

MATH 321: Mathematical Statistics

Lecture 16: Independence and the Correlation Coefficient

Chapter 4: Bivariate Distributions (4.1, 4.2, and 4.4)

Independence for random variables

Definition

- Events A and B are independent if and only if
- The definition of independence for two discrete random variables relies on this multiplication rule.

Applying the idea of independence between two events to random variables, we say that X and Y are independent random variable if and only if the events _____ and _____ are independent for all $x \in R$ and all $y \in R$.

- Definition: Let (X, Y) be a bivariate random vector with joint pdf or pmf $f(x, y)$ and marginal pdfs or pmfs $f_X(x)$ and $f_Y(y)$. Then X and Y are called **independent random variables** if, for every $x \in \mathbb{R}$ and $y \in \mathbb{R}$,

$$f(x, y) = f_X(x) \cdot f_Y(y)$$

- If X and Y are not independent, they are said to be _____.

Checking independence

- In general, if given $f(x, y)$ and checking to see if X and Y are independent (in notation: $X \perp\!\!\!\perp Y$), we have to find $f_X(x)$ and $f_Y(y)$ and then multiply and check.
- Example: Define the joint pmf of (X, Y) by

$y \backslash x$	0	1
0	1/4	1/4
1	1/4	0
2	0	1/4

- (a) Are X and Y independent?

(b) Are the events $\{X = 0\}$ and $\{Y = 0\}$ independent of each other?

- Interpreting independent random variables.

- We just saw that the random variables can be _____ even when specific _____ are _____.

This goes back to the definition of independence. In order for the entire random variables to be independent $P(X = x, Y = y) = P(X = x) \cdot P(Y = y)$ _____ pairs (x, y) .

- When X and Y are independent, observing $Y = y$ does not alter the probability model for X . Similarly, observing $X = x$ does not alter the probability model for Y .

Therefore, learning that $Y = y$ provides no information about X and learning that $X = x$ provides no information about Y .

- More examples: Determine if X and Y independent in each of the following scenarios.

1.

$$f(x, y) = 1/2 \quad \text{for } 0 \leq x \leq y \leq 2$$

$$f_X(x) = (2 - x)/2 \quad \text{for } 0 \leq x \leq 2$$

$$f_Y(y) = y/2 \quad \text{for } 0 \leq y \leq 2$$

2.

$$f(x, y) = 4xy \quad \text{for } 0 \leq x \leq 1, 0 \leq y \leq 1$$

$$f_X(x) = 2x \quad \text{for } 0 \leq x \leq 1$$

$$f_Y(y) = 2y \quad \text{for } 0 \leq y \leq 1$$

- What causes difference between the above examples?

- We just saw that the range of (X, Y) plays a crucial role in determining when random variables are independent. We can also look at the functional form of the joint pmf / pdf when checking independence.

- **Independence theorem:** X and Y are independent random variables if and only if

$$f(x, y) = g(x) \cdot h(y), \quad a \leq x \leq b, c \leq y \leq d$$

where $g(x)$ is a nonnegative function of x alone and $h(y)$ is a nonnegative function of y alone.

Note that $g(x)$ and $h(y)$ do not themselves need to be density functions.

- Checking independence by inspection.

– Let $f(x, y) = 2x$ for $0 \leq x \leq 1, 0 \leq y \leq 1$.

– Process:

1. Is the range rectangular?

2. Find any $g(x)$ and $h(y)$ such that $f(x, y) = g(x) \cdot h(y)$

** Conclusions:

Range
Separable

Range
Separable

Range
Separable

Rectangular range is a _____, but not _____ condition for independence.

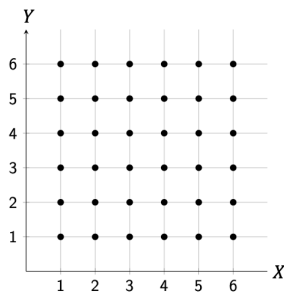
- More examples: Determine if X and Y independent in each of the following scenarios.

3. If the joint pdf is e^{xy} with a rectangle range.

4. If the joint pdf is e^{x+y} with a rectangle range.

5. If the joint pdf is $\log(x + y)$ with a rectangle range.

6. Let (X, Y) be the numbers on die 1 and die 2, respectively, when a pair of fair 6-sided dice are thrown.

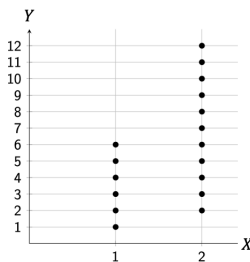


In the discrete case, “rectangular” means that the range of (X, Y) must equal the product set (aka cartesian product) of the individual ranges of X and Y :
 $\mathcal{X} \times \mathcal{Y} = (\mathcal{X}, \mathcal{Y})$

		\mathcal{X}		
		x_1	x_2	x_3
\mathcal{Y}	y_1	(x_1, y_1)	(x_2, y_1)	(x_3, y_1)
	y_2	(x_1, y_2)	(x_2, y_2)	(x_3, y_2)
	y_3	(x_1, y_3)	(x_2, y_3)	(x_3, y_3)

In other words, need _____ $f(x, y) > 0$ for all possible (x, y) combos.

7. A fair coin is tossed. If heads is tossed then one fair 6-sided die is thrown and if tails is tossed two fair 6-sided dice are thrown. Let $X = 1$ for heads and $X = 2$ for tails and let Y be the total number of dots on the dice.



8. Define the joint pmf of (X, Y) by

$y \backslash x$	-1	0	1
-1	1/18	1/9	1/6
0	1/9	0	1/6
1	1/6	1/9	1/9

NOTE: If have pmf table (and not equations) and have rectangular range \implies Need to check $f(x, y) = f_X(x) \cdot f_Y(y)$ for ALL pairs.

9. Let $f(x, y) = \frac{x+y}{21}$ for $x = 1, 2, 3$ and $y = 1, 2$.

Conditional distributions and independence

- Recall if events A and B are independent, then $P(A | B) = P(A)$ and $P(B | A) = P(B)$. Let's see if this holds for distributions as well.
- Example: An analyst is studying traffic accidents in two adjacent towns. The random variables S and T represents the waiting time between accidents in towns X and Y , respectively. The joint probability function for S and T is given by:

$$f(s, t) = e^{-(s+t)} \quad \text{for } s \geq 0 \text{ and } t \geq 0.$$

(a) Find the marginal distributions $f_S(s)$ and $f_T(t)$.

(b) Find the conditional distributions $f(s | t)$ and $f(t | s)$.

- Theorem: If X and Y are independent,

$$f(x | y) = f_X(x) \quad \text{and} \quad f(y | x) = f_Y(y)$$

- Proof:

Using independence

- We can always find the marginal distributions from the joint distribution, but the converse is not always true (so the joint distribution has more information).
- But when X and Y are _____, the joint distribution and marginal distributions contain the equal amount of information about X and Y .

That means we can also find the _____ distribution from _____ distributions.

- Example: Suppose that $X \sim \text{Exp}(\lambda = 1)$ and $Y \sim \text{Uniform}(0, 1)$ and $X \perp\!\!\!\perp Y$. Find the joint pdf of X and Y .

- Theorem: Let X and Y be independent random variables.

(a) For any $A \subset \mathbb{R}$ and $B \subset \mathbb{R}$,

$$P(X \in A, Y \in B) = P(X \in A) \cdot P(Y \in B)$$

That is, the events $\{X \in A\}$ and $\{Y \in B\}$ are independent events.

(b) Let $g(x)$ be a function only of x and $h(y)$ be a function only of y . Then

$$E[g(X) \cdot h(Y)] = E[g(X)] \cdot E[h(Y)]$$

Proof for $E(XY)$ for the discrete case:

- When applying this theorem, a bivariate question reduces to a univariate question, making it a lot simpler.

Example: Let X and Y be independent exponential ($\lambda = 1$) random variables.

1. Find $P(X \geq 4, Y < 3)$.

2. Find $E(X^2Y)$ and $E(X + Y)$.

Covariance

Introduction

- One of main purposes of studying bivariate random vectors is to study the dependence between two random variables.

Recall that the advantage of using the joint pmf / pdf over the respective marginal distributions is that it usually contains additional information about interaction between the two random variables.

We can use this joint distribution to check how much two random variables change together.

- More specifically, covariance is how we will study this dependence. Covariance is a special expected value with two very useful applications.

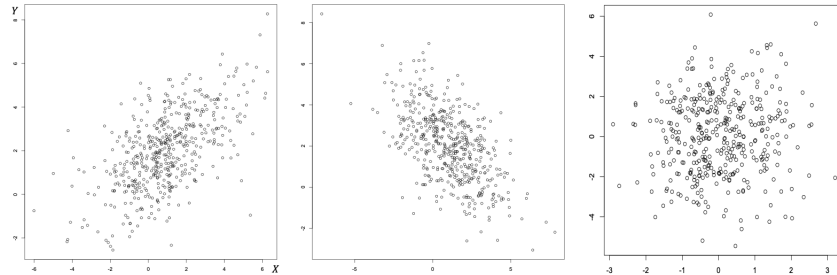
1. Finding the variance of $X + Y$.

2. Quantify linear dependence.

- There are two different scenarios that we can calculate covariance for: on a sample of data and on a probability distribution.

Covariance is easier to conceptualize if we have a sample of data. So we will start with this. But technically we are still in a probability context, so then we will shift to working with distributions (i.e. population information) rather than data points.

Visualizing dependence



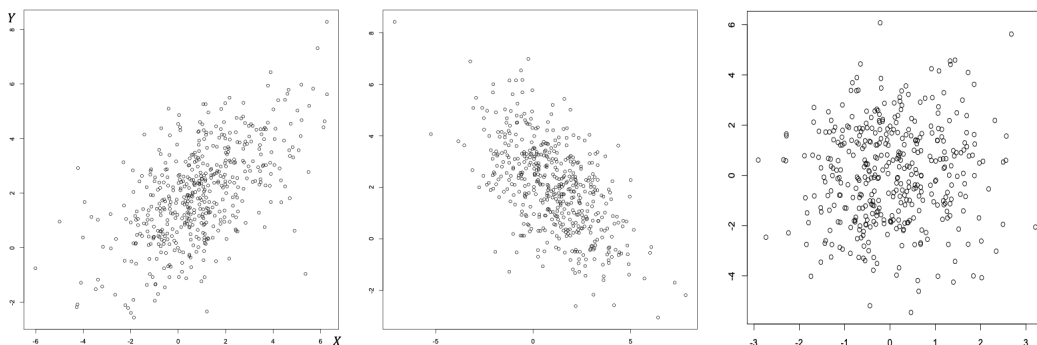
- In the first plot above:
 - Large values ($> \mu$) of X mainly correspond with _____ values of Y .
Small values ($< \mu$) of X mainly correspond with _____ values of Y .
 - In this case, the two random variables are _____ dependent / correlated.
In statistics, correlated = linear relationship.
- In the second plot above:
 - Large values of X mainly correspond with _____ values of Y .
Small values of X mainly correspond with _____ values of Y .
 - In this case, the two random variables are _____ dependent / correlated.
- In the third plot above:
 - Large values of X mainly correspond with _____ values of Y .
Small values of X mainly correspond with _____ values of Y .
 - In this case, the two random variables appear to _____ be dependent / correlated.

Quantifying dependence

- So we know how dependence shows up in a scatterplot, so then how to we quantify (measure) it?

On the plots above, we were looking at where X and Y were in relation to their respective means and then also how these related (interacted) with each other.

In statistics / modeling, interaction means multiplying terms. So here is the function that we will start with:



- If X and Y are positively dependent (plot 1), then

$$(X - \mu_X)(Y - \mu_Y)$$

will be mostly _____.

For example, height and weight. Most of people taller than the average weigh more than the average. Most of people shorter than the average weigh less than the average.

- If X and Y are negatively dependent (plot 2), then

$$(X - \mu_X)(Y - \mu_Y)$$

will be mostly _____.

For example, stress (e.g. on a mechanical part or system) and time to failure. More stress often results in shorter time to failure and less stress often leads to longer time to failure.

- If X and Y appear to not be dependent (plot 3), then

$$(X - \mu_X)(Y - \mu_Y)$$

the positives and negatives will almost _____ (spread evenly in all “quadrants”).

- So our function “results” match our visual explanation of dependence, but also want to emphasize “mostly negative and mostly positive”. Some values could have the different sign from the most of others.

We want to measure **overall tendency**. So we define the _____ of $(X - \mu_X)(Y - \mu_Y)$ as a measure of **linear dependence** between X and Y .

Definition, theorems and properties of covariance

- Definition: The **covariance** of X and Y is the number defined by

$$\text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$$

- Notes:

- Covariance is a measure of how much two random variables change together.
- If the $\text{Cov}(X, Y) > 0$, then X and Y are _____ correlated.
- If the $\text{Cov}(X, Y) < 0$, then X and Y are _____ correlated.
- If the $\text{Cov}(X, Y) = 0$, then X and Y are _____, which means there is **no linear dependence** between X and Y .
- Covariance can measure only the **linear** dependence between two random variables.

May not pick up on non-linear relationships (i.e. curvature).

- Properties / theorems of covariance:

- (i) **Calculation:** If (X, Y) is discrete, then

$$E[(X - \mu_X)(Y - \mu_Y)] =$$

Examples: Calculate $\text{Cov}(X, Y)$ for the following two joint pmfs.

(a)

$y \backslash x$	1	2
1	4/8	1/8
2	1/8	2/8

x	y	f(x,y)	(x - mu_x)(y - mu_y)	g(x,y)f(x,y)
1	1	0.5	0.141	0.0703
2	1	0.2	-0.234	-0.0469
1	2	0.125	-0.234	-0.0293
2	2	0.25	0.391	0.0977
covariance =				0.0918
mu_x =		1.375		
mu_y =		1.375		

(b)

$y \backslash x$	1	2
1	1/16	6/16
2	7/16	2/16

- (ii) Recall how with variance we had a definition and an alternate form that is easier to calculate by hand: $V(X) = E[(X - \mu_X)^2] = E(X^2) - [E(X)]^2$. We have a similar theorem for covariance:

Alternate calculation for covariance: For any random variables X and Y ,

$$\text{Cov}(X, Y) = E(XY) - E(X) \cdot E(Y)$$

Proof:

Example: Back to the investor with two asset random variables X and Y . We have previously calculated each of these pieces and thus:

The _____ covariance means that as one investment performs above average, the other tends to perform _____ average.

- (iii) **Variance is a special case of covariance:**

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

Proof:

- (iv) Order in covariance does not matter (i.e. **symmetric**).

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

Proof:

- (v) Covariance of a random variable and a constant is zero.

$$\text{If } c \text{ is a constant, then } \text{Cov}(X, c) = 0$$

Proof:

- (vi) Can factor out coefficients in covariance.

$$\text{Cov}(aX, bY) = ab \cdot \text{Cov}(X, Y)$$

Proof:

- (vii) Can factor out coefficients, but added constants disappear.

$$\text{Cov}(aX + c, bY + d) = ab \cdot \text{Cov}(X, Y)$$

Like variance, the location shift does not influence covariance, which means it does not impact the _____ between X and Y .

Example: Suppose investment X is now performing 1.2 times better than previously and they added 5 to investment Y . Find the new covariance of the two investments.

- (viii) **Distributive property** of covariance.

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

Proof:

Independence and covariance

- Theorem: If X and Y are independent random variables, then

$$\text{Cov}(X, Y) = 0$$

- Proof:

- Independent vs uncorrelated:
 - If X and Y are independent, there does not exist **any type of dependence** between X and Y .
 - If X and Y are uncorrelated (i.e. $\text{Cov}(X, Y) = 0$), there is no _____ dependence between X and Y .

But, there may be some other type of relationship.

 - Thus, independent random variables _____ be uncorrelated, but uncorrelated random variables _____ be independent.

Summary:

- Example: Let $f(x, y) = 1/3$ for $(x, y) = (0, 1), (1, 0), (2, 1)$.

(a) Are X and Y independent?

(b) Are X and Y uncorrelated?

Interpreting covariance

- We have already shown that we can determine the **direction** of the relationship based on the sign of the covariance, this is a useful interpretation.

However, we cannot use covariance to measure the **strength** of the relationship.

This is because its value depends on the scale of measurement.

- Example demonstrating this: Suppose $(X, Y) = (\text{income, savings})$ in dollars and $(X', Y') = (100X, 100Y) = (\text{income, savings})$ in cents. Further, let $\text{Cov}(X, Y) = 3$.

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

We _____ say that the linear dependence between income and savings in cents is stronger than that in dollars because _____, all we know is _____.

- Because of this, covariance cannot be used as an absolute measure of linear dependence (i.e. a single measure that tells us everything, direction and strength).

So how can we improve it?

Correlation

Definition

- Motivation: The problem can be eliminated by _____ the covariance value.
- Definition: As an absolute measure of dependence, the **correlation coefficient** of X and Y is the number defined by

$$\rho_{XY} = \text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{V(X)V(Y)}} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

- Continuing previous example:

$$\text{Corr}(X', Y') = \frac{\text{Cov}(X', Y')}{\sqrt{V(X')V(Y')}} =$$

Thus, the correlation is unit-free and _____ by change in scale or location.

- Back to investor example: It can be shown that $V(X) = 40$ and $V(Y) = 25$ and we previously found $\text{Cov}(X, Y) = -13$. Thus

Properties of correlation

- We are going to derive these!
- Covariance measures linear dependence, so let's see what happens to correlation when there is a perfect linear relationship, i.e. $Y = aX + b$.
 - Sidenote: This is called a deterministic (or functional) relationship / model as opposed to stochastic (or probabilistic or statistical), which is when there is some randomness involved.

If $Y = aX + b$,

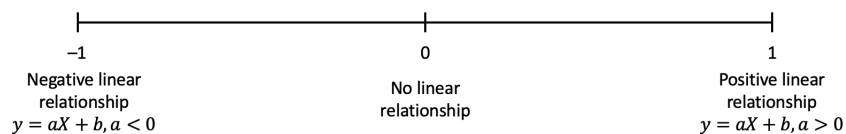
Thus when X and Y are linearly related, the correlation coefficient is _____ when the slope of the straight line is positive and _____ when the slope is negative.

- To see what might happen when X and Y are not linearly related, we can look at the extreme case when X and Y are independent and have no systematic dependence.

Obviously $\rho_{XY} =$ _____ whenever $\text{Cov}(X, Y) =$ _____. Keep in mind that variables can be uncorrelated without being independent.

- Putting these two pieces together, it can be shown that for any random variables X and Y ,

$$-1 \leq \rho_{XY} \leq 1$$



- Possible values of ρ_{XY} lie on a continuum between -1 and 1.
- Values of ρ_{XY} close to ± 1 are interpreted as an indication of a high level of linear association between X and Y .
- Values of ρ_{XY} near 0 are interpreted as implying little or no linear relationship between X and Y .

- Formally, we can state these properties in the following theorem:

Theorem: For any random variable X and Y ,

(i) $-1 \leq \rho_{XY} \leq 1$

The sign represents the direction of linear dependence between X and Y .

$|\rho_{XY}|$ represents the magnitude of dependence.

(ii) $\rho_{XY} = 1$ if and only if there exist numbers $a > 0$ and b such that $P(Y = aX + b) = 1$.

X and Y have perfect positive correlation. As X increases, Y _____.

(iii) $\rho_{XY} = -1$ if and only if there exist numbers $a < 0$ and b such that $P(Y = aX + b) = 1$.

X and Y have perfect negative correlation. As X increases, Y _____.

(iv) When $\rho_{XY} = 0$, X and Y are _____.

Examples

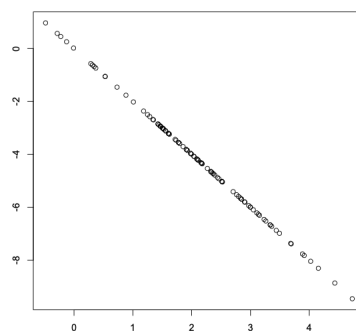
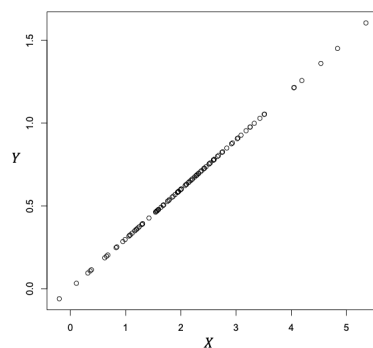
- Is the dependence between X_1 and X_2 stronger than the dependence between Y_1 and Y_2 ?

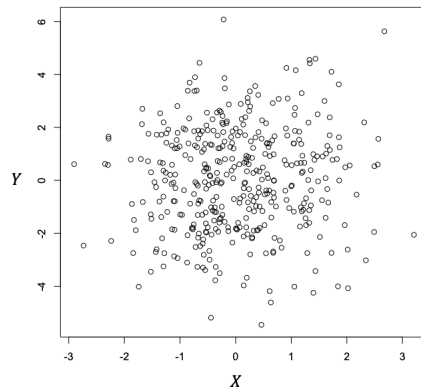
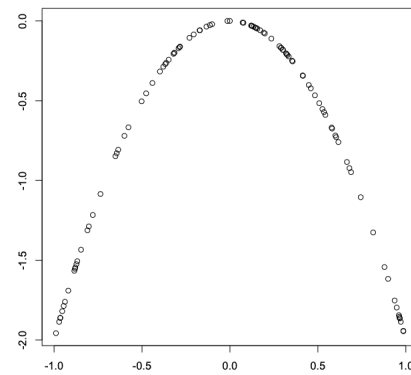
- $\text{Cov}(Y_1, Y_2) = 0.4$, $\text{Cov}(X_1, X_2) = 0.6$
- $\text{Cov}(Y_1, Y_2) = 0.4$, $\text{Cov}(X_1, X_2) = -0.6$
- $\text{Corr}(Y_1, Y_2) = 0.4$, $\text{Corr}(X_1, X_2) = 0.6$
- $\text{Corr}(Y_1, Y_2) = 0.4$, $\text{Corr}(X_1, X_2) = -0.6$

- Find ρ_{XY} :

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

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



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

Variance of $X + Y$

Derivation

- We stated at the start that $V(X + Y) \neq V(X) + V(Y)$. Now that we understand covariance, we can see why!
- Let X and Y be two random variables:

- Theorem: **Variance of a sum of two random variables**

$$V(X + Y) =$$

- Examples:

1. Back to the investor with two assets X and Y , find the $V(X + Y)$:

2. Let $f(x, y) = x + \frac{3}{2}y^2$ for $0 \leq x \leq 1, 0 \leq y \leq 1$. Find the variance of $X + Y$.

Variance of $X + Y$ when independent

- Derivation: If X and Y are independent, then
- Theorem: **Variance of a sum of two independent random variables**

If $X \perp\!\!\!\perp Y$, then $V(X + Y) =$

\implies Can only JUST add variances when independent.

- Example: Let $X \sim \text{Exp}(\lambda = 1)$ and $Y \sim \text{Exp}(\lambda = 3)$. If $X \perp\!\!\!\perp Y$, find $V(X + Y)$.