Last time

5-1.1

2 RANDOM VARIABLES X, Y

Joint p.d.f.: $\boxed{f_{XY}(x,y)}$ with $\iint f_{XY}(x,y)\,dx\,dy = 1$

Range of poss. values of X & Y

To calculate probabilities: $P((X,Y) \in R) = \iint_R f_{XY}(x,y)\,dx\,dy$

In this context we can recover the pdfs of X & Y

called marginal pdfs of X & Y:

$$f_X(x) = \int_{R_x} f_{XY}(x,y)\,dy$$

R_x = all possible y-values when $X = x$

$$f_Y(y) = \int_{R_y} f_{XY}(x,y)\,dx$$

R_y = all possible x-values when $Y = y$

Using these, you can find $\mu_X = E(X)$ and σ_X^2, μ_Y, σ_Y^2.

5-1.2

Skip 5-1.3 & all references in the rest of Chapter 5 to conditional probability (cond. pdfs)

5-1.4 Independence with r.v.s.

$$P(E\cap F) = P(E)P(F) \quad \text{when } E, F \text{ independent}$$

X and Y are independent r.v.s if "events involving only X are independent of events " " Y."

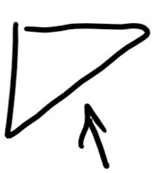
i.e. $P(X\in A, Y\in B) = P(X\in A)P(Y\in B)$

If X, Y are jointly distributed with joint pdf f_{XY}, then X and Y are independent if

$$f_{XY}(x,y) = f_X(x)\, f_Y(y)$$

Last lecture we had a joint pdf $f_{XY}(x,y) = \frac{1}{32}(x+y)$

In fact $f_X(x) = -\frac{3x^2}{64} + \frac{x}{8} + \frac{1}{4}$ (Check!)



$f_Y(y) = \frac{3y^2}{64}$ (Check!)

$0<X<Y<4$

In general <u>if</u> X & Y are independent then the domain of f_{XY} will be <u>rectangular</u>

But rectangular domain not enough to guarantee independence.

<u>Example</u> $f_{XY}(x,y) = 2x + y \quad , \; 0 < x < 1$
$-\frac{1}{2} < y < \frac{1}{2}$

Find f_X, f_Y.

$$f_X(x) = \int_{-1/2}^{1/2} 2x + y \, dy = \left[2xy + \frac{y^2}{2} \right]_{-1/2}^{1/2} = 2x. \quad (0<x<1)$$

$$f_Y(y) = \int_0^1 2x + y \, dx = \left[x^2 + yx \right]_0^1 = 1 + y. \quad \left(-\tfrac{1}{2} < y < \tfrac{1}{2}\right)$$

So $f_X(x) f_Y(y) = 2x(1+y) = 2x + 2xy \neq 2x + y$
$= f_{XY}(x,y)$.

So X & Y <u>not</u> independent.

→ SKIP 5-1.5 (More than 1 r.v.)

5.2 Covariance & Correlation

↳ How X & Y vary together (detecting interaction)

The covariance of X and Y is given by

$$\sigma_{XY} = \operatorname{cov}(X,Y) = \underbrace{E((X-\mu_X)(Y-\mu_Y))}$$

In general, $E(h(X,Y)) = \iint\limits_{\substack{\text{whole}\\ \text{region}}} h(x,y) f_{XY}(x,y)\, dx\, dy$

So $\sigma_{XY} = \iint (x-\mu_X)(y-\mu_Y) f_{XY}(x,y)\, dx\, dy$

$$\vdots$$

$$= E(XY) - \underbrace{E(X)E(Y)}_{\mu_X \mu_Y}$$

Notice if X & Y are independent, then

$$E(XY) = \iint xy\, f_{XY}(x,y)\, dx\, dy = \iint xy\, f_X(x) f_Y(y)\, dx\, dy$$

$$= \int x f_X(x)\, dx \int y f_Y(y)\, dy$$

$$= E(X)E(Y)$$

$\Rightarrow \sigma_{XY} = 0.$

If there's a positive linear relationship* between X and Y, then $\sigma_{XY} > 0$

If there's a negative linear relationship, then $\sigma_{XY} < 0$

$\sigma_{XY} = E((X-\mu_X)(X-\mu_Y))$ so if X & Y are both greater than their means, or both less than their means, $\sigma_{XY} > 0$; if one of each, $\sigma_{XY} < 0$. → * greater values of X correspond to greater values of Y

But actual value of σ_{XY} doesn't communicate the strength of the relationship.

So we have standardization:

The correlation $\text{Corr}(X,Y) = \rho_{XY} = \dfrac{\overbrace{\sigma_{XY}}^{\text{Cov}(X,Y)}}{\underbrace{\sigma_X}_{\text{s.d. of } X} \underbrace{\sigma_Y}_{\text{s.d. of } Y}}$

- no units / dimensionless
- ranges from -1 to 1

→ close to -1: then X, Y "negatively correlated"

close to $+1$: X, Y "positively correlated"

X, Y independent $\Rightarrow \sigma_{XY} = 0 \Rightarrow \rho_{XY} = 0$

But $\rho_{XY} = 0 \underset{\text{necessarily}}{\not\Rightarrow}$ X & Y independent.

Turn to additional file of notes for a worked example of a covariance/correlation calculation.

5.4 Linear Functions of Random Variables

↳ Linear combinations of r.v.s :

$$Y = c_1 X_1 + \dots + c_n X_n \quad , c_i \text{ real \#s}$$

$X_1, \dots, X_n$ r.v.s

$$\boxed{E(Y) = c_1 E(X_1) + \dots + c_n E(X_n)}$$

$$\boxed{V(Y) = c_1^2 V(X_1) + \dots + c_n^2 V(X_n) + \underbrace{\boxed{2 \underbrace{\sum\sum}_{i<j} \text{cov}(X_i, X_j)}}_{= 0}}$$

If $X_1, \dots, X_n$ are independent

So $\boxed{V(Y) = c_1^2 V(X_1) + \dots + c_n^2 V(X_n)}$

Key Examples

① Average of Random Variables $X_1, \dots, X_n$

↓ with identical mean (μ) & variance (σ^2)

$$\bar{X} = \frac{1}{n}(X_1 + \dots + X_n) = \frac{1}{n}X_1 + \frac{1}{n}X_2 + \dots + \frac{1}{n}X_n$$

By above this has

$$E(\bar{X}) = \overbrace{\frac{1}{n} \cdot n \cdot \underbrace{E(X_i)}_{\mu}}^{\frac{1}{n}E(X_1) + \dots + \frac{1}{n}E(X_n)} = \mu$$

If $X_1, \dots, X_n$ independent: $V(\bar{X}) = \frac{1}{n^2} \cdot n \cdot \underbrace{V(X_1)}_{\sigma^2} = \frac{\sigma^2}{n}$.

$$V(\bar{X}) = \frac{1}{n^2} V(X_1) + \dots + \frac{1}{n^2} V(X_n)$$

Now turn to extra file of notes for more key examples to study independently, and a worked example.