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 $Y = y$.

properties of $p(x,y)$:

- ⊙
- ⊙
- ⊙

MARGINALIZATION : a marginal distribution contains a subset of variables

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

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 $\mathbb{R}^2$

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

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

- a conditional distribution fixes a subset of variables.

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

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

INDEPENDENCE, DEPENDENCE, COVARIANCE, AND CORRELATION

① Two random variables X, Y are independent if they don't influence each other. Knowing $X = x$ conveys no info about Y

① Independence means that

Example: $p(X, Y)$ pmf

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

conditional independence : R.V.s X and Y are conditionally independent given Z if 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 $p(B_i \mid R) = 0.7$

$p(B_i \mid \neg R) = 0.1$

$p(R) = 0.25$

look at some statistical relationships related to dependence, independence

the **COVARIANCE OPERATOR** measures linear dependence between two R.V.s be aware: covariance only captures linear dependence. Just because $\mathbb{C}(X,Y)=0$ does not mean that X and Y are independent

⊙ relationship to variance

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

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