CS 70 Discrete Mathematics and Probability Theory
Summer 2023 Huang, Suzani, and Tausik

HW 5

1 Pairs of Beads

Sinho has a set of $2n$ beads ($n \geq 2$) of n different colors, such that there are two beads of each color. He wants to give out pairs of beads as gifts to all the other $n-1$ TAs, and plans on keeping the final pair for himself (since he is, after all, also a TA). To do so, he first chooses two beads at random to give to the first TA he sees. Then he chooses two beads at random from those remaining to give to the second TA he sees. He continues giving each TA he sees two beads chosen at random from his remaining beads until he has seen all $n-1$ TAs, leaving him with just the two beads he plans to keep for himself. Prove that the probability that at least one of the other TAs (*not* including Sinho himself) gets two beads of the same color is at most $\frac{1}{2}$.

Solution:

Denote A_i as the event that the ith TA gets a matching pair. We see that $\mathbb{P}[A_1] = \frac{1}{2n-1}$ since the first bead doesn't matter; what matters is that the second bead (of a possible $2n-1$) matches the first (only 1 corresponding matching bead). By symmetry, we see that $\mathbb{P}[A_i]$ is equivalent across all i. Hence, by union-bound

$$\mathbb{P}[A_1 \cup A_2 \cup \cdots \cup A_{n-1}] \leq \sum_{i=1}^{n-1} \mathbb{P}[A_i] = \frac{n-1}{2n-1} \leq \frac{n-1}{2n-2} = \frac{1}{2}$$

as desired.

2 Pairwise Independence

Note 14 Recall that the events A_1, A_2, and A_3 are *pairwise independent* if for all $i \neq j$, A_i is independent of A_j. However, pairwise independence is a weaker statement than *mutual independence*, which requires the additional condition that $\mathbb{P}[A_1 \cap A_2 \cap A_3] = \mathbb{P}[A_1]\mathbb{P}[A_2]\mathbb{P}[A_3]$.

Suppose you roll two fair six-sided dice. Let A_1 be the event that the first die lands on 1, let A_2 be the event that the second die lands on 6, and let A_3 be the event that the two dice sum to 7.

(a) Compute $\mathbb{P}[A_1]$, $\mathbb{P}[A_2]$, and $\mathbb{P}[A_3]$.

(b) Are A_1 and A_2 independent?

(c) Are A_2 and A_3 independent?

(d) Are A_1, A_2, and A_3 pairwise independent?

(e) Are A_1, A_2, and A_3 mutually independent?

Solution:

(a) We have that $\mathbb{P}[A_1] = \mathbb{P}[A_2] = \frac{1}{6}$, since we have a $\frac{1}{6}$ probability of getting a particular number on a fair die.

Since there are 6 ways in which the two dice can sum to 7 (i.e. $\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$), we have $\mathbb{P}[A_3] = \frac{1}{6}$ as well.

(b) We want to determine whether $\mathbb{P}[A_1 \cap A_2] = \mathbb{P}[A_1]\mathbb{P}[A_2]$. We already found the probabilities of A_1 and A_2 from part (a), so let's look at $\mathbb{P}[A_1 \cap A_2]$. There's only one possible outcome where the first die is a 1 and the second die is a 6, so this gives a probability of $\mathbb{P}[A_1 \cap A_2] = \frac{1}{36}$.

Since $\mathbb{P}[A_1]\mathbb{P}[A_2] = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36} = \mathbb{P}[A_1 \cap A_2]$, these two events are independent.

(c) We want to determine whether $\mathbb{P}[A_2 \cap A_3] = \mathbb{P}[A_2]\mathbb{P}[A_3]$. We already found the probabilities of A_2 and A_3 from part (a), so let's look at $\mathbb{P}[A_2 \cap A_3]$. These two events both occur if the second die lands on a 6, and the two dice sum to 7. There's only one way that this can happen, i.e. the first die must be a 1, so the intersection has probability $\mathbb{P}[A_2 \cap A_3] = \frac{1}{36}$.

Since $\mathbb{P}[A_2]\mathbb{P}[A_3] = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36} = \mathbb{P}[A_2 \cap A_3]$, these two events are independent.

(d) To see whether the three events are pairwise independent, we need to ensure that all pairs of events are independent. We've already checked that A_1 and A_2 are independent, and that A_2 and A_3 are independent, so it suffices to check whether A_1 and A_3 are independent.

Similar to the previous two parts, the intersection $A_1 \cap A_3$ means that the first die must land on a 1, and the two dice sum to 7. There's only one way for this to happen, i.e. the second die must land on a 6, so the probability is $\mathbb{P}[A_1 \cap A_3] = \frac{1}{36}$.

Since $\mathbb{P}[A_1]\mathbb{P}[A_3] = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36} = \mathbb{P}[A_1 \cap A_3]$, these two events are also independent. Since we've now shown that all possible pairs of events are independent, A_1, A_2, and A_3 are indeed pairwise independent.

(e) Mutual independence requires the additional constraint that $\mathbb{P}[A_1 \cap A_2 \cap A_3] = \mathbb{P}[A_1]\mathbb{P}[A_2]\mathbb{P}[A_3]$. We've found the individual probabilities of these events in part (a), so we only need to compute $\mathbb{P}[A_1 \cap A_2 \cap A_3]$.

Here, we must have that the first die lands on 1, the second die lands on 6, and the sum of the two dice is equal to 7. There's only one way for this to happen, i.e. the first die is a 1 and the second die is a 6, so the probability of the intersection of all three events is $\mathbb{P}[A_1 \cap A_2 \cap A_3] = \frac{1}{36}$.

However, since $\mathbb{P}[A_1]\mathbb{P}[A_2]\mathbb{P}[A_3] = \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{216} \neq \frac{1}{36} = \mathbb{P}[A_1 \cap A_2 \cap A_3]$, these three events are not mutually independent.

3 Cookie Jars

Note 15

You have two jars of cookies, each of which starts with n cookies initially. Every day, when you come home, you pick one of the two jars randomly (each jar is chosen with probability $1/2$) and eat one cookie from that jar. One day, you come home and reach inside one of the jars of cookies, but you find that is empty! Let X be the random variable representing the number of remaining cookies in non-empty jar at that time. What is the distribution of X?

Solution: Assume that you found jar 1 empty; the probability that $X = k$ and you found jar 1 empty is computed as follows.

In order for there to be k cookies remaining, you must have eaten a cookie for $2n-k$ days, and then you must have chosen jar 1 (to discover that it is empty). Within those $2n-k$ days, exactly n of those days you chose jar 1. The probability of this is $\binom{2n-k}{n}2^{-(2n-k)}$.

Furthermore, the probability that you then discover jar 1 is empty the day after is $1/2$. So, the probability that $X = k$ and you discover jar 1 empty is $\binom{2n-k}{n}2^{-(2n-k+1)}$. However, we assumed that we discovered jar 1 to be empty; the probability that $X = k$ and jar 2 is empty is the same by symmetry, so the overall probability that $X = k$ is:

$$\mathbb{P}[X = k] = \binom{2n-k}{n}\frac{1}{2^{2n-k}}, \qquad k \in \{0, \ldots, n\}.$$

4 Class Enrollment

Note 15
Note 19

Lydia has just started her CalCentral enrollment appointment. She needs to register for a geography class and a history class. There are no waitlists, and she can attempt to enroll once per day in either class or both. The CalCentral enrollment system is strange and picky, so the probability of enrolling successfully in the geography class on each attempt is p_g and the probability of enrolling successfully in the history class on each attempt is p_h. Also, these events are independent.

(a) Suppose Lydia begins by attempting to enroll in the geography class everyday and gets enrolled in it on day G. What is the distribution of G?

(b) Suppose she is not enrolled in the geography class after attempting each day for the first 7 days. What is $\mathbb{P}[G = i \mid G > 7]$, the conditional distribution of G given $G > 7$?

(c) Once she is enrolled in the geography class, she starts attempting to enroll in the history class from day $G+1$ and gets enrolled in it on day H. Find the expected number of days it takes Lydia to enroll in both the classes, i.e. $\mathbb{E}[H]$.

Suppose instead of attempting one by one, Lydia decides to attempt enrolling in both the classes from day 1. Let G be the number of days it takes to enroll in the geography class, and H be the number of days it takes to enroll in the history class.

(d) What is the distribution of G and H now? Are they independent?

(e) Let A denote the day she gets enrolled in her first class and let B denote the day she gets enrolled in both the classes. What is the distribution of A?

(f) What is the expected number of days it takes Lydia to enroll in both classes now, i.e. $\mathbb{E}[B]$?

(g) What is the expected number of classes she will be enrolled in by the end of 30 days?

Solution:

(a) $G \sim \text{Geometric}(p_g)$.

(b) Given that $G > 7$, the random variable G takes values in $\{8,9,\dots\}$. For $i = 8,9,\dots$,

$$\mathbb{P}[G = i \mid G > 7] = \frac{\mathbb{P}[G = i \wedge G > 7]}{\mathbb{P}[G > 7]} = \frac{\mathbb{P}[G = i]}{\mathbb{P}[G > 7]} = \frac{p_g(1-p_g)^{i-1}}{(1-p_g)^7} = p_g(1-p_g)^{i-8}.$$

If K denotes the additional number of days it takes to get enrolled in the geography class after day 7, i.e. $K = G - 7$, then conditioned on $G > 7$, the random variable K has the geometric distribution with parameter p_g. Note that this is the same as the distribution of G. This is known as the memoryless property of geometric distribution.

(c) We have $H - G \sim \text{Geometric}(p_h)$. Thus $\mathbb{E}[G] = 1/p_g$ and $\mathbb{E}[H-G] = 1/p_h$. And hence $\mathbb{E}[H] = \mathbb{E}[G] + \mathbb{E}[H-G] = 1/p_g + 1/p_h$.

(d) $G \sim \text{Geometric}(p_g)$, $H \sim \text{Geometric}(p_h)$. Yes they are independent.

(e) We have $A = \min\{G,H\}$ and $B = \max\{G,H\}$. We also use the following definition of the minimum:

$$\min(g,h) = \begin{cases} g & \text{if } g \le h; \\ h & \text{if } g > h. \end{cases}$$

Now, for all $k \in \{1,2,\dots\}$, $\min(G,H) = k$ is equivalent to $(G = k) \cap (H \ge k)$ or $(H = k) \cap (G > k)$. Hence,

$$\begin{aligned} \mathbb{P}[A = k] &= \mathbb{P}[\min(G,H) = k] \\ &= \mathbb{P}[(G = k) \cap (H \ge k)] + \mathbb{P}[(H = k) \cap (G > k)] \\ &= \mathbb{P}[G = k] \cdot \mathbb{P}[H \ge k] + \mathbb{P}[H = k] \cdot \mathbb{P}[G > k] \end{aligned}$$

(since G and H are independent)

$$= [(1-p_g)^{k-1} p_g](1-p_h)^{k-1} + [(1-p_h)^{k-1} p_h](1-p_g)^k$$

(since G and H are geometric)

$$\begin{aligned} &= ((1-p_g)(1-p_h))^{k-1} (p_g + p_h(1-p_g)) \\ &= (1 - p_g - p_h + p_h p_g)^{k-1} (p_g + p_h - p_g p_h). \end{aligned}$$

But this final expression is precisely the probability that a geometric RV with parameter $p_g + p_h - p_g p_h$ takes the value k. Hence $A \sim \text{Geom}(p_g + p_h - p_g p_h)$.

An alternative, slightly cleaner approach is to work with the *tail probabilities* of the geometric distribution, rather than with the usual point probabilities as above. In other words, we can work with $\mathbb{P}[A \ge k]$ rather than with $\mathbb{P}[A = k]$; clearly the values $\mathbb{P}[A \ge k]$ specify the values

$\mathbb{P}[A = k]$ since $\mathbb{P}[A = k] = \mathbb{P}[A \geq k] - \mathbb{P}[A \geq (k+1)]$, so it suffices to calculate them instead. We then get the following argument:

$$\begin{aligned}
\mathbb{P}[A \geq k] &= \mathbb{P}[\min(G,H) \geq k] \\
&= \mathbb{P}[(G \geq k) \cap (H \geq k)] \\
&= \mathbb{P}[G \geq k] \cdot \mathbb{P}[H \geq k] && \text{since } G,H \text{ are independent} \\
&= (1-p_g)^{k-1}(1-p_h)^{k-1} && \text{since } G,H \text{ are geometric} \\
&= ((1-p_g)(1-p_h))^{k-1} \\
&= (1-p_g-p_h+p_g p_h)^{k-1}.
\end{aligned}$$

This is the tail probability of a geometric distribution with parameter $p_g + p_h - p_g p_h$, so we are done.

(f) From part (e) we get $\mathbb{E}[A] = 1/(p_g + p_h - p_g p_h)$. From part (d) we have $\mathbb{E}[G] = 1/p_g$ and $\mathbb{E}[H] = 1/p_h$. We now observe that $\min\{g,h\} + \max\{g,h\} = g + h$. Using linearity of expectation we get $\mathbb{E}[A] + \mathbb{E}[B] = \mathbb{E}[G] + \mathbb{E}[H]$. Thus $\mathbb{E}[B] = 1/p_g + 1/p_h - 1/(p_g + p_h - p_g p_h)$.

(g) Let I_G and I_H be the indicator random variables of the events "$G \leq 30$" and "$H \leq 30$" respectively. Then $I_G + I_H$ is the number of classes she will be enrolled in within 30 days. Hence the answer is $\mathbb{E}[I_G] + \mathbb{E}[I_H] = \mathbb{P}[G \leq 30] + \mathbb{P}[H \leq 30] = 1 - (1-p_g)^{30} + 1 - (1-p_h)^{30}$.

5 Fishy Computations

Assume for each part that the random variable can be modelled by a Poisson distribution.

(a) Suppose that on average, a fisherman catches 20 salmon per week. What is the probability that he will catch exactly 7 salmon this week?

(b) Suppose that on average, you go to Fisherman's Wharf twice a year. What is the probability that you will go at most once in 2018?

(c) Suppose that in March, on average, there are 5.7 boats that sail in Laguna Beach per day. What is the probability there will be *at least* 3 boats sailing throughout the *next two days* in Laguna?

Solution:

(a) Let X be the number of salmon the fisherman catches per week. $X \sim \text{Poiss}(20 \text{ salmon/week})$, so

$$\mathbb{P}[X = 7 \text{ salmon/week}] = \frac{20^7}{7!}e^{-20} \approx 5.23 \cdot 10^{-4}.$$

(b) Similarly $X \sim \text{Poiss}(2)$, so

$$\mathbb{P}[X \leq 1] = \frac{2^0}{0!}e^{-2} + \frac{2^1}{1!}e^{-2} \approx 0.41.$$

(c) Let X_1 be the number of sailing boats on the next day, and X_2 be the number of sailing boats on the day after next. Now, we can model sailing boats on day i as a Poisson distribution $X_i \sim \text{Poiss}(\lambda = 5.7)$. Let Y be the number of boats that sail in the next two days. We are interested in $Y = X_1 + X_2$. We know that the sum of two independent Poisson random variables is Poisson (from Theorem 19.5 in lecture notes). Thus, we have $Y \sim \text{Poiss}(\lambda = 5.7 + 5.7 = 11.4)$.

$$\begin{aligned}
\mathbb{P}[Y \geq 3] &= 1 - \mathbb{P}[Y < 3] \\
&= 1 - \mathbb{P}[Y = 0 \cup Y = 1 \cup Y = 2] \\
&= 1 - (\mathbb{P}[Y = 0] + \mathbb{P}[Y = 1] + \mathbb{P}[Y = 2]) \\
&= 1 - \left(\frac{11.4^0}{0!} e^{-11.4} + \frac{11.4^1}{1!} e^{-11.4} + \frac{11.4^2}{2!} e^{-11.4} \right) \\
&\approx 0.999.
\end{aligned}$$

6 Dice Games

Note 20

(a) Alice and Bob are playing a game. Alice picks a random integer X between 0 and 100 inclusive, where each value is equally likely to be chosen. Bob then picks a random integer Y between 0 and X inclusive. What is $\mathbb{E}[Y]$?

(b) Alice rolls a die until she gets a 1. Let X be the number of total rolls she makes (including the last one), and let Y be the number of rolls on which she gets an even number. Compute $\mathbb{E}[Y \mid X = x]$, and use it to calculate $\mathbb{E}[Y]$.

(c) Bob plays a game in which he starts off with one die. At each time step, he rolls all the dice he has. Then, for each die, if it comes up as an odd number, he puts that die back, and adds a number of dice equal to the number displayed to his collection. (For example, if he rolls a one on the first time step, he puts that die back along with an extra die.) However, if it comes up as an even number, he removes that die from his collection.

What is the expected number of dice Bob will have after n time steps?

Solution:

(a) Let's first compute $\mathbb{E}[Y \mid X = x]$. Conditioned on $X = x$, the expected value of Y is $x/2$, so we can say that $\mathbb{E}[Y \mid X = x] = x/2$.

With total expectation, we have

$$\begin{aligned}
\mathbb{E}[Y] &= \sum_x \mathbb{E}[Y \mid X = x] \mathbb{P}[X = x] \\
&= \sum_x \frac{x}{2} \cdot \mathbb{P}[X = x] \\
&= \frac{1}{2} \sum_x x \cdot \mathbb{P}[X = x] \\
&= \frac{1}{2} \mathbb{E}[X] = \frac{1}{2} \cdot 50 = 25
\end{aligned}$$

(b) Let's compute $\mathbb{E}[Y \mid X = x]$. If Alice makes x total rolls, then before rolling a 1, she makes $x-1$ rolls that are not a 1. Since these rolls are independent, Y follows a binomial distribution with $n = x-1$ and $p = 3/5$, and $\mathbb{E}[Y \mid X = x] = \frac{3}{5}(x-1)$.

Now, we'd like to compute $\mathbb{E}[Y]$. With total expectation, we have

$$\begin{aligned}
\mathbb{E}[Y] &= \sum_x \mathbb{E}[Y \mid X = x]\mathbb{P}[X = x] \\
&= \sum_x \frac{3}{5}(x-1)\mathbb{P}[X = x] \\
&= \frac{3}{5}\sum_x x \cdot \mathbb{P}[X = x] - \frac{3}{5}\sum_x \mathbb{P}[X = x] \\
&= \frac{3}{5}\mathbb{E}[X] - \frac{3}{5}
\end{aligned}$$

Since X follows a geometric distribution with $p = 1/6$, $\mathbb{E}[X] = 6$, and

$$\mathbb{E}[Y] = \frac{3}{5}\mathbb{E}[X] - \frac{3}{5} = \frac{3}{5} \cdot 6 - \frac{3}{5} = 3.$$

(c) Let X_k be a random variable representing the number of dice after k time steps. In particular, this means that $X_0 = 1$. To compute the number of dice at step k, we first condition on $X_{k-1} = m$. Each one of the m dice is expected to leave behind 2 in its place, since there's a $\frac{1}{2}$ probability that it leaves behind 0 dice, a $\frac{1}{6}$ probability for each of 2, 4, and 6 dice, corresponding to rolling a 1, 3, and 5 respectively.

Therefore, we have $\mathbb{E}[X_k \mid X_{k-1} = m] = 2m$, so with total expectation, we have

$$\begin{aligned}
\mathbb{E}[X_k] &= \sum_m \mathbb{E}[X_k \mid X_{k-1} = m]\mathbb{P}[X_{k-1} = m] \\
&= \sum_m 2m \cdot \mathbb{P}[X_{k-1} = m] \\
&= 2\sum_m m \cdot \mathbb{P}[X_{k-1} = m] \\
&= 2\,\mathbb{E}[X_{k-1}]
\end{aligned}$$

This means that we expect to have $\mathbb{E}[X_n] = 2\,\mathbb{E}[X_{n-1}] = 2^2\,\mathbb{E}[X_{n-2}] = \cdots = 2^n\,\mathbb{E}[X_0] = 2^n$ dice.