Probability review III: Measuring dependence

Topics we'll cover

1. When are two random variables **independent**?
2. Qualitatively assessing dependence
3. Quantifying dependence: **covariance** and **correlation**

Independent random variables

Random variables X, Y are **independent** if
$\Pr(X = x, Y = y) = \Pr(X = x)\Pr(Y = y)$.

Pick a card out of a standard deck.
X = suit and Y = number.

$P(X = \heartsuit \text{ and } Y = 6) = \frac{1}{52}$

$P(X = \heartsuit) = 1/4$

$P(Y = 6) = 1/13$

$P(X = \heartsuit) * P(Y = 6) = 1/4 \cdot \frac{1}{13} = \frac{1}{52} = P(X = \heartsuit \text{ and } Y = 6)$

therefore X and Y are independent

Independent random variables

Random variables X, Y are **independent** if
$\Pr(X = x, Y = y) = \Pr(X = x)\Pr(Y = y)$.

Flip a fair coin 10 times.
X = # heads and Y = last toss.

$P(X = 10, Y = T) = 0$

Looking segments

$P(X = 10) = \frac{1}{2^{10}}$

$P(Y = T) = 1/2$

clearly $\frac{1}{2^{10}} * 1/2 \neq 0$

So X & Y in this example are clearly not independent
X & Y here are dependent

Independent random variables

Random variables X, Y are **independent** if
$\Pr(X = x, Y = y) = \Pr(X = x)\Pr(Y = y)$.

$X, Y \in \{-1, 0, 1\}$, with these probabilities:

		Y: -1	0	1
X	-1	0.4	0.16	$\boxed{0.24}$
	0	0.05	0.02	0.03
	1	0.05	0.02	0.03

X	P(x)
-1	0.4 + 0.16 + 0.24 = 0.8
0	0.05+0.02+0.03 = 0.1
1	0.05 + 0.02 + 0.03 = 0.1
+	1.0

From table
P(x=-1 and Y=1) = 0.24

Y	(Py)
-1	0.4+ 0.05+0.05 = 0.5
0	0.16 + 0.02 + 0.02 = 0.2
1	0.24+0.03+0.03 = 0.3
+	1.0

P(x=-1 and Y=-1) = 0.4

equal

P(x=0 and P(Y=-1) = 0.05

equal

To check P(x=-1) * P(Y=1) = 0.8 * 0.3 = 0.24 YES ✓
P(x=-1) * P(Y=-1) = 0.8 * 0.5 = 0.40 YES ✓
P(x=0) * P(Y=-1) = 0.1 * 0.5 = 0.05 YES ✓
and for the rest of or for a total of 9 entries

Dependence

Example: Pick a person at random, and take

$$H = \text{height}$$
$$W = \text{weight}$$

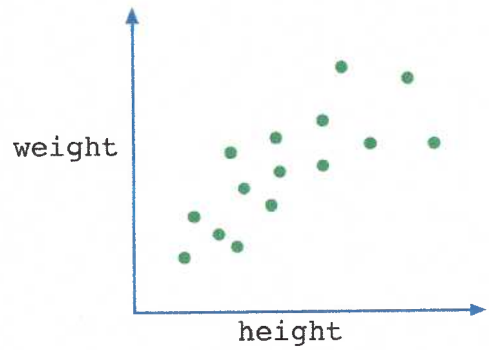
Independence would mean

$$\Pr(H = h, W = w) = \Pr(H = h)\Pr(W = w).$$

Not accurate: height and weight will be **positively correlated**.

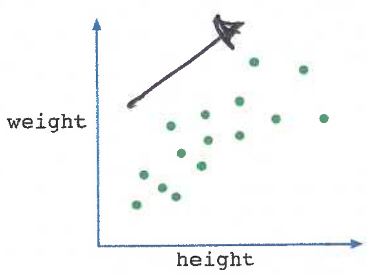
Positive correlation

H, W are **positively correlated**



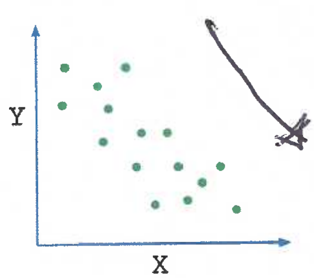
This also implies $\mathbb{E}[HW] > \mathbb{E}[H]\,\mathbb{E}[W]$.

Types of correlation

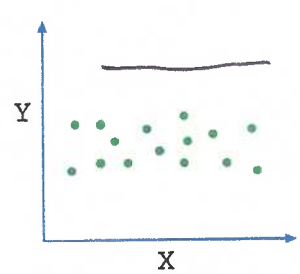


H, W **positively correlated**
This also implies

$$\mathbb{E}[HW] > \mathbb{E}[H]\,\mathbb{E}[W]$$



X, Y **negatively correlated**
$\mathbb{E}[XY] < \mathbb{E}[X]\,\mathbb{E}[Y]$

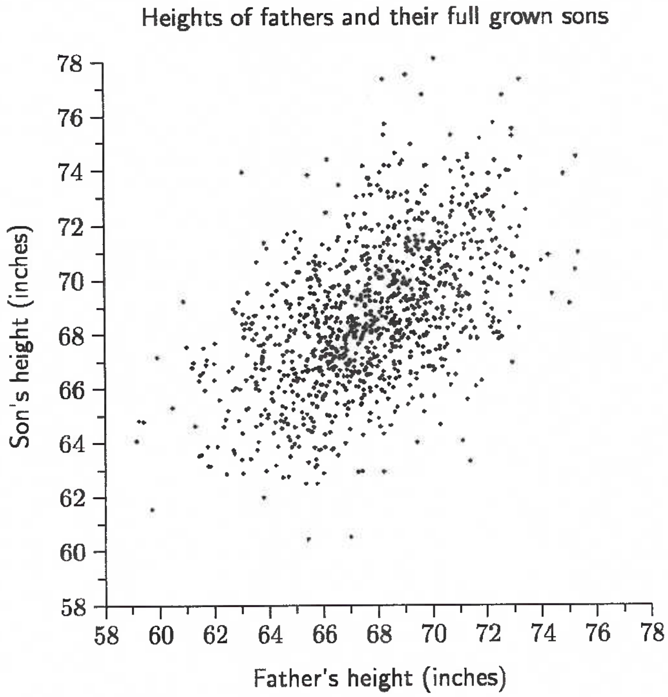


not sloping upward or downward

X, Y **uncorrelated**
$\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]$

Pearson (1903): fathers and sons

Scatter Plot



general upward trend

Correlation Coefficient r

$r \in [-1, 1]$

$-1 \Rightarrow$ perfect negative correlation

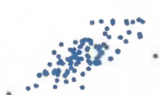
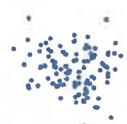
$+1 \Rightarrow$ perfect positive correlation

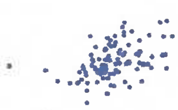
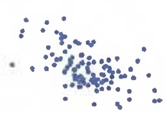
perfect linear correlation/relation

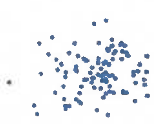
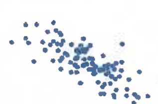
Correlation coefficient: pictures

$r = 1$  $r = 0$ 

uncorrelated when $r=0$

$r = 0.75$  $r = -0.25$ 

$r = 0.5$  $r = -0.5$ 

$r = 0.25$  $r = -0.75$ 

like a blob

anticipate heading towards negative correlation & downward slope

Looks at $E[X,Y]$ < or $= E(X) \cdot E(Y)$

Covariance and correlation

Covariance is a measure of dependence

- Covariance

$$\begin{aligned} \text{cov}(X, Y) &= \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y] \end{aligned}$$

when positive, positive correlation

when negative, negative correlation

by $\notin [-1,1]$ but ranges $[-\text{std}(X)\text{std}(Y), +\text{std}(X)\text{std}(Y)]$

Maximized when $X = Y$, in which case it is var(X).
In general, it is at most std(X)std(Y).

- Correlation

$$\text{corr}(X, Y) = \frac{\text{cov}(X, Y)}{\text{std}(X)\text{std}(Y)}$$

This is always in the range $[-1, 1]$.

If we normalize covariance by std(x)std(y), we get Pearson correlation so Correlation r will range from $[-1,1]$

If X, Y independent then cov(X, Y) = 0.
But the converse need not be true.

Independent means uncorrelated but uncorrelated does not mean independent.

Covariance and correlation: example

Find cov(X, Y) and corr(X, Y)

x	y	$\Pr(x, y)$
−1	−3	1/6
−1	3	1/3
1	−3	1/3
1	3	1/6

$$\text{Cov}(x,y) = E(XY) - E(X)E(Y) = -1 - 0 = -1$$

$$\text{Corr}(X,Y) = \frac{\text{Cov}(x,y)}{\text{std}(X)\,\text{std}(Y)} = \frac{-1}{1 \cdot 3} = -\frac{1}{3}$$

X	$P(X)$
−1	1/2
1	1/2

y	$P(y)$
−3	1/2
3	1/2

xy	P
−3	2/3
3	1/3

$E(XY) = -1$

$E(X) = 0$

$\text{Var}(X) = E(X^2) - (E(X))^2 = 1 - 0 = 1 \quad \text{std} = 1$

$E(Y) = 0$

$\text{Var}(Y) = E(y^2) - (E(Y))^2 = 3^2 - 0 = 9 \quad \text{std} = \sqrt{9} = 3$