CS 240: Reasoning Under Uncertainty (Fall 21) Nov 1

Lecture 15

Instructor: *Profs Peter J. Hass and Jie Xiong* SI Worksheet: *Juelin Liu*

1 Introduction

In this lecture, we discussed about covariance and correlation.

2 Covariance

The **covariance** between any two RVs (either discrete or continuous) X and Y is one measure of dependence that quantifies the degree to which there is a **linear relationship** between X and Y.

$$Cov(X, Y) = E(X - E(X))E(Y - E(Y)) = E(XY) - E(X)E(Y)$$

Note that X and Y are independent implying that $Cov(X, Y) = 0$ but not the other way around.
Some properties of covariance:

$$Cov(X, Y) = E(XY) - E(X)E(Y) = E(YX) - E(Y)E(X) = Cov(Y, X)$$

$$\begin{aligned} Cov(X, aY + b) &= E(X(aY + b)) - E(X)E(aY + b) \\ &= (aE(YX) + bE(X)) - (aE(Y)E(X) + bE(X)) \\ &= a(E(XY) - E(X)E(Y)) \\ &= aCov(X, Y) \end{aligned}$$

$$\begin{aligned} Cov(X, Y + Z) &= E(X(Y + Z)) - E(X)E(Y + Z) \\ &= E(XY) - E(X)E(Y) + E(XZ) - E(X)E(Z) \\ &= Cov(X, Y) + Cov(X, Z) \end{aligned}$$

$$Cov(X \sum_{i=1}^{n} Y_i) = \sum_{i=1}^{n} Cov(X, Y_i)$$

3 Correlation

The magnitude of a covariance provides little information about whether two RVs X and Y have a linear relationship. X and Y might have large covariance simply because they have large values. ex: $Cov(100X, 100Y) = 10000Cov(X, Y)$

To normalize the covariance we can use:

$$\rho(X,Y) = \frac{cov(X,Y)}{\sqrt{var(X)var(Y)}}$$

ρ is the **correlation** between X and Y, it is within the range $[-1, 1]$.

Note that the correlation with a value close to 1 or -1 implies a strong linear relationship between X and Y. A value close to 0 indicates little or no linear relationship between X and Y.

4 Problems

1. Given the following PDF, compute the covariance of X and Y.

$$f_{X,Y} = \begin{cases} 4xy, & 0 \le x \le 1, 0 \le y \le 1 \\ 0, & elsewhere \end{cases}$$

2. Given the following PDF, compute the covariance of X and Y. Are X and Y independent?

$$f_{X,Y} = \begin{cases} 6(1-y), & 0 \le x \le y \le 1 \\ 0, & elsewhere \end{cases}$$

3. Given the following PDF:

$$f_{X,Y} = \begin{cases} 1, & y-1 \le x \le 1-y,\ 0 \le y \le 1 \\ 0, & elsewhere \end{cases}$$

a) Compute the $Cov(X,Y)$
b) Are X and Y independent?
c) Find the correlation between X and Y.

5 Answers

1.

$$E(XY) = \int_0^1 \int_0^1 xy \times 4xydxdy = \int_0^1 \frac{4y^2}{3}dy = \frac{4}{9}$$

$$E(X) = \int_0^1 \int_0^1 x \times 4xydxdy = \int_0^1 \frac{4y}{3}dy = \frac{2}{3}$$

$$E(Y) = \int_0^1 \int_0^1 y \times 4xydxdy = \int_0^1 2y^2dy = \frac{2}{3}$$

$$Cov(X,Y) = E(XY) - E(X)E(Y) = \frac{4}{9} - \frac{2}{3} \times \frac{2}{3} = 0$$

2.

$$E(XY) = \int_0^1 \int_0^y xy \times 6(1-y)dxdy = \int_0^1 3y^3 - 3y^4 dy = \frac{3}{20}$$

$$E(X) = \int_0^1 \int_0^y x \times 6(1-y)dxdy = \int_0^1 3y^2 - 3y^3 dy = \frac{1}{4}$$

$$E(Y) = \int_0^1 \int_0^y y \times 6(1-y)dxdy = \int_0^1 6y^2 - 6y^3 dy = \frac{1}{2}$$

$$Cov(X,Y) = E(XY) - E(X)E(Y) = \frac{3}{20} - \frac{1}{4} \times \frac{1}{2} = \frac{1}{40}$$

They are not independent. If they are independent, the covariance has to be 0.

3. a)

$$E(XY) = \int_0^1 \int_{y-1}^{1-y} xy \times 1dxdy = \int_0^1 0dy = 0$$

$$E(X) = \int_0^1 \int_{y-1}^{1-y} x \times 1dxdy = \int_0^1 0dy = 0$$

$$E(Y) = \int_0^1 \int_{y-1}^{1-y} y \times 1dxdy = \int_0^1 2y - 2y^2 dy = \frac{1}{3}$$

$$Cov(X,Y) = E(XY) - E(X)E(Y) = 0 - 0 \times \frac{1}{3} = 0$$

b) No, they are dependent.

$$f_X = \int_0^1 1dy = 1$$

$$f_Y = \int_{y-1}^{1-y} 1dx = 2 - 2y$$

$$f_X f_Y = 2 - 2y \neq f_{X,Y} = 1$$

c) You actually don't need to compute $Var(X)$ and $Var(Y)$ in this case because the covariance is 0. But generally speaking, you need to compute them to get the correlation.

$$E(X^2) = \int_0^1 \int_{y-1}^{1-y} x^2 \times 1dxdy = \int_0^1 \frac{2}{3}(1-y)^3 dy = \frac{1}{6}$$

$$Var(X) = E(X^2) - E(X)^2 = \frac{1}{6} - 0^2 = \frac{1}{6}$$

$$E(Y^2) = \int_0^1 \int_{y-1}^{1-y} y^2 \times 1dxdy = \int_0^1 2y^2 - 2y^3 dy = \frac{1}{6}$$

$$Var(Y) = E(Y^2) - E(Y)^2 = \frac{1}{6} - \frac{1}{3} \times \frac{1}{3} = \frac{1}{18}$$

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = 0$$