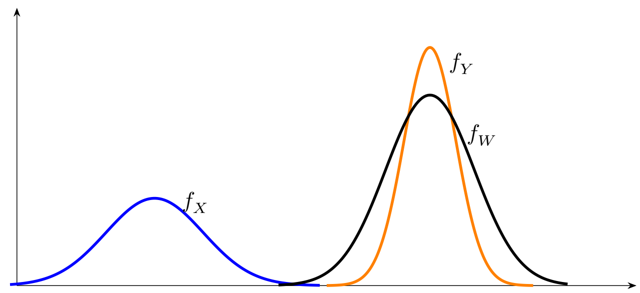
18.05 Exam 1 Solutions

Problem 0. (5 pts) **Be sure to attach your cheat sheet to your test.**

Problem 1. (10 pts: 4,6) **Concept/Quick questions**
(a) (No explanations are necessary.)
The plot shows the pdf for three independent random variables X, Y, W. All use the same horizontal and vertical scale.



Which random variable has the greatest variance?

Solution: X.(Variance measures the spread away from the mean.)

(b) *Suppose A and B are two events and $P(A) = 0.7$, $P(B) = 0.3$ and $P(A \cap B) = 0.25$. Compute each of the following*

(i) Compute $P(A \cup B)$

(ii) Compute $P(A|B)$.

Solution: (i) Inclusion exclusion: $P(A \cup B) = 0.7 + 0.3 - 0.25 = 0.75$.

(ii) $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.25}{0.3}$.

Problem 2. (15 pts: 10,5)
You create passwords as a string of 10 characters such that:

- *5 of the characters are letters (upper and lower case, i.e. 52 characters) with repetitions allowed,*
- *3 are numbers { 0,1,2,3,4,5,6,7,8,9 } with repetitions allowed, and*
- *2 are* ***distinct*** *symbols from the list of 5 symbols: { !, @, #, $, & }.*

(a) *How many passwords are there? (No need to simplify your answer.)*

Solution: First, choose the locations of the symbols $\binom{10}{2}$.
Then choose the symbols, since they have to be different and order matters, we get $5 \cdot 4$.
Then, choose the locations of the letters: $\binom{8}{5}$.
Then count the number of ways to choose 5 letters (with replacement) 52^5.
Then choose the locations of the numbers: $\binom{3}{3} = 1$.
Finally choose the numbers: 10^3.

So, the number of passwords $\binom{10}{2} \cdot \binom{8}{5} \cdot 20 \cdot 10^3 \cdot 52^5$.

(b) *With all locations for symbols, letters, or numbers in your 10 character password being equally likely, what is the probability that the two symbols are next to each other?*

Solution: Count the ways to get a password where the two symbols are adjacent:
First choose locations for the two symbols: there are 9 adjacent positions.
Then there are $5 \cdot 4$ ways to choose the sequence of two symbols.
Then choose the locations of the letters: $\binom{8}{5}$.
Then count the number of ways to choose 5 upper or lower case letters (with replacement) 52^5.
Then choose the locations of the numbers: $\binom{3}{3}$.
Then choose the numbers: 10^3.

So, $\boxed{P(\text{two adjacent symbols}) = \dfrac{9 \cdot \binom{8}{5} \cdot 5 \cdot 4 \cdot 10^3 \cdot 52^5}{\binom{10}{2} \cdot \binom{8}{5} \cdot 5 \cdot 4 \cdot 10^3 \cdot 52^5} = \dfrac{9}{\binom{10}{2}} = \dfrac{2}{10}}$

Problem 3. (25 pts: 10,5,5,5)
You have 5 four-sided and 3 six-sided dice. You put them in a cup, choose one at random, roll the die, and report the result.

Let D be the number of sides on the chosen die and let R be the result of the roll.

(a) *Make a joint probability table for D and R. Be sure to include the marginal probabilities for D and R.*

Solution: Each element of the table is simply the probability of getting a die with the indicated number of sides and then rolling the indicated number. For example,

$$P(R = 3 \text{ and } D = 6) = P(R = 3|D = 6)P(D = 6) = \frac{1}{6} \cdot \frac{3}{8} = \frac{1}{16}.$$

$R \backslash D$	4-sided	6-sided	
1	5/32	1/16	7/32
2	5/32	1/16	7/32
3	5/32	1/16	7/32
4	5/32	1/16	7/32
5	0	1/16	1/16
6	0	1/16	1/16
	5/8	3/8	

(b) *What is the probability of rolling a 3?*

Solution: This is the sum of the entries in the $R = 3$ row of the table:

$$5/32 + 1/16 = 7/32$$

(Do you see why this has to be between 1/6 and 1/4?)

(c) *Compute $P(D = 4|R = 3)$.*

Solution: We compute this as the fraction

$$\frac{P(D = 4 \text{ and } R = 3)}{P(R = 3)} = \frac{5/32}{7/32} = 5/7.$$

(d) *Are D and R independent?*

Solution: No, the joint probabilities in the table are not the products of the marginal probabilities. The easiest way to see this is to note that $P(R = 6 \text{ and } D = 4) = 0$, which does not equal $P(R = 6)P(D = 4) = 5/128$.

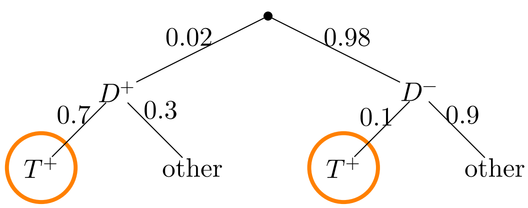
Problem 4. (10 pts)
A quick screening test for a certain disease has three outcomes: positive, negative and uncertain. Suppose it has the following percentages.

For someone with the disease: positive 70%, negative 10%, uncertain 20%.
For someone without the disease: positive 10%, negative 60%, uncertain 30%.

Suppose also, that the prevalence of the disease in the population is 2%.

What is the probability that a random person who tests positive has the disease?

Solution: We organize the problem in a tree. Here: D^+ = has disease, D^- = does not have disease; T^+ = test is positive, other = test is negative or uncertain.



$$P(D^+|T^+) = \frac{P(T^+|D^+)P(D^+)}{P(T^+)} = \frac{0.7 \cdot 0.02}{0.7 \cdot 0.02 + 0.1 \cdot 0.98} = \boxed{\frac{14}{112} = \frac{1}{8} = 0.125}.$$

Problem 5. (25 pts: 5,5,5,5,5)
Two students, Xeno and Yolanda are meeting up for lunch. They plan on a time to meet at noon. Both have class before so neither will be early. Both have class that starts at 1pm, so they will both arrive between 0 and 1 hour late. Let X be the time in hours that Xeno arrives late and let Y be the time in hours that Yolanda arrives late.

Assume that the joint pdf of these random variables is $f(x, y) = 5/4 - xy$.

(a) *Find the two marginal pdfs.*

Solution: To find the marginals we 'integrate out' the other variable.

$$f_X(x) = \int_0^1 f(x, y)\, dy = \int_0^1 5/4 - xy\, dy = \boxed{5/4 - x/2.}$$
$$f_Y(y) = \int_0^1 f(x, y)\, dx = \int_0^1 5/4 - xy\, dx = \boxed{5/4 - y/2.}$$

We could have used symmetry to deduce $f_Y(y)$ without any integration.

(b) *Are X and Y independent?*

Solution: Since the joint pdf is not the product of the marginals, i.e. $f(x, y) \neq f_X(x)f_Y(y)$, X and Y are not independent.

(c) *Find $E[X]$, $\text{Var}(X)$.* **(For these, you need to simplify the fractions.)**

Solution: We compute both $E[X]$ and $\text{Var}(X) = E[X^2] - E[X]^2$ using the marginal pdf $f_X(X)$ found in part (a).

$$E[X] = \int_0^1 x f_X(x)\,dx = \int_0^1 5x/4 - x^2/2\,dx = \frac{5}{8} - \frac{1}{6} = \boxed{\frac{11}{24}}.$$

$$E[X^2] = \int_0^1 x^2 f_X(x)\,dx = \int_0^1 5x^2/4 - x^3/2\,dx = \frac{5}{12} - \frac{1}{8} = \frac{7}{24}.$$

$$\text{Var}(X) = E[X^2] - E[X]^2 = \frac{7}{24} - \frac{11^2}{24^2} = \boxed{\frac{47}{24^2}}.$$

(d) *Compute the covariance* $\text{Cov}(X, Y)$ *and correlation* $Cor(X, Y)$.

Hint: By symmetry you know the mean and variance of Y *are the same as those for* X.

For this part, there is no need to simplify fractions.

Solution: By symmetry, we know $E[Y] = E[X] = 11/24$ and $\text{Var}(Y) = \text{Var}(X) = 47/24^2$.

We use the formula $\text{Cov}(X, Y) = E[XY] - E[X]E[Y]$.

$$E[XY] = \int_0^1 \int_0^1 xyf(x, y)\,dx\,dy = \int_0^1 \int_0^1 5xy/4 - x^2y^2\,dx = \frac{5}{16} - \frac{1}{9} = \frac{29}{144}.$$

$$\text{Cov}(X, Y) = E[XY] - E[X]E[Y] = \boxed{\frac{29}{144} - \frac{11^2}{24^2} = \frac{29}{144} - \frac{121}{144 \cdot 4} = -\frac{5}{144 \cdot 4}}$$

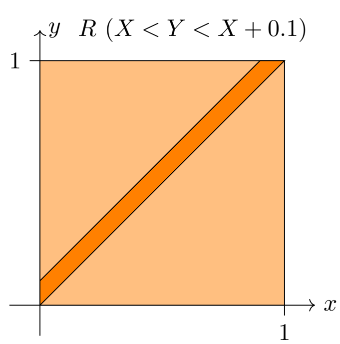
$$\text{Cor}(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} = \boxed{\frac{\text{Cov}(X, Y)}{47/24^2} = \frac{-5/24^2}{47/24^2} = \frac{-5}{47}}$$

(e) *Set up,* **but do not compute** *an expression computing the probability that Xeno and Yolanda arrive within 6 minutes (0.1 hours) of each other and that Yolanda arrives after Xeno.*

Your integral will be over a region R *in the unit square. You can leave your integral in the form* $\iint_R h(x, y)\,dx\,dy$ *and show* R *in a figure elsewhere on the page. The function* $h(x, y)$ *must be specified completely.*

Solution: The integral is $\iint_R f(x, y)\,dx\,dy = \iint_R 5/4 - xy\,dx\,dy.$

The region R is the part of the unit square where $X < Y$ and $Y - X < 0.1$. This is the strip of the triangle shown in the picture



This was not asked for, but using 18.02 we get

$$P(X < Y < X + 0.1) = \int_0^{0.9} \int_x^{x+0.1} 5/4 - xy \, dy \, dx + \int_{0.9}^1 \int_x^1 5/4 - xy \, dy \, dx$$

Problem 6. (10 pts)

A company manufactures solar panels. When homeowners install the panels, the state pays 50% of the cost. Because this subsidy is about to expire, the company wants to manufacture as many panels as it can in the next 20 days.

For a variety of reasons the number of panels it can manufacture in a day is a random variable with each day independent of the others. Suppose the daily output follows a so-called gamma distribution. The pdf of this distribution is not that complicated ($f(x) = \frac{x^4}{4! \cdot 10^{10}} e^{-x/100}$), but we'll let Wikipedia tell us the mean and variance: mean = 500, variance = $5 \cdot 10^4$.

***Estimate** the probability that they will be able to manufacture more than 10,500 panels in the next 20 days.*

Solution: Let S be the total manufactured in 20 days. The problem asks for $P(S > 10500)$.

Since S is a sum of 20 i.i.d. random variables, the central limit theorem tell us that it is approximately normal. We know that one day has mean 500 and variance $5 \cdot 10^4$. So

$$E[S] = 20 \cdot 500 = 10000 \quad \text{Var}(S) = 20 \cdot 5 \cdot 10^4 = 10^5 \quad \sigma_S = 10^3.$$

Standardizing and using the CLT we get

$$\begin{aligned} P(S > 10500) &= P\left(\frac{S - 10,000}{1000} > \frac{10,500 - 10,000}{1000}\right) \\ &\approx P(Z > 0.5) = 1 - P(Z \le 0.5) \approx 1 - 0.6915 = 0.3085 \end{aligned}$$

The decimal answer came by looking up $P(Z < 0.5) \approx 0.6915$ in the standard normal table.