Covariance and Total Expectation Intro

Covariance: measure of the relationship between two RVs

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

The sign of $\text{cov}(X,Y)$ illustrates how X and Y are related; a positive value means that X and Y tend to increase and decrease together, while a negative value means that X increases as Y decreases (and vice versa). A covariance of zero means that the two random variables are uncorrelated—there is no linear relationship between them.

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: When we want to find the expectation of a random variable X conditioned on an event A, we use the following formula:

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

This is an application of the definition of expectation. We still consider all values of X but reweigh them based on their probability of occurring together with A.

Total Expectation: For any random variable X and events $A_1, A_2, \ldots, A_n$ that partition the sample space Ω,

$$\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X \mid A_i]\,\mathbb{P}[A_i].$$

We can think of this as splitting the sample space into partitions (events) and looking at the expectation of X in each partition, weighted by the probability of that event occurring.

Often, we use another random variable to construct the partition. If Y is a random variable, then the events $Y = y_1, Y = y_2, \ldots$ partition the sample space, where $\{y_1, y_2, \ldots\}$ are all the possible values of Y. 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.

1 Covariance

Note 21

(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 $\operatorname{cov}(X_1, X_2)$? Recall that $\operatorname{cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]$.

(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 $\operatorname{cov}(X_1, X_2)$?

2 Correlation and Independence

Note 21

(a) What does it mean for two random variables to be uncorrelated?

(b) What does it mean for two random variables to be independent?

(c) Are all uncorrelated variables independent? Are all independent variables uncorrelated? If your answer is yes, justify your answer; if your answer is no, give a counterexample.

3 Dice Games

Note 21

Suppose you roll a fair six-sided die. You read off the number showing on the die, then flip that many fair coins.

(a) If the result of your die roll is i, what is the expected number of heads you see?

(b) What is the expected number of heads you see?

4 Number Game

Note 21

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 $\mathbb{E}[V]$?