Lecture 16 : Independence, Covariance and Correlation of Discrete Random Variables

Definition

Two discrete random variables X and Y defined on the same sample space are said to be independent if for nay two numbers x and y the two events $(X = x)$ and $(Y = y)$ are independent $\Leftrightarrow$

$$P((X = x) \cap (Y = y)) = P(X = x)P(Y = y)$$

$$\Leftrightarrow$$

and

$$P(X = x, Y = y) = P(X = x)P(Y = y)$$

$$\Leftrightarrow$$

$$P_{X,y}(x, y) = P_X(x)P_Y(y) \qquad (*)$$

Now (*) say the *joint pmf* $P_{X,Y}(x,y)$ is determined by the *marginal pmf*'s $P_X(x)$ and $P_Y(y)$ by taking the product.

Problem

In case X and Y are independent how do you recover the matrix (table) representing $P_{X,y}(x,y)$ from its margins?

Let's examine the table for the standard example

$X \backslash Y$	0	1	2	3	
0	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{1}{8}$	0	$\frac{1}{2}$
1	0	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{1}{8}$	$\frac{1}{2}$
	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	

Note that

$X = \sharp$ of heads on the first toss
Y = total $\sharp$ of heads in all three tosses

So we wouldn't expect X and Y to be independent (if we know $X = 1$ that restricts the values of Y.)

Lets use the formula (*)
It says the following.
Each position inside the table corresponds to two positions on the margins

1. Go to the right
2. Go Down

$X \backslash Y$	0	1	2	3	
0					$\frac{1}{2}$
1					$\frac{1}{2}$
	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	

So in the picture

1. If we go right we get $\dfrac{1}{2}$
2. If we go down we get $\dfrac{3}{8}$

If X and Y are independent then the formula (*) says the entry inside the table is obtain by multiplying 1 and 2

$X \backslash Y$	0	1	2	3	
0		$\frac{3}{16}$			$\frac{1}{2}$
1					$\frac{1}{2}$
	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	

So if X and Y wave independent then we would set

$X \backslash Y$	0	1	2	3	
0	$\frac{1}{16}$	$\frac{3}{16}$	$\frac{3}{16}$	$\frac{1}{16}$	$\frac{1}{2}$
1	$\frac{1}{16}$	$\frac{3}{16}$	$\frac{3}{16}$	$\frac{1}{16}$	$\frac{1}{2}$
	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	

$(\sharp)$

So as we expected for the basic example *X and Y are not independent.*
From (*) on page 5 we have

$X \backslash Y$	0	1	2	3
0	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{1}{8}$	0
1	0	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{1}{8}$

(*)

This is not the same as (♯).

Covariance and Correlation

In the “real world” e.g., the newspaper one often hears (reeds) that two quantities are *correlated*. This word is often taken to be synonymous with *causality*. This is not correct and the difference is extremely important even in reel life. Here are two real word examples of correlations.

1. Being rich and driving on expensive car.
2. Smoking and lung cancer.

In the first case there is no causality whereas it is critical that in the second there is.

Statisticians can observe correlations (say for 2) but not causalities.

Now for the mathematical theorem Covariance

Definition

Suppose X and Y are discrete and defined on the same sample space. Then the covariance $\mathrm{Cov}(X, Y)$ between X and Y is defined by

$$\begin{aligned}\mathrm{Cov}(X, Y) &= E((X - \mu_X)(Y - \mu_Y)) \\ &= \sum_{x,y} (x - \mu_X)(y - \mu_Y) P_{X,Y}(x, y)\end{aligned}$$

Remark

$$\text{Cov}(X, X) = E\left((X - \mu_X)^2\right) = V(X)$$

There is a shortcut formula for covariance.

Theorem (Shortcut formula)

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

Remark

If you put $X = Y$ you get the shortcut formula for the variance

$$V(X) = E(X^2) - \mu_X^2$$

Recall that X and Y are independent $\Leftrightarrow P_{X,Y}(x,y) = P_X(x)P_Y(y)$.

Theorem

X and Y are independent

$$\Rightarrow \quad \mathrm{Cov}(X,Y) = 0$$

(the reverse implication does not always hold).

Proof

$$E(XY) = \sum_{x,y} xyP_{X,Y}(x,y)$$

Proof (Cont.)

Now if X and Y are independent then

$$P_{X,Y}(x,y) = P_X(x)P_Y(y)$$

So

$$\begin{aligned} E(XY) &= \sum_{x,y} xyP_X(x)P_Y(y) \\ &= \sum_x xP_X(x) \sum_y yP_y(y) \\ &= \mu_X\mu_Y \end{aligned}$$

Hence

$$\begin{aligned} \mathrm{Cov}(X,Y) &= \mu_X\mu_Y - \mu_X\mu_Y \\ &= 0 \end{aligned}$$

□

Corelation

Let X and Y be as before and suppose $\sigma_X = \sqrt{V(X)}$ and $\sigma_Y = \sqrt{V(Y)}$ be their respective standard deviations.

Definition

The correlation, $\mathrm{Corr}(X, Y)$ *or* $\rho_{X,Y}$ *or just* ρ*, is defined by*

$$\rho_{X,Y} = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

Proposition

$$-1 \leq \rho_{X,Y} \leq 1$$

Theorem

The meaning of correlation.

1. $e_{X,Y} = 1 \Leftrightarrow Y = aX + b$ *with* $a > 0$ *"perfectly correlated"*
2. $P_{X,Y} = -1 \Leftrightarrow Y = aX + b$ *with* $a < 0$ *"perfectly anticorrelated"*
3. X *and* Y *are independent* $\Rightarrow \rho_{X,Y} = 0$ *but not conversely as we will see Pg. 18-21.*

A Good Citizen's Problem

Suppose X and Y are discrete with joint *pmf* given by that of the basic example (*)

X \ Y	0	1	2	3
0	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{1}{8}$	0
1	0	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{1}{8}$

(i) Compute $\mathrm{Cov}(X, Y)$

(ii) Compute $\rho_{X,Y}$

Solution

We first need the marginal distributions.

Solution (Cont.)

X	0	1
$P(X = x)$	$\frac{1}{2}$	$\frac{1}{2}$

So $X \sim \text{Bin}\left(1, \frac{1}{2}\right)$ so $E(X) = \frac{1}{2}$, $V(X) = \frac{1}{4}$ and $\sigma_X = \frac{1}{2}$

Y	0	1	2	3
$P(Y = y)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

So $Y \sim \text{Bin}\left(3, \frac{1}{2}\right)$ so $E(Y) = \frac{3}{2}$, $V(X) = \frac{3}{4}$ so $\sigma_Y = \frac{\sqrt{3}}{2}$.

Now we need $E(XY)$ (the hard part)

$$E(XY) = \sum_{xy} xy\, P(X = x, Y = y)$$

Trick - We are summing over entries in the matrix times xy so potentially eight terms.

Solution (Cont.)

But the four terms from first row don't contribute because $x = 0$ so $xy = 0$. Also the first term in the second row doesn't contribute since $y = 0$. So there are only three terms.

$$E(XY) = (1)(1)\left(\frac{1}{8}\right) + (1)(2)\left(\frac{2}{8}\right) + (1)(3)\left(\frac{1}{8}\right)$$
$$= \frac{1}{8}[1 + 4 + 3] = \frac{8}{8} = 1$$

So

$$\mathrm{Cov}(X, Y) = E(XY) - \mu_X\mu_Y$$
$$= 1 - \left(\frac{1}{2}\right)\left(\frac{3}{2}\right) = \frac{1}{4}$$

Solution (Cont.)

(ii) $$\rho_{X,Y} = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

$$= \frac{1/4}{\left(\frac{1}{2}\right)\left(\frac{\sqrt{3}}{2}\right)} = \frac{1/\not{4}}{\sqrt{3}/\not{4}}$$

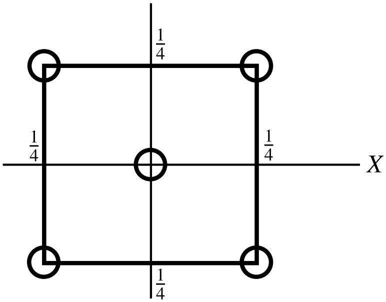
$$= \frac{-1}{\sqrt{3}} = \frac{\sqrt{3}}{3}$$

A cool counter example

We need an example to show

$$\text{Cov}(X, Y) = 0 \nRightarrow X \text{ and } Y \text{ are independent}$$

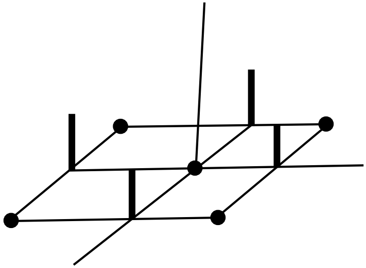
So we need to describe a *pmf*. Here is its "graph"



(♯)

What does this mean.
The corner points (with the zeroes) are $(1, 1)$, $(1, -1)$, $(-1, -1)$ and $(-1, 1)$ (clockwise)

and of course the origin.
Here is the bar graph



The vertical spikes have height $1/4$.
The matrix of the *pmf* is

$X \backslash Y$	-1	0	1	
-1	0	$\frac{1}{4}$	0	$\frac{1}{4}$
0	$\frac{1}{4}$	0	$\frac{1}{4}$	$\frac{1}{2}$
1	0	$\frac{1}{4}$	0	$\frac{1}{4}$
	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	

(*)

I have given the marginal distributions.

Here are the tables for the marginal distributions

X	-1	0	1
$P(X = x)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

$E(X) = 0$

Y	-1	0	1
$P(Y = y)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

$E(Y) = 0$

Now for the covariance.
Here is the really cool thing.
Every term in the Formula for $E(XY)$ so $E(XY)$ is the sum of nine zeroes so $E(XY) = 0$.
So

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

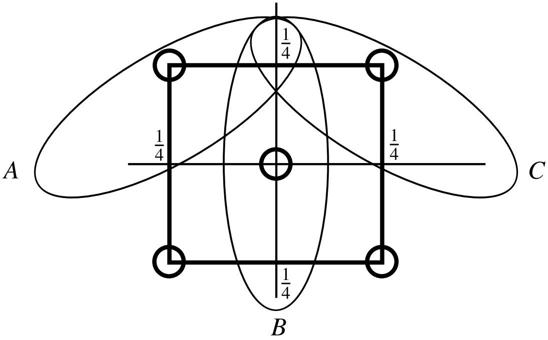
But X and Y are not independent because if we go from the outside in we get

$X \backslash Y$	-1	0	1	
-1	$1/16$	$1/8$	$1/16$	$1/4$
0	$1/8$	$1/4$	$1/8$	$1/2$
1	$1/16$	$1/8$	$1/16$	$1/4$
	$1/4$	$1/2$	$1/4$	

(**)

(*) $\neq$ (**).
So X and Y are not independent.

It turns out the picture (♯) gives us another counter example. Consider the following three events



So $A = \{(0,1), (-1,1), (-1,0)\}$

$$B = \{(0,1), (0,0), (0,-1)\}$$

$$C = \{(0,1), (1,1), (1,0)\}$$

We claim

1. A, B, C are pairwise independent but not independent. That is

$$P(A \cap B) = P(A)P(B)$$
$$P(A \cap C) = P(A)P(C)$$
$$P(B \cap C) = P(B)P(C)$$

but

$$P(A \cap B \cap C) = P(A)P(B)P(C)$$

Let's check this

$$
\begin{aligned}
P(A) &= P(\{(0,1),(-1,1),(-1,0)\} \\
&= \frac{1}{2} \\
P(B) &= P(\{(0,1),(0,0),(0,-1)\} \\
&= \frac{1}{2} \\
P(C) &= P(\{(0,1),(1,1),(1,0)\} \\
&= \frac{1}{2} \\
A \cap B &= \{(0,1)\} \\
A \cap C &= \{(0,1)\} \\
B \cap C &= \{(0,1)\}
\end{aligned}
$$

So they all of probability $\frac{1}{4}$.

$$P(A \cap B) \stackrel{?}{=} P(A)P(B)$$

$$\frac{1}{4} \qquad \left(\frac{1}{2}\right) \left(\frac{1}{2}\right)$$

yes and the some for $A \cap C$ and $B \cap C$.
But $A \cap B \cap C = \{(0,1)\}$
So

$$P(A \cap B \cap C) = P((0,1)) = \frac{1}{4}$$

But

$$P(A)P(B)P(C) = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{1}{8}$$

So

$$P(A \cap B \cap C) \neq P(A)P(B)P(C)$$

An Analogy with Vectors

Here is how I remember the properties of covariance (I learned statistics long after I learned about vectors). This analogy really comes from advanced mathematics on the notion of a “Hilbert space”.

$$\text{Random variable} \overset{\substack{\text{corresponds to}\\ \downarrow}}{\longleftrightarrow} \text{Vector } \overrightarrow{v} \text{ in } \mathbb{R}^2$$

$$\text{Cov}(X, Y) \longleftrightarrow \text{the dot product } \overrightarrow{u} \cdot \overrightarrow{v}$$

So

$$V(X) = \text{Cov}(X, X) \longleftrightarrow \overrightarrow{u} - \overrightarrow{u} = \| \overrightarrow{u} \|^2$$

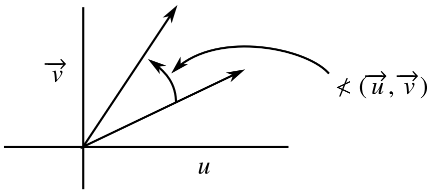
So

$$\sigma_X = \sqrt{V(X)} \longleftrightarrow \sqrt{\vec{u} \cdot \vec{u}} = \| \vec{u} \|$$

the length of the vector $\vec{u}$

Now gives two vectors in the plane the (un oriented) angle between them which I will denote $\measuredangle (\vec{u}, \vec{v}$ is the inverse cosine of $\dfrac{\vec{u} \cdot \vec{v}}{\| \vec{u} \| \| \vec{v} \|}$ that is

$$\cos(\measuredangle (\vec{u}, \vec{v})) = \frac{\vec{u} \cdot \vec{v}}{\| \vec{u} \| \| \vec{v} \|}$$



So what does this correspond to in the world of random variables

$$\frac{\overrightarrow{u} \cdot \overrightarrow{v}}{\| \overrightarrow{u} \| \| \overrightarrow{u} \|} \longleftrightarrow \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

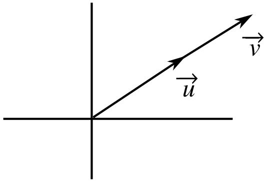
But this is just the correlation $\text{Covr}(X, Y) = \rho_{X,Y}$.

$$\rho_{X,Y} \longrightarrow \text{Cos} \not< (\overrightarrow{u}, \overrightarrow{v})$$

So what do we get from all this?

Positive Correlation

$$\cos \measuredangle (\overrightarrow{u}, \overrightarrow{v}) = 1 \Longleftrightarrow \measuredangle (\overrightarrow{u}, \overrightarrow{v}) = 0$$

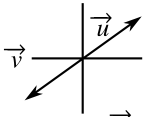


$\overrightarrow{u}$ and $\overrightarrow{v}$ lie in the same ray (half-line) so $\overrightarrow{u} = a\overrightarrow{v}$, $a > 0$.
Now what about correlation

$$\text{Corr}(X, Y) = 1 \Longleftrightarrow Y = aX + b \quad \text{with} \quad a > 0.$$

Negative Correlation

$$\cos \measuredangle (u, v) = -1 \Longleftrightarrow \measuredangle (u, v) = \Pi$$



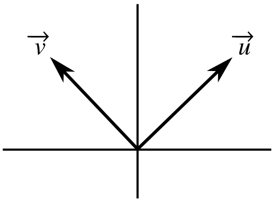
$$\Longleftrightarrow \overrightarrow{u} = a\overrightarrow{v}, a < 0$$

What about correlation

$$\mathrm{Corr}(X,Y) = -1 \iff Y = aX + b \text{ with } a < 0$$

Zero Correlation

$$\cos \measuredangle(\vec{u}, \vec{v}) = 0 \iff \vec{u} \cdot \vec{v} = 0$$
$$\iff \bar{u} \text{ and } \vec{v} \text{ are orthogonal.}$$



$$\mathrm{Corr}(X,Y) = 0 \Longleftarrow X \text{ and } Y \text{ are independent}$$

Bottom Line

Intuitively $\rho_{X,Y}$ corresponds to the cosine of the angle between the two random variables X and Y.