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

  $\to$ $P(A \mid B) = P(A)$ < by definition >

  $\to$ $P(A \cap B) = P(A)\,P(B)$ < more useful in problems >

- 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 $\{X = x\}$ and $\{Y = y\}$ 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 $y \in \mathbb{R}$,

$$P(X=x, Y=y) = P(X=x)\,P(Y=x) \implies f(x, y) = f_X(x) \cdot f_Y(y)$$

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

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	$f_Y(y)$
0	1/4	1/4	1/2
1	1/4	0	1/4
2	0	1/4	1/4
$f_X(x)$	1/2	1/2	

  (a) Are X and Y independent?

$$f(1,1) \overset{?}{=} f_X(1) * f_Y(1) \implies \text{No!}$$

$$0 \neq \tfrac{1}{4} \times \tfrac{1}{2}$$

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

$P(X=0, Y=0) \stackrel{?}{=} P(X=0) \times P(Y=0)$
$1/4 \stackrel{\checkmark}{=} 1/2 \times 1/2 \Rightarrow$ yes!

- Interpreting independent random variables.
  - We just saw that the random variables can be <u>dependent</u> even when specific <u>events</u> are <u>independent</u>.

    This goes back to the definition of independence. In order for the entire random variables to be independent $P(X = x, Y = y) = P(X = x) \cdot P(Y = y)$ <u>for ALL</u> 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.
$$f(x, y) = 1/2 \qquad \text{for } 0 \le x \le y \le 2$$
$$f_X(x) = (2 - x)/2 \qquad \text{for } 0 \le x \le 2$$
$$f_Y(y) = y/2 \qquad \text{for } 0 \le y \le 2$$

$f(x,y) \stackrel{?}{=} f_X(x) \cdot f_Y(y)$
$1/2 \neq \frac{2-x}{2} \cdot \left(\frac{y}{2}\right) \Rightarrow$ No!

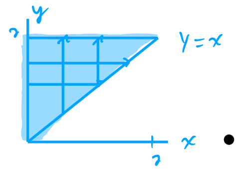
2.
$$f(x, y) = 4xy \qquad \text{for } 0 \le x \le 1,\ 0 \le y \le 1$$
$$f_X(x) = 2x \qquad \text{for } 0 \le x \le 1$$
$$f_Y(y) = 2y \qquad \text{for } 0 \le y \le 1$$

$4xy \stackrel{\checkmark}{=} 2x\,(2y) \Rightarrow$ yes!

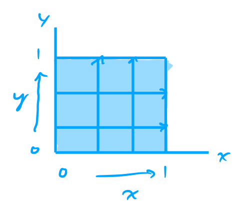
- What causes difference between the above examples?

Range of (x, y)

① $x = 0 \le x \le y$, $y = x \le y \le 2$ } obviously dependent, range of x changes based on y



② range of x is the SAME regardless of y



rectangular

- 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.

① $f(x,y) = 1/2$

② $f(x,y) = 4xy$

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

$$f(x,y) = \underbrace{g(x)\cdot h(y)}_{\text{Separable}}, \qquad \underbrace{a \le x \le b,\ c \le y \le d}_{\text{rectangular range}}$$

(i.e. not dependent on other variable)

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. $\Longrightarrow$ (any function)

- Checking independence by inspection.
  - Let $f(x,y) = 2x$ for $0 \le x \le 1$, $0 \le y \le 1$.
  - Process:
    1. Is the range rectangular? yes ✓
    2. Find any $g(x)$ and $h(y)$ such that $f(x,y) = g(x)\cdot h(y)$ yes ✓ $= 2x \cdot 1$ $\Longrightarrow X \perp\!\!\!\perp Y$

    ** Conclusions:

Range ✓ Separable ✓ $\perp\!\!\!\perp$	Range ✗ Separable ✓ Not $\perp\!\!\!\perp$	Range ✓ Separable ✗ Not $\perp\!\!\!\perp$

    Rectangular range is a <u>necessary</u>, but not <u>sufficient</u> 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.

     $(e^{x})^{y}$ is not separable ✗ $\Longrightarrow$ not $\perp\!\!\!\perp$

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

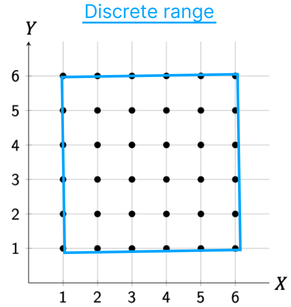
     $e^{x} e^{y}$ Separable ✓ $\Longrightarrow \perp\!\!\!\perp$

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

     Not Separable ✗ $\Longrightarrow$ not $\perp\!\!\!\perp$ $(\log(x\cdot y) = \log(x) + \log(y))$

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

Discrete range



rectangular ✓

Separable (technically) $\to$ $f(x,y) = \frac{1}{36}$
$= \frac{1}{36}\cdot 1$
$= g(x)\cdot h(y)$ ✓

$\Longrightarrow$ yes $X \perp\!\!\!\perp Y$

Contextually

$$f(x,y) = \underbrace{\tfrac{1}{6}}_{f_X(x)} \times \underbrace{\tfrac{1}{6}}_{f_Y(y)} = \tfrac{1}{36}$$

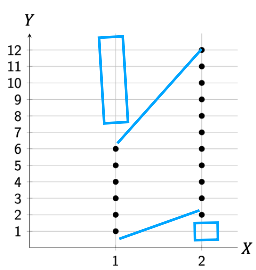
This comes from the multiplication rule

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: $\mathcal{X} \times \mathcal{Y} = (\mathcal{X}, \mathcal{Y})$

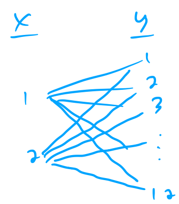
$\mathcal{Y}$ \ $\mathcal{X}$	x_1	x_2	x_3
y_1	(x_1, y_1)	(x_2, y_1)	(x_3, y_1)
y_2	(x_1, y_2)	(x_2, y_2)	(x_3, y_3)
y_3	(x_1, y_3)	(x_2, y_2)	(x_3, y_3)

In other words, need positive probabilities $f(x, y) > 0$ 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.



If product set, range should be



"fully crossed"

Missing

$(1,7), \ldots, (1,12)$ + $(2,1)$

$\Rightarrow$ If $x=1$, y can not $= 7, \ldots, 12$

$\Rightarrow$ Dependent

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

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

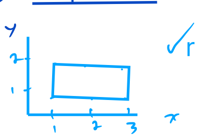
$\rightarrow$ Not rectangular, "hole" in center $\Rightarrow$ Not $\perp\!\!\!\perp$

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

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

Two ways to check independence

① Inspection



✓range

(addition)

$f(x,y) = \dfrac{x+y}{21}$ not separable ✗

$\Longrightarrow$ Not $\perp\!\!\!\perp$

② By definition

$\rightarrow$ find marginals

$\rightarrow f_X(x) = \sum_y f(x,y) = \sum_{y=1}^{2} \dfrac{x+y}{21} = \dfrac{2x+3}{21}, \quad x = 1,2,3$

$f_Y(y) = \sum_x f(x,y) = \sum_{x=1}^{3} \dfrac{x+y}{21} = \dfrac{3y+6}{21}, \quad y = 1,2$

$\rightarrow$ Check

$\rightarrow f(x,y) \stackrel{?}{=} f_X(x)\, f_Y(y)$

$\dfrac{x+y}{21} \neq \dfrac{2x+3}{21}\left(\dfrac{3y+6}{21}\right) \Longrightarrow$ Not $\perp\!\!\!\perp$

Conditional distributions and independence

- Recall if events A and B are independent, then $P(A \mid B) = P(A)$ and $P(B \mid A) = P(B)$. 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:

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

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

  $$f_S(s) = \int_0^\infty e^{-(s+t)}\,dt = -e^{-(s+t)}\Big|_0^\infty \quad F(\infty) - F(0) = e^{-s}, \quad s \geq 0$$

  $$f_T(t) = \langle \ldots \rangle = e^{-t}, \quad t \geq 0$$

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

  $$\text{for } t \geq 0 \quad f(s \mid t) = \frac{f(s,t)}{f_T(t)} = \frac{e^{-(s+\cancel{t})}}{\cancel{e^{-t}}} = e^{-s} = \underbrace{f_s(s)}_{\text{marginal of s}}$$

  $$\text{for } s \geq 0 \quad f(t \mid s) = \frac{f(s,t)}{f_s(s)} = \frac{e^{-(\cancel{s}+t)}}{\cancel{e^{-s}}} = e^{-t} = \underbrace{f_T(t)}_{\text{marginal of t}}$$

- Theorem: If X and Y are independent,

  $$\boxed{f(x \mid y) = f_X(x) \qquad \text{and} \qquad f(y \mid x) = f_Y(y)}$$

- Proof:

  If $X \perp\!\!\!\perp Y$

  $$\Rightarrow f(x \mid y) = \frac{f(x,y)}{f_Y(y)} = \frac{f_X(x)\cancel{f_Y(y)}}{\cancel{f_Y(y)}} = f_X(x) \quad \text{for every } (x,y)$$

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 independent, the joint distribution and marginal distributions contain the equal amount of information about X and Y.

  That means we can also find the joint distribution from marginal distributions.

$$f(x,y) \underset{\text{ONLY if } \perp\!\!\!\perp}{\overset{\text{can find ALWAYS}}{\rightleftarrows}} \begin{matrix} f_X(x) \\ f_Y(y) \end{matrix}$$

- 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.

$\rightarrow f_X(x) = e^{-x} \quad , x \geq 0$

$f_Y(y) = 1 \quad , 0 \leq y \leq 1$

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

$= e^{-x} \cdot 1 \quad , \begin{matrix} x \geq 0 \\ 0 \leq y \leq 1 \end{matrix}$

- 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:

$E(XY) = \sum_x \sum_y xy f(x,y)$

$= \sum_x \sum_y x y f_X(x) f_Y(y)$ < separate because $\perp\!\!\!\perp$ >

$= \left(\sum_x \left[x f_X(x) \underbrace{\sum_y y f_Y(y)}_{E(Y)}\right]\right.$ < rearranging >

$= E(Y) \sum_x x f_X(x)$

$\overset{\checkmark}{=} E(X)\, E(Y)$

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

Exponential

$F(t) = 1 - e^{-\lambda t}$

$S(t) = e^{-\lambda t}$

$P(X \geq 4, Y < 3) = P(X \geq 4)\, P(Y < 3)$ < separate because $\perp\!\!\!\perp$ >

$= e^{-4}\,(1 - e^{-3})$

≈ 0.0174

$E(X) = 1/\lambda$

$V(X) = 1/\lambda^2$

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

$E(X^2 Y) = E(X^2)\, E(Y)$, where $E(X^2) = V(X) + (E(X))^2$

$= (1 + 1)\,(1)$

$= 2$

$E(X + Y) = E(X) + E(Y)$

$= 1 + 1$

$= 2$

Reminder

$\rightarrow E[g(x) h(y)] = E[g(x)]\, E[h(y)]$ ONLY if $\perp\!\!\!\perp$

$\rightarrow E[g(x) + h(y)] = E[g(x)] + E[h(y)]$ ALWAYS

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 $X + Y$.

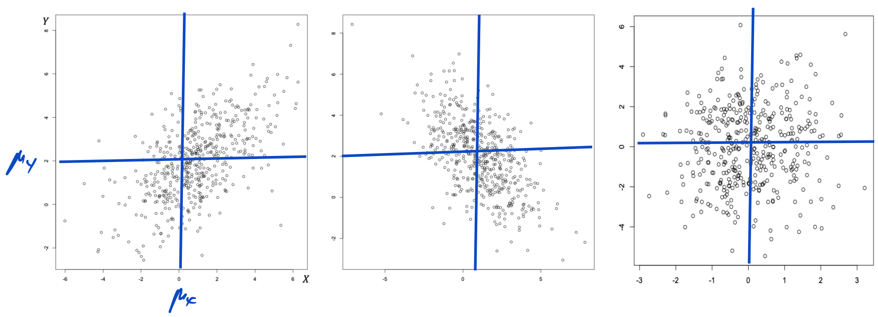
     $$V(x+y) \neq V(x) + V(y)$$

  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 __large__ values of Y.
    Small values ($< \mu$) of X mainly correspond with __small__ values of Y.
  - In this case, the two random variables are __positively__ dependent / correlated.
    In statistics, correlated = linear relationship.

- In the second plot above:
  - Large values of X mainly correspond with __small__ values of Y.
    Small values of X mainly correspond with __large__ values of Y.
  - In this case, the two random variables are __Negatively__ dependent / correlated.

- In the third plot above:
  - Large values of X mainly correspond with __small/ large__ values of Y.
    Small values of X mainly correspond with __small/ large__ values of Y.
    There is no indication of trend
  - In this case, the two random variables appear to __Not__ be dependent / correlated.

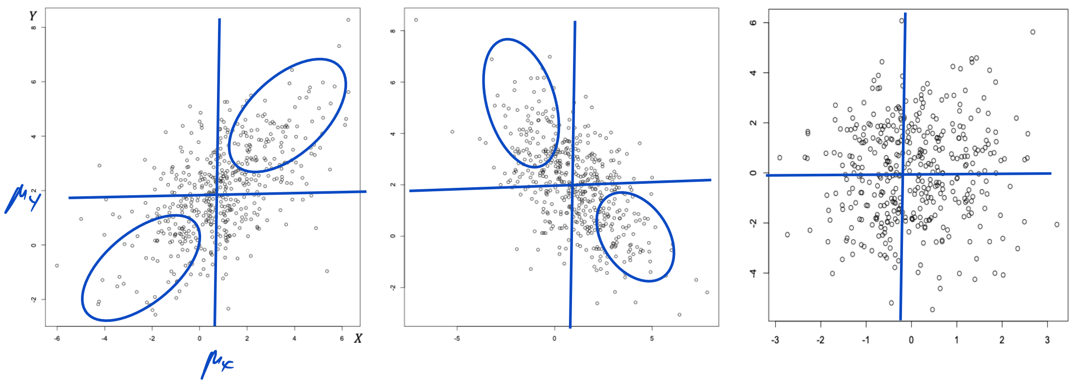
Quantifying dependence

- 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:

$$g(x,y) = \underbrace{(x-\mu_x)}_{\text{deviation from center for } x} * \underbrace{(y-\mu_y)}_{\text{for } y}$$



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

$$(X - \mu_X)(Y - \mu_Y) \longrightarrow (+)(+) \text{ OR } (-)(-) = +$$

will be mostly positive.

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

$$(X - \mu_X)(Y - \mu_Y) \longrightarrow (+)(-) \text{ OR } (-)(+) = -$$

will be mostly Negative.

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 balance out ≈ 0 (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 Expected value 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

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

- Notes:
  - Covariance is a measure of how much two random variables change together.
  - If the $\mathrm{Cov}(X,Y) > 0$, then X and Y are positively correlated.
  - If the $\mathrm{Cov}(X,Y) < 0$, then X and Y are Negatively correlated.
  - If the $\mathrm{Cov}(X,Y) = 0$, then X and Y are uncorrelated, 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

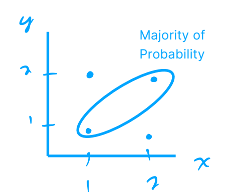
$$E[\underbrace{(X-\mu_X)(Y-\mu_Y)}_{g(x,y)}] = \sum_x \sum_y (x-\mu_x)(y-\mu_y)f(x,y)$$

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

$$\mathrm{Cov}(X,Y) = \sum_{x=1}^{2} \sum_{y=1}^{2} \left(x-\tfrac{11}{8}\right)\left(y-\tfrac{11}{8}\right)f(x,y)$$

(a)

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



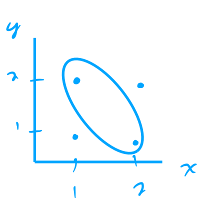
$\implies$ Expect $\mathrm{Cov} > 0$

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			

Contribute most to final value

(b)

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



$\implies$ Expect $\mathrm{Cov} < 0$

using calculator

$L_1 = x$

$L_2 = y$

$L_3 = f(x,y)$

$L_4 = (x-\mu_x)(y-\mu_y)$

$\downarrow = (L_1 - 1.5)(L_2 - 1.5625)$

1 var stats

$(L_1, L_3) \longrightarrow \bar{x} = E(x) = 1.5$

$(L_2, L_3) \longrightarrow \bar{x} = E(y) = 1.5625$

$(L_4, L_3) \longrightarrow \bar{x} = \mathrm{Cov}(X,Y) = -0.15625$

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:

$$\begin{aligned} Cov(X,Y) &= E[(X-\mu_X)(Y-\mu_Y)] \\ &= E[XY - X\mu_Y - Y\mu_X + \mu_X\mu_Y] \\ &= E(XY) - \mu_Y \underbrace{E(X)}_{=\mu_X} - \mu_X \underbrace{E(Y)}_{=\mu_Y} + \mu_X\mu_Y \\ &= E(XY) - \mu_X\mu_Y \end{aligned}$$

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

$$Cov(X,Y) = E(XY) - E(X)E(Y) = 487 - 100(5) = -13$$

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

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

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

Proof: By definition OR Alternate formula

$$\begin{aligned} Cov(X,X) &= E[(X-\mu_X)(X-\mu_X)] \\ &= E[(X-\mu_X)^2] \\ &\overset{\checkmark}{=} V(X) \end{aligned}$$

$$\begin{aligned} Cov(X,X) &= E(XX) - E(X)E(X) \\ &= E(X^2) - (E(X))^2 \\ &\overset{\checkmark}{=} V(X) \end{aligned}$$

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

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

Proof: $E[(X-\mu_X)(Y-\mu_Y)] \overset{\checkmark}{=} E[(Y-\mu_Y)(X-\mu_X)]$ $E(XY) - E(X)E(Y) \overset{\checkmark}{=} E(YX) - E(Y)E(X)$ (commutative property)

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

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

Proof:

$$\begin{aligned} &E[(X-\mu_X)(c - \underbrace{E(c)}_{=c})] \\ &= E[(X-\mu_X)(0)] \\ &= E(0) \\ &\overset{\checkmark}{=} 0 \end{aligned}$$

$$\begin{aligned} &E(Xc) - E(X)E(c) \\ &= cE(X) - cE(X) \\ &\overset{\checkmark}{=} 0 \end{aligned}$$

$V(aX)$
$Cov(aX, aX) = a^2 \, cov(X,X)$
$= a^2 V(X)$

vi) Can factor out coefficients in covariance.

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

Proof:

$\to E(aX) = aE(X) \quad + \quad E(bY) = bE(Y)$

$Cov(aX, bY) = E[(aX - a\mu_X)(bY - b\mu_Y)]$
$= ab\, E[(X-\mu_X)(Y-\mu_Y)]$
$\overset{\checkmark}{=} ab\, Cov(X,Y)$

$Cov(aX, bY) = E(aXbY) - E(aX)E(bY)$
$= ab\, E(XY) - aE(X)\, bE(Y)$
$= ab[E(X)E(Y) - E(X)E(Y)]$
$\overset{\checkmark}{=} ab\, Cov(X,Y)$

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 linear relationship 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.

$Cov(X', Y') = Cov(1.2X, Y+5) = 1.2\, \underbrace{Cov(X,Y)}_{=-13} = -15.6$

viii) **Distributive property** of covariance.

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

Proof:

$\to E(Y+Z) = E(Y) + E(Z)$

$Cov(X, Y+Z) = E[(X-\mu_X)(Y+Z-(\mu_Y+\mu_Z))]$
$= E[(X-\mu_X)((Y-\mu_Y)+(Z-\mu_Z))]$
$= E[(X-\mu_X)(Y-\mu_Y)] + E[(X-\mu_X)(Z-\mu_Z)]$
$\overset{\checkmark}{=} Cov(X,Y) + Cov(X,Z)$

$Cov(X, Y+Z) = E[X(Y+Z)] - E(X)E(Y+Z)$
$= E(XY + XZ) - E(X)[E(Y)+E(Z)]$
$= E(XY) + E(XZ) - E(X)E(Y) - E(X)E(Z)$
$= (E(XY) - E(X)E(Y)) + (E(XZ) - E(X)E(Z))$
$\overset{\checkmark}{=} Cov(X,Y) + Cov(X,Z)$

Independence and covariance

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

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

- Proof:

$Cov(X,Y) = E[\underbrace{(X-\mu_X)}_{g(X)}\underbrace{(Y-\mu_Y)}_{h(Y)}]$
$= E(X-\mu_X)\, E(Y-\mu_Y)$ < split because $\perp\!\!\!\perp$ >
$= [\underbrace{E(X)-\mu_X}_{=0}][\underbrace{E(Y)-\mu_Y}_{=0}]$
$\overset{\checkmark}{=} 0$

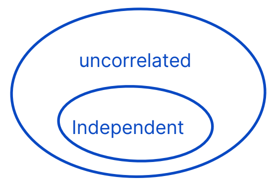
$Cov(X,Y) = E(XY) - E(X)E(Y)$
$= E(X)E(Y) - E(X)E(Y)$
$\overset{\checkmark}{=} 0$

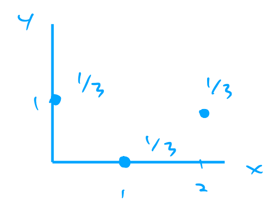
- Independent vs uncorrelated:
  - If X and Y are independent, there does not exist **any type of dependence** between X and Y.
  - If X and Y are uncorrelated (i.e. $\mathrm{Cov}(X, Y) = 0$), there is no linear dependence between X and Y.

    But, there may be some other type of relationship.
  - Thus, independent random variables must be uncorrelated, but uncorrelated random variables may not be independent.

    Summary:

    Independent $\underset{\not\Leftarrow}{\Rightarrow}$ uncorrelated





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

  (a) Are X and Y independent?

  No! Not a rectangular range

  $f(1,0) = \frac{1}{3} \neq \frac{1}{3}\left(\frac{1}{3}\right) = f_X(1)\, f_Y(0)$

  (b) Are X and Y uncorrelated?

  $\rightarrow E(X) = \sum_{x=0}^{2} x f_X(x) = \frac{1}{3}(0 + 1 + 2) = 1$

  $E(Y) = \sum_{y=0}^{1} y f_Y(y) = 0(1/3) + 1(2/3) = 2/3$

  $E(XY) = \sum_{x=0}^{2} \sum_{y=0}^{1} xy\, f(x,y) = 0(1)\frac{1}{3} + 1(0)\frac{1}{3} + 2(1)\frac{1}{3} = \frac{2}{3}$

  $\rightarrow \mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y)$

  $= \frac{2}{3} - 1\left(\frac{2}{3}\right)$

  $= 0 \implies$ uncorrelated, but not independent

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

  $$\text{Cov}(X', Y') = Cov(100X, 100Y) = 100^2 \underbrace{Cov(X,Y)}_{=3} = 30{,}000$$

  We cannot
  [redacted] say that the linear dependence between income and savings in cents is stronger than that in dollars[redacted] because magnitude is arbitrary. All we know is positive.

  $|Cov(X,Y)|$

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

  So how can we improve it?

Correlation

Definition

- Motivation: The problem can be eliminated by standardizing 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:

  $$\text{Corr}(X', Y') = \frac{\text{Cov}(X', Y')}{\sqrt{V(X')V(Y')}} = \frac{Cov(100X, 100Y)}{\sqrt{V(100X)\,V(100Y)}} = \frac{100^2\,Cov(X,Y)}{\sqrt{100^4\,V(X)\,V(Y)}} = \frac{Cov(X,Y)}{\sqrt{V(X)V(Y)}} = Corr(X,Y)$$

  Thus, the correlation is unit-free and unaffected by change in scale[redacted] 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

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{V(X)V(Y)}} = \frac{-13}{\sqrt{40(25)}} = -0.411$$

ALL items needed to find $Corr(X,Y)$

$Cov(X,Y)$

$E(XY)$, $E(X)$, $E(Y)$

$f(x,y)$, $f_X(x)$, $f_Y(y)$

$V(X)$, $V(Y)$

$E(X^2)$, $E(Y^2)$

Properties of correlation

- We are going derive these!
- Covariance measures linear dependence, so lets 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.

$$Y = \underbrace{aX+b}_{\text{deterministic}} + \underbrace{\varepsilon}_{\text{random}}, \quad \varepsilon \sim N(0,1)$$

If $Y = aX + b$,

$$\mathrm{Corr}(X, \underbrace{aX+b}_{=Y}) = \frac{\mathrm{Cov}(X, aX+b)}{\sqrt{V(X)\,\underbrace{V(aX+b)}_{=a^2 V(X)}}} = \frac{a\,\mathrm{Cov}(X,X)}{\sqrt{a^2 (V(X))^2}} = \frac{a\,\cancel{V(X)}}{|a|\,\cancel{V(X)}} = \frac{a}{|a|} = \begin{cases} 1 & \text{if } a > 0 \\ -1 & \text{if } a < 0 \end{cases}$$

Thus when X and Y are linearly related, the correlation coefficient is $+1$ when the slope of the straight line is positive and -1 when the slope is negative.

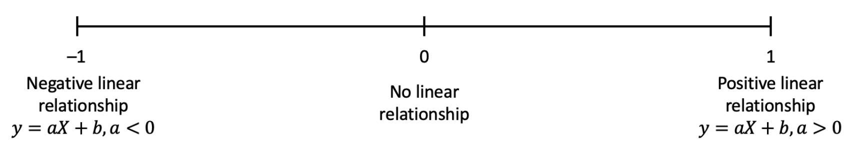
- To see what might happen when X and Y are not linearly related, we can look at the extreme case when X and Y are independent and have no systematic dependence.

$$\text{If } X \perp\!\!\!\perp Y \implies \mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y} \overset{\perp\!\!\!\perp}{=} \frac{0}{\sigma_X \sigma_Y} = 0$$

Obviously $\rho_{XY} = 0$ whenever $\mathrm{Cov}(X,Y) = 0$. 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 always increases.

(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 always decreases.

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

Examples

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

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

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

   No! Arbitrary magnitude

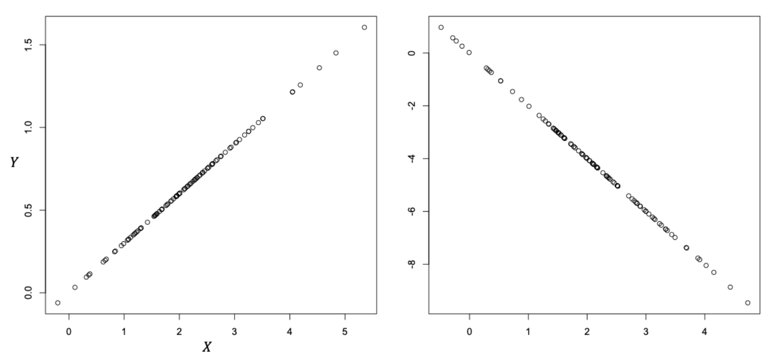
   (c) $\text{Corr}(Y_1, Y_2) = 0.4$, $\text{Corr}(X_1, X_2) = 0.6$

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

   Yes! standardized measure

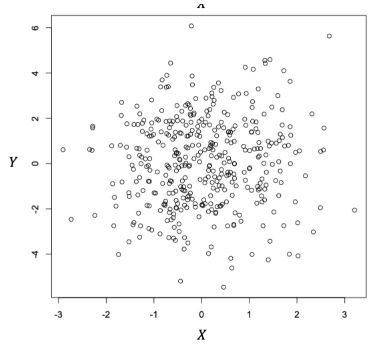
2. Find ρ_{XY}:

   (a) $\text{Corr}(X, Y) = 1$ (b) $\text{Corr}(X, Y) = -1$

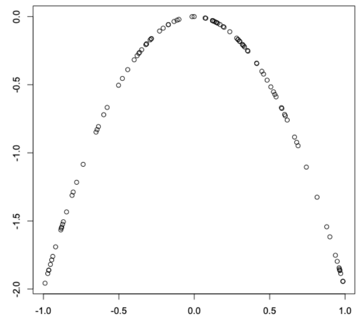


Perfect relationships are very unlikely with real data
because of random error, sampling error, etc.

(c) Corr$(X, Y) =$ 0



(d) Corr$(X, Y) =$ 0



Not a linear relationship,
but obviously very dependent

Corr(x,y)
↳ $y = x^2$

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:

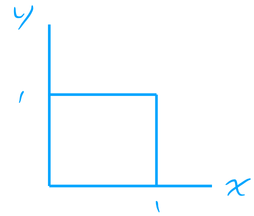
$$V(X+Y) = E[(X+Y)^2] - (E(X+Y))^2$$
$$= E(X^2 + 2XY + Y^2) - (\mu_X + \mu_Y)^2$$
$$= E(X^2) + 2E(XY) + E(Y^2) - (\mu_X^2 + 2\mu_X\mu_Y + \mu_Y^2)$$
$$= E(X^2) - \mu_X^2 + E(Y^2) - \mu_Y^2 + 2[E(XY) - \mu_X\mu_Y]$$
$$= V(X) + V(Y) + 2\,Cov(X,Y)$$

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

$$V(X+Y) = V(X) + V(Y) + 2\,Cov(X,Y)$$

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

$$V(X+Y) = V(X) + V(Y) + 2\,Cov(X,Y) = 40 + 25 + 2(-13) = 39$$



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

$\rightarrow f_X(x) = \int_0^1 f(x,y)\,dy = x + \frac{1}{2}, \quad 0 \le x \le 1$

$f_Y(y) = \int_0^1 f(x,y)\,dx = \frac{3}{2}y^2 + \frac{1}{2}, \quad 0 \le y \le 1$

$\rightarrow E(X) = \int_0^1 x f_X(x)\,dx = 7/12$ $\quad E(X^2) = \int_0^1 x^2 f_X(x)\,dx = 5/12$

$E(Y) = \int_0^1 y f_Y(y)\,dy = 5/8$ $\quad E(Y^2) = \int_0^1 y^2 f_Y(y)\,dy = 7/15$

$\rightarrow V(X) = E(X^2) - (E(X))^2 = \frac{5}{12} - \left(\frac{7}{12}\right)^2 \approx 0.076$

$V(Y) = E(Y^2) - (E(Y))^2 = \frac{7}{12} - \left(\frac{5}{8}\right)^2 \approx 0.076$

$\rightarrow E(XY) = \int_0^1 \int_0^1 xy f(x,y)\,dx\,dy = 0.354$

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

$= 0.354 - \frac{7}{12}\left(\frac{5}{8}\right)$

≈ -0.0104

$\rightarrow V(X+Y) = V(X) + V(Y) + 2Cov(X,Y)$

$= 0.076 + 0.076 + 2(-0.0104)$

$= 0.142$

Variance of $X + Y$ when independent

- Derivation: If X and Y are independent, then

$$V(X+Y) = V(X) + V(Y) + \underbrace{2Cov(X,Y)}_{=0 \text{ if } X \perp\!\!\!\perp Y}$$

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

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

$\Longrightarrow$ 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 $V(X+Y)$.

$V(X+Y) = V(X) + V(Y)$

$= 1/1^2 + 1/3^2$

$= 10/9$