CS 70	Discrete Mathematics and Probability Theory	
Spring 2025	Rao	HW 11

Due: Saturday, 4/19, 4:00 PM
Grace period until Saturday, 4/19, 6:00 PM

Sundry

Before you start writing your final homework submission, state briefly how you worked on it. Who else did you work with? List names and email addresses. (In case of homework party, you can just describe the group.)

Zachary Brandt
zbrandt@berkeley.edu

1 Coupon Collector Variance

Note 19 It's that time of the year again—Safeway is offering its Monopoly Card promotion. Each time you visit Safeway, you are given one of n different Monopoly Cards with equal probability. You need to collect them all to redeem the grand prize.

Let X be the number of visits you have to make before you can redeem the grand prize. Show that $\text{Var}(X) = n^2\left(\sum_{i=1}^{n} i^{-2}\right) - \mathbb{E}[X]$.

Solution:

We know from the notes that the expected value of X is $\mathbb{E}[X] = n \cdot \sum_{i=1}^{n} \frac{1}{n}$. The variance of X, where X is equal to $X_1 + X_2 + \cdots + X_n$, i.e., the sum of the number of boxes it takes to find the first to the n-th Monopoly Card, is then $Var(X) = Var(X_1) + Var(X_2) + \cdots + Var(X_n)$ since the individual random variables are independent of one another. The number of boxes it takes to find the first coupon has no bearing on the number of boxes it will take to find the next. Then,

$$\begin{aligned}
\mathrm{Var}(X) &= \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \cdots + \mathrm{Var}(X_n) \\
&= \left(\frac{1-p_1}{p_1^2}\right) + \left(\frac{1-p_2}{p_2^2}\right) + \ddots \\
&= \left(\frac{1-1}{1^2}\right) + \left(\frac{1-\frac{n-1}{n}}{\left(\frac{n-1}{n}\right)^2}\right) + \cdots \\
&= \left(\frac{1-1}{1^2}\right) + \left(\frac{n^2}{(n-1)^2} - \frac{n}{n-1}\right) + \left(\frac{n^2}{(n-2)^2} - \frac{n}{n-2}\right) + \cdot \\
&= \left(\frac{n^2}{n^2} + \frac{n^2}{(n-1)^2} + \frac{n^2}{(n-2)^2} + \dots\right) - \left(\frac{n}{n} + \frac{n}{n-1} + \frac{n}{n-2} + \cdots\right) \\
&= n^2 \cdot \left(\frac{1}{1} + \frac{1}{(2)^2} + \cdots + \frac{1}{(n)^2}\right) - n \cdot \left(\frac{1}{1} + \frac{1}{2} + \cdots + \frac{1}{n}\right) \\
&= n^2 \sum_{i=1}^{n} \frac{1}{i^2} - n \sum_{i=1}^{n} \frac{1}{i} \\
&= n^2 \sum_{i=1}^{n} \frac{1}{i^2} - \mathbb{E}[X]
\end{aligned}$$

2 Diversify Your Hand

Note 15
Note 16

You are dealt 5 cards from a standard 52 card deck. Let X be the number of distinct values in your hand. For instance, the hand (A, A, A, 2, 3) has 3 distinct values.

(a) Calculate $\mathbb{E}[X]$. (Hint: Consider indicator variables X_i representing whether i appears in the hand.)

(b) Calculate $\mathrm{Var}(X)$. The answer expression will be quite involved; you do not need to simplify anything.

Solution:

(a) The expected value of X, where $X = X_1 + X_2 + \cdots + X_1 3$, is

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{i=1}^{13} \mathbb{E}(X_i) \\
&= \sum_{i=1}^{13} \mathbb{P}[X_i = 1] \\
&= \sum_{i=1}^{13} \left(1 - \frac{\binom{48}{5}}{\binom{52}{5}}\right) \\
&= 13\left(1 - \frac{\binom{48}{5}}{\binom{52}{5}}\right)
\end{aligned}$$

(b) The variance of X is then

$$\begin{aligned}
& \mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \\
\text{Not independent} \quad & = \sum_{i=1}^{13} \mathbb{E}[X_i^2] + \sum_{i \neq j}^{13} \mathbb{E}[X_i X_j] - (\mathbb{E}[X])^2 \\
\text{12 ways per card} \quad & = \sum_{i=1}^{13} \mathbb{E}[X_i^2] + 13 \times 12 \times \mathbb{P}[X_i X_j = 1] - (\mathbb{E}[X])^2 \\
& = \sum_{i=1}^{13} \mathbb{E}[X_i^2] + 13 \cdot 12(1 - \mathbb{P}[X_i = 0] - \mathbb{P}[X_j = 0] + \mathbb{P}[X_i = 0, X_j = 0]) - (\mathbb{E}[X])^2 \\
& = 13\left(1 - \frac{\binom{48}{5}}{\binom{52}{5}}\right) + 156\left(1 - 2\frac{\binom{48}{5}}{\binom{52}{5}} + \frac{\binom{44}{5}}{\binom{52}{5}}\right) - \left(13\left(1 - \frac{\binom{48}{5}}{\binom{52}{5}}\right)\right)^2
\end{aligned}$$

3 Double-Check Your Intuition Again

Note 16

(a) You roll a fair six-sided die and record the result X. You roll the die again and record the result Y.

  (i) What is $\mathrm{cov}(X+Y, X-Y)$?

  (ii) Prove that $X+Y$ and $X-Y$ are not independent.

For each of the problems below, if you think the answer is "yes" then provide a proof. If you think the answer is "no", then provide a counterexample.

(b) If X is a random variable and $\mathrm{Var}(X) = 0$, then must X be a constant?

(c) If X is a random variable and c is a constant, then is $\mathrm{Var}(cX) = c\,\mathrm{Var}(X)$?

(d) If A and B are random variables with nonzero standard deviations and $\mathrm{Corr}(A,B) = 0$, then are A and B independent?

(e) If X and Y are not necessarily independent random variables, but $\mathrm{Corr}(X,Y) = 0$, and X and Y have nonzero standard deviations, then is $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$?

(f) If X and Y are random variables then is $\mathbb{E}[\max(X,Y)\min(X,Y)] = \mathbb{E}[XY]$?

(g) If X and Y are independent random variables with nonzero standard deviations, then is

$$\mathrm{Corr}(\max(X,Y), \min(X,Y)) = \mathrm{Corr}(X,Y)?$$

Solution:

(a) For part (i), the covariance is equal to zero, as shown below:

$$\begin{aligned}
\operatorname{cov}(X+Y, X-Y) &= \operatorname{cov}(X,X) - \operatorname{cov}(X,Y) + \operatorname{cov}(Y,X) - \operatorname{cov}(Y,Y) \\
&= \operatorname{cov}(X,X) - \operatorname{cov}(Y,Y) \\
&= \operatorname{Var}(X) - \operatorname{Var}(Y) \\
&= 0
\end{aligned}$$

For part (ii), consider $X+Y=1$ but $X-Y=0$, which cannot be the case since if the two random variables take on an equal number they cannot sum to 1. So, the probability of this event is 0. However, the probabilities of the individual events are both non-zero, so $\mathbb{P}[X+Y=1] \times \mathbb{P}[X-Y=0] \neq \mathbb{P}[X+Y=1, X-Y=0]$, and the events are not independent.

(b) Yes, if X is a random variable and $\operatorname{Var}(X)=0$, then X is constant. If the variance is equal to zero, $\operatorname{Var}(X) = \mathbb{E}[(X-\mathbb{E}(X))^2] = 0 \to X - \mathbb{E}[X] = 0 \to X = \mathbb{E}[X]$, and since $\mathbb{E}[X]$ is constant, so must X.

(c) No, consider a random variable X and $c=2$, then

$$\operatorname{Var}(2X) = \mathbb{E}[4X^2] - (\mathbb{E}[2X])^2 = 4\,\mathbb{E}[X^2] - 4(\mathbb{E}[X])^2 = 4\operatorname{Var}(X) \neq 2\operatorname{Var}(X)$$

(d) No, not necessarily, if $\operatorname{Corr}(X,Y)=0$ and the variables have non-zero standard deviations, then $\operatorname{cov}(X,Y)=0$, which does not necessarily imply that the variables are independent. For example, consider the variables $X+Y$ and $X-Y$ from part (a).

(e) Yes, if $\operatorname{Corr}(X,Y)=0$ and the variables have non-zero standard deviations, then $\operatorname{cov}(X,Y) = 0$. In the general case for calculating the variance of two random variables $\operatorname{Var}(X,Y) = \operatorname{Var}(X) + \operatorname{Var}(Y) - 2\operatorname{cov}(X,Y)$, but since $\operatorname{cov}(X,Y)=0$, then it reduces to $\operatorname{Var}(X,Y) = \operatorname{Var}(X) + \operatorname{Var}(Y)$.

(f) Yes, because in all cases $X > Y$, $X < Y$, and $X = Y$, both sides of the equality are equal. If $X = Y$, then $\mathbb{E}[XY] = \mathbb{E}[XX]$ and similarly both the min and max functions both return the same value which then also becomes $\mathbb{E}[XX]$. If the two random variables are not equal, then necessarily the minimum and maximum functions either return X,Y or Y,X, respectively to multiply XY in the $\mathbb{E}$ function.

(g) This question essentially asks if $\operatorname{cov}(\max(X,Y), \min(X,Y)) = \operatorname{cov}(X,Y)$, since the denominators on both sides of the equality from the correlation formulas are the same and have non-zero standard deviations. The answer is yes, since from the previous part we have shown that $\mathbb{E}[\max(X,Y)\min(X,Y)] = \mathbb{E}[XY]$, and also $\mathbb{E}[\max(X,Y)]\,\mathbb{E}[\min(X,Y)] = \mathbb{E}[X]\,\mathbb{E}[Y]$, with the same logic as before. If the random variables are equal, then the equality will be true since the minimum and maximum functions output the same value. If they are not the same then the minimum function will return one value, and the maximum necessarily returns the other.

4 Dice Games

Note 20

(a) 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 $\mathbb{E}[Y]$.

(b) 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.

Compute the expected number of dice Bob will have after n time steps. (Hint: compute the value of $\mathbb{E}[X_k \mid X_{k-1} = m]$ to derive a recursive expression for X_k, where X_i is the random variable representing the number of dice after i time steps.)

Solution:

(a) The expected value of the random variable Y is 3. The first variable is a geometrically distributed random variable with a probability of success of $\frac{1}{6}$. The second variable is a binomially distributed random variable where the number of trials is the value that X takes on, minus the last one, with a probability of success of $\frac{3}{5}$ (there are only 5 numbers to consider since it can't be 1). The expected value of a binomially distributed variable like Y is $n \times p$ (consider a series of indicator random variables), therefore, $\mathbb{E}[Y \mid X = x] = (x-1) \times \frac{3}{5}$. To find the expected value of Y in general, the expected value of $Y \mid X = x$ gives the answer:

$$\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y \mid X = x]] = \mathbb{E}[(X-1) \times \frac{3}{5}] = (\mathbb{E}[X] - 1) \times \frac{3}{5} = (6-1) \times \frac{3}{5} = 3$$

.

(b) I instead consider a random variable "$+X_i$", which represents the added number of dice from one particular die. $+X_i$, which I will now refer to as X_i, has the following distribution:

$$X_i = \begin{cases} -1, & \text{w.p. } \frac{1}{2} \\ +1, & \text{w.p. } \frac{1}{6} \\ +3, & \text{w.p. } \frac{1}{6} \\ +5, & \text{w.p. } \frac{1}{6}. \end{cases}$$

The expected value for any X_i is then

$$\mathbb{E}[X_i] = \frac{1}{2}(-1) + \frac{1}{6}(1+3+5) = -\frac{1}{2} + \frac{3}{2} = 1.$$

For a given k then,

$$\mathbb{E}[+X_k \mid X_{k-1} = m] = \mathbb{E}[X_1 + X_2 + \cdots + X_m] = \mathbb{E}[X_1] + \mathbb{E}[X_2] + \cdots + \mathbb{E}[X_m] = m.$$

Therefore, the expected value of X_k (not $+X_k$) is the number of dice already in existence, m, plus the additional dice, m, so $\mathbb{E}[X_k \mid X_{k-1} = m] = m + m = 2m$. The expected value of X_n follows a recursive expansion, $\mathbb{E}[X_n] = 2\,\mathbb{E}[X_{n-1}] = 2^2\,\mathbb{E}[X_{n-2}] = \cdots = 2^n\,\mathbb{E}[X_0] = 2^n \times 1 = 2^n$.

5 LLSE and Graphs

Note 20

Consider a graph with n vertices numbered 1 through n, where n is a positive integer ≥ 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 $\mathbb{E}[D_1]$ and $\mathbb{E}[D_2]$.

(b) Compute $\text{Var}(D_1)$.

(c) Compute $\text{cov}(D_1, D_2)$.

(d) Using the information from the first three parts, what is $L(D_2 \mid D_1)$?

Solution:

(a) D_1 can be represented as a series of $n-1$ indicator random variables, each representing the existence of an edge between the vertex and the other vertices. So, $D_1 = X_2 + X_3 + \cdots + X_n$. Then, the expected value is $\mathbb{E}[D_1] = \mathbb{E}[X_2 + X_3 + \cdots + X_n] = \mathbb{E}[X_2] + \mathbb{E}[X_3] + \cdots + \mathbb{E}[X_n] = (n-1)p$. $\mathbb{E}[D_2]$ takes on the same value.

(b) The variance of D_1 is $\text{Var}(X_2 + X_3 + \cdots + X_n) = \text{Var}(X_2) + \text{Var}(X_3) + \cdots + \text{Var}(X_n) = (n-1)(p-p^2)$.

(c) The covariance of D_1 and D_2 boils down to the variance of X_2, as seen below:

$$\begin{aligned}
\text{cov}(D_1, D_2) &= \text{cov}(X_2 + X_3 + \cdots + X_n, Y_1 + Y_3 + \cdots + Y_n) \\
&= \text{cov}(X_2 + X_3 + \cdots + X_n, Y_1 + Y_3 + \cdots + Y_n) \\
&= \sum_{i \neq 1}^{n} \sum_{j \neq 2}^{n} \text{cov}(X_i, Y_i) \\
&= \text{cov}(X_2, Y_1) \quad \text{Variables otherwise independent} \\
&= \text{cov}(X_2, X_2) \quad X_2 \text{ and } Y_1 \text{ represent same state} \\
&= \text{Var}(X_2) \\
&= p - p^2
\end{aligned}$$

(d) The linear least squares estimate of D_2 given D_1 is

$$\begin{aligned}
\hat{D}_2 &= \mathbb{E}[D_2] + \frac{\text{cov}(D_1, D_2)}{\text{Var}(D_1)}(D_1 - \mathbb{E}[D_1]) \\
&= (n-1)p + \frac{p-p^2}{(n-1)(p-p^2)}(D_1 - (n-1)p) \\
&= (n-1)p + \frac{D_1}{n-1} - p
\end{aligned}$$

6 Balls in Bins Estimation

Note 20

We throw $n > 0$ balls into $m \geq 2$ bins. Let X and Y represent the number of balls that land in bin 1 and 2 respectively.

(a) Calculate $\mathbb{E}[Y \mid X]$. [*Hint*: Your intuition may be more useful than formal calculations.]

(b) What is $L[Y \mid X]$ (where $L[Y \mid X]$ is the best linear estimator of Y given X)? [*Hint*: Your justification should be no more than two or three sentences, no calculations necessary! Think carefully about the meaning of the conditional expectation.]

(c) Unfortunately, your friend is not convinced by your answer to the previous part. Compute $\mathbb{E}[X]$ and $\mathbb{E}[Y]$.

(d) Compute $\mathrm{Var}(X)$.

(e) Compute $\mathrm{cov}(X,Y)$.

(f) Compute $L[Y \mid X]$ using the formula. Ensure that your answer is the same as your answer to part (b).

Solution:

(a) The expected value of Y, the number of balls in bin 2, given some amount of balls in bin 1, is $\mathbb{E}[Y \mid X] = \frac{n-X}{m-1}$, since there are $n - X$ balls remaining, and each have a $\frac{1}{m-1}$ probability of going in bin 2.

(b) $L[Y \mid X]$ is just $\mathbb{E}[Y \mid X]$ or $\frac{n-X}{m-1}$ since this previous answer is already linear in X. Additionally, no other function can do a better job at predicting given some value of X than $\mathbb{E}[Y \mid X]$.

(c) The expected value of both X and Y is $\frac{n}{m}$ since there are n balls and each has a $\frac{1}{m}$ probability of going into any of the bins.

(d) The variance of X is $\mathrm{Var}(X) = n \cdot \frac{1}{m} - n \cdot \frac{1}{m^2} = n(m^{-1} - m^{-2})$.

(e) The covariance of X and Y is

$$\begin{aligned}
\mathrm{cov}(X,Y) &= \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y] \\
&= \binom{n}{2}\frac{1}{m^2} - \frac{n^2}{m^2} \\
&= \frac{n(n-1)}{m^2} - \frac{n^2}{m^2} \\
&= \frac{(-n)}{m^2}
\end{aligned}$$

The second step is so because for $\mathbb{E}[XY]$ there are $\binom{n}{2}$ distinct pairs of balls and each bin placement has a $\frac{1}{m}$ and then $\frac{1}{m}$ probability of happening (as indicator random variables).

(f) The linear least squares estimator of Y given X is then

$$\begin{aligned}
\hat{Y} &= \mathbb{E}[Y] + \frac{\operatorname{cov}(X,Y)}{\operatorname{Var}(X)}(X - \mathbb{E}[X]) \\
&= \frac{n}{m} + \frac{\frac{(-n)}{m^2}}{n(m^{-1} - m^{-2})}(X - \frac{n}{m}) \\
&= \frac{n}{m} + \frac{-1}{m-1}(X - \frac{n}{m}) \\
&= \frac{n}{m} + \frac{-X}{m-1} + \frac{n}{m^2 - m}
\end{aligned}$$