Prelims Probability & Statistics

Hilary Term Collection

January 12th, 2023

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	Total
Points:	25	25	25	25	100
Score:					

☺☺☺ **Good Luck!** ☺☺☺

---

1. All parts of this question are taken *verbatim* from your assigned practice problems.

[2] (a) Let X be a variable with support set $\{-1, +1\}$ and $p(-1) = 1/3$. Write down the CDF of X.

Solution: $F(x) = \begin{cases} 0, & x < -1 \\ 1/3, & -1 \leq x < 1 \\ 1, & x \geq 1 \end{cases}$

[2] (b) If X is a continuous random variable and a, b are constants, how do we calculate $\mathbb{P}(a \leq X \leq b)$?

Solution: $\mathbb{P}(a \leq X \leq b) = \int_a^b f(x)\,dx$ where f is the pdf of X.

[2] (c) Let A and B be two mutually exclusive events, each with positive probability. Are they independent? Explain briefly.

Solution: If A and B are mutually exclusive, then $\mathbb{P}(A \cap B) = 0$. By the definition of conditional probability $\mathbb{P}(A|B) = \mathbb{P}(A \cap B)/\mathbb{P}(B) = 0$, where this expression is well-defined because $\mathbb{P}(B) > 0$. Since $\mathbb{P}(A)$ is also greater than zero, this means that $\mathbb{P}(A) \neq \mathbb{P}(A|B)$. Learning that B has occurred changes the probability of A, therefore A and B are *not independent*.

[2] (d) Let X and Y be random variables such that $\text{Var}(X) = 2$, $\text{Var}(Y) = 1$, and $\text{Cov}(X, Y) = 0$. Calculate $\text{Var}(X - Y)$.

Solution: $\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y) - 2\text{Cov}(X, Y) = 3$

[5] (e) 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:

(i) $E(5X + 2)/10 = [5E(X) + 2]/10$

Solution: TRUE by the linearity of expectation.

(ii) $E(10/X) = 10/E(X)$

Solution: FALSE. By linearity $E(10/X) = 10E(1/X)$ so the statement would be true if and only if $1/E(X) = E(1/X)$. Note that $f(x) = 1/x$

is a *nonlinear* function. By direct calculation,

$$E(1/X) = \sum_{\text{all } x} \frac{1}{x} \cdot p(x) = \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{4} \cdot \frac{1}{2} = \frac{1}{4} + \frac{1}{8} = \frac{3}{8}$$

$$1/E(X) = \left[\sum_{\text{all } x} x \cdot p(x)\right]^{-1} = \left[2 \cdot \frac{1}{2} + 4 \cdot \frac{1}{2}\right]^{-1} = \frac{1}{1+2} = \frac{1}{3}$$

so we see that $1/E(X)$ does not equal $E(1/X)$ in this case. Another way to answer this part is by appealing to Jensen's inequality. Since the RV X is positive, we only need to worry about the positive "branch" of the hyperbolic function $f(x) = 1/x$. For $x > 0$ this function is strictly convex. Since X is not constant, Jensen's inequality is strict in this case and we obtain $f(E[X]) < E[f(X)]$. And indeed: $1/3 < 3/8$.

(iii) $E(X^2) = [E(X)]^2$

Solution: FALSE. By direct calculation:

$$(E[X])^2 = \left[\sum_{\text{all}x} x \cdot p(x)\right]^2 = \left[2 \cdot \frac{1}{2} + 4 \cdot \frac{1}{2}\right]^2 = (3)^2 = 9$$

$$E[X^2] = \sum_{\text{all } x} x^2 \cdot p(x) = 2^2 \cdot \frac{1}{2} + 3^2 \cdot \frac{1}{2} = \frac{4}{2} + \frac{9}{2} = \frac{13}{2}$$

and these are not equal. Another way to solve this is by appealing to the strict version of Jensen's inequality, which applies here as in the previous part since the function in question is not affine and X is not constant. But my *favorite* way to solve this is by noting that $E(X^2) = [E(X)]^2$ is equivalent to $E(X^2) - [E(X)]^2 = 0$. But this statement says that $\text{Var}(X) = 0$. The only way for a random variable to have a zero variance is if it is a constant, i.e. if it does not in fact *vary*. We see that X *does* vary, so its variance cannot be zero, so the equality cannot hold.

6 (f) Let X and U be random variables with $\mathbb{E}(U|X) = 0$ and $\mathbb{E}(X) = 5$. Let a and b be constants, and further define $Y = a + bX + U$.

(i) Show that $\mathbb{E}(U) = 0$.

Solution: By the law of iterated expectations $\mathbb{E}(U) = \mathbb{E}[\mathbb{E}(U|X)] = \mathbb{E}[0] = 0$.

(ii) Calculate $\mathbb{E}(Y|X = 0)$.

Solution:

$$\begin{aligned}\mathbb{E}(Y|X=0) &= \mathbb{E}[a+bX+U|X=0] = a+b\times 0+\mathbb{E}(U|X=0)\\ &= a+0=a\end{aligned}$$

(iii) Suppose that $\mathbb{E}(U) = 0$ but $\mathbb{E}(U|X)$ is unknown. Would this make any difference to your answers in parts (b) and (c) above?

Solution: Part (b) is unchanged but part (c) becomes

$$\mathbb{E}(Y|X=0) = a + \mathbb{E}(U|X=0).$$

This will not equal a unless $\mathbb{E}(U|X=0) = 0$.

6 (g) Let $X \sim N(0,1)$. Calculate the following quantities, using the standard normal table at the end of this document if needed.

(i) $\mathbb{P}(X=0)$

Solution: 0

(ii) $\mathbb{P}(X \leq -1)$

Solution: 0.1587

(iii) $\mathbb{P}(-1.2 \leq X \leq 1.2)$

Solution: $1 - 2 \times 0.1151 = 0.7698$

(iv) the value of t such that $\mathbb{P}(X \leq t) = 0.5$

Solution: 0

(v) the value of t such that $\mathbb{P}(X > t) = 0.8$

Solution: -0.84–0.85

(vi) the value of $t > 0$ such that $\mathbb{P}(-t \leq X \leq t) = 0.9$

Solution: 1.64–1.65

2. In the land of Wendimoor, one in fifty people have a disease called *pararibulitis* that causes painful and disturbing hallucinations. Fortunately this condition is easily treated by a daily dose of *Subtracterall*. Because Subtracterall has some mildly unpleasant side effects, the physicians of Wendimoor only prescribe it for patients who

test positive for pararibulitis. The test has an accuracy of 80%. In other words, a person with pararibulitis has an 80% of testing positive and a person without pararibulitis has an 80% chance of testing negative.

15 (a) Amanda is a randomly-chosen citizen of Wendimoor. We test her for *pararibulitis* and the test comes back positive. What is the probability that she has the disease? In your solution, let T be the event that Amanda **tests** positive and D be the event that she has the **disease**.

Solution: By Bayes' Theorem:

$$\mathbb{P}(D|T) = \frac{\mathbb{P}(T|D)\mathbb{P}(D)}{\mathbb{P}(T)}$$

And by the law of total probability,

$$\begin{aligned}\mathbb{P}(T) &= \mathbb{P}(T|D)\mathbb{P}(D) + \mathbb{P}(T|D^c)\mathbb{P}(D^c)\\ &= 4/5 \times 1/50 + 1/5 \times 49/50\\ &= 4/250 + 49/250\\ &= 53/250\end{aligned}$$

Therefore: $\mathbb{P}(D|T) = (4/250)/(53/250) = 4/53 \approx 0.075$. Just for fun, here's a much faster way of solving the problem by calculating the *odds*:

$$\frac{\mathbb{P}(D|T)}{\mathbb{P}(D^c|T)} = \frac{\mathbb{P}(T|D)}{\mathbb{P}(T|D^c)} \cdot \frac{\mathbb{P}(D)}{\mathbb{P}(D^c)} = \frac{4/5}{1/5} \cdot \frac{1/50}{49/50} = \frac{4}{49}.$$

Expressed as a probability, this becomes $4/(49+4) = 4/53$ as above.

6 (b) Half of people with pararibulitis notice that they have symptoms and visit a physician to get tested. At the same time, one in ten of people who *do not* have pararibulitis *mistakenly believe* that they have symptoms and visit a physician to get tested. Anyone who visits a physician and tests positive gets a prescription; no one else does. Among people with pararibulitis, what fraction have a prescription? Among people who *do not* have pararibulitis, what fraction have a prescription? Let S be the event that someone is prescribed **Subtracterall** and V be the even that she **visits** a physician to get tested.

Solution: By the law of total probability,

$$\mathbb{P}(S|D) = \mathbb{P}(S|D,V)\mathbb{P}(V|D) + \mathbb{P}(S|D,V^c)\mathbb{P}(V^c|D).$$

To simplify this expression, first notice that $\mathbb{P}(S|D,V^c) = 0$. In words: if you don't visit a physician to test get tested, you can't get a prescription even if you have the disease. Next, notice that $\mathbb{P}(S|D,V) = \mathbb{P}(T|D) = 4/5$. In words: if you visit a physician to get tested and you have the disease,

the chance that you will get a prescription equals the chance that you test positive. Therefore:

$$\mathbb{P}(S|D) = \mathbb{P}(T|D)\mathbb{P}(V|D) = 4/5 \times 1/2 = 2/5.$$

Similarly,

$$\begin{aligned}\mathbb{P}(S|D^c) &= \mathbb{P}(S|D^c, V)\mathbb{P}(V|D^c) + \mathbb{P}(S|D^c, V^c)\mathbb{P}(V^c|D^c) \\ &= \mathbb{P}(T|D^c)\mathbb{P}(V|D^c) = 1/5 \times 1/10 = 1/50.\end{aligned}$$

4 (c) Continuing from part (b), Amanda argues that pararibulitis is *under-diagnosed*: most people who have the disease are not prescribed the Subtracterall. Todd argues that pararibulitis is *over-diagnosed*: most people with a prescription for Subtracterall do not in fact have pararibulitis. Who's right? Explain briefly.

Solution: They're both right! Surprisingly, it's possible for a disease to be *both* over-diagnosed and under-diagnosed. From our solution to the previous part, we know that only 40% of people with pararibulitis are prescribed Subtracterall; most of the people with this disease are untreated. At the same time, Todd is right that most people who take Subtracterall do not have pararibulitis. By Bayes' Theorem

$$\mathbb{P}(D|S) = \frac{\mathbb{P}(S|D)\mathbb{P}(D)}{\mathbb{P}(S)} = \frac{2/5 \times 1/50}{\mathbb{P}(S)}$$

and by the law of total probability,

$$\begin{aligned}\mathbb{P}(S) &= \mathbb{P}(S|D)\mathbb{P}(D) + \mathbb{P}(S|D^c)\mathbb{P}(D^c) \\ &= 2/5 \times 1/50 + 1/50 \times 49/50 \\ &= 1/50 \times 69/50\end{aligned}$$

Therefore, $\mathbb{P}(D|S) = (2/5)/(69/50) = 20/69 \approx 0.29$.

As above, it's simpler to solve this using odds:

$$\frac{\mathbb{P}(D|S)}{\mathbb{P}(D^c|S)} = \frac{\mathbb{P}(S|D)}{\mathbb{P}(S|D^c)} \cdot \frac{\mathbb{P}(D)}{\mathbb{P}(D^c)} = \frac{2/5}{1/50} \cdot \frac{1/50}{49/50} = \frac{20}{49}$$

from which we immediately see that it's much less likely than not that a person who takes Subtracterall has pararibulitis. (We don't even need to convert to a probability to see this because 20 is much smaller than 49.)

3. Let $X_1 \sim$ Bernoulli(1/2) independently of $X_2 \sim$ Bernoulli(2/3) and define $Y = 2X_1$ and $Z = X_2 - X_1$.

15 (a) Express the joint pmf of Y and Z in tabular form. Please put the realizations of Y in the *rows* of the table and realizations of Z in the *columns*.

Solution:

		Z		
		-1	0	1
Y	0	0	1/6	1/3
	2	1/6	1/3	0

$p_{YZ}(0,-1) = 0$ since $Y = 0$ implies $X_1 = 0$ so Z is at least 0.

$p_{YZ}(0,0) = 1/2 \times 1/3 = 1/6$ since this corresponds to $X_1 = X_2 = 0$.

$p_{YZ}(0,1) = 1/2 \times 2/3 = 1/3$ since this corresponds to $X_1 = 0, X_2 = 1$.

$p_{YZ}(2,-1) = 1/2 \times 1/3 = 1/6$ since this corresponds to $X_1 = 1, X_2 = 0$.

$p_{YZ}(2,0) = 1/2 \times 2/3 = 1/3$ since this corresponds to $X_1 = X_2 = 1$.

$p_{YZ}(2,1) = 0$ since $Y = 2$ implies $X_1 = 1$ so Z is at most 0.

3 (b) What is the marginal pmf of Y?

Solution: Since X_1 is a Bernoulli(1/2) random variable and $Y = 2X_1$, the support of Y is $\{0, 2\}$, $p_Y(0) = 1/2$ and $p_Y(2) = 1/2$. This agrees with the values we obtain by summing over the rows of the table in the previous part.

3 (c) What is the marginal pmf of Z?

Solution: The support of Z is $\{-1, 0, 1\}$. Summing over the columns of the table we find that $p_Z(-1) = 1/6$, $p_Z(0) = 1/2$ and $p_Z(1) = 1/3$.

4 (d) Calculate the covariance between Y and Z. Are Y and Z independent?

Solution: We proceed via the shortcut formula:

$$\text{Cov}(Y, Z) = \mathbb{E}[YZ] - \mathbb{E}[Y]E[Z]$$

From the marginal pmfs,

$$\begin{aligned} \mathbb{E}[Y] &= 0 \times 1/2 + 2 \times 1/2 = 1 \\ \mathbb{E}[Z] &= -1 \times 1/6 + 0 \times 1/2 + 1 \times 1/3 = 1/6 \end{aligned}$$

Calculating $\mathbb{E}[YZ]$ is easy since there is only *one* non-zero term in the sum: $2 \times -1 \times 1/6 = -1/3$. Thus, we have

$$\text{Cov}(Y, Z) = -1/3 - 1/6 \times 1 = -1/2.$$

Because they have a nonzero covariance, Y and Z cannot be independent.

4. Let Y be a continuous random variable with support $[0,1]$ and pdf $f(y) = Cy^3(1-y)$.

6 (a) Calculate the value of the constant C in the pdf of Y.

Solution: Since pdfs must integrate to one over their support,

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty} f(y)\,dy = C\int_0^1 (y^3 - y^4)\,dy = C\left(\frac{y^4}{4} - \frac{y^5}{5}\right)\Big|_0^1 \\ &= C\left(\frac{1}{4} - \frac{1}{5}\right) = C\left(\frac{5-4}{20}\right) = \frac{C}{20} \end{aligned}$$

Hence, $C = 20$.

7 (b) Calculate the CDF of Y.

Solution: Using our calculations from the previous part, we need only change the upper limit of integration and substitute 20 for C to show that

$$F(y_0) = \int_{-\infty}^{y_0} f(y)\,dy = 20\left(\frac{y^4}{4} - \frac{y^5}{5}\right)\Big|_0^{y_0} = 5y_0^4 - 4y_0^5$$

for $y_0 \in [0,1]$. Thus, the full expression for the CDF of Y is

$$F(y_0) = \begin{cases} 0, & y_0 < 0 \\ 5y_0^4 - 4y_0^5, & 0 \le y_0 < 1 \\ 1, & y_0 \ge 1 \end{cases}$$

6 (c) Calculate the expected value of Y.

Solution:

$$\begin{aligned} E[Y] &= 20\int_0^1 y(y^3 - y^4)\,dx = 20\int_0^1 (y^4 - y^5)\,dx = 20\left(\frac{y^5}{5} - \frac{y^6}{6}\right)\Big|_0^1 \\ &= 20\left(\frac{1}{5} - \frac{1}{6}\right) = 20\left(\frac{6-5}{30}\right) = \frac{2}{3} \end{aligned}$$

6 (d) Calculate the variance of Y.

Solution: By the shortcut formula, we only need to calculate $E[Y^2]$ since we know $E[Y]^2 = 4/9$ from the preceding part. We have,

$$E[X^2] = 20\int_0^1 (y^5 - y^6)\,dy = 20\left(\frac{y^6}{6} - \frac{y^7}{7}\right)\Big|_0^1 = 20\left(\frac{1}{6} - \frac{1}{7}\right) = 10/21$$

Hence $Var(Y) = 10/21 - 4/9 = (30-28)/63 = 2/63$.