Problem 1: (1 points) If X is a random variable with $E(X) = 2$, what is $E(2X - 1)$?
(You do not need to show your work or justify your answers for this problem.)

Choose one:

- ● 3
- ○ 4
- ○ 8
- ○ Not enough information

By linearity of expectation,
$\mathbb{E}(2x-1) = 2\mathbb{E}[x] - 1 = 2\cdot 2 - 1 = 3.$

Problem 2: (1 points) Let X be the outcome of a single fair die roll. What is $p_X(5)$?
(You do not need to show your work or justify your answers for this problem.)

Choose one:

- ○ 3.5
- ○ 5/6
- ● 1/6
- ○ Not enough information

By definition of PMF from class, $p_X(5) = P(X=5) = \frac{1}{6}$.

Problem 3: (3 points) If A, B, and C are mutually independent events with $P(A) > 0$ and $P(B) > 0$ and $P(C) > 0$, which of the following are always true?
(You do not need to show your work or justify your answers for this problem.)

Choose ALL that apply:

- ■ $P(A \cap B) = P(A) \cdot P(B)$
- □ $P(A \cup B^c) = P(A) + P(B^c)$
- ■ $P(A|B \cap C) = P(A)$

The definition of mutual independence is that A, B, & C are mutually independent if

$P(A\cap B) = P(A)P(B)$ (1)
$P(B\cap C) = P(B)P(C)$ (2)
$P(A\cap C) = P(A)P(C)$ (3)
$P(A\cap B\cap C) = P(A)P(B)P(C)$. (4)

Therefore, the first is true (line 1 above). The third is true because $P(A|B\cap C) = \frac{P(A\cap B\cap C)}{P(B\cap C)} = \frac{P(A)P(B)P(C)}{P(B)P(C)} = P(A)$

↑ lines (2) and (4)

The second is false because by inclusion-exclusion
$P(A\cup B^c) = P(A) + P(B) - P(A\cap B^c)$.
Because A & B independent means A & B^c independent,
$P(A\cup B^c) = P(A) + P(B) - P(A)P(B^c)$.
So this is false when $P(A), P(B^c) > 0$.

Problem 4: (7 points) You decide you want to build a spam filter to delete some of your many spam emails. You make the following observations:

- **50%** of all your emails are spam
- **20%** of your spam emails contain the phrase "CONSTRUCTION ALERT"
- Only **.1%** of your non-spam emails contain the phrase "CONSTRUCTION ALERT"

Let S be the event that a randomly selected email is spam, and let C be the event that it contains the phrase "CONSTRUCTION ALERT." So your observations can be written as: $P(S) = .5$ and $P(C|S) = .2$ and $P(C|S^c) = .001$. Find the following probabilities.

(a) (3 points) $P(C)$ (the overall probability that an email will contain the phrase "CONSTRUCTION ALERT")

"S" spam
"C" alert

$P(S) = .5$
$P(C|S) = .2$
$P(C|S^c) = .001$

$P(C) = P(C|S)P(S) + P(C|S^c)P(S^c)$

$P(C)$?

$P(C) = (.2)(.5) + (.001)(.5)$

Answer: $(.2)(.5) + (.001)(.5)$

(b) (4 points) $P(S|C)$ (the probability that an email is spam given that it contains the phrase "CONSTRUCTION ALERT")

$P(S|C) = \frac{P(S\cap C)}{P(C)} = \frac{0.5\times 0.2}{0.1005} = \frac{0.1}{0.1005}$

Answer: $P(S|C) = \frac{0.1000}{0.1005}$

Problem 5: (8 points) Let X be a random variable that takes the values -2, 0, 2, and 3 (and no other values), with the following probabilities:

$P(X = -2) = 1/2$
$P(X = 0) = 1/12$
$P(X = 2) = 1/4$
$P(X = 3) = 1/6$

(a) (2 points) Find $E(X)$.

$E(X) = -2(\frac{1}{2}) + 0\cdot\frac{1}{12} + 2\cdot\frac{1}{4} + 3(\frac{1}{6})$

Answer: $E(X) = -1 + \frac{1}{2} + \frac{1}{2} = 0$

(b) (4 points) Let $Y = X^2$. Find the probability mass function of Y.

~~PMF of Y:~~ Because $Y = X^2$, So $Y = (-2)^2, 0^2, 2^2$ and 3^2
$= 4, 0, 4, 9$.

Because $P(X=-2) = 1/2$, and $P(X=2) = 1/4$
So $P(Y=4) = 1/2 + 1/4 = 3/4$
And $P(Y=0) = P(X=0) = 1/12$
$P(Y=9) = P(X=3) = 1/6$

So $p_Y(k) = \begin{cases} 3/4 & \text{if } k=4 \\ 1/12 & \text{if } k=0 \\ 1/6 & \text{if } k=9 \\ 0 & \text{if else.} \end{cases}$

Problem 5 (continued)
(Copied from the previous page for reference)
$P(X = -2) = 1/2$
$P(X = 0) = 1/12$
$P(X = 2) = 1/4$
$P(X = 3) = 1/6$

(c) (2 points) Find $F_Y(1)$, where $Y = X^2$.

$F_Y(1) = P(Y \le 1)$
$= P(Y=0)$
$= \frac{1}{12}$

A jar contains yellow jelly beans and pink jelly beans. You randomly draw jelly beans one at a time from the jar **without replacement** until you get a pink one (pink ones are your favorite). What is the expected number of yellow jelly beans you draw until the first pink one?

There was much confusion on this question, so we will give three different solutions to it. The first one is the intended solution, the second is closest to what many students put on the exam, and the third one is too pretty to not give.

0.1 Solution 1

We claim the expected number of yellow jelly beans drawn before the first pink one is exactly 5. We will use indicators and linearity of expectation to prove it. The key step with any use of indicator variables is defining our event appropriately. So, we will assume that all 119 jelly beans each have a unique number $1, \ldots, 119$. For convenience, we assume that the yellow jelly beans have indices $1, \ldots, 100$ (this will make the notation easier). Then we will pick a uniformly random ordering on these 119 jelly beans. Now for the key step. For each yellow jelly bean i, we define the event A_i as follows:

$$A_i := \text{Every pink jelly bean appears after } i \text{ in the ordering}$$

Clearly, if A_i holds for some yellow jelly bean, that jelly bean must be one of the jelly beans that we draw before picking our first pink jelly bean. More formally, if X is the number of yellow jelly beans drawn before picking the first pink one, then

$$X = 1_{A_1} + \cdots + 1_{A_{100}}$$

(remember that the yellow jelly beans are the first 100 indices in the ordering). Thus we can use linearity of expectation to find

$$\mathbb{E}[X] = \mathbb{E}[1_{A_1} + \cdots + 1_{A_{100}}] = \mathbb{E}[1_{A_1}] + \cdots + \mathbb{E}[1_{A_{100}}]$$

Furthermore, we picked a uniformly random ordering on the jelly beans, so the events A_i all have the same probability. So our desired expectation is

$$\mathbb{E}[1_{A_1}] + \cdots + \mathbb{E}[1_{A_{100}}] = 100\mathbb{E}[1_{A_1}] = 100\mathbb{P}[A_1]$$

as the expected value of an indicator variable is the probability of the event indicated. So all we have to do is find the probability that A_1 holds, i.e. that yellow jelly bean number 1 appears before every pink jelly bean in a uniformly random ordering.

To that end, we observe that A_1 only depends on the relative positions of 1 and the pink jelly beans, not on the positions of the other jelly beans. More formally we can say that A_1 is independent of the ordering of the other yellow jelly beans. (Note that the *absolute* position of jelly bean 1 in the ordering is very much dependent on the positions of the other yellow jelly beans, and only the *relative* position of jelly bean 1 to the pink jelly beans is independent of the positions of the other yellow jelly beans). So we can ignore the positions of the other yellow jelly beans entirely, and just focus on the ordering of the yellow jelly bean 1 and the 19 pink jelly beans. In order for jelly bean 1 to appear before all the pink jelly beans, it must occur first in the relative ordering on these 20 jelly beans. Hence,

$$\mathbb{P}\left[A_1\right] = \frac{\#\{\text{relative orderings in which 1 appears first}\}}{\#\{\text{relative orderings}\}} = \frac{1 * 19!}{20!} = \frac{1}{20}$$

as we can pick any ordering on the pink jelly beans after placing the single jelly bean at the first position in the relative ordering. Therefore,

$$\mathbb{E}\left[X\right] = 100\mathbb{P}\left[A_1\right] = \frac{100}{20} = 5$$

0.2 Solution 2

Instead of using indicator variables, we can directly apply the definition of expectation. **Don't do this**.

If you want to see the gory details, here they are. Define X to be the number of yellow jelly beans drawn before picking the first pink jelly bean. Since we are drawing without replacement, X can take on any integer value from 0 (we pick a pink bean on the first draw), to 100(we pick all 100 yellow beans before picking a pink one). Then by the definition of expectation,

$$\mathbb{E}\left[X\right] = \sum_{k=0}^{100} k * \mathbb{P}\left[X = k\right]$$

. We will first So now we have to find $\mathbb{P}\left[X = k\right]$.

To estimate it, we observe that on our first draw we have 100 ways to pick a yellow jelly bean out of 119 ways to pick a jelly bean overall. Then on our second draw, we have removed one jelly bean, so we have 99 ways to pick a yellow jelly bean out of 118 ways overall. The pattern continues so that on our kth draw we have $100-k+1$ ways to draw a yellow jelly bean out of $119-k+1$ ways to draw a jelly bean overall *as long as we have only drawn yellow jelly beans until this draw*. Finally, on draw $k+1$, we must pick a yellow jelly bean, which we can do in 19 ways out of $119 - k$ ways overall. Therefore, we can express $\mathbb{P}\left[X = k\right]$ as

$$\mathbb{P}\left[X = k\right] = \frac{100}{119}\frac{99}{118}\cdots\frac{100-k+1}{119-k+1}\frac{19}{199-k} = \frac{19}{199-k}\prod_{i=1}^{k}\frac{100-i+1}{119-i+1}$$

.

We can fill in our expectation formula.

$$\sum_{k=0}^{100} k*\mathbb{P}\left[X = k\right] = \sum_{k=0}^{100} k\frac{19}{119-k}\prod_{i=1}^{k}\frac{100-i+1}{119-i+1} = 19\sum_{k=1}^{100}\frac{k}{119-k}\prod_{i=1}^{k}\frac{100-i+1}{119-i+1}$$

where we observe that the term in the sum is 0 when $k = 0$. Now comes the unpleasant part; we have to simplify this beast[1].

In fact, we will not simplify this series directly[2]. Instead, we will use the bonus problem from homework 3 which states that if Y is a random variable taking only nonnegative integer values, then

$$\mathbb{E}[Y] = \sum_{i=1}^{\infty} \mathbb{P}[Y \geq i]$$

. In our case, X only takes the values 0 to 100, so X certainly qualifies for this condition. Thus we can write

$$\mathbb{E}[X] = \sum_{k=1}^{\infty} \mathbb{P}[X \geq k] = \sum_{k=1}^{100} \mathbb{P}[X \geq k]$$

as we know that X can be at most 100. By similar reasoning as above,

$$\mathbb{P}[X \geq k] = \prod_{i=1}^{k} \frac{100 - i + 1}{119 - i + 1}$$

(notice we do not have the the $\frac{19}{119-k}$ term. This is the reason why we did not simplify directly) So now we have the (slightly more manageable) sum

$$\sum_{k=1}^{100} \prod_{i=1}^{k} \frac{100 - i + 1}{119 - i + 1}$$

We begin by noticing that the product at value k is simply one additional term multiplied by the product at the value $k - 1$. Hence, the sum *telescopes* in the following manner:

$$\sum_{k=1}^{100} \prod_{i=1}^{k} \frac{100 - i + 1}{119 - i + 1} = \frac{100}{119}\left(1 + \frac{99}{118}\left(1 + \dots \frac{3}{22}\left(1 + \frac{2}{21}\left(1 + \frac{1}{20}\right)\right)\dots\right)\right)$$

. This is not the same as the product of 100 terms, but instead is a nested product with 100 pairs of parentheses. Notice that $1 + \frac{1}{20} = \frac{21}{20}$, so the innermost term of the sum is $\frac{2}{21}\frac{21}{20} = \frac{2}{20}$. The next term is then $\frac{3}{22}\left(1 + \frac{2}{20}\right) = \frac{3}{2}\frac{22}{20} = \frac{3}{20}$. This pattern also continues until we reach the outermost term of $\frac{100}{119}\frac{119}{20} = \frac{100}{20} = 5$. Thus we will draw 5 yellow jelly beans in expectation before drawing the first pink bean.

[1] Otherwise known a hypergeometric series

[2] If you can figure out how to do simplify it directly, please let us know. We would be very interested to see how.

0.3 Solution 3

Finally, we have a very elegant solution which is similar in spirit to the first solution, but has a few tricks of its own. Instead of placing a random ordering on the 119 jelly beans at the start, we begin by placing them in a circular ordering and then picking a random point at which to break the circle and form an ordering of the remaining beans. More formally, we consider a circle with 120 spaces on it, and we place uniformly at random the 110 yellow beans, the 19 pink beans, and 1 white jelly bean which will serve to mark the break-point for the ordering. Now we can construct an ordering of the yellow and pink jelly beans by finding the white jelly bean, removing it, and then moving around the circle clockwise. We define $B_1, \ldots, B_{20}$ to be the number of yellow jelly beans between the consecutive occurrences of white/pink jelly beans. In particular, B_i is the number of yellow jelly beans between the $i-1$st pink bean (or the white bean if $i = 1$) and the ith pink bean (or the white bean if $i = 20$). Notice that every yellow jelly bean must be counted by exactly one B_i. Thus,

$$B_1 + \cdots + B_{20} = 100$$

no matter which circular ordering we pick. It follows that

$$\mathbb{E}[B_1] + \cdots + \mathbb{E}[B_{20}] = \mathbb{E}[100] = 100$$

. Notice the random variables B_i all have the same distribution as each circular ordering is equally likely. Thus,

$$100 = \mathbb{E}[B_1] + \cdots + \mathbb{E}[B_{20}] = 20\mathbb{E}[B_1]$$

from which it follows that $\mathbb{E}[B_1] = 5$. Thus in expectation, we draw 5 yellow jelly beans before picking the first pink jelly bean.