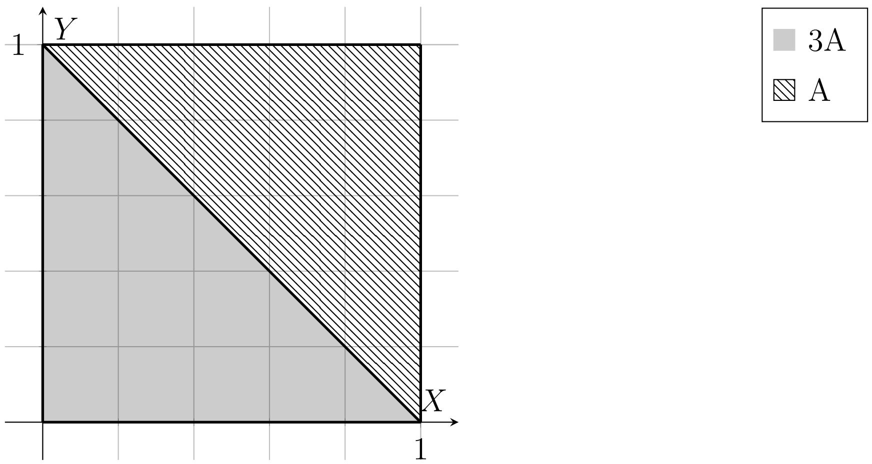
(b) What is $\mathrm{Var}(Y)$?

(a) $\mathbb{E}[Y] = \frac{2}{5} \cdot 4 + \frac{2}{5} \cdot \frac{9}{2} + \frac{1}{5} \cdot 5 = \frac{22}{5} = 4.4$

(b) Utilizing law of total variance:

$$\begin{aligned}
\mathrm{Var}(Y) &= \mathbb{E}[\mathrm{Var}(Y|X)] + \mathrm{Var}(\mathbb{E}[Y|X]) \\
&= \frac{2}{5} \cdot 1 + \frac{2}{5} \cdot \frac{1}{2} + \frac{1}{5} \cdot \frac{1}{2} + \frac{2}{5}(4 - 4.4)^2 + \frac{2}{5}(4.5 - 4.4)^2 + \frac{1}{5}(5 - 4.4)^2 \\
&= .4 + .2 + .1 + .064 + .004 + .072 \\
&= .84
\end{aligned}$$

3 Graphical Density [3 + 3 + 5 + 7 + 8 points]



(a) Are X and Y independent? Remember to justify your answer.

(b) What is the value of A?

(c) Compute $f_X(x)$.

(d) Compute $\mathbb{E}[Y|X = x]$. You may leave your answer as fraction of terms containing x, but you may not have an integral.

(e) What is $\mathbb{E}[X - Y|X + Y]$?

(a) No. For example, when $X = 0$, the expected value of Y is $\frac{1}{2}$, while when $X = \frac{1}{2}$ the expected value of Y is less than $\frac{1}{2}$ since there is more probability mass in the bottom triangle.

(b) Total area must integrate to 1, so $\frac{1}{2}(3A) + \frac{1}{2}A = 1 \implies A = \frac{1}{2}$.

(c) The vertical line at $X = x$ breaks up into two pieces from each triangle:

$$\begin{aligned} f_X(x) &= \int_0^1 f_{X,Y}(x,y)dy \\ &= \int_0^{1-x} \frac{3}{2}dy + \int_{1-x}^1 \frac{1}{2}dy \\ &= \frac{3}{2}(1-x) + \frac{1}{2}x \\ &= \frac{3}{2} - x \end{aligned}$$

(d) Since $f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$, we can compute this directly by the definition of expectation.

$$\begin{aligned} \mathrm{E}[Y|X] &= \int_0^{1-x} y\frac{3A}{3/2-x}dy + \int_{1-x}^1 y\frac{A}{3/2-x}dy \\ &= \frac{3A(1-x)^2}{3-2x} + \frac{A(1-(1-x)^2)}{3-2x} \\ &= \frac{3-4x+2x^2}{2(3-2x)}. \end{aligned}$$

Alternate Solution: We can split this expectation into the case when y falls in the $3A$ region (equivalently, when it falls below $1-x$), and when y falls in the A region. When it falls below $1-x$, its expectation will be $(1-x)/2$, whereas when it falls above $1-x$, its expectation will be $(1+1-x)/2 = (2-x)/2$. It remains to figure out with what probability y falls below $1-x$. Let B be the (constant) density $f_{Y|X}(y|x)$ when y is above $1-x$. Then $3B$ is the density below $1-x$, and in order to integrate to 1 we must have

$$3B(1-x) + Bx = 1 \implies B = \frac{1}{3-2x}.$$

Then we have

$$\mathrm{E}[Y|X] = \frac{1-x}{2}\cdot\frac{3(1-x)}{3-2x} + \frac{2-x}{2}\cdot\frac{x}{3-2x},$$

which, after simplifying, yields the same result, namely:

$$\mathrm{E}[Y|X] = \frac{3-4x+2x^2}{2(3-2x)}.$$

(e) We see that given $X+Y=c$, which is a line parallel to the diagonal line in the graph, the values of $X-Y$ (which is perpendicular to $X+Y$) is uniformly distributed, centered around 0. So $E[X-Y|X+Y]=0$. Another way to see this is that $E[X|X+Y] = E[Y|X+Y] = (X+Y)/2$, so by linearity of expectation $E[X-Y|X+Y]=0$.

4 Bowling [3 + 5 + 7 + 9 points]

In the game of bowling, you roll a ball toward a set of 10 pins with the goal of knocking them down. You get two rolls to knock down as many pins as possible, and the pins are not reset after the first roll. If there are m pins standing at the beginning of your roll, you knock down each of them independently with probability 1/2. Let B_1 and B_2 denote the number of pins you knock down on the first and second rolls, respectively, so that the total number of pins knocked down is $B = B_1 + B_2$.

(a) What is the distribution of B conditioned on $\{B_1 = b_1\}$?

(b) Use Markov's inequality to give a simple upper bound on the probability that you knock down all 10 pins (i.e., $\Pr\{B \geq 10\} \leq$?)

(c) Suppose $B = 2$. What is the probability that $B_1 = 1$?

(d) Compute the MGF $M_B(t)$.

(a) $B - b_1 \sim \text{Binom}(10 - b_1, \frac{1}{2})$

(b) Let's first calculate the expectation of B:

$$\mathbb{E}[B] = \mathbb{E}[B_1 + B_2] = \mathbb{E}[B_1 + \mathbb{E}[B_2|B_1]] = \mathbb{E}[B_1 + \frac{1}{2}(10 - B_1)] = \frac{15}{2}$$

Then, $\Pr(B \geq 10) \leq \frac{\mathbb{E}[B]}{10} = \frac{3}{4}$.

(c) Using Bayes rule:

$$\begin{aligned}
\Pr[B_1|B=2] &= \frac{\Pr[B=2|B_1=1]\Pr[B_1=1]}{\Pr[B=2]} \\
&= \frac{\Pr[B=2|B_1=1]\Pr[B_1=1]}{\Pr[B_1=0 \cap B_2=2] + \Pr[B_1=1 \cap B_2=1] + \Pr[B_1=2 \cap B_2=0]} \\
&= \frac{9\cdot(1/2)\cdot(1/2)^8\cdot 10\cdot(1/2)\cdot(1/2)^9}{(1/2)^{10}\binom{10}{2}(1/2)^{10} + 90(1/2)^{19} + \binom{10}{2}(1/2)^{10}(1/2)^8} \\
&= \frac{90/2}{45\cdot(1/4) + 90/2 + 45} \\
&= 4/9.
\end{aligned}$$

Alternate Solution: We can also notice that $B \sim \text{Binom}(10, \frac{3}{4})$ as the probability a pin is knocked over is one minus the probability it survives both balls, independent of all other pins. This simplifies the calculation of the denominator:

$$\begin{aligned}\Pr[B_1|B=2] &= \frac{\Pr[B=2|B_1=1]\Pr[B_1=1]}{\Pr[B=2]} \\ &= \frac{90\cdot(1/2)^{19}}{9\cdot 45\cdot(1/2)^{20}} \\ &= \frac{4}{9}\end{aligned}$$

(d) Again, if we have the insight that B is binomial the calculation becomes simple:

$$M_B(t) = (\tfrac{1}{4} + \tfrac{3}{4}e^t)^{10}$$

Other approaches, such as using $M_B(t) = \mathbb{E}[e^{tB_1}\mathbb{E}[e^{tB_2}|B_1]]$ or trying to directly calculate the expectation are hard to simplify. Note that we cannot say $M_B(t) = M_{B_1}(t)M_{B_2}(t)$ as B_1 and B_2 are not independent.

5 r/wallstreetbets [3 + 4 + 10 + 7 points]

Your friends, Reina and Han, each bought a share of GME stock. You also hold a share, and your strategy is to sell it when the first of your two friends closes their position (sells their GME share). Assume that the amount of time in days before Reina and Han sell their share are independent and can be modeled as $\text{Exp}(\lambda_1)$ and $\text{Exp}(\lambda_2)$ RVs, respectively (note that time is continuous).

(a) What is the probability that Reina closes her position before Han?

(b) If you bought at a time when Reina and Han were both holding their shares, what is the expected number of days before you will sell your share?

(c) Upon consideration, you decide that the previous strategy is too risky. Instead, you will sell whenever the first of your friends closes their position, or after holding your share for three days. What is the expected number of days before you sell your share?

(Hint: The tail sum formula can also be applied in the continuous case: $\mathbb{E}[X] = \int_0^\infty P(X > t)dt$ for a non-negative random variable X, but this is not the only way to solve it.)

(d) Suppose you make $h(t) = (e^{(\lambda_1+\lambda_2)t} - 1)$ dollars if you sell after $t \geq 0$ days. What is your expected gain if you employ the strategy of part (c)?

a) Let $X_1 \sim Exp(\lambda_1)$ represent the time that Reina sells, and $X_2 \sim Exp(\lambda_2)$ represent the time that Han sells. Then, the probability X_1 is the min of two independent exponentials is:

$$\Pr[X_1 = \min(X_1, X_2)] = \tfrac{\lambda_1}{\lambda_1+\lambda_2}$$

b) Defining $X = \min(X_1, X_2)$, we know that $X \sim Exp(\lambda_1 + \lambda_2)$. Thus the expected value is:

$$\mathbb{E}[\min(X_1, X_2)] = \mathbb{E}[X] = \tfrac{1}{\lambda_1+\lambda_2}$$

c) Defining $X = \min(X_1, X_2)$, we need to find $\mathbb{E}[\min(X, 3)]$.

$$\begin{aligned}
\mathbb{E}[\min(X,3)] &= 3\Pr[X > 3] + \mathbb{E}[X|X \leq 3]\Pr[X \leq 3] \\
&= 3\Pr[X > 3] + \mathbb{E}[X] - \mathbb{E}[X|X > 3]\Pr[X > 3] \\
&= 3\Pr[X > 3]) + \mathbb{E}[X] - (3 + \mathbb{E}[X])\Pr[X > 3] \\
&= 3\Pr[X > 3] + \mathbb{E}[X] - 3P(X > 3) - \mathbb{E}[X]\Pr[X > 3] \\
&= \mathbb{E}[X] - \mathbb{E}[X]\Pr[X > 3] \\
&= \frac{1}{\lambda_1+\lambda_2} - \frac{e^{-3(\lambda_1+\lambda_2)t}}{\lambda_1+\lambda_2} \\
&= \frac{1 - e^{-3(\lambda_1+\lambda_2)}}{\lambda_1+\lambda_2}
\end{aligned}$$

Alternate Solution: Using the tail sum formula, we have:

$$\begin{aligned}
\mathbb{E}[\min(X_1,X_2,3)] &= \int_0^\infty \Pr[\min(X_1,X_2,3) > t]dt \\
&= \int_0^\infty \Pr[X_1 > t]\Pr[X_2 > t]\Pr[3 > t]dt \\
&= \int_0^3 \Pr[X_1 > t]\Pr[X_2 > t]dt \\
&= \int_0^3 e^{-\lambda_1 t}e^{-\lambda_2 t}dt \\
&= \int_0^3 e^{-(\lambda_1+\lambda_2)t}dt \\
&= -\frac{1}{\lambda_1+\lambda_2}e^{-(\lambda_1+\lambda_2)t}\Big|_0^3 \\
&= \frac{1 - e^{-3(\lambda_1+\lambda_2)}}{\lambda_1+\lambda_2}
\end{aligned}$$

d) Let $X = \min(X_1, X_2)$. For simplicity, let's consider $h'(t) = h(t) + 1$. Our expected gain is

$\mathbb{E}[h(t)] = \mathbb{E}[h'(t)] - 1$. Evaluating the integral:

$$\begin{aligned}
\mathbb{E}[h'(X)] &= \int_0^3 h'(t)\Pr[X = t]dt + \Pr[X > 3]h'(3) \\
&= \int_0^3 (\lambda_1 + \lambda_2)e^{-(\lambda_1+\lambda_2)t}e^{(\lambda_1+\lambda_2)t}dt + e^{-3(\lambda_1+\lambda_2)}e^{3(\lambda_1+\lambda_2)} \\
&= \int_0^3 (\lambda_1 + \lambda_2)dt + 1 \\
&= (\lambda_1 + \lambda_2)t\Big|_0^3 + 1 \\
&= 3\lambda_1 + 3\lambda_2 + 1
\end{aligned}$$

So the expected gain is $3(\lambda_1 + \lambda_2)$.

Note that we can't say $\mathbb{E}[h(t)] = h(\mathbb{E}[t])$, as h is a nonlinear function.