

Variance Intro

Note 20

Variance: denoted by $\text{Var}(X)$; measure of how much X deviates from its mean, i.e. its spread.

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

Properties: for random variables X, Y and constant a ,

- $\text{Var}(aX) = a^2 \text{Var}(X)$
- $\text{Var}(X + a) = \text{Var}(X)$
- If X, Y independent, then $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$

Variance of sum of (not necessarily independent) indicator variables: Let X_1, \dots, X_n be indicator variables for events A_1, \dots, A_n , respectively (i.e., $X_i = 1$ if event A_i occurs, and 0 otherwise). The variance of the sum $X = X_1 + \dots + X_n$ can be calculated as:

$$\text{Var}(X) = \mathbb{E}[(X_1 + \dots + X_n)^2] - \mathbb{E}[X_1 + \dots + X_n]^2 = \sum_{i=1}^n \mathbb{E}[X_i^2] + \sum_{i \neq j} \mathbb{E}[X_i X_j] - \left(\sum_{i=1}^n \mathbb{E}[X_i] \right)^2$$

Note that the term $\sum_{i \neq j} \mathbb{E}[X_i X_j]$ is equivalent to $2 \sum_{i < j} \mathbb{E}[X_i X_j]$.

$\mathbb{E}[X_i^2] = \mathbb{E}[X_i] = \mathbb{P}[A_i]$ since $X_i^2 = X_i$ for indicator variables, and $\mathbb{E}[X_i X_j] = \mathbb{P}[A_i \cap A_j]$.

1 Student Life

Note 20

In an attempt to avoid having to do laundry often, Marcus comes up with a system. Every night, he designates one of his shirts as his dirtiest shirt. In the morning, he randomly picks one of his shirts to wear. If he picked the dirtiest one, he puts it in a dirty pile at the end of the day (a shirt in the dirty pile is not used again until it is cleaned, and the dirty pile is not considered as one of the n locations).

When Marcus puts his last shirt into the dirty pile, he finally does his laundry, and again designates one of his shirts as his dirtiest shirt (laundry isn't perfect) before going to bed. This process then repeats.

- (a) If Marcus has n shirts, what is the expected number of days between laundry events? Your answer should be a function of n involving no summations.

- (b) Now, instead of organizing his shirts in his dresser, he throws his shirts randomly onto one of n different locations in his room (one shirt per location), designates one of his shirts as his dirtiest shirt, and one location as the dirtiest location.

In the morning, if he happens to pick the dirtiest shirt, *and* the dirtiest shirt was in the dirtiest location, then he puts the shirt into the dirty pile at the end of the day. He does not throw any future shirts into that location and also does not consider it as a candidate for future dirtiest locations (as it is too dirty).

What is the expected number of days between laundry events now? Again, your answer should be a function of n involving no summations.

2 Dice Variance

Note 20

- (a) Let X be a random variable representing the outcome of the roll of one fair 6-sided die. What is $\text{Var}(X)$?

(b) Let Z be a random variable representing the average of n rolls of a fair 6-sided die. What is $\text{Var}(Z)$?

3 Elevator Variance

Note 20

A building has n upper floors numbered $1, 2, \dots, n$, plus a ground floor G . At the ground floor, m people get on the elevator together, and each person gets off at one of the n upper floors uniformly at random and independently of everyone else.

- (a) If N is the number of floors the elevator does not stop at, express N as a sum of indicator random variables and compute $\mathbb{E}[N]$.
- (b) Write N^2 in terms of the indicators you defined in part (a) and compute $\mathbb{E}[N^2]$.
- (c) Using your answers to the previous parts, compute $\text{Var}(N)$.