CS 70 Discrete Mathematics and Probability Theory
Spring 2026 Sinclair, Song

HW 12

1 Coupon Collector Variance

Note 15
Note 17

It's that time of the year again (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

$$\text{Var}(X) = n^2\left(\sum_{i=1}^{n}\frac{1}{i^2}\right) - \mathbb{E}[X]$$

.

Solution:

Note that this is the coupon collector's problem, but now we have to find the variance. Let X_i be the number of additional visits needed to collect the ith unique Monopoly card, given that we have already collected $i-1$ unique Monopoly cards. Then $X = \sum_{i=1}^{n} X_i$ and each X_i is geometrically distributed with $p = (n-i+1)/n$. Moreover, the random variables themselves are independent, since each time you collect a new card, you are starting from a clean slate.

$$\begin{aligned}
\text{Var}(X) &= \sum_{i=1}^{n}\text{Var}(X_i) && \text{(as the } X_i \text{ are independent)}\\
&= \sum_{i=1}^{n}\frac{1-(n-i+1)/n}{[(n-i+1)/n]^2} && \text{(variance of a geometric r.v. is } (1-p)/p^2)\\
&= \sum_{i=1}^{n}\frac{1-i/n}{(i/n)^2} && \text{(substituting } i \to n-i+1, \text{ which still ranges from 1 to } n)\\
&= \sum_{i=1}^{n}\frac{n(n-i)}{i^2}\\
&= \sum_{i=1}^{n}\frac{n^2}{i^2} - \sum_{i=1}^{n}\frac{n}{i}\\
&= n^2\left(\sum_{i=1}^{n}\frac{1}{i^2}\right) - \mathbb{E}[X] && \text{(using the coupon collector problem expected value).}
\end{aligned}$$

2 Double-Check Your Intuition Again

Note 17

(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 $\text{cov}(X+Y, X-Y)$?

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

The problems below are not related to the scenario above. For each of them, 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 $\text{Var}(X) = 0$, then must X be a constant?

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

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

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

The two subparts below are **optional** and will not be graded but are recommended for practice.

(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

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

Solution:

(a) (i) Using bilinearity of covariance, we have

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

where we use that $\text{cov}(X,Y) = \text{cov}(Y,X)$ to get the second equality.

(ii) Observe that $\mathbb{P}[X+Y=7, X-Y=0] = 0$ because if $X-Y=0$, then the sum of our two dice rolls must be even. However, both $\mathbb{P}[X+Y=7]$ and $\mathbb{P}[X-Y=0]$ are nonzero, so $\mathbb{P}[X+Y=7, X-Y=0] \neq \mathbb{P}[X+Y=7] \cdot \mathbb{P}[X-Y=0]$.

(b) Yes. If we write $\mu = \mathbb{E}[X]$, then $0 = \text{Var}(X) = \mathbb{E}[(X-\mu)^2]$ so $(X-\mu)^2$ must be identically 0 since perfect squares are non-negative. Thus $X = \mu$.

(c) No. We have $\text{Var}(cX) = \mathbb{E}[(cX - \mathbb{E}[cX])^2] = c^2\,\mathbb{E}[(X - \mathbb{E}[X])^2] = c^2\,\text{Var}(X)$ so if $\text{Var}(X) \neq 0$ and $c \neq 0$ or $c \neq 1$ then $\text{Var}(cX) \neq c\,\text{Var}(X)$. This does prove that $\sigma(cX) = |c|\sigma(X)$ though.

(d) No. Let $A = X+Y$ and $B = X-Y$ from part (a). Since A and B are not constants then part (b) says they must have nonzero variances which means they also have nonzero standard deviations. Part (a) says that their covariance is 0 which means they are uncorrelated, and that they are not independent.

Recall from lecture that the converse is true though.

(e) Yes. If $\text{Corr}(X,Y) = 0$, then $\text{cov}(X,Y) = 0$. We have $\text{Var}(X+Y) = \text{cov}(X+Y, X+Y) = \text{Var}(X) + \text{Var}(Y) + 2\,\text{cov}(X,Y) = \text{Var}(X) + \text{Var}(Y)$.

(f) Yes. For any values x, y we have $\max(x,y)\min(x,y) = xy$. Thus, $\mathbb{E}[\max(X,Y)\min(X,Y)] = \mathbb{E}[XY]$.

(g) No. You may be tempted to think that because $(\max(x,y), \min(x,y))$ is either (x,y) or (y,x), then $\text{Corr}(\max(X,Y), \min(X,Y)) = \text{Corr}(X,Y)$ because $\text{Corr}(X,Y) = \text{Corr}(Y,X)$. That reasoning is flawed because $(\max(X,Y), \min(X,Y))$ is not always equal to (X,Y) or always equal to (Y,X) and the inconsistency affects the correlation. It is possible for X and Y to be independent while $\max(X,Y)$ and $\min(X,Y)$ are not.

For a concrete example, suppose X is either 0 or 1 with probability $1/2$ each and Y is independently drawn from the same distribution. Then $\text{Corr}(X,Y) = 0$ because X and Y are independent. Even though X never gives information about Y, if you know $\max(X,Y) = 0$ then you know for sure $\min(X,Y) = 0$.

More formally, $\max(X,Y) = 1$ with probability $3/4$ and 0 with probability $1/4$, and $\min(X,Y) = 1$ with probability $1/4$ and 0 with probability $3/4$. This means

$$\mathbb{E}[\max(X,Y)] = 1 \cdot \frac{3}{4} + 0 \cdot \frac{1}{4} = \frac{3}{4}$$

and

$$\mathbb{E}[\min(X,Y)] = 1 \cdot \frac{1}{4} + 0 \cdot \frac{3}{4} = \frac{1}{4}.$$

Thus,

$$\begin{aligned}\text{cov}(\max(X,Y), \min(X,Y)) &= \mathbb{E}[\max(X,Y)\min(X,Y)] - \frac{3}{16} \\ &= \frac{1}{4} - \frac{3}{16} = \frac{1}{16} \neq 0\end{aligned}$$

We conclude that $\text{Corr}(\max(X,Y), \min(X,Y)) \neq 0 = \text{Corr}(X,Y)$.

3 Dice Games

Note 17

(a) Alice rolls a fair six-sided 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 fair six-sided 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 fair six-sided 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. As your base case, $E[X_0] = 1$.)

Solution:

(a) 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

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

(b) 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

$$\mathbb{E}[X_k \mid X_{k-1} = m] = m\left(\frac{1}{6}(0) + \frac{1}{6}(2) + \frac{1}{6}(0) + \frac{1}{6}(4) + \frac{1}{6}(0) + \frac{1}{6}(6)\right) = 2m$$

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.

4 Fishy Computations

Note 18

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 2024?

(c) Suppose 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. $X \sim \text{Poisson}(20 \text{ salmon/week})$, so

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

(b) Similarly $Y \sim \text{Poisson}(2)$, so

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

(c) Let S_1 be the number of sailing boats on the next day, and S_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 $S_i \sim \text{Poisson}(\lambda = 5.7)$. Let Z be the number of boats that sail in the next two days. We are interested in $Z = S_1 + S_2$. We know that the sum of two independent Poisson random variables is Poisson. Thus, we have $Z \sim \text{Poisson}(\lambda = 5.7 + 5.7 = 11.4)$.

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

5 Geometric and Poisson

Note 18

Let $X \sim \text{Geometric}(p)$ and $Y \sim \text{Poisson}(\lambda)$ be independent random variables. Compute $\mathbb{P}[X > Y]$. Your final answer should not have summations.

Hint: Use the total probability rule.

Solution: We condition on Y so we can use the nice property of geometric random variables that $\mathbb{P}[X > k] = (1-p)^k$. This gives

$$\begin{aligned}
\mathbb{P}[X > Y] &= \sum_{y=0}^{\infty} \mathbb{P}[X > Y \mid Y = y] \cdot \mathbb{P}[Y = y] \\
&= \sum_{y=0}^{\infty} (1-p)^y \cdot \frac{e^{-\lambda}\lambda^y}{y!} \\
&= e^{-\lambda p} e^{\lambda p} \sum_{y=0}^{\infty} \frac{e^{-\lambda}(\lambda(1-p))^y}{y!} \\
&= e^{-\lambda p} \sum_{y=0}^{\infty} \frac{e^{-\lambda(1-p)}(\lambda(1-p))^y}{y!} \\
&= e^{-\lambda p}
\end{aligned}$$

To simplify the last summation, we observe that the sum could be interpreted as the sum of the probabilities for a Poisson$(\lambda(1-p))$ random variable, which is equal to 1. Alternatively, you can use the Taylor series $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$ to simplify the sum.

6 Balls and Bins

Note 17

Throw n balls into m bins, where m and n are positive integers. Let X be the number of bins with exactly one ball. Compute $\text{Var}(X)$. Your final answer should not contain any summations.

Solution: Let X_i be the indicator that bin i has exactly one ball, for each $i = 1, \ldots, m$. Since $X = \sum_i X_i$, we can use the computational formula for variance:

$$\begin{aligned}
\text{Var}(X) &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 \\
&= \mathbb{E}\left[\left(\sum_{i=1}^{m} X_i\right)^2\right] - \left(\mathbb{E}[\sum_{i=1}^{m} X_i]\right)^2 \\
&= \mathbb{E}\left[\sum_{i \neq j} X_i X_j + \sum_{i=1}^{m} X_i^2\right] - \left(\sum_{i=1}^{m} \mathbb{E}[X_i]\right)^2 \\
&= \sum_{i \neq j} \mathbb{E}[X_i X_j] + \sum_{i=1}^{m} \mathbb{E}[X_i] - \left(\sum_{i=1}^{m} \mathbb{E}[X_i]\right)^2,
\end{aligned}$$

where the last line followed from linearity of expectation and recognizing that $X_i^2 = X_i$, since it can only take on the values 0 or 1 and either of those values squared is itself, thus $X_i^2 = X_i$.

Since the number of balls that fall into bin i is distributed as Binomial$(n, \frac{1}{m})$, we have

$$\begin{aligned}\mathbb{E}[X_i] &= \mathbb{P}[1 \text{ ball in bin } i] \\ &= \binom{n}{1} \cdot \left(\frac{1}{m}\right)^1 \left(1 - \frac{1}{m}\right)^{n-1} \\ &= \frac{n}{m}\left(1 - \frac{1}{m}\right)^{n-1}\end{aligned}$$

For $j \in \{1, \ldots, n\}$ not equal to i, we can notice that $X_i X_j$ can only take on two values: 0 and 1. This means that $\mathbb{E}[X_i X_j] = \mathbb{P}[X_i X_j = 1] = \mathbb{P}[X_i = 1, X_j = 1]$, or the probability that exactly 1 ball falls into bin i and exactly one ball falls into bin j. This turns out to be

$$\begin{aligned}\mathbb{E}[X_i X_j] &= \mathbb{P}[X_i X_j = 1] \\ &= \underbrace{\binom{n}{1}}_{\substack{\text{choose ball} \\ \text{for bin } i}} \underbrace{\binom{n-1}{1}}_{\substack{\text{choose ball} \\ \text{for bin } j}} \underbrace{\left(\frac{1}{m}\right)^1}_{\substack{\text{chosen ball} \\ \text{goes in bin } i}} \underbrace{\left(\frac{1}{m}\right)^1}_{\substack{\text{chosen ball} \\ \text{goes in bin } j}} \underbrace{\left(1 - \frac{2}{m}\right)^{n-2}}_{\substack{\text{other balls} \\ \text{not in bins } i \text{ or } j}} \\ &= \frac{n(n-1)}{m^2}\left(1 - \frac{2}{m}\right)^{n-2}.\end{aligned}$$

Noting that $\sum_{i \neq j}$ has $m(m-1)$ terms, and the rest of the sums have m terms, we find

$$\mathrm{Var}(X) = m(m-1) \cdot \frac{n(n-1)}{m^2}\left(1 - \frac{2}{m}\right)^{n-2} + m \cdot \frac{n}{m}\left(1 - \frac{1}{m}\right)^{n-1} - m^2\left[\frac{n}{m}\left(1 - \frac{1}{m}\right)^{n-1}\right]^2.$$