Covariance and Correlation

CS 70, Summer 2019

Lecture 22, 7/31/19

Last Time...

- Variance measures **deviation from mean**
- Variance is additive for **independent RVs**
- Use **linearity of expectation** and **indicator variables**

Today:

- Proof that var. is additive for ind. RVs
- Talk about **covariance and correlation**
- Some RV practice (if time)

Product of RVs

Let X be a RV with values in A.
Let Y be a RV with values in B.

What does the distribution of XY look like?

What if X, Y are **independent**?

For Independent RVs...

If X and Y are independent, we can show that:

$$\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$$

Variance is Additive for Ind. RVs

If X and Y are independent, we can show that:

$$\text{Var}(X + Y) = \text{Var}[X] + \text{Var}[Y]$$

Can You Multiply?

If X and Y are **independent**, then:

$$\mathbb{E}[XY] =$$

Is the **converse** true?

Covariance

Measures "how independent" two RVs are. Let $\mathbb{E}[X] = \mu_1$, and let $\mathbb{E}[Y] = \mu_2$.

$$\text{Cov}(X, Y) =$$

Alternate Form:

$$\text{Cov}(X, Y) =$$

Example: Coin Flips I

I flip two **fair** coins.
Let X count the number of heads, and let Y be an indicator for the first coin being a head.

First, what is XY?

Example: Coin Flips I

What is $\text{Cov}(X, Y)$?

Properties of Covariance I

If X and Y are independent, then:

$$\text{Cov}(X, Y) =$$

Is the **converse** true?

Properties of Covariance II

What happens when we take $\text{Cov}(X, X)$?

Can use **either definition** of covariance!

Properties of Covariance III

Covariance is **bilinear.**

$$\text{Cov}(a_1X_1 + a_2X_2, Y) =$$

$$\text{Cov}(X, b_1Y_1 + b_2Y_2) =$$

Practice: Bilinearity!

Simplify $\text{Cov}(3X + 4Y, 5X - 2Y)$.

Example: Coin Flips II

Same setup: Two fair flips. X is number of heads, and Y is indicator for the first coin heads.

Recall: $\text{Cov}(X, Y) =$

Now, let Y' be an indicator for the first coin being a tail. How does the covariance change?

Properties of Covariance IV

For **any** two RVs X, Y:

$$\text{Var}(X + Y) =$$

Break

If you could eliminate one food so that **no one would eat it ever again**, what would you pick to destroy?

Correlation

For **any** two RVs X, Y that are **not constant**:

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

Sanity Check!

What is $\text{Corr}(X, X)$?

What is $\text{Corr}(X, -X)$?

What is $\text{Corr}(X, Y)$ for X, Y independent?

Example: Coin Flips II

I flip two **fair** coins.
Let X count the number of heads, and let Y be an indicator for the first coin being a head.

What is $\text{Corr}(X, Y)$?

Size of Correlation?

For any RVs X and Y that are **not constants**:

$$-1 \leq \text{Corr}(X, Y) \leq 1$$

Proof: Define new RVs using X and Y:

$\tilde{X} =$
$\tilde{Y} =$

Size of Correlation?

(Continued:)

RV Practice: Two Roads

There are two paths from Soda to VLSB.
I usually choose a path uniformly at random.
minutes spent on Path 1 is a Geometric(p_1) RV.
minutes spent on Path 2 is a Geometric(p_2) RV.

Today, it took me 6 minutes to walk from Soda to VLSB. Given this, what is the probability that I chose Path 1?

RV Practice: Two Roads

Continued:

RV Practice: RandomSort

I have cards labeled $1, 2, \ldots, n$. They are shuffled.

I want them in order. I sort them in a naive way. I start with all cards in an "unsorted" pile. I draw cards from the unsorted pile uniformly at random until I get card 1. I place card 1 in a "sorted" pile, and continue, this time looking for card 2.

RV Practice: RandomSort

What is the expected number of draws I need?

What is the variance of the number of draws?

RV Practice: Packets

Packets arrive from sources A and B.
I fix a time interval. Over this interval, the number of packets from A and B have $\text{Poisson}(\lambda_A)$ and $\text{Poisson}(\lambda_B)$ distributions, and are **independent**.

What is the distribution of the total number of packets I receive in this time interval?

RV Practice: Packets

What is the probability that over this interval, I receive **exactly** 2 packets?

What is the expected number of packets I receive over this interval?

What is the variance of this number?

Summary

- Covariance and correlation measure **how independent** two RVs are.
- Variance can be **expressed** and **manipulated** in terms of covariance.
- Independent RVs have **zero covariance** and **zero correlation**. However, **the converse is not true!**