Prelims Probability & Statistics

Hilary Term Collection

January 2024

Instructions: You have 90 minutes to complete this closed-book, closed-notes collection. Graphing/programmable calculators are not permitted, but scientific or ordinary four-function calculators are allowed. Show your work for full credit. A table of the standard normal distribution is attached to this document.

Question:	1	2	3	4	5	6	7	8	Total
Points:	5	5	5	5	10	10	30	30	100
Score:									

☺☺☺ **Good Luck!** ☺☺☺

Part A

Answer each of the following short questions, nearly all of which are taken *verbatim* from your assigned review and problem set questions.

5 1. Let $X \sim \mathrm{N}(\mu = -2, \sigma^2 = 25)$. Without consulting a table, what is the approximate value of $\mathbb{P}(-12 \leq X \leq 8)$?

Solution:

$$\mathbb{P}(-12 \leq X \leq 8) = \mathbb{P}\left[\frac{-12-(-2)}{5} \leq X \leq \frac{8-(-2)}{5}\right] = \mathbb{P}(-2 \leq Z \leq 2)$$

where $Z \sim \mathrm{N}(0,1)$. Therefore the desired probability is approximately 0.95.

5 2. Let X be a discrete random variable that is equally likely to take on the values 2 and 4 and never takes on any other values. Prove or disprove the following claims, either by using a property of expected value that you have learned, or by direct calculation

(a) $\mathbb{E}(X+10) = \mathbb{E}(X) + 10$

(b) $\mathbb{E}(X/10) = \mathbb{E}(X)/10$

(c) $\mathbb{E}(10/X) = 10/\mathbb{E}(X)$

(d) $\mathbb{E}(X^2) = \mathbb{E}[(X)]^2$

(e) $\mathbb{E}\left(\frac{5X+2}{10}\right) = \frac{5\mathbb{E}(X)+2}{10}$

Solution: Parts (a), (b), (c) and (e) are true by the linearity of expectation. Part (b) is false by direct calculation: $10/\mathbb{E}(X) = 10/3 \approx 3.33$ while

$$\mathbb{E}(10/X) = \frac{1}{2} \times \frac{10}{2} + \frac{1}{2} \times \frac{10}{4} = \frac{1}{2} \times 30/4 = \frac{15}{4} = 3.75.$$

Part (d) is likewise false by direct calculation: $[\mathbb{E}(X)]^2 = 3^2 = 9$ while

$$\mathbb{E}(X^2) = \frac{1}{2} \times 2^2 + \frac{1}{2} \times 4^2 = 10.$$

5 3. Let X and Y be independent random variables with $X \in \{-1, 1\}$ and $p_X(1) = 0.5$ while $Y \in \{0, 1\}$ and $p_Y(1) = 0.8$. Write out the joint pmf p_{XY} of (X, Y) in tabular form, with X in the rows of the table and Y in the columns.

Solution: Because X and Y are independent, their joint pmf equals the product of their marginal pmfs: $p_{XY}(x, y) = p_X(x)p_Y(y)$ so we have:

		Y	
		0	1
X	-1	0.1	0.4
	1	0.1	0.4

5 4. Let X and Y be random variables such that $\mathbb{E}[Y] \neq 0$. Show that

$$\frac{\mathbb{E}[X]}{\mathbb{E}[Y]} - \mathbb{E}\left[\frac{X}{Y}\right] = \frac{\mathrm{Cov}(X/Y, Y)}{\mathbb{E}[Y]}.$$

Solution: By the "shortcut rule" for covariance

$$\begin{aligned}\mathrm{Cov}(X/Y, Y) &= \mathbb{E}\left[\frac{X}{Y} \cdot Y\right] - \mathbb{E}\left[\frac{X}{Y}\right]\mathbb{E}[Y] \\ &= \mathbb{E}[X] - \mathbb{E}\left[\frac{X}{Y}\right]\mathbb{E}[Y].\end{aligned}$$

Dividing both sides by $\mathbb{E}[Y]$ gives the desired result.

10 5. Let A and B be two events. Use the axioms of probability to show that

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

Solution: When calculating $\mathbb{P}(A \cup B)$ we need to avoid "double-counting" the event $\mathbb{P}(A \cap B)$ since it belongs to both A and B. Define the shorthand $C \setminus D$ to mean $C \cap D^C$. You can think of it as $C \setminus D$ as what's left over after "subtracting" any elements of D from C. A Venn diagram shows how $A \cup B$ can be divided into three mutually exclusive events:

$$A \cup B = (A \setminus B) \cup (A \cap B) \cup (B \setminus A).$$

Therefore, by the third axiom of probability,

$$\mathbb{P}(A \cup B) = \mathbb{P}(A \setminus B) + \mathbb{P}(A \cap B) + \mathbb{P}(B \setminus A).$$

Notice that, in order to obtain the desired result, it would suffice to show that $\mathbb{P}(A \setminus B) = \mathbb{P}(A) - \mathbb{P}(A \cap B)$ and $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A \cap B)$. To see

why, substitute them into the RHS of our expression for $\mathbb{P}(A \cup B)$. Indeed these equations for $\mathbb{P}(A \setminus B)$ and $\mathbb{P}(B \setminus A)$ hold. First write

$$A = (A \setminus B) \cup (A \cap B), \quad B = (B \setminus A) \cup (A \cap B).$$

This is the same idea that we used above when we divided $A \cup B$ into three mutually exclusive events. Here we divide each of A and B into a pair of mutually exclusive events. Therefore, by the third axiom of probability

$$\mathbb{P}(A) = \mathbb{P}(A \setminus B) + \mathbb{P}(A \cap B), \quad \mathbb{P}(B) = \mathbb{P}(B \setminus A) + \mathbb{P}(A \cap B).$$

Re-arranging gives the equalities for $\mathbb{P}(A \setminus B)$ and $\mathbb{P}(B \setminus A)$. We're done!

10 6. I flip 20 fair coins, and discard any that come up heads. I then flip the remaining coins a second time and count up the number of tails T that I obtain. Use the linearity of expectation to calculate $\mathbb{E}(T)$.

Solution: Let $X_i = 1$ if the if the i^{th} coin comes up tails on *both* its first and second toss, zero otherwise. Then $T = \sum_{i=1}^{20} X_i$ and $\mathbb{E}[T] = \sum_{i=1}^{20} \mathbb{E}[X_i]$ by linearity. Since each of the coins is fair, $\mathbb{E}[X_i] = \mathbb{E}[X_1]$ for all i. Thus $\mathbb{E}[T] = 20\mathbb{E}[X_1]$. Finally, since

$$\begin{aligned}\mathbb{E}[X_1] &= 0 \times [1 - \mathbb{P}(\text{Tails, Tails})] + 1 \times \mathbb{P}(\text{Tails, Tails}) \\ &= \mathbb{P}(\text{Tails, Tails}) \\ &= 1/2 \times 1/2 = 1/4\end{aligned}$$

we obtain $\mathbb{E}[T] = 20/4 = 5$.

Part B

Answer both of these questions. They are somewhat more challenging than those from Part A, and ask you to apply what you've learned to more complicated situations.

7. Stephen has two urns, each of which contains three balls. The "blue urn" contains two blue balls and one red ball; the "red urn" contains two red balls and one blue ball. Neither urn has any distinguishing characteristics. Without looking inside to see which balls they contain, there is no way to tell the urns apart. To determine whether Frank, Natalie and Ruth should be allowed to continue teaching undergraduates, Stephen decides to assess their understanding of conditional probability. He flips a fair coin to choose randomly between the two urns. Then he lets Frank, Natalie and Ruth make one draw from the urn apiece. They're not allowed to peek inside the urn, and they have to **replace their ball** after making the draw. They're also not allowed to show each other which ball they drew. After each tutor makes their draw, they state aloud which urn they think Stephen has chosen.

[5] (a) Frank goes first: he draws a blue ball so he guesses that Stephen has chosen the blue urn. Based on the information available to him, what is the probability that Frank's guess is correct?

Solution: By Bayes' Theorem:

$$\mathbb{P}(\text{Blue Urn}|\text{Blue Ball}) = \frac{\mathbb{P}(\text{Blue Ball}|\text{Blue Urn})\mathbb{P}(\text{Blue Urn})}{\mathbb{P}(\text{Blue Ball})}$$

and by the Law of Total Probability:

$$\begin{aligned}\mathbb{P}(\text{Blue Ball}) &= \mathbb{P}(\text{Blue Ball}|\text{Blue Urn})\mathbb{P}(\text{Blue Urn}) + \mathbb{P}(\text{Blue Ball}|\text{Red Urn})\mathbb{P}(\text{Red Urn}) \\ &= 2/3 \times 1/2 + 1/3 \times 1/2.\end{aligned}$$

Therefore,

$$\mathbb{P}(\text{Blue Urn}|\text{Blue Ball}) = \frac{2/3 \times 1/2}{2/3 \times 1/2 + 1/3 \times 1/2} = \frac{2/3}{2/3 + 1/3} = 2/3.$$

[10] (b) Natalie goes second. She knows that Frank is rational: since he guessed "blue urn" this means he must have seen a blue ball. She reaches into the urn and draws a red ball. Show that, based on all the information available to Natalie, the probability that Stephen chose the blue urn is 1/2.

Solution: To keep the notation from getting out of control, let U_B be the event that Stephen chose the blue urn, U_R be the event that Stephen chose the red urn, F_B be the event that Frank draws a blue ball, and N_R be the

event that Natalie draws a red ball. By Bayes' Theorem

$$\mathbb{P}(U_B|F_B \cap N_R) = \frac{\mathbb{P}(F_B \cap N_R|U_B)\mathbb{P}(U_B)}{\mathbb{P}(F_B \cap N_R)}.$$

Since Frank's draw is independent of Natalie's given the urn–they're different people making different draws from the same urn at different points in time– the Law of Total Probability gives

$$\begin{aligned}\mathbb{P}(F_B \cap N_R) &= \mathbb{P}(F_B \cap N_R|U_B)\mathbb{P}(U_B) + \mathbb{P}(F_B \cap N_R|U_R)\mathbb{P}(U_R)\\ &= 2/3 \times 1/3 \times 1/2 + 1/3 \times 2/3 \times 1/2\\ &= 1/9 + 1/9\end{aligned}$$

and therefore we obtain

$$\mathbb{P}(U_B|F_B \cap N_R) = \frac{1/9}{1/9 + 1/9} = 1/2.$$

10 (c) Natalie guesses "blue urn." Ruth knows that both Frank and Natalie are rational. Since Frank guessed "blue urn", he must have drawn a blue ball. But she can't be sure which color ball Natalie drew. Natalie always guesses the urn that has the higher probability based on the information available to her. In the event of a tie, she guesses randomly. Since Natalie guessed "blue urn" Ruth realizes that there are two possibilities: either Natalie drew a blue ball, or she drew a red ball and randomly guessed "blue urn." Ruth reaches into the urn and draws a red ball. What should Ruth guess? Find the probability that her guess is correct, based on the information available to her.

Solution: This one is tricky! Start by asking what Ruth knows for sure: she knows that Frank drew a blue ball and that she herself drew a red one. By *exactly the same reasoning* as in the preceding part, Ruth's posterior probability of U_B is $1/2$, based on her own draw and what she knows about Frank's draw. Notice that this is *exactly the same* as the prior probability that Stephen chose the blue urn. Since Ruth's draw and Frank's draw precisely "cancel each other out" in terms of the information they provide, we can ignore them when solving the problem. So we pretend that Frank and Ruth didn't draw any balls. Ruth knows that Natalie drew a ball and guessed "blue urn". Ruth also knows that Natalie's decision rule is to guess "blue urn" if she draws a blue ball, and to guess randomly if she draws a red ball. Abbreviate "Natalie guessed blue" by NGB. By Bayes' Theorem:

$$\mathbb{P}(U_B|\text{NGB}) = \frac{\mathbb{P}(\text{NGB}|U_B)\mathbb{P}(U_B)}{\mathbb{P}(\text{NGB})}.$$

By the Law of Total Probability,

$$\mathbb{P}(\text{NGB}) = \mathbb{P}(\text{NGB}|U_B)\mathbb{P}(U_B) + \mathbb{P}(\text{NGB}|U_R)\mathbb{P}(U_R).$$

We know that $\mathbb{P}(U_B) = \mathbb{P}(U_R) = 1/2$, so Bayes' Theorem simplifies to:

$$\mathbb{P}(U_B|\text{NGB}) = \frac{\mathbb{P}(\text{NGB}|U_B)}{\mathbb{P}(\text{NGB}|U_B) + \mathbb{P}(\text{NGB}|U_R)}.$$

It remains to calculate $\mathbb{P}(NGB|U_B)$ and $\mathbb{P}(\text{NGB}|U_R)$. Each of these requires a *second* application of the Law of Total Probability:

$$\begin{aligned}\mathbb{P}(\text{NGB}|U_B) &= \mathbb{P}(\text{NGB}|U_B, N_B)\mathbb{P}(N_B|U_B) + \mathbb{P}(\text{NGB}|U_B, N_R)\mathbb{P}(N_R|U_B)\\ &= 1 \times \mathbb{P}(N_B|U_B) + 1/2 \times \mathbb{P}(N_R|U_B)\\ &= 2/3 + 1/2 \times 1/3 = 5/6\end{aligned}$$

$$\begin{aligned}\mathbb{P}(\text{NGB}|U_R) &= \mathbb{P}(\text{NGB}|U_R, N_B)\mathbb{P}(N_B|U_R) + \mathbb{P}(\text{NGB}|U_R, N_R)\mathbb{P}(N_R|U_R)\\ &= 1 \times \mathbb{P}(N_B|U_R) + 1/2 \times \mathbb{P}(N_R|U_R)\\ &= 1/3 + 1/2 \times 2/3 = 4/6.\end{aligned}$$

Therefore, $\mathbb{P}(U_B|\text{NGB}) = \dfrac{5/6}{4/6 + 5/6} = 5/9$. Since this is bigger than $1/2$, Ruth should guess "blue urn".

5 (d) Now suppose that Frank, Natalie, and Ruth had some way to surreptitiously inform one another about which balls each of them had drawn, without Stephen realizing. As before, Frank draws a blue, Natalie red, and Ruth red. But now Ruth knows exactly which ball each of them drew. What should Ruth guess? Find the probability that her guess is correct based on the information available to her in this alternative scenario. **Note that you can solve this part *without* solving part (c).**

Solution: As above, Frank and Ruth's draws "cancel out" so we can ignore them: we're in exactly the same situation as at the start of the problem, where we have a prior probability of 1/2 for each urn. But now Ruth also knows for sure that Natalie drew a red ball. So by the same reasoning as in part (a), with red replacing blue, the posterior probability that Stephen chose the red urn is 2/3. Ruth should guess red.

8. Sam and Caroline have \$100 and need to make big returns *fast*. Elon offers them the chance to play the following game. First Sam and Caroline must choose the fraction φ of the \$100 that they want to bet. This can be any real number between 0 and 1 inclusive, so they can bet *nothing*, *everything*, or anything in between. Then Elon makes a random draw X_1 with probability density function $f(x) = 2x$ for $x \in [0, 1]$, zero otherwise. If $X_1 > 1/2$, Sam and Caroline win: they keep their initial \$100 and Elon gives them an additional $\varphi \times$ \$100. If $X_1 \leq 1/2$, Sam and Caroline lose: they have to pay $\varphi \times$ \$100 to Elon. The game then proceeds to *round two*: Elon makes a *second* random draw X_2 that is independent of X_1 but has the same pdf. The rules in round two are the same as before, except that Sam and Caroline now bet a fraction φ of however much money they have *after* round one. Depending on whether they win or lose in the first round, this could be more or less than \$100. The game ends after round two. Note that the same value of φ is used in each round.

5 (a) Calculate $p \equiv \mathbb{P}(X_1 > 1/2) = \mathbb{P}(X_2 > 1/2)$.

Solution:

$$p \equiv \mathbb{P}(X_1 > 1/2) = \int_{1/2}^{\infty} f(x)\,dx = \int_{1/2}^{1} 2x\,dx = x^2\big|_{1/2}^{1} = 1 - (1/2)^2 = 3/4.$$

5 (b) Let R_1 be a random variable that takes the value 1 if $X_1 > 1/2$ and 0 otherwise. Define R_2 analogously for X_2. Write down the probability mass function of the random variable $S \equiv R_1 + R_2$.

Solution: Since R_1 and R_2 are binary, the support set of their sum, S, is $\{0, 1, 2\}$. The only way for S to equal 0 is if $R_1 = R_2 = 0$. This occurs with probability $(1-p)^2 = 1/16$ by independence. Similarly, the only way for S to equal 2 is if $R_1 = R_2 = 1$. This occurs with probability $p^2 = 9/16$ by independence. The only other possible realization of S is 1. By the complement rule, this realization occurs with probability $1 - 1/16 - 9/16 = 6/16$. Therefore:

$$p_S(0) = 1/16, \quad p_S(1) = 3/8, \quad p_S(2) = 9/16.$$

5 (c) Sam wants to choose $\varphi = 1$ and risk *everything* in both rounds of the game. In expectation, how much money will Sam and Caroline have at the end of the game if they follow this plan?

Solution: We can use the probabilities from the previous part to calculate the expectation. When $\varphi = 1$, Sam and Caroline end up with \$0 unless $S = 2$, in which case they end up with \$400. (If they win twice, they double

their money twice.) Since $p_S(2) = 9/16$, the expectation is

$$\$0 \times 7/16 + \$400 \times 9/16 = 225$$

5 (d) Caroline thinks Sam is crazy: setting $\varphi = 1$ means there's a good chance they'll go bust, ending up with \$0 at the end of the game. Caroline proposes setting $\varphi = 1/2$ instead. If they follow Caroline's strategy, what how much money will they have in expectation at the end of the game? What is the chance of going bust under Caroline's strategy?

Solution: To keep things organized, it's helpful to make a table showing how much money Sam and Caroline will have at the end of the game depending on the realizations of R_1 and R_2. We've already calculated two of the required probabilities above. The only ones we didn't calculate are $\mathbb{P}(R_1 = 0, R_2 = 1)$ and $\mathbb{P}(R_1 = 1, R_2 = 0)$. Since R_1 and R_2 are iid, both of these are simply $3/4 \times 1/4 = 3/16$.

R_1	R_2	$\mathbb{P}(\cdot)$	\$ At End of Game
0	0	1/16	$1/2 \times 1/2 \times \$100 = \$25$
0	1	3/16	$1/2 \times 3/2 \times \$100 = \$75$
1	0	3/16	$3/2 \times 1/2 \times \$100 = \$75$
1	1	9/16	$3/2 \times 3/2 \times \$100 = \$225$

Notice that we didn't actually *need* the probabilities in the second and third row separately: each of them has the same payoff so we could just as well have written the table in terms of the values of S as follows:

S	$\mathbb{P}(\cdot)$	\$ At End of Game
0	1/16	$1/2 \times 1/2 \times \$100 = \$25$
1	6/16	$1/2 \times 3/2 \times \$100 = \$75$
1	9/16	$3/2 \times 3/2 \times \$100 = \$225$

Caroline's strategy ensures that you will *never* go bust. If you follow it, in expectation you will end up with

$$\$25 \times 1/16 + \$75 \times 6/16 + \$225 \times 9/16 = \$2500/16 = \$156.25.$$

10 (e) Let $W(\varphi)$ be the amount of money that Sam and Caroline have at the end of the game as a function of the fraction φ that they choose to bet. It turns out that $W(\varphi) = \$100 \times (1+\varphi)^S(1-\varphi)^{2-S}$. Briefly explain why this is so, and then use this expression to show that $\varphi = 1/2$ maximizes $\mathbb{E}[\log W(f)]$, where log denotes the natural logarithm. You don't have to check the second-order condition.

Solution: The random variable S counts up how many times Sam and Caroline *win* while $2 - S$ counts up how many times they *lose*. Each time they

win, their initial \$100 is multiplied by $(1+\varphi)$; each time they lose it is multiplied by $(1-\varphi)$. This is exactly what the formula does. Now we'll solve the optimization problem. Taking logs,

$$\begin{aligned}\mathbb{E}[\log W(\varphi)] &= \mathbb{E}\left[\log\left\{\$100 \times (1+\varphi)^S(1-\varphi)^{2-S}\right\}\right] \\ &= \mathbb{E}\left[\log(\$100) + S\log(1+\varphi) + (2-S)\log(1-\varphi)\right] \\ &= \log(\$100) + \mathbb{E}[S]\log(1+\varphi) + \mathbb{E}[2-S]\log(1-\varphi)\end{aligned}$$

by the linearity of expectation, since f, 2 and \$100 are constants. Differentiating with respect to φ gives the following first order condition:

$$\frac{\mathbb{E}[S]}{1+\varphi} - \frac{\mathbb{E}[2-S]}{1-\varphi} = 0$$

Re-arranging,

$$\mathbb{E}[S](1-\varphi) = \mathbb{E}[2-S](1+\varphi)$$

Using the calculations from part (a),

$$\mathbb{E}[S] = 0 \times 1/16 + 1 \times 6/16 + 2 \times 9/16 = 24/16 = 3/2.$$

(Alternatively, S is a Binomial$(2, 3/4)$ RV so its expectation is $2\times 3/4 = 3/2$.) Hence, by the linearity of expectation $\mathbb{E}[2-S] = 1/2$. Substituting these values into the first order condition gives $3/2 \times (1-\varphi) = 1/2 \times (1+\varphi)$. Re-arranging, we see that the optimal value of φ is indeed $1/2$.