Two random variables

- An RV by itself is not that interesting
- In this course: **relationships** between two or more RVs
- Example: Do **college** graduates **earn** more than non-college graduates?
- Other examples: see intro slides

Concepts

Tools for describing multiple random variables:

- **Joint**, **marginal** distribution
- **Conditional** distribution/expectation/variance
- **Independence**, covariance and correlation
- **Rules** for expectations of multiple RVs

Example (SW, 2.2)

Commute	Rain ($X = 0$)	No rain ($X = 1$)	Total
Long ($Y = 20$)	0.15	0.07	[0.22]
Short ($Y = 40$)	0.15	0.63	[0.78]
Total	[0.30]	[0.70]	[1.00]

Table: 2.2 in SW, p. 26

Joint, marginal, conditional

Two random variables, X and Y

- Joint:
  - $P(X = x, Y = y)$,
  - probability that it rains and I have a long commute
- Marginal
  - $P(X = x) = \sum_{i=1}^{k} P(X = x, Y = y_i)$
  - probability that it rains
- Conditional
  - $P(Y = y | X = x) = \frac{P(X=x, Y=y)}{P(X=x)}$
  - probability of having a long commute **given** that it rains

Conditional mean and variance

- Mean and variance for conditional distribution
- The **conditional expectation** of X given $Y = y$ is the expectation of X if you already have the information that $Y = y$
- Same for the **conditional variance**

Conditional expectation: example

- M is the number of times your computer crashes
- A denotes computer is old ($A = 0$) or new ($A = 1$)

	M				
	0	1	2	3	4
$A = 0$	0.35	0.065	0.05	0.025	0.01
$A = 1$	0.45	0.035	0.01	0.005	0.00

- Q: If I **know** the computer is old, how many crashes do I **expect**?

Conditional expectation: formula

- The answer to the question is

$$E[M|A=0] = \sum_{k=1}^{K} M_k P(M_k|A=0)$$

- Compare this to the **unconditional** expectation

$$E[M] = \sum_{k=1}^{K} M_k P(M_k)$$

- Replaced **marginal** by **conditional** probability distribution

Covariance, correlation

- Two more ways to measure the relationship between two r.v.'s are the **covariance** and **correlation**
- The **covariance** and **correlation** measure whether two variables are moving together, with respect to the mean
- Formulas: overhead

Interpretation

- The sign of the **covariance** is easy to interpret
- The magnitude of the covariance is hard to interpret
- The sign and magnitude of the **correlation** are easy to interpret
- The correlation is always between -1 and 1
- The closer to 1 (-1), the more strongly positively (negatively) correlated the variables are

Mean and variance, multiple RV

- Most important rules to remember from “The Mean and Variance of Sums of Random Variables”, pp. 32–35:
  - $E[X + Y] = E[X] + E[Y]$
  - $Var[X + Y] = Var[X] + Var[Y] + 2Cov(X, Y)$
- If the covariance between X and Y is zero ($\sigma_{XY} = 0$),

$$Var[X + Y] = Var[X] + Var[Y]$$