TAM 598 Lecture 5 :

COLLECTIONS OF RANDOM VARIABLES

Announcements:

- HW 1 covers lectures 1-4 ; due on Feb 12
- to be submitted via CANVAS

Joint Probability Mass Function

consider two discrete random variables X and Y. The joint probability mass function $f_{XY}(x,y)$ gives us the probability that $X=x$ and $Y=y$.

$$p(x,y) = p(X=x, Y=y) = f_{XY}(x,y) := p(\{\omega : X(\omega)=x, Y(\omega)=y\})$$

properties of $p(x,y)$:

- non-negative $\quad p(x,y) \geq 0$
- normalized $\quad \sum_x \sum_y p(x,y) = 1$
- marginalization: if you marginalize out one of the random variables, you get the other's pmf

$$p(x) = \sum_y p(x,y) \qquad p(y) = \sum_x p(x,y)$$

MARGINALIZATION : a marginal distribution contains a subset of variables

eg) pmf of X, Y

	x_1	x_2	X		x_5	
$Y =$ y_1	0.06	0.14	0.03	0.25	0.00	0.48
Y	0.01	0.02	0.05	0.06	0.09	0.24
y_3	0.06	0.15	0.00	0.04	0.02	0.28
	0.14	0.31	0.08	0.35	0.11	$p(Y)$
	$p(X=x_1)$				$p(X=x_5)$	

"What is the distribution I would get for X if I ignored Y?"

for $p(x,y)$, two marginals are available : $p(X), p(Y)$

$$p(Y) = \sum_x p(x,y)$$

$$p(X) = \sum_y p(x,y)$$

JOINT PROBABILITY DENSITY FUNCTION

- applies now to two continuous random variables X, Y
- $f_{XY}(x,y)$ is the function that yields the probability that (X,Y) belong to any Borel subset A of $\mathbb{R}^2$

$$P\left((X,Y) \in A \right) = \iint_A \underbrace{f_{XY}(x,y)}_{\text{or } p(x,y)}\, dx\, dy$$

- $f_{XY}(x,y)$ is also non-negative and normalized. And, we can again marginalize out one variable at a time:

$$p(x) = \int_{-\infty}^{\infty} p(x,y)\, dy \qquad p(y) = \int_{-\infty}^{\infty} p(x,y)\, dx$$

CONDITIONING A RANDOM VARIABLE ON ANOTHER $p(x|y)$

- a conditional distribution fixes a subset of variables. "What distribution would I get for X if I set $Y=y$?"

- for both discrete and continuous, divide the joint by the marginal.

$$p(A|B) = \frac{p(A,B)}{p(B)}$$

A, B are logical sentences. This is just the product rule from Lecture 2.

let $A = (X=x)$, $B = (Y=y)$

$$\boxed{p(x|y) = \frac{p(x,y)}{p(y)}}$$

- we can also say

$$p(x|y) = \frac{p(x,y)}{\int p(x',y)\,dx'}$$

* knowing the joint pdf, can find any conditional

eg) pmf of X, Y

	X				
	0.06	0.14	0.03	0.25	0.00
Y	0.01	0.02	0.05	0.06	0.09
	0.06	0.15	0.00	0.04	0.02

say we want $p(X \mid Y = y_2)$

0.24

$p(Y=y_2)$

$\times \ \frac{1}{0.24} \quad \frac{1}{0.24} \quad \frac{1}{0.24} \quad \frac{1}{0.24} \quad \frac{1}{0.24}$

$\boxed{0.05 \mid 0.10 \mid 0.21 \mid 0.25 \mid 0.39} \longrightarrow p(X \mid Y = y_2)$

$p(x_1 \mid y_2)$

EXPECTATION OF TWO RANDOM VARIABLES

let X, Y be two random variables. Let $Z = g(X,Y)$ be a third random variable. Then the expectation of Z is

$$\mathbb{E}[Z] = \mathbb{E}[g(X,Y)]$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\, p(x,y)\, dx\, dy$$

if $Z = X+Y$ $\quad\quad \mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$

INDEPENDENCE, DEPENDENCE, COVARIANCE, AND CORRELATION

- Two random variables X, Y are independent if they don't influence each other. Knowing $X = x$ conveys no info about Y

- Independence means that the joint probability distribution factorizes

$$\boxed{P(X,Y) = P(X)\,P(Y)} \qquad \text{we write} \quad \boxed{X \perp Y}$$

This also implies that the two conditionals are equal to their marginals:

$$p(x|y) = \frac{p(x,y)}{p(y)} = \frac{p(x)\,p(y)}{p(y)} = p(x)$$

$$P(X|Y) = P(X) \qquad \text{and} \qquad P(Y|X) = P(Y)$$

- Independence also tells us that the expectation of XY also factorizes

$$E[XY] = E[X]\,E[Y]$$

Example: $p(X, Y)$ pmf

		X			
Y	0.03	0.03	0.03	0.05	0.07
	0.04	0.05	0.04	0.07	0.10
	0.07	0.08	0.06	0.12	0.16

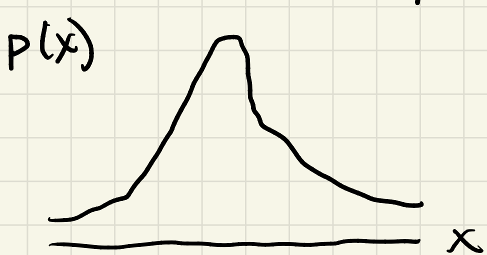
3x5 matrix, elements that sum to 1

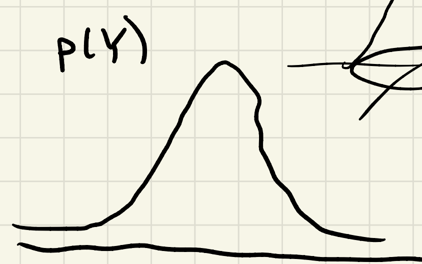
If independent, only need 8 variables to obtain the full matrix

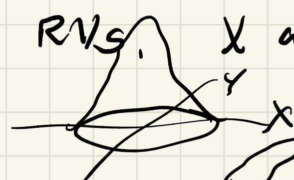
	0.14	0.16	0.13	0.24	0.33	$p(X)$
0.21	0.14×0.21					
0.29						
0.50						
$p(Y)$						

this is a "rank 1" matrix

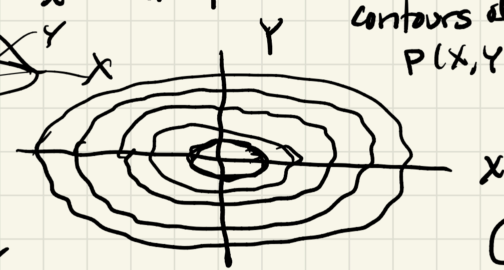
Example: two independent continuous RVs X and Y







contours of $p(X,Y)$



conditional independence : R.V.s X and Y are conditionally independent given Z if

$$p(X, Y \mid Z = z) = p(X \mid Z = z)\, p(Y \mid Z = z)$$

for all $Z = z$.

example: customers buy umbrellas at a store

B_i = event that customer i buys an umbrella $\quad i = 1, 2, \ldots$

R = event that it is raining today

say $p(B_i \mid R) = 0.7$
$p(B_i \mid \neg R) = 0.1$
$p(R) = 0.25$

marginalize:

then $p(B_i) = p(B_i \mid R)\,p(R) + p(B_i \mid \neg R)\,p(\neg R)$

$$= (0.7)(0.25) + (0.1)(0.75)$$

$$= \underline{\underline{0.25}}$$

consider the joint probability $p(B_1, B_2)$

$$\underline{p(B_1, B_2)} = p(B_1, B_2, R) + p(B_1, B_2, \neg R) \quad \text{marginalization}$$

$$= p(B_1, B_2 \mid R)\,p(R) + p(B_1, B_2 \mid \neg R)\,p(\neg R) \quad \text{sum rule}$$

used $B_1|R$ and $B_2|R$ are independent

$$= p(B_1 \mid R)\,p(B_2 \mid R)\,p(R) + p(B_1 \mid \neg R)\,p(B_2 \mid \neg R)\,p(\neg R)$$

$$= (0.7)(0.7)(0.25) + (0.1)(0.1)(0.75) = \underline{\underline{0.13}}$$

Note: $p(B_1, B_2) \neq p(B_1)\,p(B_2) = (0.25)^2$

look at some statistical relationships related to dependence, independence

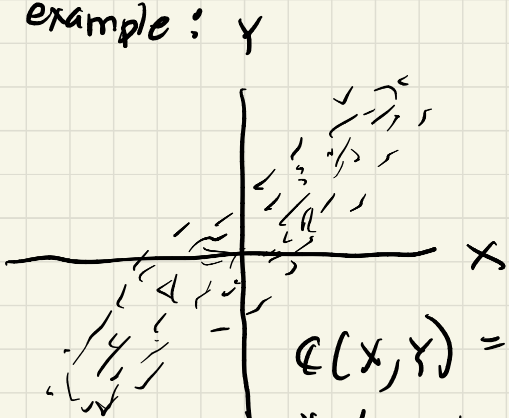
the **COVARIANCE OPERATOR** measures <u>linear</u> dependence between two R.V.s

$$C[X,Y] = \mathbb{E}\left[(X - \mathbb{E}(X))(Y - \mathbb{E}[Y]) \right]$$

$$= \mathbb{E}\left[XY - X\mathbb{E}[Y] - Y\mathbb{E}[X] + \mathbb{E}[X]\mathbb{E}[Y] \right]$$

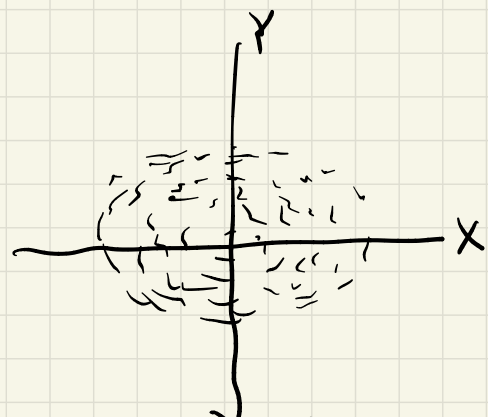
$$= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

example:



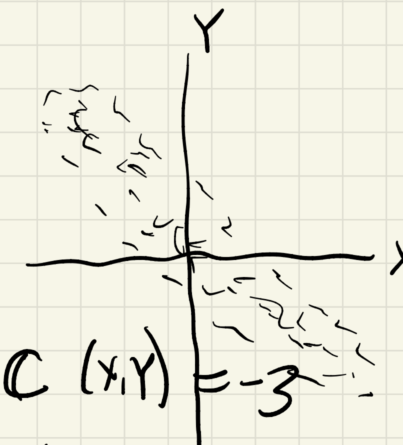
$C(X,Y) = 2$

X, Y positively correlated



$C(X,Y) = 0$

no line dep

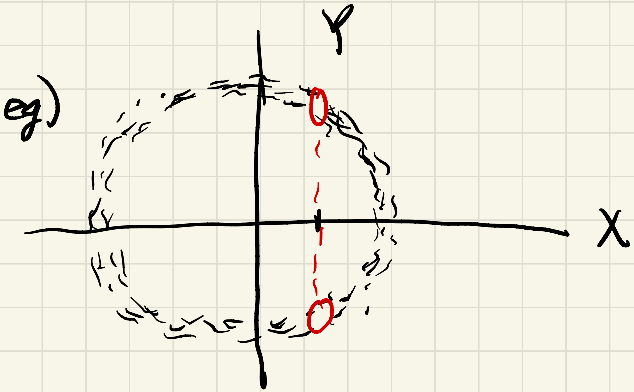


$C(X,Y) = -3$

X, Y negatively correlated

be aware: covariance only captures linear dependence. Just because $\mathbb{C}(X,Y)=0$ does not mean that X and Y are independent

eg)



- knowing X gives info about Y and v.v $\Rightarrow$ not independent
- but $\mathbb{C}(X,Y)=0$ because the dependence is not linear

properties of covariance:

- symmetric $\mathbb{C}(X,Y) = \mathbb{C}(Y,X)$
- linear $\mathbb{C}(aX+bY,Z) = a\,\mathbb{C}(X,Z) + b\,\mathbb{C}(Y,Z)$
- relationship to variance

$$\mathbb{V}(X+Y) = \mathbb{V}(X) + \mathbb{V}(Y) + 2\,\mathbb{C}(X,Y)$$

⊙ relationship to variance

$$\mathbb{V}[X+Y] = \mathbb{V}[X] + \mathbb{V}[Y] + 2\,\mathbb{C}[X,Y]$$

pf: $\mathbb{V}[X+Y] = \mathbb{E}\left[\left((X+Y) - \mathbb{E}(X+Y)\right)^2\right]$

$$= \mathbb{E}\left[(X+Y-\mathbb{E}[X]-\mathbb{E}[Y])^2\right]$$

$$= \mathbb{E}\left[\underline{(X-\mathbb{E}[X])^2} + \underline{(Y-\mathbb{E}[Y])^2}\right.$$

$$\left. + 2\,\underline{\mathbb{E}\left[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])\right]}\right]$$

$$= \mathbb{V}[X] + \mathbb{V}[Y] +$$

$$2\,\mathbb{C}(X,Y)$$

The PEARSON CORRELATION COEFFICIENT:

- covariance is sensitive to units, and its generally unknown what range it can take
- correlation coefficient addresses this problem by normalizing the covariance by each of the std deviations

$$\text{corr}(X,Y) = \rho_{X,Y} = \frac{\mathbb{C}(X,Y)}{\sqrt{\mathbb{V}[X]\mathbb{V}[Y]}}$$

$$\text{corr}(X,Y) \in [-1,1]$$

ie) define new RV's

$$X' = \frac{X}{\sqrt{\mathbb{V}[X]}} \quad \text{and} \quad Y' = \frac{Y}{\sqrt{\mathbb{V}[Y]}}$$

correlation is scale free

$$\text{corr}(X,Y) = \rho_{X,Y} = \mathbb{C}[X',Y']$$