Guerrilla Section 4

1. **Indicator Variables**

   (a) After throwing n balls into m bins at random, what is the expected number of bins that contains exactly k balls?

   (b) Alice and Bob each draw k cards out of a deck of 52 distinct cards with replacement. Find k such that the expected number of cards drawn by both Alice and Bob is 1.

   (c) How many people do you need in a room so that you expect that there is going to be a shared birthday on a Monday of the year (assume 52 Mondays in a year and 365 days in a year)?

Solution:

(a) Let $X_i = 1$ if bin i contains exactly k balls and $X_i = 0$ otherwise.
$E[X_i] = \binom{n}{k}(\frac{1}{m})^k(\frac{m-1}{m})^{n-k} = \binom{n}{k}(\frac{(m-1)^{n-k}}{m^n})$
$E[X] = \sum_{i=1}^{m}\binom{n}{k}(\frac{(m-1)^{n-k}}{m^n}) = \binom{n}{k}(\frac{(m-1)^{n-k}}{m^{n-1}})$

(b) Let $X_i = 1$ if card i is chosen by both Alice and Bob and $X_i = 0$ otherwise.
After drawing k cards, the probability of drawing a particular card is $\frac{k}{52}$ so
$E[X_i] = (1-(\frac{51}{52})^k)*(1-(\frac{51}{52})^k)$
$E[X] = \sum_{i=1}^{52}(1-(\frac{51}{52})^k)^2 = 52*(1-(\frac{51}{52})^k)^2$
Setting $E[X] = 1$, we have $k = 7.69 \approx 8$

(c) For $i < j$, let $X_{ij} = 1$ if i, j share a birthday and $X_{ij} = 0$ otherwise. Then, the total number of shared birthdays is $X = \sum_{i=1}^{k-1}\sum_{j=i+1}^{k} X_{ij}$, where k is the total number of people in the room. There is $\frac{52}{365}$ chance that person i has a birthday on a Monday and $\frac{1}{365}$ chance that person j has same birthday as person i so
$E[X] = \sum_{i=1}^{k-1}\sum_{j=i+1}^{k}\frac{52}{365}\frac{1}{365}$
$= \frac{52}{365^2}\sum_{i=1}^{k-1}\sum_{j=i+1}^{k} 1$
$= \frac{52}{365^2}\sum_{i=1}^{k-1}(k-i)$
$= \frac{52}{365^2}[k(k-1) - \frac{k(k-1)}{2}]$
$= \frac{52}{365^2}[\frac{k(k-1)}{2}]$
We want $E[X] = 1$ so $k \approx 72$

2. **Conditional Expectation**

   (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 $E[Y]$?

(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 $E[Y|X]$, and use it to calculate $E[Y]$.

(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 E[Y|X]. Conditioned on $X = x$, the expected value of Y is $\frac{x}{2}$, so we can say that $E[Y|X] = \frac{X}{2}$. Taking the expectation of both sides gives us that $E[E[Y|X]] = E[\frac{X}{2}]$, or that $E[Y] = \frac{1}{2}E[X] = \frac{1}{2}(50) = 25$.

(b) Let's compute $E[Y|X = x]$. 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 $p = \frac{3}{5}$, and $E[Y|X = x] = \frac{3}{5}(x-1)$. Therefore, $E[Y|X] = \frac{3}{5}(X-1)$.

Now, we'd like to compute E[Y]. Taking the expectation of both sides gives us that $E[E[Y|X]] = \frac{3}{5}E[X] - \frac{3}{5}$. Since X follows a geometric distribution with $p = \frac{1}{6}$, $E[X] = 6$, and $E[Y] = \frac{3}{5}E[X] - \frac{3}{5} = \frac{3}{5}(6) - \frac{3}{5} = 3$.

(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, $E[X_k|X_{k-1} = m] = 2m$, and $E[X_k|X_{k-1}] = 2X_{k-1}$. This means that after n time steps, we expect there to be $E[X_n] = 2^n E[X_0] = 2^n$ dice.

3. **LLSE and Graphs**

Consider a graph with n vertices numbered 1 through n, where $n \geq 2$. For each pair of distinct vertices, we add an undirected edge between them independently with probability p. Let D_1 be the random variable representing the degree of vertex 1, and let D_2 be the random variable representing the degree of vertex 2.

(a) Compute $E[D_1]$ and $E[D_2]$.

(b) Compute $Var(D_1)$.

(c) Compute $Cov(D_1, D_2)$.

(d) Using the information from the first three parts, what is LLSE($D_2|D_1$)?

Solution: Throughout this problem, let X_{ij} be an indicator random variable for whether the edge between vertex i and vertex j exists. Note that $X_{ij} = X_{ji}$.

(a) By linearity of expectation, $E[D_1] = E[\sum_{i=2}^{n} X_{1j}] = \sum_{i=2}^{n} E[X_{1j}] = (n-1)E[X_{ij}] = (n-1)p$. By symmetry, $E[D_2] = (n-1)p$ also.

(b) Solution 1: Write the variance of D_1 as a sum of covariances.

$$\begin{aligned} Var(D_1) &= Cov(\sum_{i=2}^{n} X_{1i}, \sum_{i=2}^{n} X_{1i}) \\ &= (n-1)Var(X_{12}) + ((n-1)^2 - (n-1))Cov(X_{12}, X_{13}) \\ &= (n-1)p(1-p) + 0 \\ &= (n-1)p(1-p). \end{aligned}$$

In the third line, we used the fact that X_{1i} and X_{1j} are independent if $i \neq j$, so their covariance is zero.

Solution 2: Compute the variance directly.

$$\begin{aligned} Var(X) &= E[X^2] - (E[X])^2 \\ &= E[(\sum_{i=2}^{n} X_{1i})^2] - (n-1)^2 p^2 \\ &= (n-1)E[X_{12}^2] + ((n-1)^2 - (n-1))E[X_{12}X_{13}] - (n-1)^2 p^2 \\ &= (n-1)p + (n^2 - 3n + 2)p^2 - (n-1)^2 p^2 \\ &= (n-1)p + (n-1)(n-2)p^2 - (n-1)^2 p^2 \\ &= (n-1)p(1 + (n-2)p - (n-1)p) \\ &= (n-1)p(1-p) \end{aligned}$$

(c) We can write $Cov(D_1, D_2) = Cov(\sum_{i=2}^{n} X_{1i}, \sum_{i=1, i\neq 2}^{n} X_{2i}) = \sum_{i=2}^{n} \sum_{j=1, j\neq 2}^{n} Cov(X_{1i}, X_{2j})$. Note that all pairs of X_{1i}, X_{2j} are independent except for when $i = 2$ and $j = 1$, so all terms in the sum are zero except for $Cov(X_{12}, X_{21})$, and our covariance is just equal to $Cov(X_{12}, X_{21}) = Var(X_{12}) = p(1-p)$.

(d) Since $LLSE(D_2|D_1) = E[D_2] + \frac{Cov(D_1, D_2)}{Var(D_1)}(D_1 - E[D_1])$, we plug in our values from the first three parts to get that $LLSE(D_2|D_1) = (n-1)p + \frac{p(1-p)}{(n-1)p(1-p)}(D_1 - (n-1)p) = (n-1)p + \frac{1}{n-1}(D_1 - (n-1)p) = \frac{1}{n-1}D_1 + (n-2)p$.

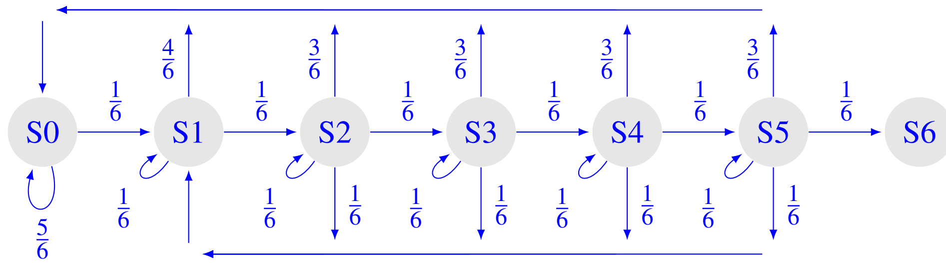
4. **Markov Chains** Suppose we have a traditional 6-sided dice. We would like to roll a monotonically increasing sequence of values from 1 to 6 such that every value rolled, s_n, is at least as large as the previous value, s_{n-1} and at most as large as $s_{n-1} + 1$. For example, $1, 1, 2, 2, 2, 3, 3, 4, 5, 5, 5, 6$, would be a valid sequence.

   (a) Find the expected number of rolls one must take in order to reach such a sequence

(b) What is the expected number of rolls if we had an n-sided dice and tried to roll a sequence monotonically increasing sequence from 1 to n

Solution:

(a) In this situation, we would have 7 possible states for each of the possible values rolled. S0 would represent the state where we have yet to roll a 1 to begin the sequence.



Given this Markov chain, we can create the following transition matrix

$$\frac{1}{6}\begin{bmatrix} 5 & 1 & 0 & 0 & 0 & 0 & 0 \\ 4 & 1 & 1 & 0 & 0 & 0 & 0 \\ 3 & 1 & 1 & 1 & 0 & 0 & 0 \\ 3 & 1 & 0 & 1 & 1 & 0 & 0 \\ 3 & 1 & 0 & 0 & 1 & 1 & 0 \\ 3 & 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 6 \end{bmatrix}$$

Using this matrix we can find the stationary distribution of this game as

$$\begin{bmatrix} 0.79168 & 0.16667 & 0.03333 & 0.00667 & 0.00133 & 0.000267 & 0.0000444 \end{bmatrix}$$

The reciprocals of these values would therefore be the expected number of rolls to reach each state. Therefore, in order to reach 6, it would take $\boxed{22500}$ rolls.

Another method for tackling is problem is as follows. Starting at S0, the probability of rolling a 1 is $\frac{1}{6}$, it takes an expected 6 rolls to go from S0 to S1. $Rolls(\text{S1}) = 6$

When a number is rolled, it takes $\frac{6}{5}$ rolls in order to roll something else. Thus, to go from S1 to S2, there is a $\frac{1}{5}$ chance that it takes $\frac{6}{5}$ rolls. In this case, we simply roll a 2. If we do not roll a 2, that means that the sequence must restart and we must first roll a 1 and then a 2. Therefore, there is a $(\frac{1}{5})(\frac{4}{5})$ chance that it takes $6+2(\frac{6}{5}$ rolls. Continuing the pattern, there is a $(\frac{1}{5})(\frac{4}{5})^2$ chance that it takes $(2)6+3(\frac{6}{5}$ rolls. The expected number of rolls to reach S2 would therefore be:

$$\sum_{k=0}^{\infty}\frac{1}{5}(\frac{4}{5})^k(6k+\frac{6}{5}(k+1)) = 4(6)+5(\frac{6}{5}) = 30$$

Thus $Rolls(\text{S2}) = 30+6 = 36$

More generally, the expected number of rolls to progress in state is:

$$\sum_{k=0}^{\infty}\frac{1}{5}(\frac{4}{5})^k(\mathbf{P}k+\frac{6}{5}(k+1))$$

Where **P** represents the number of extra rolls taken if a rolled value is not a desired outcome.

The transition from S2 to S3 represents the common case that applies to S3, S4, and S5. In this case, if the roll outcome is not 3, there is a $\frac{1}{4}$ chance that a 1 was rolled. In this event, we can go directly back to S1 and skip S0. This would save us an average of 6 rolls. If a 1 is not rolled, then we must go back to S0. Therefore, missing the desired outcome of 3 in S2 yields the $36-\frac{3}{2}$ extra rolls because:

$$\frac{1}{4}(36-6)+\frac{3}{4}36 = 36-\frac{3}{2}$$

Now using the same reasoning as before, the expected number of rolls to go from S2 to S3 becomes:

$$\sum_{k=0}^{\infty}\frac{1}{5}(\frac{4}{5})^k((36-\frac{3}{2})k+\frac{6}{5}(k+1)) = 4(36-\frac{3}{2})+5(\frac{6}{5}) = 4(36)$$

$$Rolls(S3) = 36+4(36) = 5(36) = 180$$

Therefore, the number of rolls to reach S3 is five times the number of rolls to reach S2. This pattern of multiplying by 5 continues in the transition from S3 to S4, S4 to S5, and S5 to S6.

Thus, $Rolls(\text{S6}) = 5Rolls(\text{S5}) = 5^2Rolls(\text{S4}) = 5^3Rolls(\text{S3}) = 5^4Rolls(\text{S2}) = 5^4(36) = \boxed{22500}$

5. **Markov Chains 2** Suppose you are flipping a standard coin (One Head and One Tail) until you get the same side 3 times (Heads, Heads, Heads) or (Tails, Tails, Tails) in a row

(a) Construct an markov chain that describes the situation with a start state and end state

(b) Given that you have flipped a (Tails, Heads) so far how many expected number of flips?

(c) What is the expected number of flips from the start state?

Solution:

(a) The appropriate markov chain has 6 states: Start, H1, H2, T1, T2, and End.
For starting node, there is an outgoing edge to H1 and T1, each with equal probability of $\frac{1}{2}$
For H1, there is an outgoing edge to H2 and T1, each with equal probability of $\frac{1}{2}$
For H2, there is an outgoing edge to End and T1, each with equal probability of $\frac{1}{2}$
For T1, there is an outgoing edge to H1 and T2, each with equal probability of $\frac{1}{2}$
For T2, there is an outgoing edge to H1 and End, each with equal probability of $\frac{1}{2}$

(b) If you got a Tails and then a Heads, you are currently in the H1 state. Thus, we must calculate the expected number of flips to end from H1. Thus we will do this with a system of equations. Since we are not trying to solve for the starting state, we have 5 unknowns that depend on 5 linearly independent equations.
$\beta(H_1) = 1 + 0.5\beta(T_1) + 0.5\beta(H_2)$
$\beta(H_2) = 1 + 0.5\beta(E) + 0.5\beta(T_1)$
$\beta(T_1) = 1 + 0.5\beta(T_2) + 0.5\beta(H_1)$
$\beta(T_2) = 1 + 0.5\beta(E) + 0.5\beta(H_1)$
$\beta(E) = 0$
If we solve this system of equations, we get $\beta(H_1) = 6, \beta(H_2) = 4, \beta(T_1) = 6, \beta(T_2) = 4$.

(c) $\beta(S) = 1 + 0.5\beta(H_1) + 0.5\beta(T_1) = 1 + 0.5 * 6 + 0.5 * 6 = 7$

6. **Darts (Again!)** Alvin is playing darts. His aim follows an exponential distribution; that is, the probability that the dart is x distance from the center is $P[X = x] = e^{-x}$. The board's radius is 4 units.

(a) What is the probability the dart will stay within the board?

(b) Say you know Alvin will always make it within the board. What is the probability he is within 1 unit from the center?

(c) If Alvin is within 1 unit from the center, he scores 4 points, if he is within 2 units, he scores 3, etc. In other words, Alvin scores $\lfloor 5 - x \rfloor$, where x is the distance from the center. What is Alvin's expected score after one throw?

Solution:

(a) The CDF of an exponential is $P[X \le x] = 1 - e^{-x}$. Therefore,

$$P[X \le 4] = 1 - e^{-4}$$

(b) We are given that the dart must be within the board, which means that the dart is at least 4 units away from the center. We can use the definition of conditional expectation:

$$\begin{aligned} P[X \le 1 | X \le 4] &= \frac{P[X \le 1 \cap X \le 4]}{P[X \le 4]} \\ &= \frac{P[X \le 1]}{P[X \le 4]} \\ &= \frac{1 - e^{-1}}{1 - e^{-4}} \end{aligned}$$

(c)

$$\begin{aligned} E[score] &= \int_0^1 4e^{-x} dx + \int_1^2 3e^{-x} dx + \int_2^3 2e^{-x} dx + \int_3^4 e^{-x} dx \\ &= 4(-e^{-1} + 1) + 3(-e^{-2} + e^{-1}) + 2(-e^{-3} + e^{-2}) + (-e^{-4} + e^{-3}) \\ &= 4 - e^{-1} - e^{-2} - e^{-3} - e^{-4} \end{aligned}$$