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 any two numbers x and y the two events $(X = x)$ and $(Y = y)$ are independent $\Longleftrightarrow$

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

and

$$\Longleftrightarrow \quad P(X=x, Y=y) = P(X=x)P(Y=y)$$

$$\Longleftrightarrow$$

$$p_{X,Y}(x,y) = p_X(x)\, p_Y(y) \quad (*)$$

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\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 = # of heads on the first toss

Y = total # 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 position on the margins

1. Go to the right
2. Go down

x\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 $\frac{1}{2}$
2. If we go down we get $\frac{3}{8}$

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

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

So if X and Y were independent then we would get

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

(#)

So as we expected for the basic example X and Y are not independent.

From (*) on page 5 we have

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

(*)

This is not the same as (#).

Covariance and Correlation

In the "reel world" e.g. the newspaper one often hears (reeds) thet two quantities are correlated. This word is often teken to be synonymors with causality. This is not correct and the difference is extremely important even in reel life. Here are two reel word examples of correletions

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

In the first core there is no causality whereas it is criticel thet in the second there is.

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

Now for the mathematical theory

Covariance

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

$$Cov(X,Y) = E((X-\mu_X)(Y-\mu_Y))$$

$$= \sum_{x,y} (x-\mu_X)(y-\mu_Y)p_{X,Y}(x,y)$$

Remark

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

There is a shortcut formula for covariance.

Theorem (shortcut formula)

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

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

Theorem

X and Y are independent

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

(the reverse implication does not always hold).

Proof

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

Now if X and Y are independent then

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

so

$$E(XY) = \sum_{x,y} xy\, P_X(x)\, P_Y(y)$$

$$= \sum_{x} x\, P_X(x) \sum_{y} y\, P_Y(y)$$

$$= \mu_X \, \mu_Y$$

Hence

$$\mathrm{Cov}(X,Y) = \mu_X \mu_Y - \mu_X \mu_Y$$

$$= 0$$

Corelation

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

Definition

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

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

Proposition

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

Theorem — The meaning of correlation

1. $\rho_{X,Y} = 1 \iff Y = aX + b$ with $a > 0$
   "perfectly correlated"
2. $\rho_{X,Y} = -1 \iff Y = aX + b$ with $a < 0$
   "perfectly anticorrelated"
3. X and Y are independent

   $\Longrightarrow \rho_{X,Y} = 0$ but not converseley.

   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 (*)

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

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

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

Solution

We first need the marginal distributions

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

so $X \sim Bin(1, \frac{1}{2})$ 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 $Y \sim Bin(3, \frac{1}{2})$ so $E(Y) = \frac{3}{2}$, $V(Y) = \frac{3}{4}$

so $\sigma_Y = \frac{\sqrt{3}}{2}$

<u>Now we need $E(XY)$</u> (the hard part)

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

Trick – we are summing over entries in the matrix times xy so potentially <u>eight</u> terms

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 <u>three</u> terms

$$E(XY) = (1)(1)\left(\tfrac{1}{8}\right) + (1)(2)\left(\tfrac{2}{8}\right) + (1)(3)\left(\tfrac{1}{8}\right)$$

$$= \frac{1}{8}\left[1+4+3\right] = \frac{8}{8} = 1$$

So

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

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

$$= \frac{1/4}{\left(\frac{1}{2}\right)\left(\frac{\sqrt{3}}{2}\right)} = \frac{\frac{1}{\cancel{4}}}{\frac{\sqrt{3}}{\cancel{4}}}$$

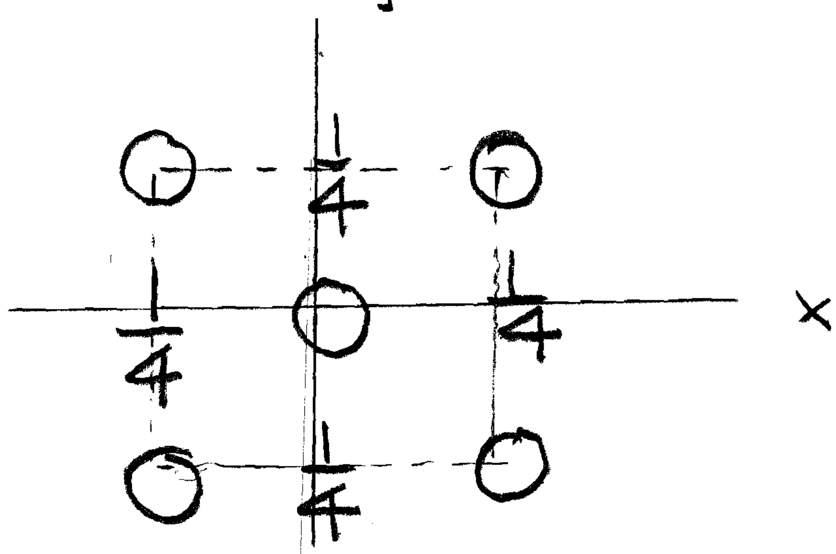
$$= \frac{1}{\sqrt{3}} = \frac{\sqrt{3}}{3}$$

A cool counterexample

We need an example to show

$$\mathrm{Cov}(X,Y) = 0 \not\Longrightarrow X \text{ and } Y \text{ are independent}$$

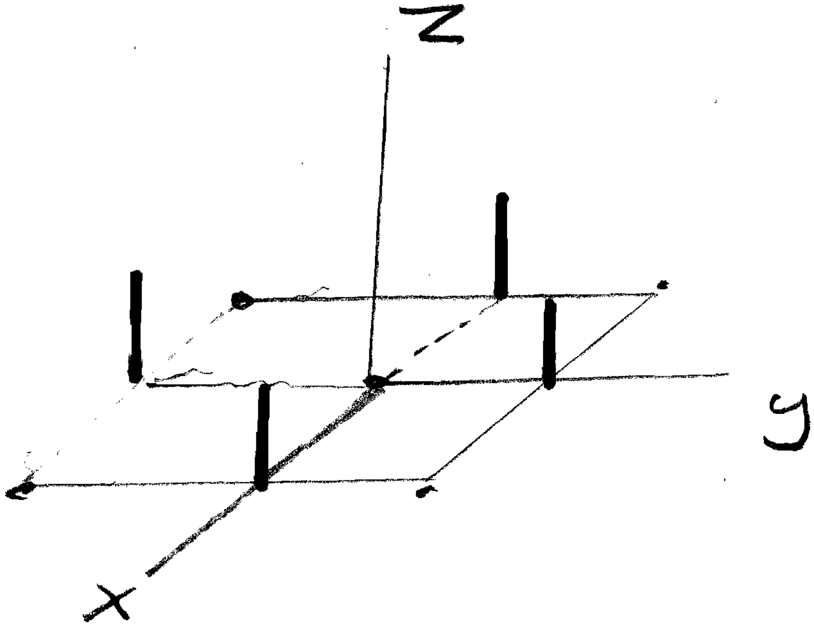
So we need to describe a pmf.
Here is its "graph"



(#)

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 $\frac{1}{4}$.
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

$$Cov(X,Y) = E(XY) - E(X)E(Y) = 0 - (0)(0)$$

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

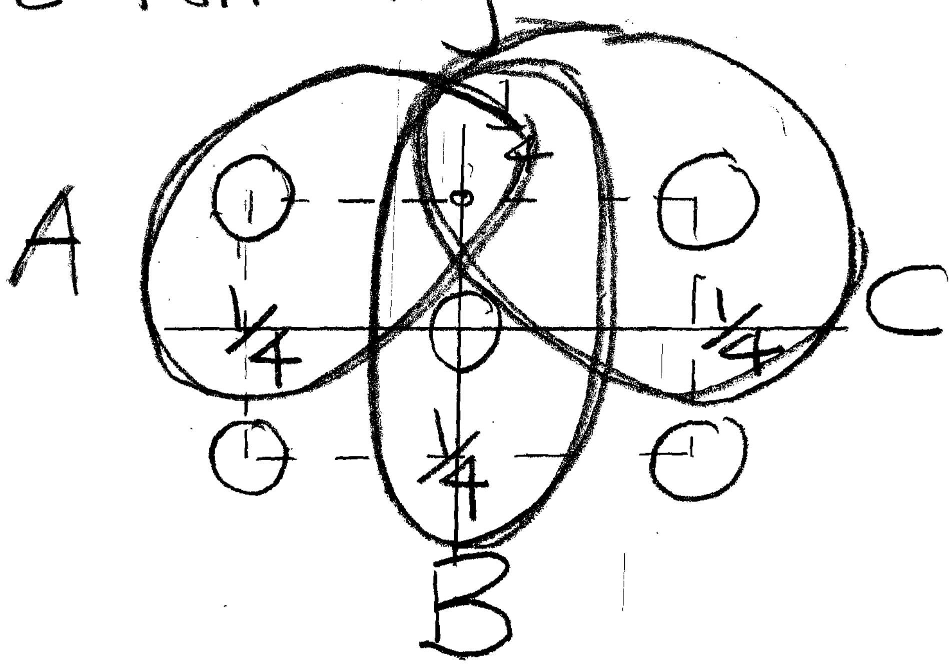
X \ Y	−1	0	1	
−1	1/16	1/8	1/16	1/4
0	1/8	1/4	1/8	1/2
1	1/16	1/8	1/16	1/4
	1/4	1/2	1/4	

(**)

(*) ≠ (**)

so X and Y are not independent

It turns out the picture (#) gives us another counterexample. 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

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

$$= 1/2$$

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

$$= \frac{1}{2}$$

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

$$= \frac{1}{2}$$

$$A \cap B = \{(0,1)\}$$

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

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

So they all of probability $1/4$

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

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

yes and the same 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".

corresponds to

Random variable $\longleftrightarrow$ Vector $\vec{v}$ in $\mathbb{R}^2$

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

so

$$V(X) = \mathrm{Cov}(X,X) \longleftrightarrow \vec{u}\cdot\vec{u} = \|\vec{u}\|^2$$

So,

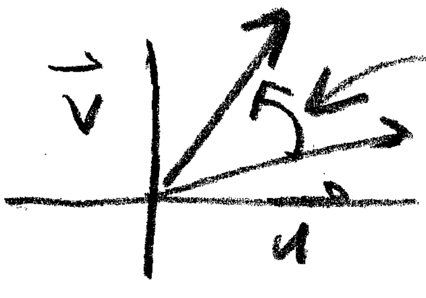
$$\sigma_X = \sqrt{V(X)} \longleftrightarrow \sqrt{\vec{u}\cdot\vec{u}} = \|\vec{u}\|$$

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

Now given two vectors in the plane the (unoriented) 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

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



$\measuredangle(\vec{u},\vec{v})$

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

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

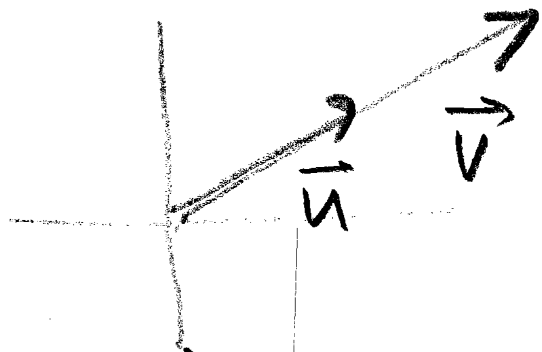
But this is just the correlation $Corr(X,Y) = \rho_{X,Y}$.

$$\rho_{X,Y} \longleftrightarrow \cos \measuredangle(\vec{u},\vec{v})$$

So what do we get from all this?

Positive Correlation

$$\cos \measuredangle(\vec{u},\vec{v}) = 1 \iff \measuredangle(\vec{u},\vec{v}) = 0$$



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

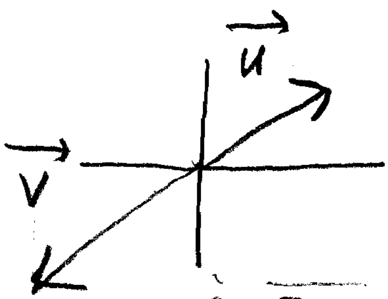
Now what about correlation

$$\mathrm{Corr}(X,Y) = 1 \iff Y = aX + b$$

with $a > 0$.

Negative Correlation

$$\cos \measuredangle(u,v) = -1 \iff \measuredangle(u,v) = \pi$$



$$\iff \vec{u} = a\vec{v},\ a < 0$$

What about correlation

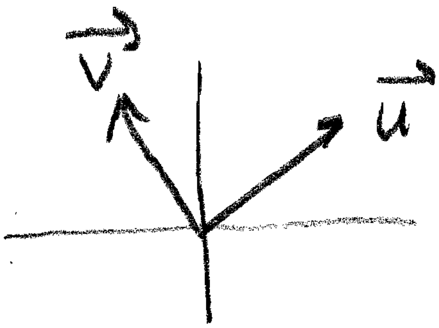
$$\mathrm{Corr}(X,Y) = -1 \iff Y = aX + b$$

with $a<0$

Zero Correlation

$$\cos\angle(\vec{u},\vec{v}) = 0 \iff \vec{u}\cdot\vec{v} = 0$$

$\iff$ $\vec{u}$ and $\vec{v}$ 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