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 one of the random variables values, 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_1		0.06	0.14	0.03	0.25	0.00
	Y	0.01	0.02	0.05	0.06	0.09
y_3		0.06	0.15	0.00	0.04	0.02

$p(Y)$: 0.48, 0.24, 0.28

0.14 0.31 0.08 0.35 0.11 $p(X)$

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

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

$$\left.\begin{aligned} p(x) &= \sum_y p(x,y) \\ p(y) &= \sum_x p(x,y) \end{aligned}\right\} \text{sum out the other variables}$$

eg samples from X and Y together but somebody came and threw away all the Ys

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 $\mathbb{R}^2$

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

- $f_{XY}(x,y)$ 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 Lect. 2

let
$$\left.\begin{aligned} A &= (X=x) \\ B &= (Y=y) \end{aligned}\right\} \qquad \boxed{p(x|y) = \frac{p(x,y)}{p(y)}}$$

- we can also say

$$p(x|y) = \frac{p(x,y)}{\int_{-\infty}^{\infty} p(x',y)\,dx'}$$

← knowing the joint pdf, can find the conditional of one given the other eg) pmf of X, Y

$$p(X \mid Y = y_2)$$

	x_1	x_2	X		x_5
y_1	0.06	0.14	0.03	0.25	0.00
Y	0.01	0.02	0.05	0.06	0.09
y_3	0.06	0.15	0.00	0.04	0.02

$\rightarrow 0.24$

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

0.05	0.10	0.21	0.25	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 $Z = X + Y$, $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$

INDEPENDENCE, DEPENDENCE, COVARIANCE, AND CORRELATION

joint probs/pdfs are interesting only when they capture dependence between variables

- 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 distribution factorizes

$$\boxed{P(X, Y) = P(X)\,P(Y)}$$

we write $\boxed{X \perp Y}$

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

$$p(X|Y) = p(X) \quad \text{and} \quad p(Y|X) = p(Y)$$

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

- Independence also tells us that the expectation of X, Y also factorizes

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

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 sum to 1

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

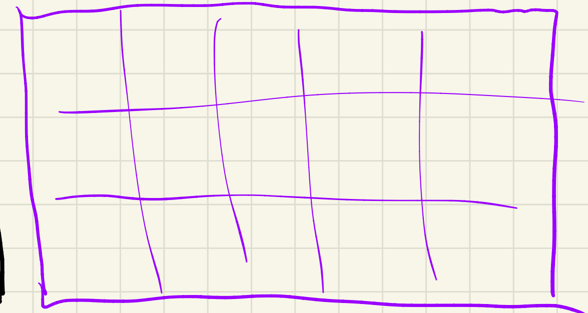
0.14	0.16	0.13	0.24	0.33

$p(X)$ marg

$=$

0.21
0.29
0.50

$p(Y)$ marginal

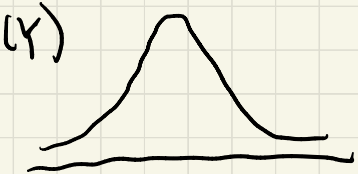


this is a "rank 1" matrix

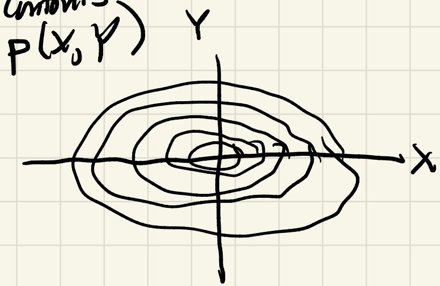
Example: two independent continuous RVs

$p(X) =$

$p(Y)$



contours $p(X, Y)$



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

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

comes up a lot when reasoning about latent structure in data

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

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

$= 0.25$

consider joint probability

marginalize

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

sum rule

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

$B_1 \mid R$ and $B_2 \mid R$ are indep (cond indep)

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

$$= (0.7)(0.7)(0.25) + (0.1)(0.1)(0.75) = 0.13$$

B_1 and B_2 are not indep but they are cond. indep given the knowledge that its raining

Note $p(B_1, B_2) \neq p(B_1)\, p(B_2) = (0.25)(0.25) = 0.0625$

look at some statistical relationships related to dependence, independence

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

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

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

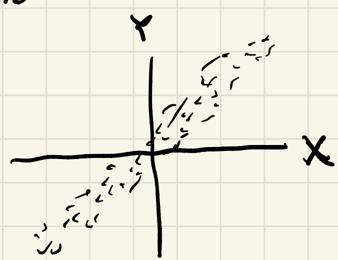
$$= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

do two RV's deviate from their means in a coincident way

both > mean or less mean at same time, cov > 0.

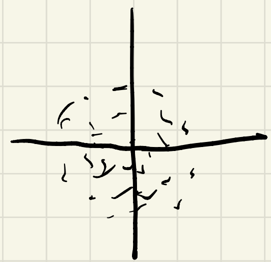
one >, one < then neg

example:



$C(X,Y) = 2$

X,Y positively correlated



$C(X,Y) = 0$

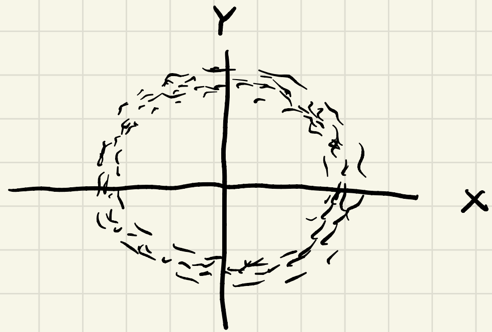
no linear dependence

$C(X,Y) = -3$

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

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

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

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

interpret $\mathbb{C}(X,Y) > 0$: fluctuations amplify

The PEARSON CORRELATION COEFFICIENT :

(height - m, ft : weight)

- 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

$$\boxed{\operatorname{corr}(x,y) = \rho_{x,Y} = \frac{\mathbb{C}(x,Y)}{\sqrt{\mathbb{V}[x]\,\mathbb{V}[Y]}}} \in [-1,1]$$

ie) define new RVs $X' = X \big/ \sqrt{\mathbb{V}[X]}$ and $Y' = Y \big/ \sqrt{\mathbb{V}[Y]}$

has $\mathbb{V}[x'] \in [-1,1]$ . . . .

then $\rho_{XY} = \mathbb{C}(X', Y')$

so correlation is scale free.