CMPE 320: Probability, Statistics, and Random Processes

# Lecture 21: Covariance and correlation

Spring 2018

Seung-Jun Kim



## Announcement

- Don't forget to put in your course evaluation!
- Special review session: Monday 11am-1pm. E-mail to the TA if you want to attend.

# Introduction

- "Covariance" provides a quantitative measure of the strength and direction of the relationship between 2 RVs

- It plays important role in variety of contexts, especially when one wants to predict the value of one RV based on the value of the other



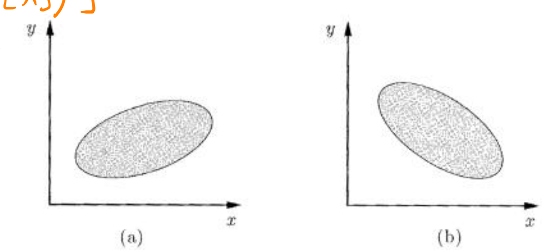
# Covariance and uncorrelatedness

- Covariance of two RVs X and Y is defined as

$$\boxed{\mathrm{cov}(X, Y) = E[(X - E[X])(Y - E[Y])]}$$

$\mathrm{var}(X) = E[(X - E[X])^2]$

- Positive (negative) covariance indicates that the values of X – E[X] and Y – E[Y] tend to the same (opposite) sign.




- When cov(X,Y) = 0, X and Y are said to be *uncorrelated*

# Some formulas for covariance

- cov(X,Y) = E[XY] - E[X] E[Y] If X, Y are uncorrelated, $E[XY] = E[X]\,E[Y]$
  $cov(X,Y) = E[(X-E[X])(Y-E[Y])] = E[XY - X\,E[Y] - E[X]\,Y + E[X]\,E[Y]]$
  $= E[XY] - E[X]E[Y] - \cancel{E[X]E[Y]} + \cancel{E[X]}E[Y] = E[XY] - E[X]\,E(Y]$
- cov(X,X) = var(X)
  $var(X) = E[(X-E(X))^2] = E[(X-E(X))(X-E(X))]$
- cov(X, aY + b) = a cov(X,Y)
- cov(X, Y + Z) = cov(X,Y) + cov(X,Z)

UMBC CMPE 320 Seung-Jun Kim

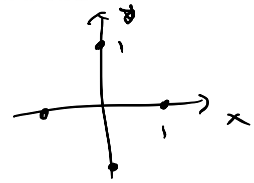
# Independence implies uncorrelated

- If X and Y are independent, they are uncorrelated
  We saw before that if X and Y are independent, $E[XY] = E[X]E[Y]$
  but this means that X and Y are uncorrelated
- This does not mean the converse is true. That is, even if X and Y are uncorrelated, they are not necessarily independent.

**Example 4.13.** The pair of random variables $(X, Y)$ takes the values $(1, 0)$, $(0, 1)$, $(-1, 0)$, and $(0, -1)$, each with probability $1/4$

(a) Are X and Y uncorrelated?

(b) Are X and Y independent?




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

$E[X] = 1 \cdot \frac{1}{4} + (-1) \cdot \frac{1}{4} + 0 \cdot \frac{1}{2} = 0 \qquad E[Y] = 0$

$\text{cov}(X, Y) = E[XY] = 0 \quad \Rightarrow \quad X, Y$ : uncorrelated

(b) Check $p_{X|Y}(x|y) = p_X(x)$

$$p_{X|Y}(x|1) = \begin{cases} 1, & \text{if } x = 0 \\ 0, & \text{otherwise} \end{cases}$$

$$p_X(x) = \begin{cases} \frac{1}{4}, & \text{if } x = 1 \text{ or } -1 \\ \frac{1}{2}, & \text{if } x = 0 \end{cases}$$

$\Rightarrow p_{X|Y} \neq p_X$

Therefore, X, Y are not independent



# Correlation coefficient

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

- $-1 \leq \rho \leq 1$ always holds ➔ $\rho$ is a normalized version of covariance
- If $\rho > 0$ (or $\rho < 0$), the values of X – E[X] and Y – E[Y] tend to have the same (opposite) sign.
- The magnitude $|\rho|$ provides a normalized measure of how strongly X and Y are correlated

If $|\rho| = 1 \iff Y - E[Y] = c(X - E[X])$

**Example 4.14.** Consider $n$ independent tosses of a coin with probability of a head equal to $p$. Let $X$ and $Y$ be the numbers of heads and of tails, respectively, and let us compute the correlation coefficient of $X$ and $Y$.

$$X + Y = n$$

$$-)\ E[X] + E[Y] = n$$

$$X - E[X] + (Y - E[Y]) = 0 \Rightarrow X - E[X] = -(Y - E[Y])$$

$$\mathrm{cov}(X,Y) = E[(X-E[X])(Y-E[Y])] = E[-(X-E[X])^2] = -\mathrm{var}(X)$$
$$= -\mathrm{var}(Y)$$

$$\rho(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{var}(X)\,\mathrm{var}(Y)}} = \frac{-\mathrm{var}(X)}{\sqrt{\mathrm{var}(X)^2}} = -1$$

# Variance of sum of RVs

- If X and Y are RVs with finite variance,

$$\mathrm{var}(X+Y) = \mathrm{var}(X) + \mathrm{var}(Y) + 2\mathrm{cov}(X,Y)$$

Recall that when $X, Y$ are independent, $\mathrm{var}(X+Y) = \mathrm{var}(X) + \mathrm{var}(Y)$.
This is true whenever $\mathrm{cov}(X,Y) = 0$, that is, they are uncorrelated

- If $X_1, X_2, \ldots, X_n$ are RVs with finite variance,

$$\mathrm{var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{var}(X_i) + \sum_{\{(i,j)|i\neq j\}} \mathrm{cov}(X_i, X_j)$$

 Seung-Jun Kim

**Example 4.15.** Consider the hat problem discussed in Section 2.5, where $n$ people throw their hats in a box and then pick a hat at random. Let us find the variance of $X$, the number of people who pick their own hat. $(n > 1)$

$X_i$ : RV with value 1 if $i$-th person picks her own hat

$X = X_1 + X_2 + \cdots + X_n$

$X_i$ is Bernoulli RV with $p = \frac{1}{n}$

$$\text{var}(X) = \text{var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \text{var}(X_i) + \sum_{\substack{i,j=1 \\ i \neq j}}^{n} \text{cov}(X_i, X_j)$$

$$E[X_i] = \frac{1}{n}, \quad \text{var}(X_i) = E[X_i^2] - (E[X_i])^2 = \frac{1}{n} - \frac{1}{n^2} = \frac{1}{n}\left(1 - \frac{1}{n}\right)$$

$$\text{cov}(X_i, X_j) = E[X_i X_j] - E[X_i]E[X_j] = \underbrace{P(X_i = 1, X_j = 1)}_{= P(X_i=1)P(X_j=1 \mid X_i=1)} - \frac{1}{n^2}$$

$$= \frac{1}{n} \cdot \frac{1}{n-1} - \frac{1}{n^2} = \frac{1}{n^2(n-1)}$$

$$\text{var}(X) = n \cdot \frac{1}{n}\left(1 - \frac{1}{n}\right) + (n^2 - n) \cdot \frac{1}{n^2(n-1)} = 1$$