

14.30 EXAM 1 - SUGGESTED ANSWERS

TA: Tonja Bowen Bishop

Question 1

A.

(i) *False/Uncertain.* For A and B that exhaust S , $((A \cup B) \cap (A \cap B)) = A \cap B$. This will only be equal to the empty set if A and B are disjoint.

(ii) *False/Uncertain.* $\Pr(A \cap B) = \Pr(A) \times \Pr(B)$ if and only if A and B are independent, by the definition of independence. Since we only know that A and B are disjoint, they may or may not be independent.

(iii) *True.* Because the range of Y depends on the value of X , the conditional pdf of $Y|X$ cannot be equal to its marginal pdf.

(iv) *False/Uncertain.* While this statement will be true for a discrete random variable, for a continuous random variable it is possible to have $f_X(x) > 1$ at some values of x . In this case, we only require that $\int f_X(x) dx = 1$.

(v) *False/Uncertain.* $F_X(a) = F_Y(a)$ for all $a \in \mathfrak{R}$ tells us that X and Y have the same distribution. But because X and Y are both random variables, they may or may not be equal to each other. For example, if we throw a die and let X be the number that comes up, and then throw it again and let Y be the number that comes up, both random variables will have the same distribution, but there is no guarantee that they will be equal in value.

(vi) *True.* Because X and Y are independent, we know that $f_{X,Y}(x,y) = f_X(x) f_Y(y)$. So $F_{X,Y}(x,y) = \int_{-\infty}^x \int_{-\infty}^y f_{X,Y}(t,s) ds dt = \int_{-\infty}^x \int_{-\infty}^y f_X(t) f_Y(s) ds dt = \int_{-\infty}^x f_X(t) dt \int_{-\infty}^y f_Y(s) ds = F_X(x) F_Y(y)$.

B. It is easiest to start by calculating $f_X(x)$. From the graph we can see that for $x \leq 0.5$ or $x \geq 2$, $f_X(x) = 0$, and for $0.5 < x \leq 1$, $f_X(x) = 1$ (the slope of the cdf over this range). Finally, we know that $f_X(x) = kx$ for $1 < x < 2$, and that $\int_1^2 kx dx = 0.5$, since we know that $\Pr(X \leq 1) = 0.5$.

So it must be that $k = \frac{1}{3}$. Thus we have $f_X(x) = \begin{cases} 1 & 0.5 < x \leq 1 \\ \frac{1}{3}x & 1 < x < 2 \\ 0 & \text{elsewhere} \end{cases}$.

Then we can integrate to find $F_X(x) = \begin{cases} 0 & x \leq 0.5 \\ x - \frac{1}{2} & 0.5 < x \leq 1 \\ \frac{1}{3} + \frac{x^2}{6} & 1 < x < 2 \\ 1 & X \geq 2 \end{cases}$. Note that

in general, we cannot determine whether or not the inequalities are strict. For the graph, see the attached figure.

Question 2

(i) We multiply together the number of ways the students from each school can be selected: $\binom{15}{3}\binom{30}{2}\binom{36}{3}$.

(ii) Now we ignore the information on school affiliation. We have 81 students, of whom 35 are women. So there are $\binom{81}{8}$ possible ways to choose the committee, and $\binom{46}{6}\binom{35}{2}$ ways to select a committee with exactly two women. So we have

$$\Pr(2 \text{ women}) = \frac{\binom{46}{6}\binom{35}{2}}{\binom{81}{8}}$$

(iii) We can select the first group in $\binom{15}{5}\binom{30}{10}\binom{36}{12}$ ways, and conditional on that selection, there are $\binom{10}{5}\binom{20}{10}\binom{24}{12}$ to choose the second group, and then all remaining students are in the third group. However, we must be careful to avoid multiple counting of divisions that put the same students together into differently ordered groups (i.e. putting a set X of students in group 1, a set Y in group 2, and a set Z in group 3 vs. putting X in group 2, Y in group 3, and Z in group 1), since the wording of the question does not suggest that we should care about the ordering of the groups. So the total number of different ways we can separate the students into three groups is

$$\frac{\binom{15}{5}\binom{30}{10}\binom{36}{12}\binom{10}{5}\binom{20}{10}\binom{24}{12}}{3!}$$

Question 3

(i) If the claim is true, we can think about the test as a binomial experiment, where X is the number of students out of 20 that improve their score on the second test (each student is a stage). We want to figure the probability that exactly 16 of 20 improve their scores, given that for each person, the probability of improvement is equal and independent of the improvement of other students. What is this probability? The score could improve for one of two reasons: either the pill was effective for that person, or the pill was not effective, and the person just happened to do better on the second test. So the probability of an improved test score when the claim is true is $(0.8) + (0.2)(0.5) = 0.9$

$$\begin{aligned} \Pr(X = 16|True) &= \binom{20}{16} (0.9)^{16} (0.1)^4 \\ &\approx 0.08978 \end{aligned}$$

(ii) We have the same setup, but now an improvement in test scores only occurs randomly. Thus the probability of improvement is 0.5, and so we have

$$\begin{aligned}\Pr(X = 16|Not\ True) &= \binom{20}{16} (0.5)^{16} (0.5)^4 \\ &\approx 0.00462\end{aligned}$$

(iii) Now we will use Bayes rule to find $\Pr(True|X = 16)$:

$$\begin{aligned}\Pr(True|X = 16) &= \frac{\Pr(X = 16|True) \Pr(True)}{\Pr(X = 16|True) \Pr(True) + \Pr(X = 16|Not\ True) \Pr(Not\ True)} \\ &= \frac{\binom{20}{16} (0.9)^{16} (0.1)^4 \left(\frac{1}{2}\right)}{\binom{20}{16} (0.9)^{16} (0.1)^4 \left(\frac{1}{2}\right) + \binom{20}{16} (0.5)^{16} (0.5)^4 \left(\frac{1}{2}\right)} \\ &= \frac{(0.9)^{16} (0.1)^4}{(0.9)^{16} (0.1)^4 + (0.5)^{16} (0.5)^4} \\ &\approx 0.95105\end{aligned}$$

(iv) For this question, we use the law of total probability based on Dr. Testing's new belief about the veracity of the firm's claim. Note that we are now interested in the probability that the student's math ability will actually improve, rather than just her test score.

$$\begin{aligned}\Pr(improve) &= \Pr(improve|True) \Pr(True) + \Pr(improve|Not\ True) \Pr(Not\ True) \\ &\approx (0.8)(0.95105) + (0)(1 - 0.95105) \\ &\approx 0.76084\end{aligned}$$

Question 4

(i) We find k by setting the integral over the space of possible (x, y) values equal to one and solving:

$$\begin{aligned}k \int_{10}^{40} \int_{10}^{40} (x^2 + xy) \, dy \, dx &= 1 \\ k \int_{10}^{40} x \left(\int_{10}^{40} (x + y) \, dy \right) \, dx &= 1 \\ k \int_{10}^{40} (30x^2 + 750x) \, dx &= 1 \\ k(1192500) &= 1 \\ k &\approx 8.39 \times 10^{-7}\end{aligned}$$

(ii) We want to find the probability that both x and y are between 20 and 30 psi. So we just need to integrate over this range.

$$\Pr(X, Y \in (20, 30)) = k \int_{20}^{30} \int_{20}^{30} (x^2 + xy) dy dx$$

(iii) Now we want the probability that $X > Y$, and again, we integrate over the relevant range within the sample space.

$$\Pr(X > Y) = k \int_{10}^{40} \int_{10}^x (x^2 + xy) dy dx$$

Or, equivalently

$$\Pr(X > Y) = k \int_{10}^{40} \int_y^{40} (x^2 + xy) dx dy$$

(iv) We first find the marginal distribution of X .

$$\begin{aligned} f_X(x) &= \int_{10}^{40} f_{X,Y}(x, y) dy \\ &= kx \int_{10}^{40} (x + y) dy \\ &= k(30x^2 + 750x) \end{aligned}$$

Next, to find the conditional distribution of $X|Y$, we will need the marginal distribution of Y .

$$\begin{aligned} f_Y(y) &= \int_{10}^{40} f_{X,Y}(x, y) dx \\ &= k \int_{10}^{40} (x^2 + xy) dx \\ &= k(21000 + 750y) \end{aligned}$$

Now, the conditional distribution.

$$\begin{aligned} f_{X|Y}(x|y) &= \frac{f_{X,Y}(x, y)}{f_Y(y)} \\ &= \frac{k(x^2 + xy)}{k(21000 + 750y)} \\ &= \frac{x^2 + xy}{21000 + 750y} \end{aligned}$$

Thus, knowing the pressure in the left tire will change our assessment of the probability distribution of the pressure in the right tire. In other words, X and Y are not independent.

(v) We need to integrate over the area within the sample space that satisfies both $X + Y \leq 40$ and $|X - Y| \leq 1$, or in other words, we are

interested in the area bounded by the lines $y = x + 1$, $y = x - 1$, $y = 40 - x$, $y = 10$, and $x = 10$, as shown in the attached figure. We will need to divide the area into three parts, and to do so, we need to find the coordinates of the points where various lines intersect. Point (a) is then intersection of the lines $y = x - 1$ and $y = 10$, so the coordinates of point (a) are $(11, 10)$. Similarly, point (b) is the intersection of the lines $y = x + 1$ and $x = 10$, so it has coordinates $(10, 11)$. Point (c) is the intersection of $y = x + 1$ and $y = 40 - x$, so it's coordinates are $(19.5, 20.5)$, and point (d) is the intersection of $y = x - 1$ and $y = 40 - x$ at $(20.5, 19.5)$. Now, the probability can be calculated by dividing the area of interest with vertical lines at $x = 11$ and $x = 19.5$, as follows:

$$\begin{aligned} \Pr((X + Y \leq 40) \cap (|X - Y| \leq 1)) &= \int_{10}^{11} \int_{10}^{x+1} k(x^2 + xy) \, dy \, dx + \\ &\int_{11}^{19.5} \int_{x-1}^{x+1} k(x^2 + xy) \, dy \, dx + \\ &\int_{19.5}^{20.5} \int_{x-1}^{40-x} k(x^2 + xy) \, dy \, dx \end{aligned}$$

Or, alternatively, we can separate with horizontal lines at $y = 11$ and $y = 19.5$:

$$\begin{aligned} \Pr((X + Y \leq 40) \cap (|X - Y| \leq 1)) &= \int_{10}^{11} \int_{10}^{y+1} k(x^2 + xy) \, dx \, dy + \\ &\int_{11}^{19.5} \int_{y-1}^{y+1} k(x^2 + xy) \, dx \, dy + \\ &\int_{19.5}^{20.5} \int_{y-1}^{40-y} k(x^2 + xy) \, dx \, dy \end{aligned}$$