Variance and Covariance Intro

Note 17

The **Law of the Unconscious Statistician** (LOTUS): for RV X and function f on the range of X:

$$\mathbb{E}[f(X)] = \sum_k f(k) \cdot \mathbb{P}[X = k].$$

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 are independent, then $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$

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

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

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

Covariance: measure of the relationship between two random variables X, Y:

$$\text{cov}(X,Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y].$$

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

- $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2\,\text{cov}(X,Y)$
- $\text{cov}(X,X) = \text{Var}(X)$
- $\text{cov}(X,Y) = \text{cov}(Y,X)$
- Bilinearity: $\text{cov}(X+Y,Z) = \text{cov}(X,Z) + \text{cov}(Y,Z)$ and $\text{cov}(aX,Y) = a\,\text{cov}(X,Y)$

Conditional Expectation: The expectation of a random variable X conditioned on an event A:

$$\mathbb{E}[X \mid A] = \sum_x x \cdot \mathbb{P}[(X = x) \mid A].$$

Often, we use $Y = y$ as the given event. In this case, $\mathbb{E}[X \mid Y = y]$ is a function of Y: it takes inputs $y \in Y$ and outputs $f(y) = \mathbb{E}[X \mid Y = y]$. So $f(Y) = \mathbb{E}[X \mid Y]$ is itself a random variable.

Law of Total Expectation: for random variables X, Y (where $\mathbb{E}[|X|] < \infty$):

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X \mid Y]]$$

1 Dice Variance

Note 17

(a) Let X be a random variable representing the outcome of one roll of a 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 $\text{Var}(Z)$?

2 Indicator Expectation

Note 16
Note 17

Let A_1, A_2 be two events where $\mathbb{P}[A_1] = p_1, \mathbb{P}[A_2] = p_2, \mathbb{P}[A_1 \cap A_2] = p_3$. Let X_1, X_2 be indicator variables for A_1 and A_2, respectively.

(a) What is $\mathbb{E}[X_1^2]$? (*Hint*: think about what values X_1^2 can take on, with what probability.)

(b) What is $\mathbb{E}[X_1 X_2]$?

3 Elevator Variance

Note 17

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) Let S be the number of floors the elevator skips. Express S as a sum of indicator random variables and compute $\mathbb{E}[S]$.

(b) Write S^2 in terms of the indicators you defined in part (a) and compute $\mathbb{E}[S^2]$.

(c) Using your answers to the previous parts, compute $\mathrm{Var}(S)$.

4 Covariance

Note 17

(a) 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 events of the first and second ball being red, respectively. What is $\text{cov}(X_1, X_2)$?

(b) Now, we have two bags A and B, with 5 red and 5 blue balls each. Draw a ball uniformly at random from A, record its color, and then place it in B. Then draw a ball uniformly at random from B and record its color. Let X_1 and X_2 be indicator random variables for the events of the first and second draws being red, respectively. What is $\text{cov}(X_1, X_2)$?

5 Number Game

Note 17

Sinho and Vrettos are playing a game where they each choose an integer uniformly at random from $[0, 100]$, then whoever has the larger number wins (in the event of a tie, they replay). However, Vrettos doesn't like losing, so he's rigged his random number generator such that it instead picks randomly from the integers between Sinho's number and 100. Let S be Sinho's number and V be Vrettos' number.

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

(b) What is $\mathbb{E}[V \mid S = s]$, where s is any constant such that $0 \leq s \leq 100$?

(c) What is $\mathbb{E}[V]$?