Discussion 5D

Tarang Srivastava - CS70 Summer 2020

Mini Review

Lecture Highlights: Variance

$$Var(X) = E[X^2] - E[X]^2$$

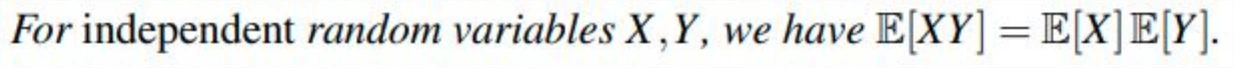
$$Var(X + \overset{\text{constant}}{\overset{\downarrow}{c}}) = Var(X)$$

$$Var(cX) = c^2 \underline{Var(X)}$$

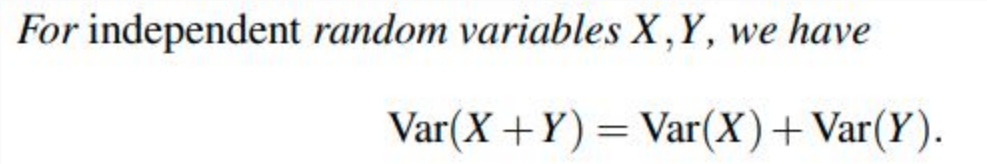
$$\mathrm{Cov}(X,Y) = \mathbb{E}[(X-\mu_X)(Y-\mu_Y)] = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y],$$

For independent *random variables* X, Y, *we have* $\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]$.

Lecture Highlights: Independent R.V.

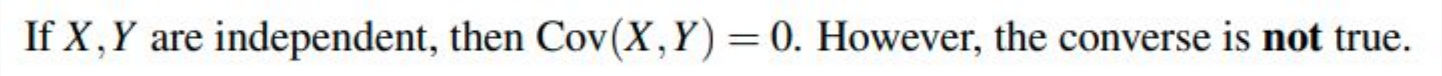


For independent *random variables* X,Y, *we have* $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.



For independent *random variables* X,Y, *we have*

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y).$$



If X,Y are independent, then $\mathrm{Cov}(X,Y) = 0$. However, the converse is **not** true.

Lecture Highlights: Covariance and Correlation

$$\mathrm{Cov}(X,Y) = \mathbb{E}[(X-\mu_X)(Y-\mu_Y)] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y],$$

$$\mathrm{Cov}(X,X) = \mathrm{Var}(X).$$

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X,Y).$$

The square root $\sigma(X) := \sqrt{\mathrm{Var}(X)}$ *is called the* standard deviation *of* X.

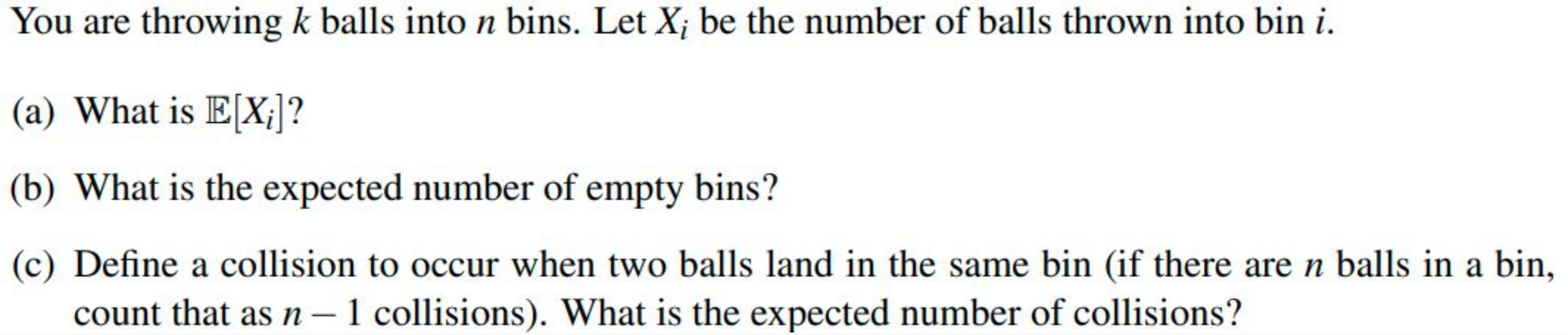
$$\mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma(X)\sigma(Y)}.$$

For any pair of random variables X *and* Y *with* $\sigma(X) > 0$ *and* $\sigma(Y) > 0$,

$$-1 \leq \mathrm{Corr}(X,Y) \leq +1.$$

Question 1

Question 1: Ball in Bins



You are throwing k balls into n bins. Let X_i be the number of balls thrown into bin i.

(a) What is $\mathbb{E}[X_i]$?

(b) What is the expected number of empty bins?

(c) Define a collision to occur when two balls land in the same bin (if there are n balls in a bin, count that as $n-1$ collisions). What is the expected number of collisions?

Question 1: Ball in Bins

You are throwing k balls into n bins. Let X_i be the number of balls thrown into bin i.

(a) What is $\mathbb{E}[X_i]$?

$$B_i = \begin{cases} 1 & \text{iF ball goes into } i \\ 0 & \text{if not.} \end{cases}$$

B_i is the indicator for the i^{th} ball going into Bin i

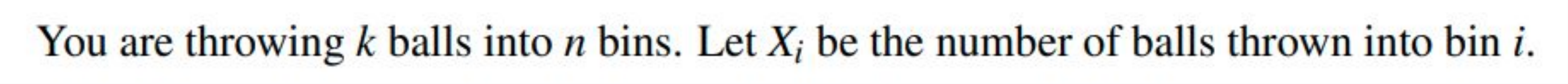
$$E[X_i] = E[B_1] + E[B_2] + \cdots + E[B_u]$$

$$\frac{1}{n} \quad \frac{1}{n} \quad + \cdots \quad \frac{1}{n} = \frac{k}{n}$$

(a) We will use linearity of expectation. Note that the expectation of an indicator variable is just the probability the indicator variable = 1. (Verify for yourself that is true).

$$\mathbb{E}[X_i] = \mathbb{P}[\text{ball 1 falls into bin } i] + \mathbb{P}[\text{ball 2 falls into bin } i] \cdots = \frac{1}{n} + \cdots + \frac{1}{n} = \frac{k}{n}.$$

Question 1: Ball in Bins



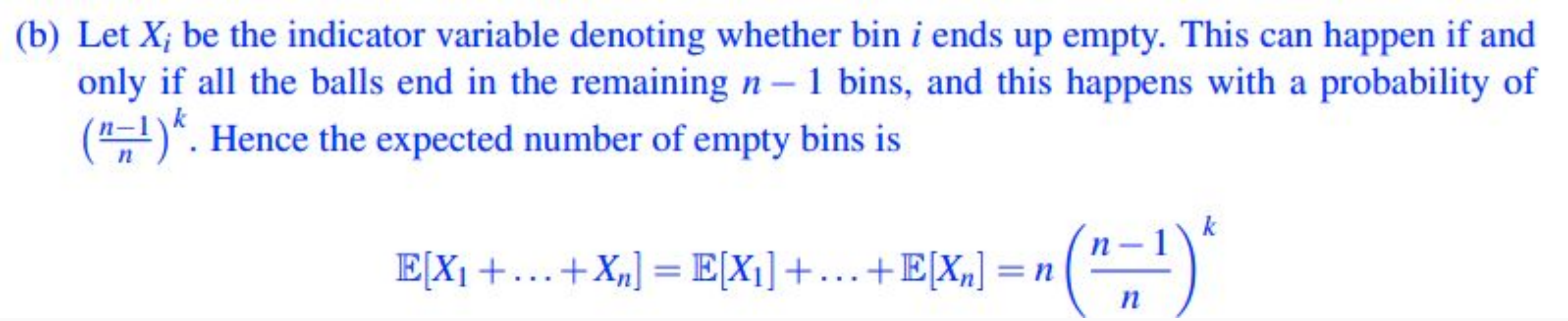
You are throwing k balls into n bins. Let X_i be the number of balls thrown into bin i.

(b) What is the expected number of empty bins?

Y_1 is if bin 1 is empty

↳ $p = \left(\frac{n-1}{n}\right)^k$

$$E[Y] = \left(\frac{n-1}{n}\right)^k \cdot n = n \cdot \left(\frac{n-1}{n}\right)^k$$



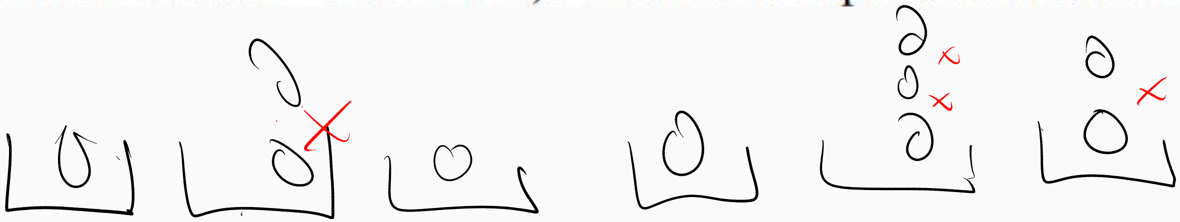
(b) Let X_i be the indicator variable denoting whether bin i ends up empty. This can happen if and only if all the balls end in the remaining $n-1$ bins, and this happens with a probability of $\left(\frac{n-1}{n}\right)^k$. Hence the expected number of empty bins is

$$\mathbb{E}[X_1 + \ldots + X_n] = \mathbb{E}[X_1] + \ldots + \mathbb{E}[X_n] = n\left(\frac{n-1}{n}\right)^k$$

Question 1: Ball in Bins

You are throwing k balls into n bins. Let X_i be the number of balls thrown into bin i.

(c) Define a collision to occur when two balls land in the same bin (if there are n balls in a bin, count that as $n-1$ collisions). What is the expected number of collisions?



$k = 9$

$n = 6$

$n_{Filled} = 5$

$10 - 6 = 4$

$$E(C) = E(k) - E(n - n_{empty})$$

$$k - n - E[n_{empty}]$$

$$k - n - n\left(\frac{n-1}{n}\right)^{n}$$

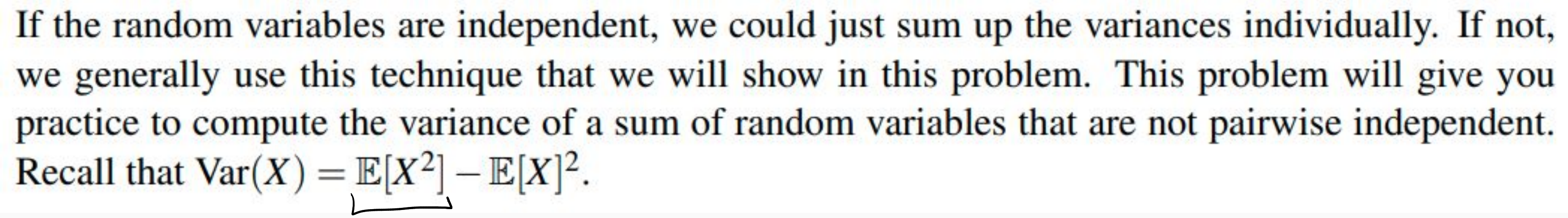
(c) The number of collisions is the number of balls minus the number of occupied bins, since the first ball of every occupied bin is not a collision.

$$\mathbb{E}[\text{collisions}] = k - \mathbb{E}[\text{occupied bins}] = k - n + \mathbb{E}[\text{empty locations}]$$

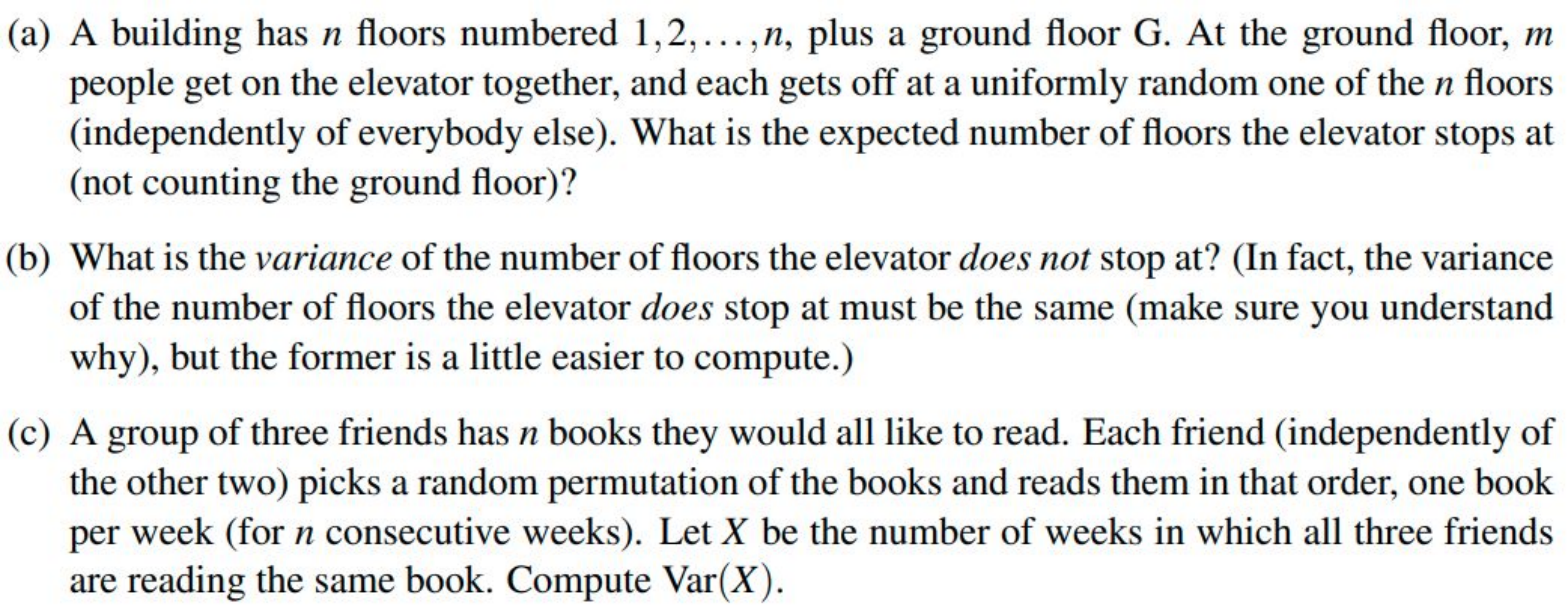
$$= k - n + n\left(1 - \frac{1}{n}\right)^{k}$$

Question 2

Question 2: Variance



If the random variables are independent, we could just sum up the variances individually. If not, we generally use this technique that we will show in this problem. This problem will give you practice to compute the variance of a sum of random variables that are not pairwise independent. Recall that $\mathrm{Var}(X) = \underbrace{\mathbb{E}[X^2]}_{} - \mathbb{E}[X]^2$.



(a) A building has n floors numbered $1, 2, \ldots, n$, plus a ground floor G. At the ground floor, m people get on the elevator together, and each gets off at a uniformly random one of the n floors (independently of everybody else). What is the expected number of floors the elevator stops at (not counting the ground floor)?

(b) What is the *variance* of the number of floors the elevator *does not* stop at? (In fact, the variance of the number of floors the elevator *does* stop at must be the same (make sure you understand why), but the former is a little easier to compute.)

(c) A group of three friends has n books they would all like to read. Each friend (independently of the other two) picks a random permutation of the books and reads them in that order, one book per week (for n consecutive weeks). Let X be the number of weeks in which all three friends are reading the same book. Compute $\mathrm{Var}(X)$.

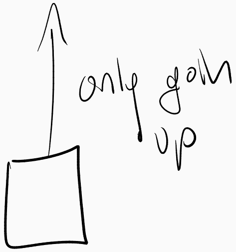
Question 2: Variance

Recall that $\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$.

(a) A building has n floors numbered $1, 2, \ldots, n$, plus a ground floor G. At the ground floor, m people get on the elevator together, and each gets off at a uniformly random one of the n floors (independently of everybody else). What is the expected number of floors the elevator stops at (not counting the ground floor)?

X : # of Floors we stop at

I_i : if we stopped at floor i → indicator



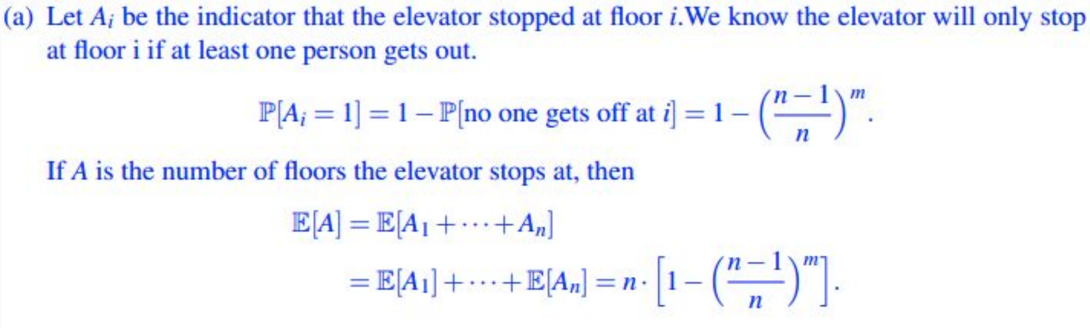
$$X = I_1 + I_2 + \cdots + I_{\boxed{n}} \text{ — n floors}$$

$$E[I_i] = 1 - \left(\begin{matrix}\text{no one} \\ \text{gets off}\end{matrix}\right)$$

$$E[X] = \sum_{i=1}^{n} E[I_i] = \boxed{n\cdot} E[I_i]$$

$$E[I_i] = \boxed{\left[1 - \left(\frac{n-1}{n}\right)^m\right]}$$

$$= \cancel{0 \cdot P(I_i = 0)} + 1 \cdot P(I_i = 1)$$



(a) Let A_i be the indicator that the elevator stopped at floor i.We know the elevator will only stop at floor i if at least one person gets out.

$$\mathbb{P}[A_i = 1] = 1 - \mathbb{P}[\text{no one gets off at } i] = 1 - \left(\frac{n-1}{n}\right)^m.$$

If A is the number of floors the elevator stops at, then

$$\mathbb{E}[A] = \mathbb{E}[A_1 + \cdots + A_n]$$
$$= \mathbb{E}[A_1] + \cdots + \mathbb{E}[A_n] = n \cdot \left[1 - \left(\frac{n-1}{n}\right)^m\right].$$

$$\boxed{n\left[1 - \left(\frac{n-1}{n}\right)^m\right]}$$

Question 2: Variance

Recall that $\mathrm{Var}(X) = \boxed{\mathbb{E}[X^2]} - \mathbb{E}[X]^2$.

(b) What is the *variance* of the number of floors the elevator *does not* stop at? (In fact, the variance of the number of floors the elevator *does* stop at must be the same (make sure you understand why), but the former is a little easier to compute.)

X : # Floor dont stop at

I_i : if we dont stop at floor i

$X = I_1 + I_2 + \dots + I_n$

$E[I_i] = \left(\frac{n-1}{n}\right)^m$

By linearity of expectation

$$E[X] = \sum_{i=1}^{n} E[I_i] = n \cdot \left(\frac{n-1}{n}\right)^m \Rightarrow E[X]^2 = \boxed{n^2 \left(\frac{n-1}{n}\right)^{2m}}$$

$(n-1)^n$

$$E[X^2] = E[X \cdot X] = E[(I_1 + \dots + I_n)(I_1 + \dots + I_n)]$$

$$= \sum_{i=1}^{n} E[I_i I_i] = E[I_i] \qquad \sum_{i \neq j} E[\overset{0,1}{I_i}\,\overset{0,1}{I_j}]$$

how many terms are there? $n(n-1)$

$P[I_i I_i = 1]$

$[I_i^2 = 1]$

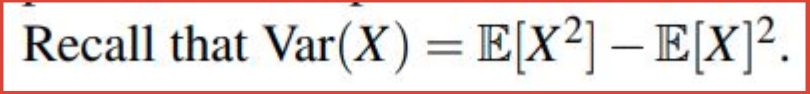
$P[I_i \cdot I_j = 1]$

$P[I_i = 1 \wedge I_j = 1] = \left(\frac{n-2}{n}\right)^m$

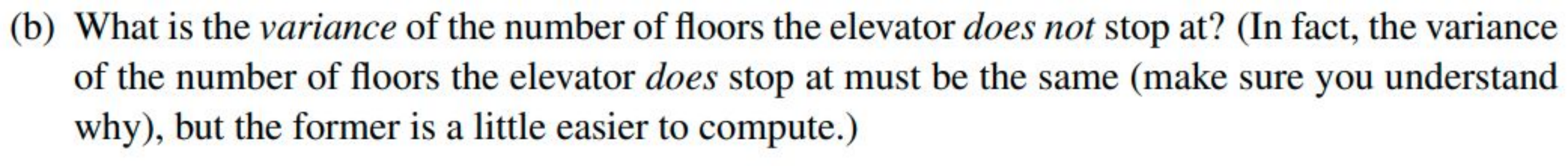
$$E[X^2] = n \cdot \left(\frac{n-1}{n}\right)^m + n(n-1)\left(\frac{n-2}{n}\right)^m \left(\frac{n-1}{n}\right)^m$$

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

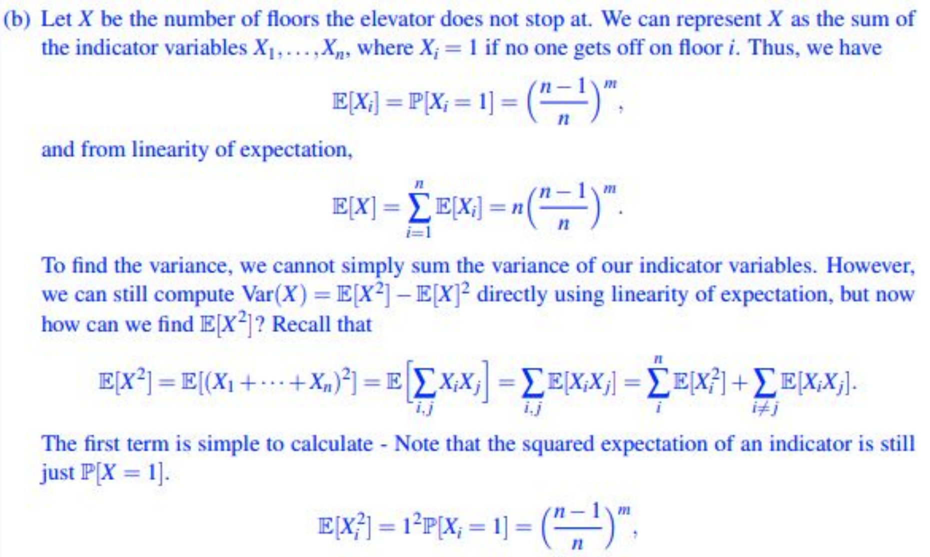
Question 2: Variance



Recall that $\mathrm{Var}(X)=\mathbb{E}[X^2]-\mathbb{E}[X]^2$.



(b) What is the *variance* of the number of floors the elevator *does not* stop at? (In fact, the variance of the number of floors the elevator *does* stop at must be the same (make sure you understand why), but the former is a little easier to compute.)



(b) Let X be the number of floors the elevator does not stop at. We can represent X as the sum of the indicator variables $X_1,\dots,X_n$, where $X_i=1$ if no one gets off on floor i. Thus, we have

$$\mathbb{E}[X_i]=\mathbb{P}[X_i=1]=\left(\frac{n-1}{n}\right)^m,$$

and from linearity of expectation,

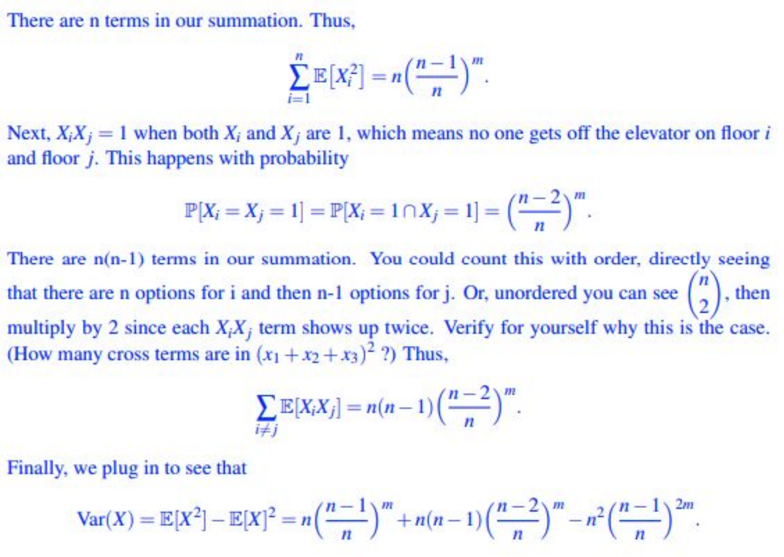
$$\mathbb{E}[X]=\sum_{i=1}^{n}\mathbb{E}[X_i]=n\left(\frac{n-1}{n}\right)^m.$$

To find the variance, we cannot simply sum the variance of our indicator variables. However, we can still compute $\mathrm{Var}(X)=\mathbb{E}[X^2]-\mathbb{E}[X]^2$ directly using linearity of expectation, but now how can we find $\mathbb{E}[X^2]$? Recall that

$$\mathbb{E}[X^2]=\mathbb{E}[(X_1+\cdots+X_n)^2]=\mathbb{E}\left[\sum_{i,j}X_iX_j\right]=\sum_{i,j}\mathbb{E}[X_iX_j]=\sum_{i}^{n}\mathbb{E}[X_i^2]+\sum_{i\neq j}\mathbb{E}[X_iX_j].$$

The first term is simple to calculate - Note that the squared expectation of an indicator is still just $\mathbb{P}[X=1]$.

$$\mathbb{E}[X_i^2]=1^2\mathbb{P}[X_i=1]=\left(\frac{n-1}{n}\right)^m,$$



There are n terms in our summation. Thus,

$$\sum_{i=1}^{n}\mathbb{E}[X_i^2]=n\left(\frac{n-1}{n}\right)^m.$$

Next, $X_iX_j=1$ when both X_i and X_j are 1, which means no one gets off the elevator on floor i and floor j. This happens with probability

$$\mathbb{P}[X_i=X_j=1]=\mathbb{P}[X_i=1\cap X_j=1]=\left(\frac{n-2}{n}\right)^m.$$

There are n(n-1) terms in our summation. You could count this with order, directly seeing that there are n options for i and then n-1 options for j. Or, unordered you can see $\binom{n}{2}$, then multiply by 2 since each X_iX_j term shows up twice. Verify for yourself why this is the case. (How many cross terms are in $(x_1+x_2+x_3)^2$?) Thus,

$$\sum_{i\neq j}\mathbb{E}[X_iX_j]=n(n-1)\left(\frac{n-2}{n}\right)^m.$$

Finally, we plug in to see that

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

Question 2: Variance

Recall that $\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$.

(c) A group of three friends has n books they would all like to read. Each friend (independently of the other two) picks a random permutation of the books and reads them in that order, one book per week (for n consecutive weeks). Let X be the number of weeks in which all three friends are reading the same book. Compute $\mathrm{Var}(X)$.

F_1 :	b_1	b_4	b_3	b_2
F_2	b_3	b_4	b_2	b_1
F_3	b_3	b_4	b_2	b_1

I_i = on the ith week to all read the same book

$$E[I_i] = 1 \cdot \frac{1}{n} \cdot \frac{1}{n} = \frac{1}{n^2}$$

$$E[X] = \sum_{i=1}^{n} E[I_i] = n \cdot \frac{1}{n^2} = \boxed{\frac{1}{n}}$$

$$\boxed{E[X]^2 = \frac{1}{n^2}}$$

$$X = I_1 + I_2 + \dots + I_n$$

$$E[X^2] = E[X \cdot X] = E[(I_1 + \dots + I_n)(I_1 + \dots + I_n)]$$

how many terms $n \cdot (n-1)$

$$\sum_{i \neq j} E[I_i I_j]$$

$$P(I_i = 1 \wedge I_j = 1)$$

ith week $\frac{1}{n} \cdot \frac{1}{n}$ jth $\frac{1}{n-1} \frac{1}{n-1} = \boxed{\frac{1}{n^2} \cdot \frac{1}{(n-1)^2}}$

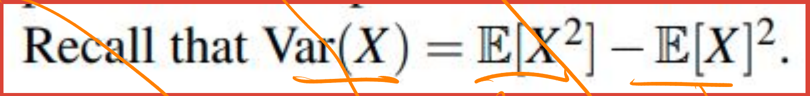
$$\sum_{i=1}^{n} E[I_i I_i] = E[I_i]$$

$$\frac{1}{n^2}$$

$$n \cdot \frac{1}{n^2} = \boxed{\frac{1}{n}}$$

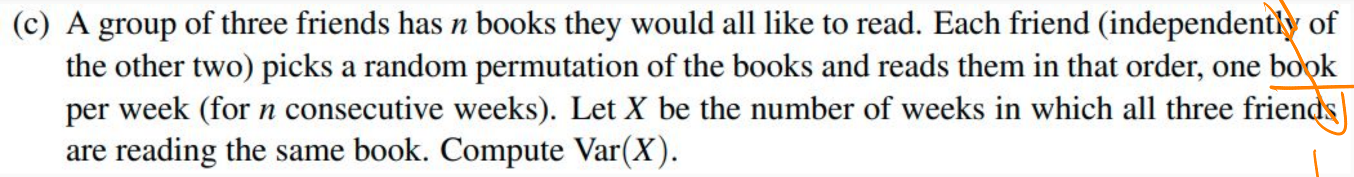
$$\frac{1}{n(n-1)} \leftarrow n \cdot (n-1) \cdot \frac{1}{n^2} \cdot \frac{1}{(n-1)^2}$$

Question 2: Variance

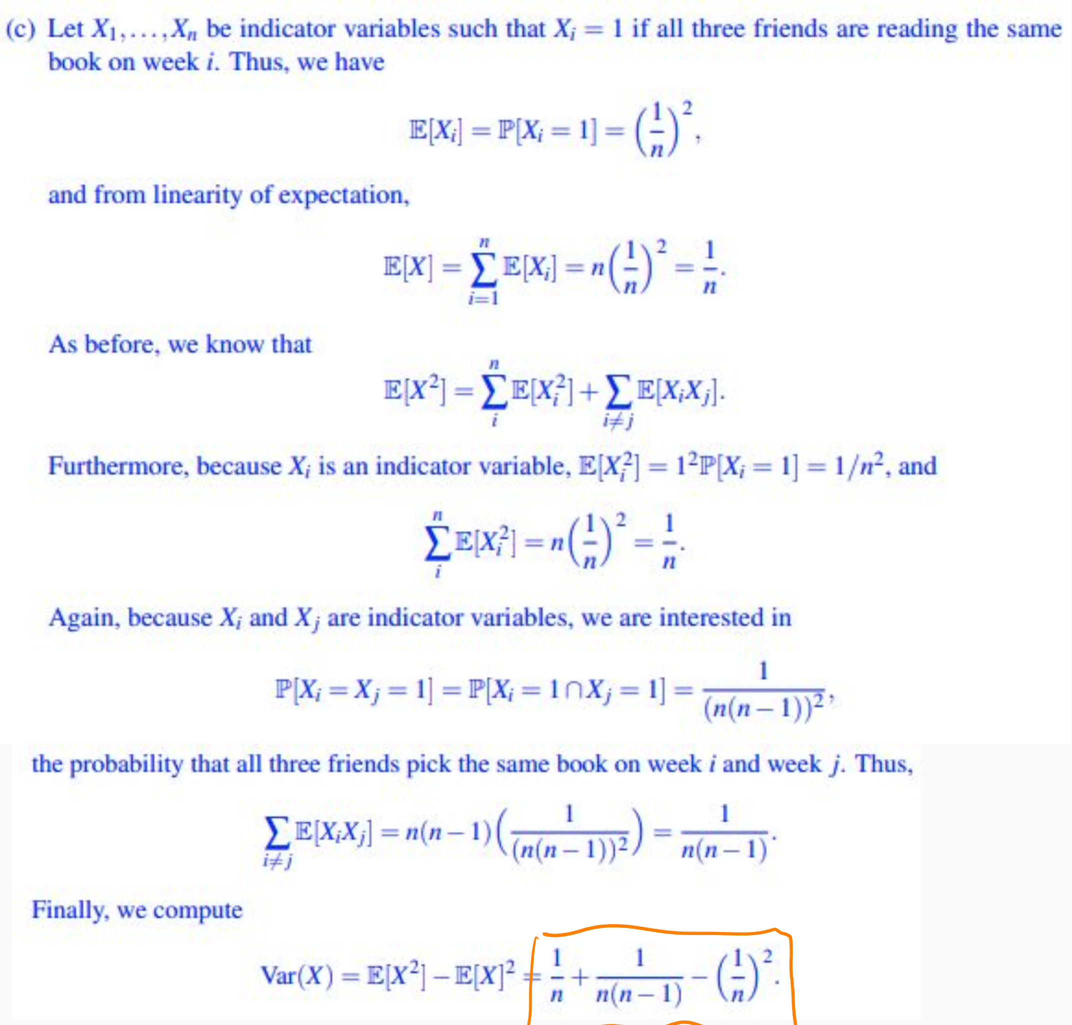


Recall that $\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$.

$\frac{1}{n} + \frac{1}{n(n-1)} - \frac{1}{n^2}$



(c) A group of three friends has n books they would all like to read. Each friend (independently of the other two) picks a random permutation of the books and reads them in that order, one book per week (for n consecutive weeks). Let X be the number of weeks in which all three friends are reading the same book. Compute $\mathrm{Var}(X)$.



(c) Let $X_1, \ldots, X_n$ be indicator variables such that $X_i = 1$ if all three friends are reading the same book on week i. Thus, we have

$$\mathbb{E}[X_i] = \mathbb{P}[X_i = 1] = \left(\frac{1}{n}\right)^2,$$

and from linearity of expectation,

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

As before, we know that

$$\mathbb{E}[X^2] = \sum_{i}^{n} \mathbb{E}[X_i^2] + \sum_{i \neq j} \mathbb{E}[X_i X_j].$$

Furthermore, because X_i is an indicator variable, $\mathbb{E}[X_i^2] = 1^2 \mathbb{P}[X_i = 1] = 1/n^2$, and

$$\sum_{i}^{n} \mathbb{E}[X_i^2] = n\left(\frac{1}{n}\right)^2 = \frac{1}{n}.$$

Again, because X_i and X_j are indicator variables, we are interested in

$$\mathbb{P}[X_i = X_j = 1] = \mathbb{P}[X_i = 1 \cap X_j = 1] = \frac{1}{(n(n-1))^2},$$

the probability that all three friends pick the same book on week i and week j. Thus,

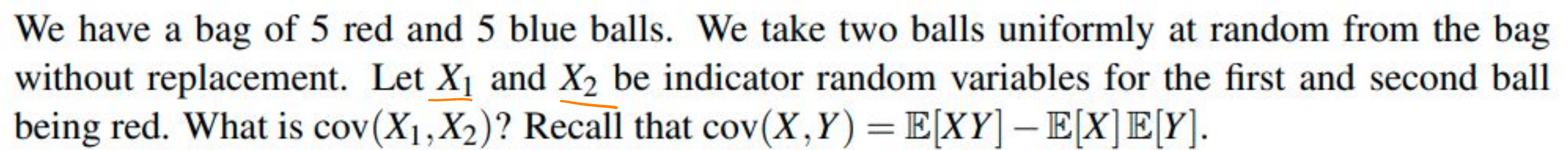
$$\sum_{i \neq j} \mathbb{E}[X_i X_j] = n(n-1)\left(\frac{1}{(n(n-1))^2}\right) = \frac{1}{n(n-1)}.$$

Finally, we compute

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

Question 3

Question 3



We have a bag of 5 red and 5 blue balls. We take two balls uniformly at random from the bag without replacement. Let X_1 and X_2 be indicator random variables for the first and second ball being red. What is $\mathrm{cov}(X_1, X_2)$? Recall that $\mathrm{cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$.

Solution :

Computational

$$E[X_1] = \frac{5}{10}$$

reccomended

Concrete

for first ball

pick red

pick blue

$$E[X_2] = \text{Abstract } \boxed{\frac{5}{10}} \checkmark \qquad \frac{5}{10}\cdot\frac{4}{9} + \frac{5}{10}\,\frac{5}{9}$$

$$\frac{20}{90} + \frac{25}{90} = \frac{45}{90} = \frac{1}{2} \checkmark$$

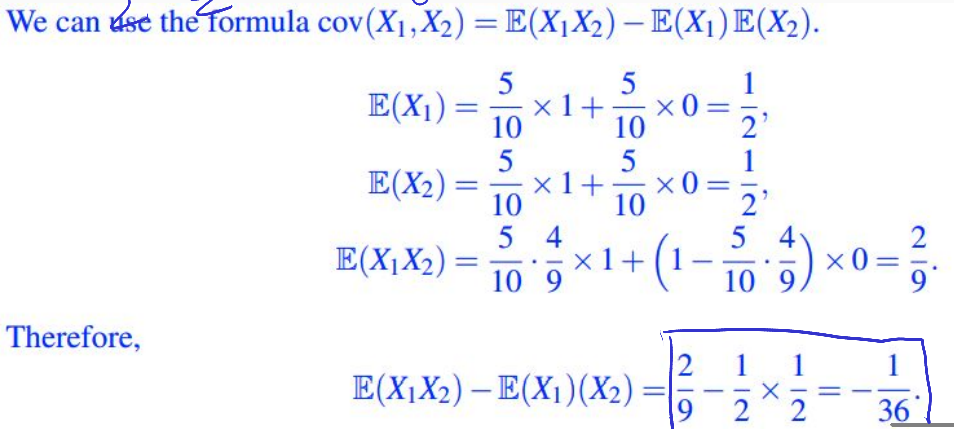
$$E[X_1 \cdot X_2]$$

$$\mathrm{Cov}(X_1, X_2) = \frac{2}{9} - \frac{1}{2}\cdot\frac{1}{2} = -\frac{1}{36}$$

$$P(\overset{1,0}{X_1} \cdot \overset{1,0}{X_2} = 1)$$

$$p(X_1 = 1 \wedge X_2 = 1)$$

$$\frac{5}{10}\cdot\frac{4}{9} = \frac{20}{90} = \frac{2}{9}$$



We can use the formula $\mathrm{cov}(X_1,X_2) = \mathbb{E}(X_1X_2) - \mathbb{E}(X_1)\mathbb{E}(X_2)$.

$$\mathbb{E}(X_1) = \frac{5}{10}\times 1 + \frac{5}{10}\times 0 = \frac{1}{2},$$

$$\mathbb{E}(X_2) = \frac{5}{10}\times 1 + \frac{5}{10}\times 0 = \frac{1}{2},$$

$$\mathbb{E}(X_1X_2) = \frac{5}{10}\cdot\frac{4}{9}\times 1 + \left(1 - \frac{5}{10}\cdot\frac{4}{9}\right)\times 0 = \frac{2}{9}.$$

Therefore,

$$\mathbb{E}(X_1X_2) - \mathbb{E}(X_1)(X_2) = \boxed{\frac{2}{9} - \frac{1}{2}\times\frac{1}{2} = -\frac{1}{36}.}$$

Question 3