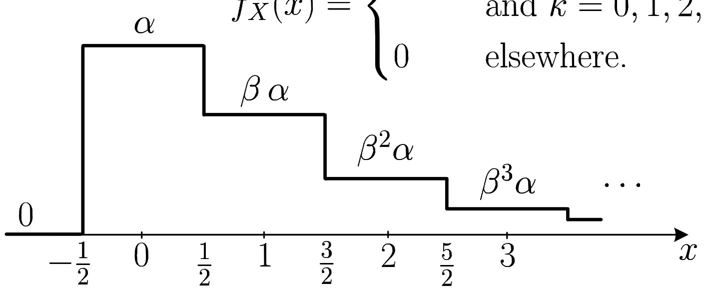
EECS 70 Discrete Math and Probability
Spring 2020 UC Berkeley

Midterm 3

1. Stepping Stones (22 Points)

Consider a continuous random variable X whose PDF is illustrated in the figure below, where $0 < \alpha$ and $0 < \beta < 1$.

$$f_X(x) = \begin{cases} \beta^k \alpha & \text{for } x \in [k - 1/2, k + 1/2), \\ & \text{and } k = 0, 1, 2, 3, \ldots \\ 0 & \text{elsewhere.} \end{cases}$$



In one or more parts of this problem, you may or may not find it useful to know that for any $\gamma \in \mathbb{R}$ such that $|\gamma| < 1$, the following identities hold:

$$\sum_{k=0}^{\infty} \gamma^k = \frac{1}{1-\gamma} \quad \text{and} \quad \sum_{k=0}^{\infty} k\gamma^k = \frac{\gamma}{(1-\gamma)^2}$$

(a) (3 pts) **Determine** $\Pr(X \geq 0)$**. Justify your answer for full credit.** Express your answer in terms of in terms of α, β, or both.

(HINT: What is $\Pr(X \geq 0)$ *in terms of* $\Pr(X < 0)$*?)*

Solutions: We recognize that $\Pr(X \geq 0) = 1 - \Pr(X < 0)$. Looking at the PDF $f_X(x)$, we note that $\Pr(X < 0) = \int_{-\infty}^{0} f_X(x)dx = \alpha/2$ by inspection since that is the area under the given curve up to 0. Therefore,

$$\Pr(X \geq 0) = 1 - \Pr(X < 0) = 1 - \frac{\alpha}{2}.$$

(b) (5 pts) **Determine α in terms of β. Justify your answer for full credit.**

(HINT: What is the relevant condition that any valid PDF must satisfy?)

Solutions: The total area under the PDF $f_X(x)$ must be one. Since each of the rectangular steps has a base of length 1, the total area is:

$$\sum_{k=0}^{\infty} \beta^k \alpha = 1, \quad \text{so } \frac{\alpha}{1-\beta} = 1, \quad \text{so} \quad \alpha = 1 - \beta.$$

(c) (8 pts) Consider a new random variable Y that represents rounding X to the nearest integer. That is, $Y = k$ whenever $X \in [k-\frac{1}{2}, k+\frac{1}{2})$. **Determine $\mathbb{E}[Y]$. Show all your work for full credit.** If you're unsure of your expression for α in the previous part, you may leave your expression for the mean in terms of α and β.

(HINT: For natural numbers k, what is $\Pr\big(X \in [k-\frac{1}{2}, k+\frac{1}{2})\big)$*? This can help you make progress towards being able to use the identities given above.)*

Solutions: By taking the area under the PDF for the range $[k-\frac{1}{2}, k+\frac{1}{2})$, we see that $\Pr\big(X \in [k-\frac{1}{2}, k+\frac{1}{2})\big) = \beta^k\alpha$, which implies that $\Pr(Y = k) = \beta^k\alpha$. Since Y can take on any non-negative integer value, we find its expectation as follows:

$$\mathbb{E}[Y] = \sum_{k=0}^{\infty} k \Pr(Y = k) = \sum_{k=0}^{\infty} k\beta^k\alpha = \alpha \sum_{k=0}^{\infty} k\beta^k = \alpha \cdot \frac{\beta}{(1-\beta)^2}$$

If you plug in $\alpha = 1-\beta$ from the previous part, this becomes $\mathbb{E}[Y] = \dfrac{\beta}{1-\beta}$.

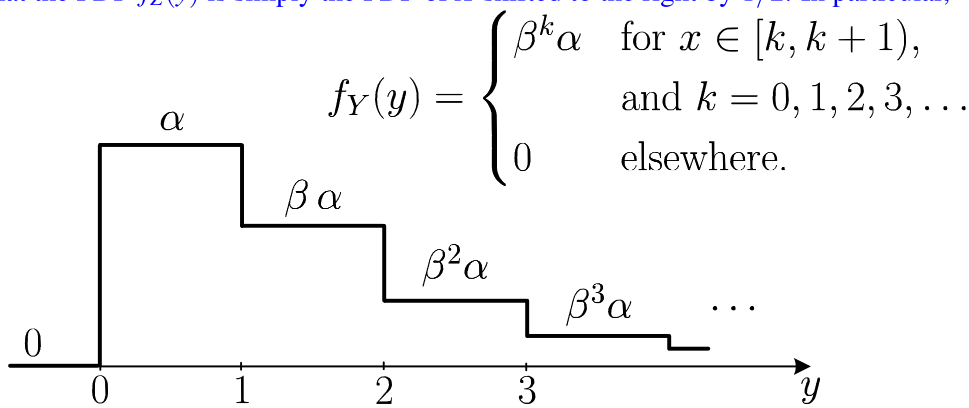
You could also have noticed that this random variable Y is a geometric random variable with 1 subtracted from it. So $\frac{1}{1-\beta} - 1 = \frac{1-(1-\beta)}{1-\beta} = \frac{\beta}{1-\beta}$.

(d) (6 pts) **Show that**:

$$\Pr(X \geq b) \leq \frac{\mathbb{E}[X] + \frac{1}{2}}{b + \frac{1}{2}}, \quad \text{for all} \quad b > -\frac{1}{2}.$$

(HINT: Let $c = b + \frac{1}{2}$ *and rewrite what you want to show in terms of c. Does this remind you of anything? Can* $Z = X + \frac{1}{2}$ *ever be negative?)*

Solutions: Define a new random variable Z such that $Z = X + 1/2$. Clearly, Z is nonnegative. We note that the PDF $f_Z(y)$ is simply the PDF of X shifted to the right by $1/2$. In particular,



Therefore,

$$\mathbb{E}[Z] = \mathbb{E}[X + 1/2] = \mathbb{E}[X] + 1/2$$

Since Z is a nonnegative random variable, we can invoke Markov's Inequality:

$$\Pr(Z \geq c) \leq \frac{\mathbb{E}[Z]}{c} \quad \text{for all } c > 0$$

$$\Pr(X + 1/2 \geq c) \leq \frac{\mathbb{E}[X] + 1/2}{c}$$

$$\Pr(X \geq c - 1/2) \leq \frac{\mathbb{E}[X] + 1/2}{c}.$$

Defining $b = c - 1/2$, we have $b > -\frac{1}{2}$ and $c = b + 1/2$, so

$$\Pr(X \geq b) \leq \frac{\mathbb{E}[X] + 1/2}{b + 1/2} \quad \text{for all} \quad b > -\frac{1}{2}.$$

2. Disks on a Chess Board (12 pts)

Let n be a positive integer. Say we have $4n$ disks. We want to place them each on a chess board of size $2n \times 2n$ tiles, one at a time. Assume that each tile may host any number of disks, and that when placing each disk, all $(2n)^2$ tiles are equally likely to be selected for disk placement. Different disks are placed independently — say by rolling a $(2n)^2$-sided fair die.

Let the random variable X denote the total number of disks that are placed on the first row of tiles after all disks have been placed.

Let the random variable Y denote the total number of disks that are placed on the second row of tiles after all disks have been placed.

Because there is nothing different about the first row and the second row, X and Y are identically distributed.

Determine $\mathbb{E}[X]$ and Var(X) in terms of n.

Solutions: Write X as a sum of indicators $\sum_{i=1}^{4n} X_i$, where X_i is 1 if the i-th disk was placed in the first row and 0 otherwise. We note that $X_i \sim \text{Bernoulli}(p)$, where $p = \frac{2n}{(2n)^2} = \frac{1}{2n}$ is the probability that the i-th disk was placed in the first row. So $\mathbb{E}[X_i] = \frac{1}{2n}$ and $\text{Var}(X) = (\frac{1}{2n})(1 - \frac{1}{2n})$.

So by linearity of expectation:

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^{4n} X_i\right] = \sum_{i=1}^{4n} \mathbb{E}[X_i] = 4n \cdot \frac{1}{2n} = 2$$

And since all the X_i are independent (each disk placement doesn't affect the placement of any other disk), we can simply sum the indicators' variances as well:

$$\text{Var}(X) = \text{Var}\left(\sum_{i=1}^{4n} X_i\right) = \sum_{i=1}^{4n} \text{Var}(X_i) = 4n \cdot \frac{1}{2n}\left(1 - \frac{1}{2n}\right) = 2 - \frac{1}{n}$$

Alternate solution: One could notice that $X \sim \text{Binom}(4n, \frac{1}{2n})$ and get the expectation and variance for free. Each placement of a disk is like an independent trial, where "success" is defined as being placed in the first row.

Note: It is interesting to note that as n gets large, the X random variable here approaches a Poisson(2) random variable. Notice how the variance is approaching 2 as well as the mean.

However, X and Y are not independent since the number of disks in the first row clearly constrains the number of disks that can be in the second row.

Determine the joint distribution of X and Y, i.e. $\Pr[X = x, Y = y]$ for all $x, y \in \{0, 1, \ldots, 4n\}$. Show all your work for full credit.

(HINT: It might be helpful to write X as a sum of appropriate indicator random variables for the first part of this.)

Solutions: We split $\Pr[X = x, Y = y]$ into $\Pr[X = x]\Pr[Y = y|X = x]$ and calculate each of the terms. As X is $\text{Binom}(4n, \frac{1}{2n})$, we have $\Pr[X = x] = \binom{4n}{x}(\frac{1}{2n})^x(1 - \frac{1}{2n})^{4n-x}$. When conditioned on the event $X = x$, Y is also binomially distributed, but with only $4n - x$ trials and probability of success being $\frac{1}{2n-1}$ (the first row is out of play now). So $\Pr[Y = y|X = x] = \binom{4n-x}{y}(\frac{1}{2n-1})^y(1 - \frac{1}{2n-1})^{4n-x-y}$ and the joint probability is:

$$\binom{4n}{x}\left(\frac{1}{2n}\right)^x\left(1 - \frac{1}{2n}\right)^{4n-x}\binom{4n-x}{y}\left(\frac{1}{2n-1}\right)^y\left(1 - \frac{1}{2n-1}\right)^{4n-x-y}, \text{ for } x + y \le 4n, \text{ 0 otherwise}$$

Alternate solution: To calculate $\Pr[X = x, Y = y]$, we can realize that there are $\binom{4n}{x}$ to choose the x disks to go in the first row, and $\binom{4n-x}{y}$ ways to choose the y disks to go on the second row (out of the remaining $4n - x$ which were not chosen to go on the first row). Once we've assigned disks to go in the first and second row, there is a $(\frac{1}{2n})^{x+y}$ probability that the $x+y$ disks we have assigned actually do end up in their assigned row (either the first or second row), and a $(1 - \frac{2}{2n})^{4n-(x+y)}$ probability that the remaining disks don't get placed in the first or second row. So we have:

$$\Pr[X = x, Y = y] = \binom{4n}{x}\binom{4n-x}{y}\left(\frac{1}{2n}\right)^{x+y}\left(1 - \frac{1}{n}\right)^{4n-(x+y)}$$

which can be shown to be equivalent to our original solution. This is essentially akin to a derivation for a modified binomial distribution where there are two different variants of "success".

Yet another alternate solution: First we can find the probability that a total of $x+y$ disks were placed on the first and second row in total, which is $\binom{4n}{x+y}(1/n)^{x+y}(1-1/n)^{4n-(x+y)}$ (this is just $\Pr\left(\text{Binom}(4n, \frac{2}{2n}) = x+y\right)$). Then we can find the conditional probability that exactly x disks (out of the $x+y$ we are given to be in the first two rows) landed in the first row and the remaining y landed in the second. If a disk lands in the first or second row, it had equal probability of landing in one or the other, so this probability is $\binom{x+y}{x}(1/2)^{x+y}$ (this is just $\Pr(\text{Binom}(x+y, 1/2) = x)$). So we get:

$$\Pr[X = x, Y = y] = \binom{4n}{x+y}\left(\frac{1}{n}\right)^{x+y}\left(1 - \frac{1}{n}\right)^{4n-(x+y)}\binom{x+y}{x}\left(\frac{1}{2}\right)^{x+y}$$

Note: Again, it is interesting to note that as n gets large, the random variables X and Y approach independent Poisson(2) random variables. We can actually show that our above expression for the joint probability $\Pr[X = x, Y = y]$ approaches that of a joint Poisson distribution. If X and Y were independent Poisson(2), then we would have:

$$\Pr[X = x, Y = y] = \Pr[X = x]\Pr[Y = y] = \frac{2^x}{x!}e^{-2} \cdot \frac{2^y}{y!}e^{-2} = \frac{1}{x!y!} \cdot 2^{x+y}e^{-4}$$

Now we take our originally derived expression for $\Pr[X = x, Y = y]$, apply Stirling's approximation, and show that as n grows large this probability converges to the above joint Poisson probability.

$$\begin{aligned}
\Pr[X = x, Y = y] &= \binom{4n}{x}\binom{4n-x}{y}\left(\frac{1}{2n}\right)^{x+y}\left(1 - \frac{1}{n}\right)^{4n-(x+y)} \\
&= \frac{(4n)!}{x!y!(4n-(x+y))!} \cdot \left(\frac{1}{2n}\right)^{x+y}\left(1 - \frac{1}{n}\right)^{4n-(x+y)} \\
&\approx \frac{1}{x!y!}\frac{\sqrt{2\pi(4n)}(4n/e)^{4n}}{\sqrt{2\pi(4n-(x+y))}((4n-(x+y))/e)^{4n-(x+y)}}\left(\frac{1}{2n}\right)^{x+y}\left(1 - \frac{1}{n}\right)^{4n-(x+y)} \\
&= \frac{1}{x!y!} \cdot \frac{\sqrt{4n}}{\sqrt{4n-(x+y)}} \cdot \frac{(4n)^{4n}}{(4n-(x+y))^{4n-(x+y)}} \cdot e^{-(x+y)}\left(\frac{1}{2n}\right)^{x+y}\left(1 - \frac{1}{n}\right)^{4n-(x+y)} \\
&= \frac{1}{x!y!} \cdot \frac{\sqrt{4n}}{\sqrt{4n-(x+y)}} \cdot (4n-(x+y))^{x+y} \cdot \left(\frac{1}{1 - \frac{x+y}{4n}}\right)^{4n} \cdot e^{-(x+y)}\left(\frac{1}{2n}\right)^{x+y}\frac{((1-\frac{1}{n})^n)^4}{(1-\frac{1}{n})^{x+y}} \\
&\xrightarrow[n\to\infty]{} \frac{1}{x!y!} \cdot 1 \cdot (4n)^{x+y} \cdot \frac{1}{e^{-(x+y)}} \cdot e^{-(x+y)}\left(\frac{1}{2n}\right)^{x+y}\frac{e^{-4}}{(1-0)^{x+y}} \\
&= \frac{1}{x!y!} \cdot 2^{x+y}e^{-4}
\end{aligned}$$

Here we use the famous limit: $\left(1 + \frac{x}{n}\right)^n \xrightarrow[n\to\infty]{} e^x$.

3. Waiting in Line at Disneyland (19 pts)

Rohan, Sally, and Tom go to Disneyland and each choose to wait in line for different rides. Each person joins their respective line at the same time.

- Rohan chooses Radiator Spring Racers, which has a wait time R uniformly distributed between 0.0 and 1.0.
- Sally chooses Space Mountain, which has a wait time S that follows an exponential PDF with parameter $\lambda = 2$.
- Tom chooses Toy Story Mania, which has a wait time T that follows the distribution defined by the following PDF:

$$f_T(t) = \begin{cases} 2-2t & \text{if } t \in [0,1] \\ 0 & \text{elsewhere.} \end{cases}$$

Note that R, S, and T are all *continuous* and mutually independent random variables that have units of hours.

(a) (6 pts) **Determine $\mathbb{E}[T]$ and give an explicit expression for the CDF $F_T(t) = \Pr(T \leq t)$. Show all your work for full credit.** Be sure that your answer for $F_T(t)$ is well-defined for all real number inputs.

Solutions: By definition:

$$\mathbb{E}[T] = \int_{-\infty}^{\infty} t f_T(t)\, dt = \int_0^1 (2t - 2t^2)\, dx = \left(t^2 - \frac{2}{3}t^3\right)\Big|_0^1 = \frac{1}{3}$$

$$F_T(t) = \int_{-\infty}^{t} f_T(x)\, dx = \begin{cases} 0 & \text{if } t < 0 \\ \int_0^t (2-2x)\, dx = 2t - t^2 & \text{if } t \in [0,1] \\ 1 & \text{if } t > 1 \end{cases}$$

(b) (5 pts) Let A be the event that the wait time for Space Mountain is greater than h hours. **Determine $f_{S|A}(s)$, the conditional PDF of S given A, in terms of s and h. Show all your work for full credit.** Be sure that your answer is well-defined for all real number inputs s and h.

Also answer the following: **If Sally has already waited h hours, how much more time is she expected to wait before getting on the ride?** Justify your answer.

Solutions: By the memoryless property of exponential distributions, the amount of *additional* time Sally waits is still exponentially distributed with parameter $\lambda = 2$. So we have

$$f_{S|A}(s) = f_S(s-h) = \begin{cases} 2e^{-2(s-h)} & \text{if } s \geq h \\ 0 & \text{otherwise} \end{cases}$$

For this same reason, the expected further wait time after h hours have passed is still $\mathbb{E}[S] = 1/\lambda = \frac{1}{2}$ hours.

Note: A solution which explicitly found $\Pr(A) = e^{-2h}$ and divided the PDF of S (which is $2e^{-2s}$ at $s \geq h$) by that probability also received full credit, as long as you recognized that the conditional probability density for $s < h$ has to be zero. Then you could also integrate over this conditional PDF to find the answer to the second question.

(c) (8 pts) Let W be the random variable which represents the wait time (in hours) of the person (out of Rohan, Sally, and Tom) who got onto their respective ride *last*. **Explicitly determine $F_W(w)$, the CDF of W. Show all your work for full credit.** Be sure that your answer is well-defined for all real number inputs w.

(HINT: In addition to what you found in part (a), find $F_R(r)$ and $F_S(s)$, the CDFs of R and S, respectively. Doing so will earn partial credit as well as guide you towards the answer.)

Solutions: First we can explicitly state the CDFs of R and S (which are uniform and exponential random variables, respectively):

$$F_R(r) = \begin{cases} 0 & \text{if } t < 0 \\ r & \text{if } t \in [0,1] \\ 1 & \text{if } t > 1 \end{cases} \qquad F_S(s) = \begin{cases} 1 - e^{-2s} & \text{if } t \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Then we can realize that $W = \max\{R,S,T\}$, and that for W to be at most w, then R, S, and T must all also be at most w. Using this, and the CDF of T as found in part (a), we can calculate the CDF:

$$\begin{aligned} F_W(w) &= \Pr(W \leq w) = \Pr(\max\{R,S,T\} \leq w) \\ &= \Pr(R \leq w \cap S \leq w \cap T \leq w) \\ &= \Pr(R \leq w)\Pr(S \leq w)\Pr(T \leq w) \quad \text{by independence of } R,\ S, \text{ and } T \\ &= F_R(w)F_S(w)F_T(w) \\ &= \begin{cases} 0 & \text{if } w < 0 \\ w(1-e^{-2w})(2w-w^2) & \text{if } w \in [0,1] \\ 1-e^{-2w} & \text{if } w > 1 \end{cases} \end{aligned}$$

4. Transparent Coupon Collecting (10 pts)

Lalitha decides to go collecting coupons from transparent cereal boxes. There are 10 unique kinds of coupons and Lalitha wishes to collect one of each kind. Each shop has 5 cereal boxes in stock, and each box has exactly one coupon in it. Each box's coupon is independently randomly chosen, with the 10 kinds of coupons being equally likely. Lalitha can see which coupons are in which boxes before picking which one box to buy. Due to the current restrictions in place, she may buy at most one cereal box per shop and can't visit the same shop more than once. Still, Lalitha must collect all kinds of coupons, and wishes to *minimize* the number of shops visited.

What is the expected number of shops Lalitha visits? Show all your work for full credit. You may leave your answer as a summation.

(HINT: If Lalitha already has k distinct coupon types before walking into a shop, what is the probability that Lalitha does not find a new coupon type in the shop?)

Solutions: Following the hint, if Lalitha has k distinct coupons, then each cereal box has a $\frac{k}{10}$ chance of being one she already owns. This results in the current store having a $(\frac{k}{10})^5$ probability of being all boxes with coupons she already owns. Hence, with probability $p = 1-(\frac{k}{10})^5$, the store Lalitha visits will have a new coupon for her to collect. As with the standard Coupon Collector problem, we set $S = X_0 + X_2 + \cdots + X_9$. However, the X_k will now be distributed as a geometric distribution with $p_k = 1-(\frac{k}{10})^5$. Now by using linearity of expectation and plugging in the expectations of the geometric random variables as $1/p_k$, we get:

$$\mathbb{E}[S] = \mathbb{E}\left[\sum_{k=0}^{9} X_k\right] = \sum_{k=0}^{9} \mathbb{E}[X_k] = \sum_{k=0}^{9} \frac{1}{1-(\frac{k}{10})^5}$$

5. Trios (22 pts)

Suppose there are n students at a Homework Party. A TA walks around and gives each student a wristband that is either red, blue, or green, with equal probability for each. The students' wristband colors are independent from one another.

For this question, a "trio" is defined as an unordered *set* of three students. (For instance, when $n = 4$, if we think of the students as being A, B, C, D; then there are four different trios: ABC, ABD, ACD, BCD.)

A "same-color trio" is a trio in which all three members of the trio receive the same wristband color.

Let S be the random variable representing the number of same-color trios among n students, where $n \geq 6$ throughout this problem.

(a) (5 pts) **What is $\mathbb{E}[S]$ in terms of n? Show all your work for full credit.**
(HINT: Write S in terms of a sum of indicator random variables. What does each indicator random variable represent? How many possible trios of students are there, in terms of n? What is the probability that the specific trio ABC is a same-color trio? Answering these will earn partial credit as well as guide you towards the answer.)

Solutions: We have $\binom{n}{3}$ possible trios. A trio becomes a same-color trio with probability $1/9$. To see why, fix the color of the first student. Then each of the second and third students has a matching wristband's color with probability $1/3$. You can also think about this in counting terms: the first person has three choices, but the other two has only one each. Of all $3^3 = 27$ choices, a same-color trio happens with probability $(3 \cdot 1 \cdot 1)/27 = 1/9$.

If we let label all the trios from 1 to $\binom{n}{3}$ and let X_i be an indicator random variable that takes on a value of 1 when trio i is same-color and 0 otherwise and $S = \sum_{i=1}^{\binom{n}{3}} X_i$, then by Linearity of Expectation, the expected number of same-color trios is just $\frac{\binom{n}{3}}{9}$.

(b) (5 pts) With respect to any given trio (e.g. ABC), we can classify any other trio into one of three groups, based on the number of members it shares with the given trio. Let t_k be the number of trios that share k members with the given trio, where $k \in \{0, 1, 2\}$. (No need to define $t_3 = 1$ because there is only one trio that shares 3 members with a given trio, since three members define a unique trio.) Clearly, if t is the total number possible trios, $t = t_0 + t_1 + t_2 + 1$.

In terms of the total number of people n, determine an expression for each of t_0, t_1, and t_2. Justify your answer for full credit.

Solutions: If we want no overlaps, three entirely new people from the $n-3$ people not in the original trio need to be chosen: $t_0 = \binom{n-3}{3}$.

If we want one overlap, we have $\binom{3}{1}$ choices on which person belongs to both trios, and $\binom{n-3}{2}$ ways to choose the other two people: $t_1 = \binom{3}{1}\binom{n-3}{2} = 3 \cdot \binom{n-3}{2}$.

If we want two overlaps, we have $\binom{3}{2}$ choices on which two people belong to both trios, and $\binom{n-3}{1}$ ways to choose the other two people: $t_2 = \binom{3}{2}\binom{n-3}{1} = 3 \cdot (n-3)$.

(c) (2 pts) Assume we know that the trio ABC is a same-color trio. **Given that, what is the probability that the trio ABD is also a same-color trio?** Justify your answer for full credit.

Solutions: We know that A and B already have the same color, so all that remains is finding the probability that D also has that exact same color, which is $1/3$ since all three colors are equally likely for D.

(d) (10 pts) **What is $\mathbb{E}[S^2]$ in terms of n? Show all your work for full credit.** Whenever possible, use the terms t_0, t_1, and t_2 from part (b) instead of plugging in the actual expressions.

(HINT: You're going to have to expand out S^2 and then group terms by category. After you do that, the previous two parts can be helpful.)
(HINT 2: This part can take some time to work out. Unless you're confident, we suggest that you make sure you've attempted all problems before doing this part.)

Solutions: Define S as a sum of indicators X_i as we did in part (a) so $S = \sum_i X_i$. Now, we just need to expand out the desired quantity following the hint and using linearity of expectation:

$$\begin{aligned}
\mathbb{E}[S^2] &= \mathbb{E}\left[\left(\sum_i X_i\right)^2\right] &(1)\\
&= \mathbb{E}\left[\left(\sum_i X_i\right)\left(\sum_j X_j\right)\right] &(2)\\
&= \mathbb{E}\left[\sum_i \sum_j X_i X_j\right] &(3)\\
&= \sum_i \sum_j \mathbb{E}[X_i X_j] &(4)\\
&= \sum_i \left(\sum_{j=i} \mathbb{E}[X_i X_j] + \sum_{j \text{ shares 2 people with } i} \mathbb{E}[X_i X_j]\right.\\
&\quad \left. + \sum_{j \text{ shares 1 person with } i} \mathbb{E}[X_i X_j] + \sum_{j \text{ shares 0 people with } i} \mathbb{E}[X_i X_j]\right) &(5)
\end{aligned}$$

Here, we can look at each of those terms and compute the expectation. There's exactly one trio that is the same as the original trio i and so $\sum_{j=i} \mathbb{E}[X_i X_j] = \mathbb{E}[X_i^2] = \mathbb{E}[X_i] = \frac{1}{9}$ since X_i is an indicator random variable so $X_i^2 = X_i$.

There are t_2 different trios in the second summation term of (5) and each of them behave like the previous problem part with ABC and ABD. This means that $\sum_{j \text{ shares 2 people with } i} \mathbb{E}[X_i X_j] = t_2 \mathbb{E}[X_{ABC} X_{ABD}] = t_2 \mathbb{E}[X_{ABCD}]$ since for $X_{ABC} X_{ABD}$ to be 1, it has to be that all four of them have the same color. But that happens with probability $\frac{3}{3^4} = \frac{3}{81} = \frac{1}{27}$. This means that this entire term becomes $\frac{3t_2}{81}$.

There are t_1 different trios in the third summation term of (5) and each of them behaves like ABC and ADE. This means that $\sum_{j \text{ shares 1 person with } i} \mathbb{E}[X_i X_j] = t_1 \mathbb{E}[X_{ABC} X_{ADE}] = t_1 \mathbb{E}[X_{ABCDE}] = t_1 \frac{3}{3^5} = \frac{t_1}{81}$.

For the last term, there are t_0 different trios in the third summation term of (5) and each of them behaves like ABC and DEF. Since they share no members, the two corresponding indicator random variables are independent of each other. This means that $\sum_{j \text{ shares 0 people with } i} \mathbb{E}[X_i X_j] = t_0 \mathbb{E}[X_{ABC} X_{DEF}] = t_0 (\frac{1}{9})(\frac{1}{9}) = \frac{t_0}{81}$.
Putting all these together, we get that

$$\begin{aligned}
\mathbb{E}[S^2] &= \sum_i \left(\frac{1}{9} + \frac{3}{81} t_2 + \frac{1}{81} t_1 + \frac{1}{81} t_0\right) &(6)\\
&= \binom{n}{3} \left(\frac{1}{9} + \frac{3}{81} t_2 + \frac{1}{81} t_1 + \frac{1}{81} t_0\right). &(7)
\end{aligned}$$

Note: It is not required to answer the question, but it is interesting to compare this to the case where the trios were i.i.d. in terms of whether they were same-colored or not. Call these hypothetical random variables $\widetilde{X}_i$ and let $\widetilde{S}_n = \sum_i \widetilde{X}_i$. In that case, the expression for $\mathbb{E}[\widetilde{S}^2]$ would be $\binom{n}{3}(\frac{1}{9} + \frac{1}{81}t_2 + \frac{1}{81}t_1 + \frac{1}{81}t_0)$ which only differs from (7) by an $\binom{n}{3}\frac{2}{81}t_2$ — the expression in (7) is bigger by this much. Is this a little or is this a lot?

In the hypothetical i.i.d. case, we know that the variance $\text{Var}[\widetilde{S}_n] = \binom{n}{3}\text{Var}[\widetilde{X}_{ABC}] = \binom{n}{3}(\frac{1}{9}\cdot\frac{8}{9}) = \binom{n}{3}\frac{8}{81}$. Since $\text{Var}[\widetilde{S}_n] = \mathbb{E}[\widetilde{S}_n^2] - (\mathbb{E}[\widetilde{S}_n])^2 = \mathbb{E}[\widetilde{S}_n^2] - (\binom{n}{3}\frac{1}{9})^2$, we can alternatively understand $\mathbb{E}[\widetilde{S}_n^2] = \frac{\binom{n}{3}}{81}(8 + \binom{n}{3})$. Now for comparison, $\mathbb{E}[S_n^2] = \mathbb{E}[\widetilde{S}_n^2] + \binom{n}{3}\frac{2}{81}t_2 = \mathbb{E}[\widetilde{S}_n^2] + \binom{n}{3}\frac{6}{81}(n-3)$. This means that the variance of S_n is bigger than the variance of $\widetilde{S}_n$ by a seemingly massive amount $\binom{n}{3}\frac{6}{81}(n-3)$. (Why seemingly massive? Because this is much bigger than the variance of $\widetilde{S}_n$ itself. It isn't negligible.) But is this really that big?

To understand this, we need to think about when variance matters. It matters while proving laws of large numbers by means of Chebyshev's inequality. So, we should think about what a weak law of large numbers calculation involving the S_n would look like. We would like to say that in a large HW party, approximately $\frac{1}{9}$ of the trios would be same-colored with high probability. For the appropriate Chebyshev's inequality style proof of a weak law of large numbers, we would need to divide the variance of S_n by the *square* of the number of trios — and this is where we notice that the effect of this additional amount of variance is just $\frac{\binom{n}{3}\frac{6}{81}(n-3)}{(\binom{n}{3})^2} = \frac{\frac{6}{81}(n-3)}{\binom{n}{3}} = \frac{36}{81}\frac{n-3}{n(n-1)(n-2)}$ which is going to zero quadratically in n. (Notice that this is slower than how fast $\binom{n}{3}$ grows with n, which is cubic.) So, this additional amount of variance is not actually that big given its context.

Calculations like this are why understanding the proofs of weak laws of large numbers, etc. are so important — they let us extend our understanding beyond the i.i.d. case which is very important in practice for engineers analyzing and designing systems.

6. Meow at the Market (18 pts)

Wendy and Earnest lost their cat Meow at the flea market. They remember Meow's last position, which was at the candy stall, and know that in the i-th minute after Meow went missing, Meow moved east by X_i meters, where $X_i \sim \text{Poisson}(2)$ and all X_i are independent and identically distributed (i.i.d.). It then makes sense to consider $S_n = \sum_{i=1}^{n} X_i$, which can be interpreted as the distance Meow is from the stand after n minutes.

Recall that for a Poisson(λ) random variable, the mean is λ and the variance is also λ.

(a) (4 pts) **Determine $\mathbb{E}[S_n]$, and Var(S_n). Justify your answers for full credit.**

Solutions: Note that for all X_i, the mean and variance are both λ.

$$\mathbb{E}[S_n] = \mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbb{E}[X_i] = n \cdot 2$$
$$\text{Var}(S_n) = \text{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \text{Var}(X_i) = n \cdot 2$$

The first line is by linearity of expectation. The second line is due to the fact that the variance of a sum of independent random variables is the sum of the variances.

For each of the remaining parts, assume that $n = 100$ (i.e. 100 minutes have elapsed).

To find Meow, Wendy and Earnest construct a search interval which is centered around the position located $\mathbb{E}[S_{100}]$ meters east of the candy stall.

(b) (6 pts) Using Chebyshev's inequality, **provide the tightest upper bound on the width W of the interval such that we can guarantee Meow is within the search interval with probability at least $\frac{7}{8}$. Show all your work for full credit.**

Solutions:

$$\Pr\left(|S_{100} - \mathbb{E}[S_{100}]| \geq \frac{W}{2}\right) \leq \frac{\text{Var}(S_{100})}{(W/2)^2}$$
$$\text{We want } \Pr\left(|S_{100} - \mathbb{E}[S_{100}]| \geq \frac{W}{2}\right) \leq 1 - \frac{7}{8} = \frac{1}{8}$$
$$\text{So, we set } \frac{\text{Var}(S_{100})}{(W/2)^2} = \frac{200}{W^2/4} = \frac{1}{8}$$
$$\Rightarrow W = 80$$

So an 80m wide search interval is safe.

(c) (8 pts) Say Wendy and Earnest want to change the width of their interval but decide to keep it centered at the same location: $\mathbb{E}[S_{100}]$ meters east of the candy stall. They want their chances of finding Meow in the search interval to be approximately 97%. Using the Central Limit Theorem and the standard normal CDF table (on the next page) as a means of approximation, **what width W should they choose for their search interval? Show all your work for full credit.**

Solutions: Let $Z = \dfrac{S_n - \mathbb{E}[S_n]}{\sqrt{\text{Var}(S_n)}} = \dfrac{S_n - 200}{10 \cdot \sqrt{2}}$.

The CLT states that $Z \sim \mathcal{N}(0,1)$. So, we can find some interval $[-z, +z]$ around $\mathbb{E}[Z]$, that corresponds to a probability of .97, and then translate it back to be in terms of our original random variable to get an interval in terms of meters around $\mathbb{E}[S_n]$

$$Pr(|Z| \leq z) = 0.97 \Rightarrow \Phi(z) - \Phi(-z) = 2 \cdot \Phi(z) - 1 = 0.97 \Rightarrow \Phi(z) = 0.985$$

By looking at the table, this width corresponds to $z = 2.17$ normalized standard deviations.
Converting back in terms of deviations about S_n, we get:

$$\left| \frac{S_n - 200}{10 \cdot \sqrt{2}} \right| \leq 2.17$$

$$-2.17 \leq \frac{S_n - 200}{10 \cdot \sqrt{2}} \leq 2.17$$

$$200 - 2.17 \cdot 10 \cdot \sqrt{2} \leq S_n \leq 200 + 2.17 \cdot 10 \cdot \sqrt{2}$$

This corresponds to a width of $W = 2 \cdot 2.17 \cdot 10 \cdot \sqrt{2} = 43.4\sqrt{2} \approx 61.4$ meters.
Notice that this width is actually smaller than the upper-bound that Chebyshev's inequality was giving us, despite the fact that we are demanding a much higher probability of 97 percent. This is representative of the looseness of Chebyshev's inequality.

Introduction to Probability, 2nd Ed, by D. Bertsekas and J. Tsitsiklis, Athena Scientific, 2008

	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

The standard normal table. The entries in this table provide the numerical values of $\Phi(y) = \mathbf{P}(Y \leq y)$, where Y is a standard normal random variable, for y between 0 and 3.49. For example, to find $\Phi(1.71)$, we look at the row corresponding to 1.7 and the column corresponding to 0.01, so that $\Phi(1.71) = .9564$. When y is negative, the value of $\Phi(y)$ can be found using the formula $\Phi(y) = 1 - \Phi(-y)$.

Contributors:

- Babak Ayazifar.
- Jonathan Lin.
- Amin Ghafari.
- Albert Zhan.
- Yiming Ding.
- Khoa Tran.
- Anant Sahai.
- Khalil Sarwari.

