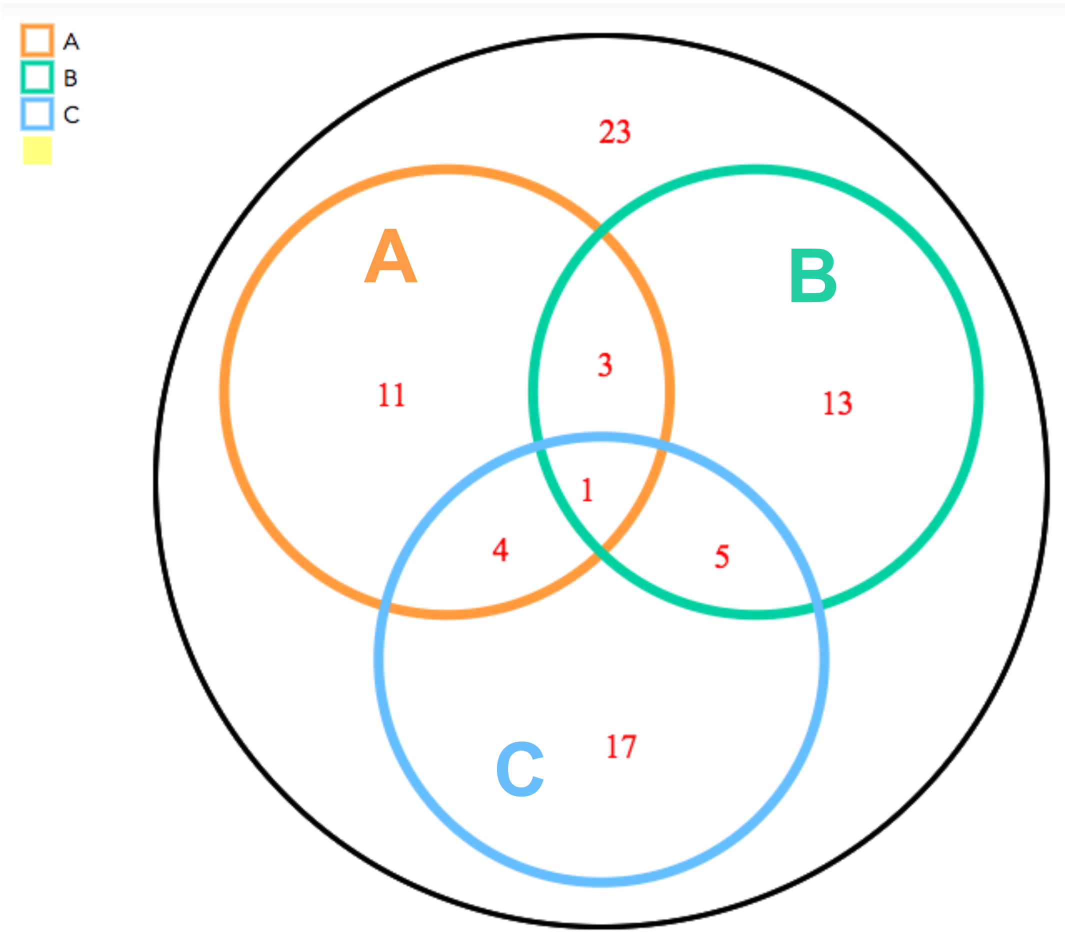
10-606 Probability

Abhinav Maurya

Events



$$N = 77$$

$$P(A) = \frac{19}{77}$$

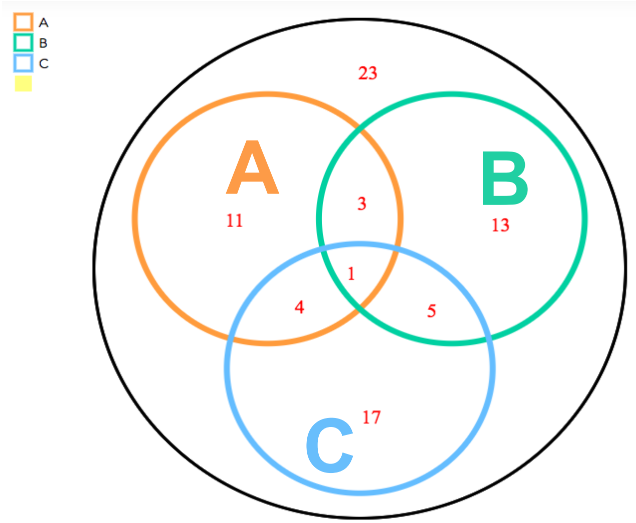
$$P(A \cap B) = \frac{4}{77}$$

$$P(A \cap B \cap C) = \frac{1}{77}$$

$$P(A \cap !B \cap C) = \frac{4}{77}$$

$$P(A|B \cup C) = \frac{8}{43}$$

Random Variables



- Event A indicates stock A goes up
- Event !A indicates the opposite
- Boolean random variable A and !A can indicate the stock movement too.

- Random Variables
  - Measurement in an experiment
- Types
  - Univariate: e.g. X=output of dice roll
  - Multivariate: e.g. $X=(X_1,X_2)$=(width of box, height of box)
  - Discrete: e.g. X=output of dice roll
  - Continuous: e.g. X=width of box
  - Hybrid: e.g. X=reading of a censored sensor

Joint Probability Table

		Intelligence		
		low	*high*	
Grade	*A*	0.07	0.18	0.25
	B	0.28	0.09	0.37
	C	0.35	0.03	0.38
		0.7	0.3	1

Intelligence=I	Grade=G	Probability(Intelligence=I, Grade=G)
low	A	0.07
low	B	0.28
low	C	0.35
high	A	0.18
high	B	0.09
high	C	0.03

Joint Probabilities

		Intelligence		
		low	*high*	
	A	0.07	0.18	0.25
Grade	*B*	0.28	0.09	0.37
	C	0.35	0.03	0.38
		0.7	0.3	1

$$P(Grade = A, Intelligence = low) = 0.07$$

$$P(A = a, B = b, C = c, \dots)$$

Marginal Probabilities

		Intelligence: *low*	*high*	
Grade	*A*	0.07	0.18	0.25
	B	0.28	0.09	0.37
	C	0.35	0.03	0.38
		0.7	0.3	1

$$P(Grade = A) = P(Grade = A, Intelligence = low) + P(Grade = A, Intelligence = high) = 0.25$$

Sum/Marginalization Rule

$$P(A = a) = \sum_{b \in \mathcal{B}} P(A = a, B = b)$$

Marginal Probabilities

		Intelligence low	Intelligence high	
Grade	A	0.07	0.18	0.25
	B	0.28	0.09	0.37
	C	0.35	0.03	0.38
		0.7	0.3	1

$$\begin{aligned} P(Intelligence = high) = {} & P(Grade = A, Intelligence = high) \\ & + P(Grade = B, Intelligence = high) \\ & + P(Grade = C, Intelligence = high) = 0.3 \end{aligned}$$

Sum/Marginalization Rule

$$P(A = a) = \sum_{b \in \mathcal{B}} P(A = a, B = b)$$

Conditional Probabilities

		Intelligence *low*	*high*	
Grade	*A*	0.07	0.18	0.25
	B	0.28	0.09	0.37
	C	0.35	0.03	0.38
		0.7	0.3	1

$$P(Grade = A \mid Intelligence = low) = \frac{P(Grade = A, Intelligence = low)}{P(Intelligence = low)}$$

$$= \frac{0.07}{0.7} = 0.1$$

Conditioning Rule

$$P(A = a | B = b) = \frac{P(A = a, B = b)}{P(B = b)}$$

Conditional Probabilities

		Intelligence		
		low	*high*	
Grade	*A*	0.07	0.18	0.25
	B	0.28	0.09	0.37
	C	0.35	0.03	0.38
		0.7	0.3	1

$$P(Grade = C \mid Intelligence = high) = \frac{P(Grade = C, Intelligence = high)}{P(Intelligence = high)}$$

$$= \frac{0.03}{0.3} = 0.1$$

Conditioning Rule

$$P(A = a|B = b) = \frac{P(A = a, B = b)}{P(B = b)}$$

Conditional Probability Table

		Intelligence		
		low	*high*	
Grade	*A*	0.07	0.18	0.25
	B	0.28	0.09	0.37
	C	0.35	0.03	0.38
		0.7	0.3	1

		Intelligence | *Grade*		
		low	*high*	
Grade	*A*	**0.28**	**0.72**	**1.00**
	B	**0.76**	**0.24**	**1.00**
	C	**0.92**	**0.08**	**1.00**

- Same shape as joint probability table
- But shows $P(A = a|B = b)$ instead of $P(A = a, B = b)$
- Sum of all conditional probabilities under a condition equal to 1.

Conditional Probability Table

		Intelligence low	*Intelligence* high	
Grade	*A*	0.07	0.18	0.25
	B	0.28	0.09	0.37
	C	0.35	0.03	0.38
		0.7	0.3	1

		Intelligence low	*Intelligence* high
Grade* | *Intel.	*A*	**0.10**	**0.60**
	B	**0.40**	**0.30**
	C	**0.50**	**0.10**
		1.00	**1.00**

- Same shape as joint probability table
- But shows $P(A = a|B = b)$ instead of $P(A = a, B = b)$
- Sum of all conditional probabilities under a condition equal to 1.

Chain Rule

- Sum/Marginalization Rule

$$P(A = a) = \sum_{b \in \mathcal{B}} P(A = a, B = b)$$

- Conditioning Rule

$$P(A = a | B = b) = \frac{P(A = a, B = b)}{P(B = b)}$$

- Product/Chain Rule

$$P(A = a, B = b) = P(A = a | B = b) \cdot P(B = b)$$

		Intelligence	
		low	*high*
Grade	*A*	**0.10**	**0.60**
| Intel.	*B*	**0.40**	**0.30**
	C	**0.50**	**0.10**
		1.00	**1.00**

		Intelligence	
		low	*high*
Grade	*A*	???	
	B		
	C		
		0.7	0.3

Finally the Bayes Rule

- Sum/Marginalization Rule

$$P(A=a) = \sum_{b \in \mathcal{B}} P(A=a, B=b)$$

- Conditioning Rule

$$P(A=a|B=b) = \frac{P(A=a, B=b)}{P(B=b)}$$

- Product/Chain Rule

$$P(A=a, B=b) = P(A=a|B=b) \cdot P(B=b)$$

- Bayes Rule

$$\begin{aligned} P(A=a|B=b) &= \frac{P(A=a, B=b)}{P(B=b)} \\ &= \frac{P(B=b|A=a) \cdot P(A=a)}{P(B=b)} \\ &= \frac{P(B=b|A=a) \cdot P(A=a)}{\sum_{a \in \mathcal{A}} P(B=b|A=a) \cdot P(A=a)} \end{aligned}$$

Let's do a simple inference with the Bayes rule

- H~Headache, F~Flu
- P(H) = 1/10
- P(F) = 1/40
- P(H | F) = 1/2

$$P(F \text{ and } H) = P(H \mid F) \times P(F) = \frac{1}{2} \times \frac{1}{40} = \frac{1}{80}$$

$$P(F \mid H) = \frac{P(F \text{ and } H)}{P(H)} = \frac{1/80}{1/10} = \frac{1}{8}$$

Another Bayes rule example

- Two coins – first coin is Head+Tail, the second coin has both heads.
- Randomly picked one of the coins with equal probability.
- Tossed it and observed 10 heads.
- What is the probability that you picked the first coin?

$$P(FC|10H) = \frac{P(10H|FC)\cdot P(FC)}{P(10H)}$$

$$P(FC|10H) = \frac{P(10H|FC)\cdot P(FC)}{P(10H|FC)\cdot P(FC)+P(10H|SC)\cdot P(SC)}$$

Conditioning Caveat

- All rules hold when conditioning on a particular event or on particular values of random variables

- Marginalization Rule

$$P(A = a | C = c, D = d) = \sum_{b \in \mathcal{B}} P(A = a, B = b | C = c, D = d)$$

- Conditioning Rule

$$P(A = a | B = b, C = c, D = d) = \frac{P(A = a, B = b | C = c, D = d)}{P(B = b | C = c, D = d)}$$

- Chain Rule

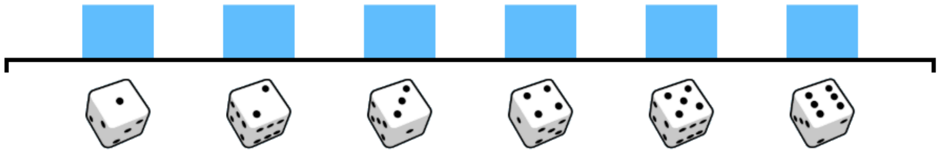
$$P(A = a, B = b | C = c) = P(A = a | B = b, C = c) \cdot P(B = b | C = c)$$

- Bayes Rule

$$P(A = a | B = b, C = c) = \frac{P(B = b | A = a, C = c) \cdot P(A = a | C = c)}{\sum_{a \in \mathcal{A}} P(B = b | A = a, C = c) \cdot P(A = a | C = c)}$$

Expectation

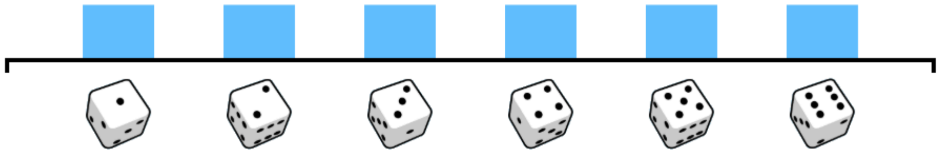
$$E[X] = \sum xP(x)$$



$$E[X] = 3.5$$

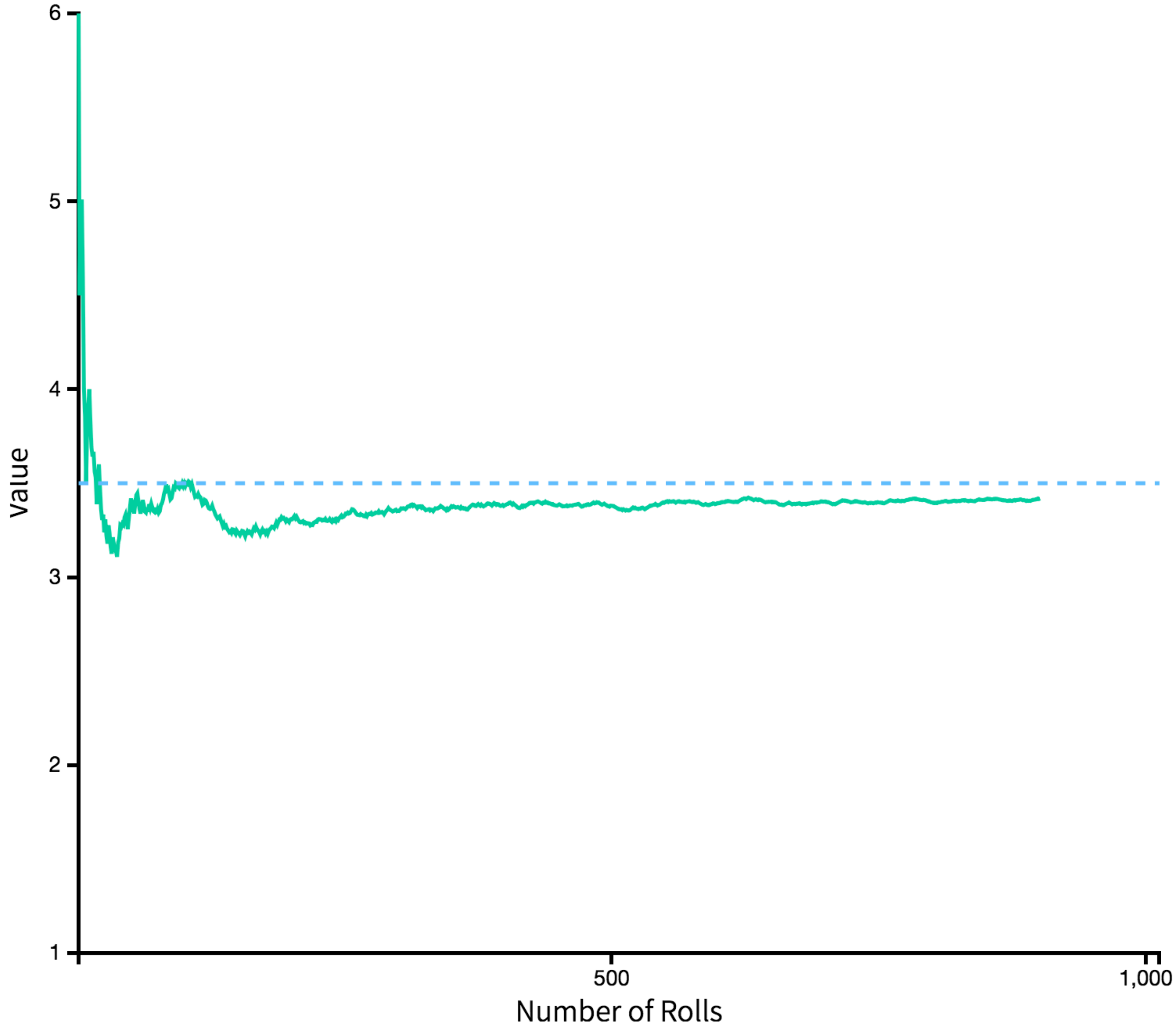
Variance

$$Var[X] = E[(X - E[X])^2]$$



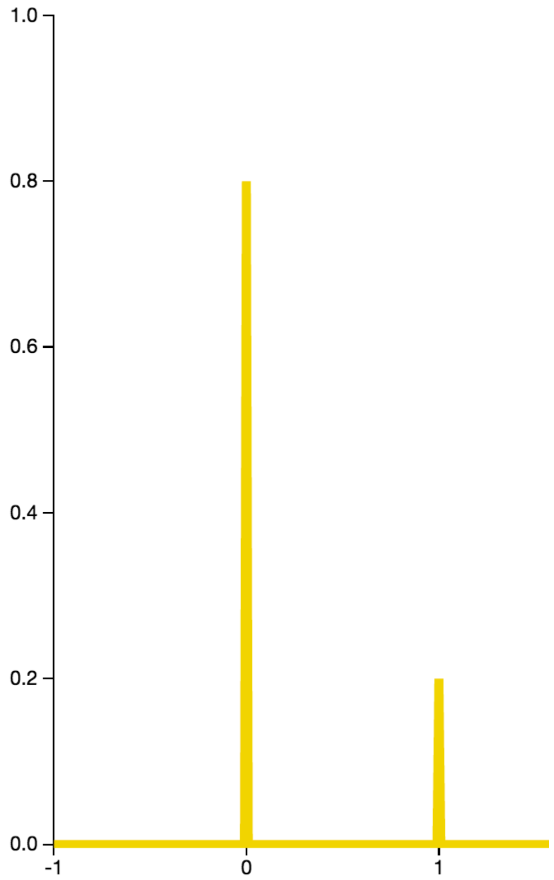
$$Var[X] = \frac{1}{6}\{(1 - 3.5)^2 + (2 - 3.5)^2 + \cdots\} = \frac{35}{12}$$

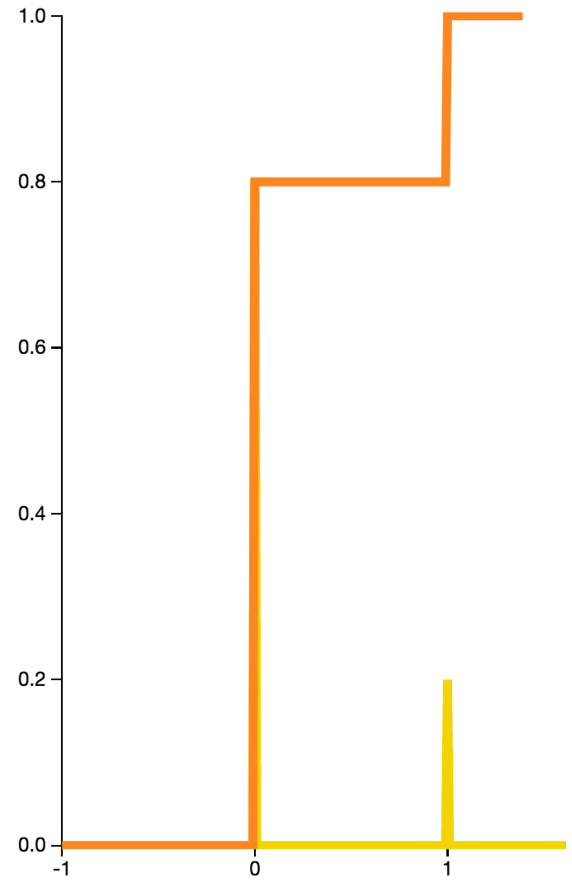
Central Limit Theorem



Discrete Random Variables

$$P(X = x) = f(x)$$
$$P(X < x) = F(x)$$

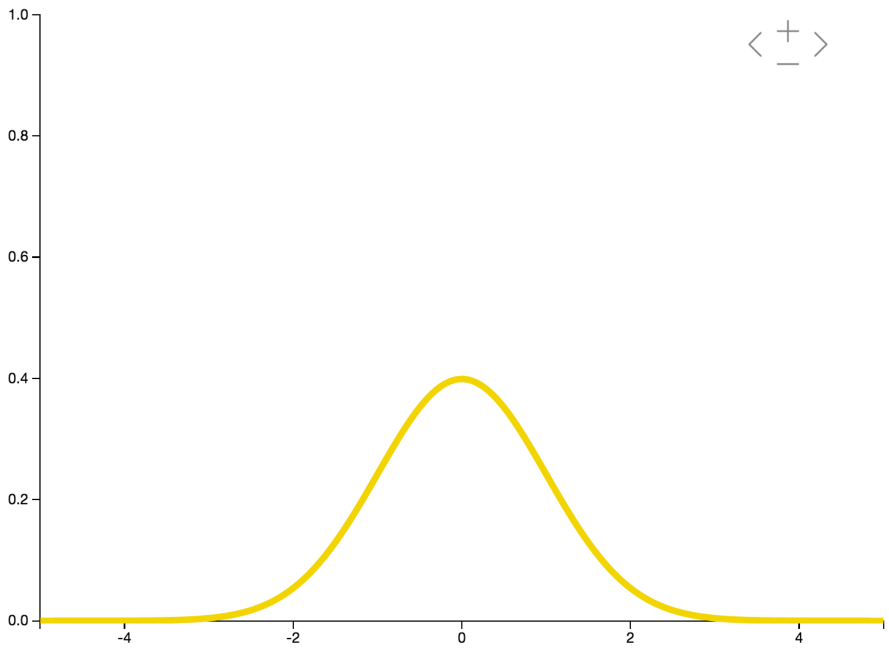




Continuous Random Variables

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

$$P(X < x) = F(x)$$



Some resources

- Seeing Theory: https://seeing-theory.brown.edu/
- Chapter 2 of Koller and Friedman's Probabilistic Graphical Models
- Larry Wasserman's All of Statistics

Thanks!

Questions?