06: Random Variables

David Varodayan
January 17, 2020
Adapted from slides by Lisa Yan

Probability of events

E or F
$P(E \cup F)$

De Morgan's

E and F
$P(EF)$

Mutually exclusive?

Just add!

$P(E) + P(F)$

Inclusion-Exclusion Principle

$P(E) + P(F) - P(E \cap F)$

Independent?

Just multiply!

$P(E)P(F)$

Chain Rule

$P(E)P(F|E)$
$P(F)P(E|F)$

Hash table fun

- m strings are hashed (unequally) into a hash table with n buckets.
- Each string hashed is an independent trial w.p. p_i of getting hashed into bucket i.

1. E = bucket 1 has ≥ 1 string hashed into it.

What is $P(E)$?

Define S_i = string i is hashed into bucket 1
S_i^C = string i is not hashed into bucket 1

$P(S_i) = p_1$
$P(S_i^C) = 1 - p_1$

$$P(E) = P(S_1 \cup S_2 \cup \cdots \cup S_m)$$

$$= 1 - P\left((S_1 \cup S_2 \cup \cdots \cup S_m)^C\right)$$ Complement

$$= 1 - P\left(S_1^C S_2^C \cdots S_m^C\right)$$ De Morgan's Law

$$= 1 - P\left(S_1^C\right)P\left(S_2^C\right) \cdots P(S_m^C) = 1 - \left(P\left(S_1^C\right)\right)^m$$ S_i independent trials

$$= 1 - (1 - p_1)^m$$

The fun never stops with hash tables

- m strings are hashed (unequally) into a hash table with n buckets.
- Each string hashed is an independent trial w.p. p_i of getting hashed into bucket i.

1. E = bucket 1 has $\geq$ 1 string hashed into it.
2. E = at least 1 of buckets 1 to k has $\geq$ 1 string hashed into it.
3. E = each of buckets 1 to k has $\geq$ 1 string hashed into it.

What is $P(E)$?

Define F_i = bucket i has at least one string in it

The fun never stops with hash tables

- m strings are hashed (unequally) into a hash table with n buckets.
- Each string hashed is an independent trial w.p. p_i of getting hashed into bucket i.

1. E = bucket 1 has $\geq$ 1 string hashed into it.
2. E = at least 1 of buckets 1 to k has $\geq$ 1 string hashed into it.
3. E = each of buckets 1 to k has $\geq$ 1 string hashed into it.

What is $P(E)$?

Define F_i = bucket i has at least one string in it

$$P(E) = P(F_1 F_2 \cdots F_k)$$

$$= 1 - P\left((F_1 F_2 \cdots F_k)^C\right) \quad \text{Complement}$$

$$= 1 - P\left(F_1^C \cup F_2^C \cup \cdots \cup F_k^C\right) \quad \text{De Morgan's Law}$$

$$= 1 - P\left(\bigcup_{i=1}^{k} F_i^c\right) = 1 - \sum_{r=1}^{k} (-1)^{(r+1)} \sum_{i_1 < \cdots < i_r} P\left(F_{i_1}^c F_{i_2}^c \ldots F_{i_r}^c\right)$$

where $P\left(F_{i_1}^c F_{i_2}^c \ldots F_{i_r}^c\right) = (1 - p_{i_1} - p_{i_2} \ldots - p_{i_r})^m$

Today's plan

Conditional Independence

Random Variables

PMFs and CDFs

Expectation

Conditional Paradigm

For any events A, B, and E, you can condition consistently on E, and all formulas still hold:

Axiom 1	$0 \leq P(A\|E) \leq 1$
Corollary 1 (complement)	$P(A\|E) = 1 - P(A^C\|E)$
Commutativity	$P(AB\|E) = P(BA\|E)$
Chain Rule	$P(AB\|E) = P(B\|E)P(A\|BE)$
Bayes' Theorem	$P(A\|BE) = \frac{P(B\|AE)P(A\|E)}{P(B\|E)}$

Conditional Independence

Independent events E and F ⟺ $P(EF) = P(E)P(F)$
$P(E|F) = P(E)$

Two events A and B are defined as conditionally independent given E if:

$$P(AB|E) = P(A|E)P(B|E)$$

An equivalent definition:

A. $P(A|B) = P(A)$
B. $P(A|BE) = P(A)$
C. $P(A|BE) = P(A|E)$

Netflix and Condition

Review

Let E = a user watches Life is Beautiful.

Let F = a user watches Amelie.

What is $P(E)$?

$$P(E) \approx \frac{\text{\# people who have watched movie}}{\text{\# people on Netflix}} = \frac{10{,}234{,}231}{50{,}923{,}123} \approx 0.20$$

What is the probability that a user watches Life is Beautiful, given they watched Amelie?

$$P(E|F) = \frac{P(EF)}{P(F)} = \frac{\text{\# people who have watched both}}{\text{\# people who have watched Amelie}} \approx 0.42$$

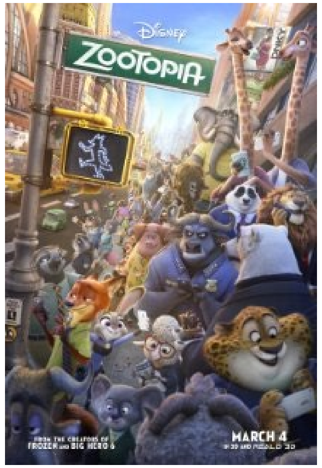
Netflix and Condition

Review

Let E be the event that a user watches the given movie.
Let F be the event that the same user watches Amelie.

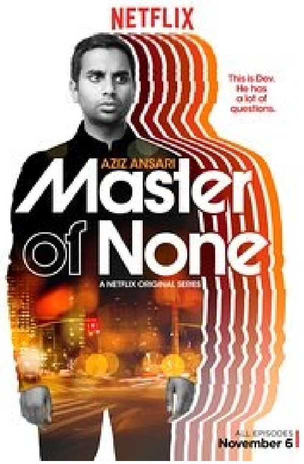
$P(E) = 0.19$

$P(E|F) = 0.14$



$P(E) = 0.32$

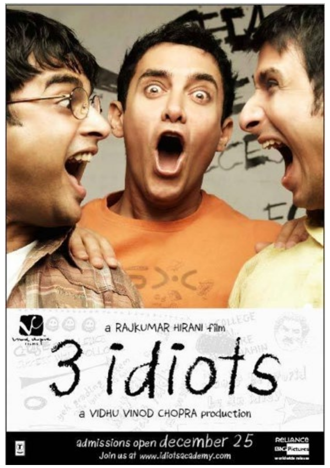
$P(E|F) = 0.35$



$P(E) = 0.20$

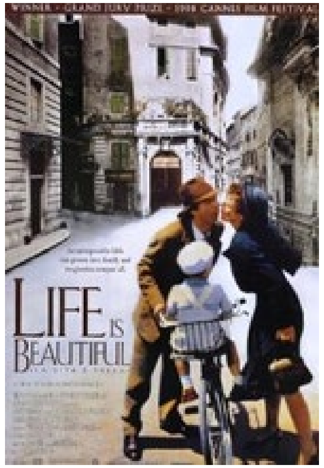
$P(E|F) = 0.20$

Independent!



$P(E) = 0.09$

$P(E|F) = 0.72$



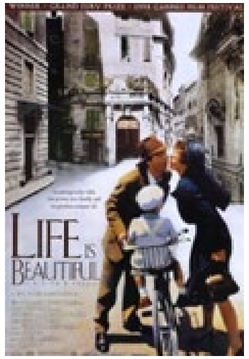
$P(E) = 0.20$

$P(E|F) = 0.42$

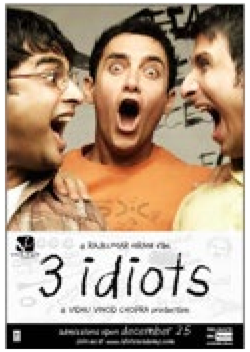
Netflix and Condition

Watched:

E_1

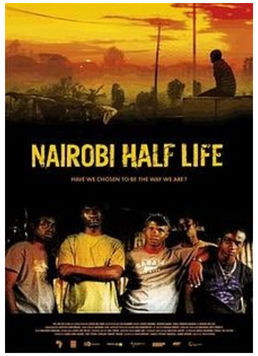


E_2



E_3

Will they watch?



E_4

What if $E_1 E_2 E_3 E_4$ are not independent? (e.g., all international emotional comedies)

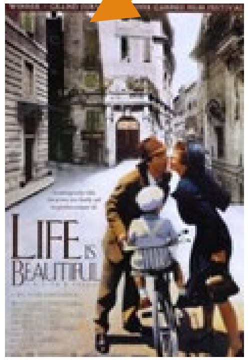
$$P(E_4 | E_1 E_2 E_3) = \frac{P(E_1 E_2 E_3 E_4)}{P(E_1 E_2 E_3)}$$

Netflix and Condition

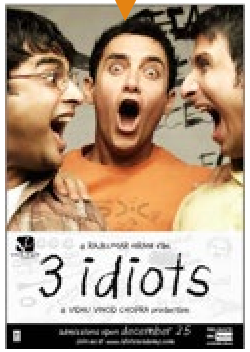
K: likes international emotional comedies

Watched:

E_1

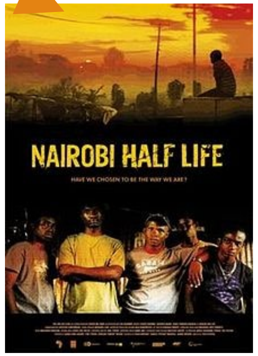


E_2



E_3

Will they watch?



E_4

What if $E_1 E_2 E_3 E_4$ are *conditionally independent* K?

$$P(E_4 | E_1 E_2 E_3) = \frac{P(E_1 E_2 E_3 E_4)}{P(E_1 E_2 E_3)}$$

$$P(E_4 | E_1 E_2 E_3 K) = P(E_4 | K)$$

Netflix and Condition

K: likes international emotional comedies

E_1 E_2 E_3 E_4

$E_1 E_2 E_3 E_4$ are dependent

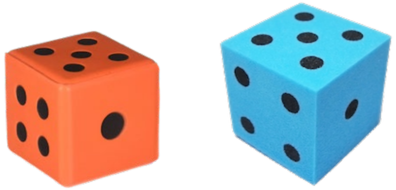
$E_1 E_2 E_3 E_4$ are conditionally independent given K

Dependent events can become conditionally independent.

Not-so-independent dice

Cond. independent E and F given G	⟺	$P(EF\|G) = P(E\|G)P(F\|G)$ $P(E\|FG) = P(E\|G)$

Roll two 6-sided dice, yielding values D_1 and D_2.



Let event E: $D_1 = 1$
event F: $D_2 = 6$
event G: $D_1 + D_2 = 7$

$G = \{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\}$

1. Are E and F independent?

$P(E) = 1/6$ $\quad P(F) = 1/6$ $\quad P(EF) = 1/36$

2. Are E and F independent given G?

Independent events can become conditionally dependent.

The beauty of conditional independence

Generalized Chain Rule:

$$P(E_1 E_2 E_3 \ldots E_n F) =$$

$$P(F)P(E_1|F)P(E_2|E_1 F)P(E_3|E_1 E_2 F) \ldots P(E_n|E_1 E_2 \ldots E_{n-1} F)$$

If $E_1, E_2, \ldots, E_n$ are all conditionally independent given F:

$$P(E_1 E_2 E_3 \ldots E_n F) = P(F)P(E_1|F)P(E_2|F) \cdots P(E_n|F)$$

More on this in a future lecture!

Conditional independence is a Big Deal

Conditional independence is a practical, real-world way of decomposing hard probability questions.

"Exploiting conditional independence to generate fast probabilistic computations is one of the main contributions CS has made to probability theory."

–Judea Pearl wins 2011 Turing Award,
"For fundamental contributions to artificial intelligence through the development of a calculus for probabilistic and causal reasoning"

Independence relationships can change with conditioning.

A and B independent	does NOT necessarily mean	A and B independent given E.

Today's plan

Conditional Independence

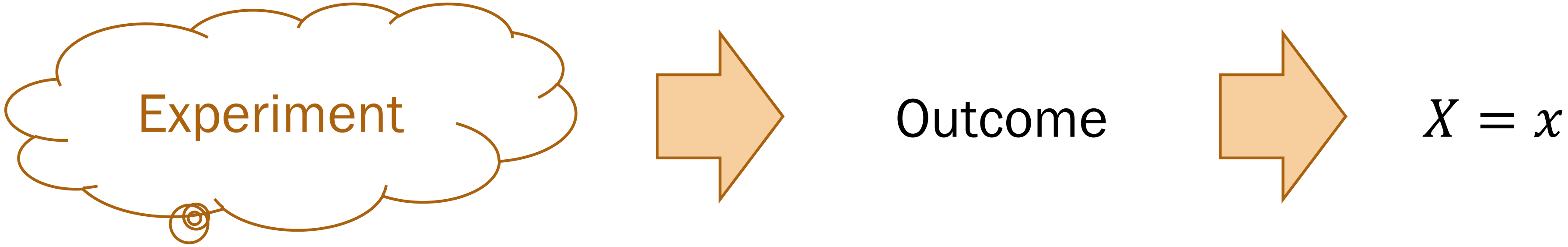
Random Variables

PMFs and CDFs

Expectation

Random Variable

A random variable is a real-valued function defined on a sample space.



Example:

3 coins are flipped.
Let X = # of heads.
X is a random variable.

1. What is the value of X for the outcomes:
- (T,T,T)?
- (H,H,T)?

2. What is the event (set of outcomes) where $X = 2$?

3. What is $P(X = 2)$?

Random variables are NOT events!

It is confusing that random variables and events use the same notation.

- **Random variables $\neq$ events.**
- We can define an event to be a particular assignment of a random variable.

Example:

3 coins are flipped.
Let X = # of heads.
X is a random variable.

$X = x$	$P(X = x)$	Set of outcomes	Possible event E
$X = \mathbf{0}$	1/8	{(T, T, T)}	Flip 0 heads
$X = \mathbf{1}$	3/8	{(H, T, T), (T, H, T), (T, T, H)}	Flip exactly 1 head
$X = \mathbf{2}$	3/8	{(H, H, T), (H, T, H), (T, H, H)}	The event where $X = \mathbf{2}$
$X = \mathbf{3}$	1/8	{(H, H, H)}	Flip 0 tails
$X \geq 4$	0	{ }	Flip 4 or more heads

Example random variable

Consider 5 flips of a coin which comes up heads with probability p.

- Each coin flip is an independent trial.
- Recall $P(2 \text{ heads}) = \binom{5}{2} p^2 (1-p)^3, \;\; P(3 \text{ heads}) = \binom{5}{3} p^3 (1-p)^2$

Let $Y = \#$ of heads on 5 flips.

1. What is the range of Y?
   In other words, what are the values that Y can take on with non-zero probability?

2. What is $P(Y = k)$, where k is in the range of Y?

Today's plan

Conditional Independence

Random Variables

PMFs and CDFs

Expectation

Probability Mass Function (PMF)

Y

random variable
(e.g., # of heads in
5 coin flips,
unbiased coin)

$Y = 2$

event

$P(Y = 2)$

probability
(number b/t 0 and 1)

variable

$P(Y = k)$

function on k with
range 0 and 1

Discrete RVs and Probability Mass Functions

A random variable X is discrete if its range has countably many values.

- $X = x$, where $x \in \{x_1, x_2, x_3, \ldots\}$

The probability mass function (PMF) of a discrete random variable is

$$P(X = x) = \underbrace{p(x)}_{\text{shorthand notation}} = \underbrace{p_X(x)}$$

shorthand notation

- Probabilities must sum to 1: $\sum_{i=1}^{\infty} p(x_i) = 1$

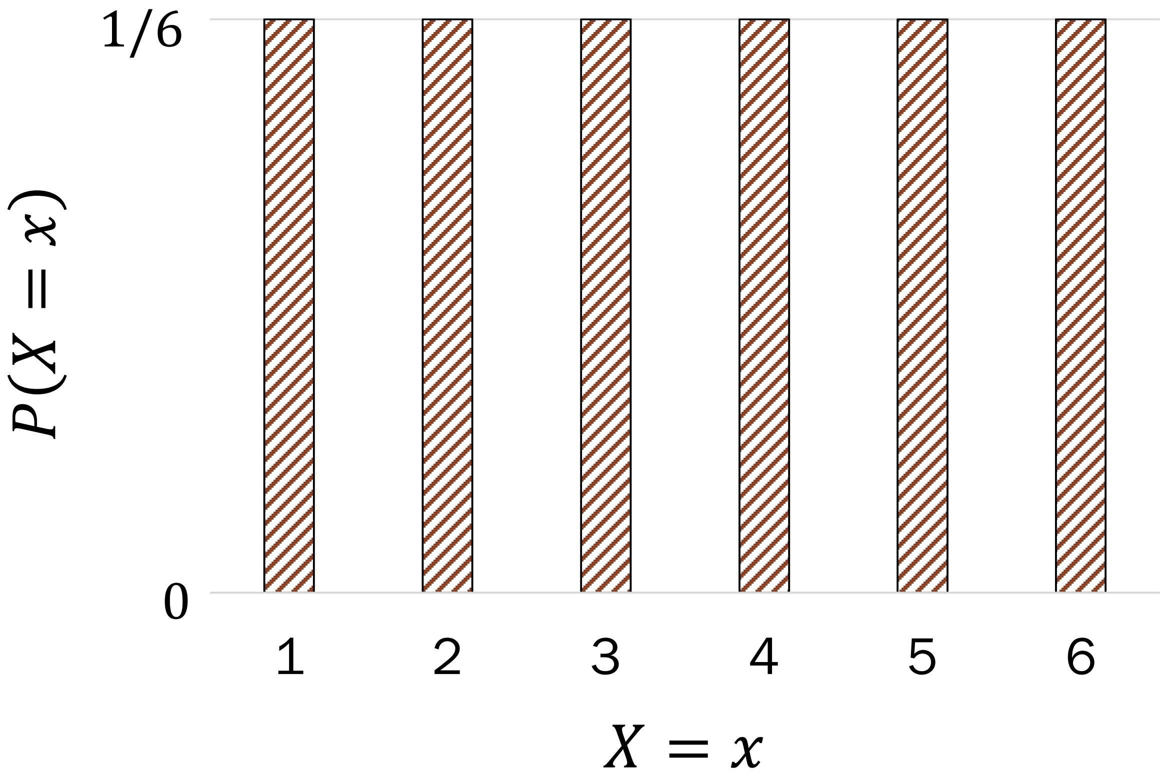
This last bullet is a good way to verify any PMF you create.

PMF for a single 6-sided die

Let X be a random variable that represents the result of a single dice roll.

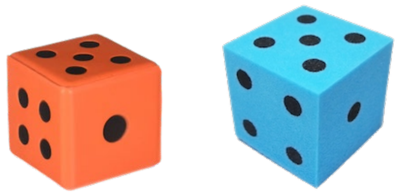
- Range of X : $\{1, 2, 3, 4, 5, 6\}$
- Therefore X is a discrete random variable.
- PMF of X:

$$p(x) = \begin{cases} 1/6 & x \in \{1, \ldots, 6\} \\ 0 & \text{otherwise} \end{cases}$$



PMF for the sum of two dice

Let Y be a random variable that represents the sum of two independent dice rolls.

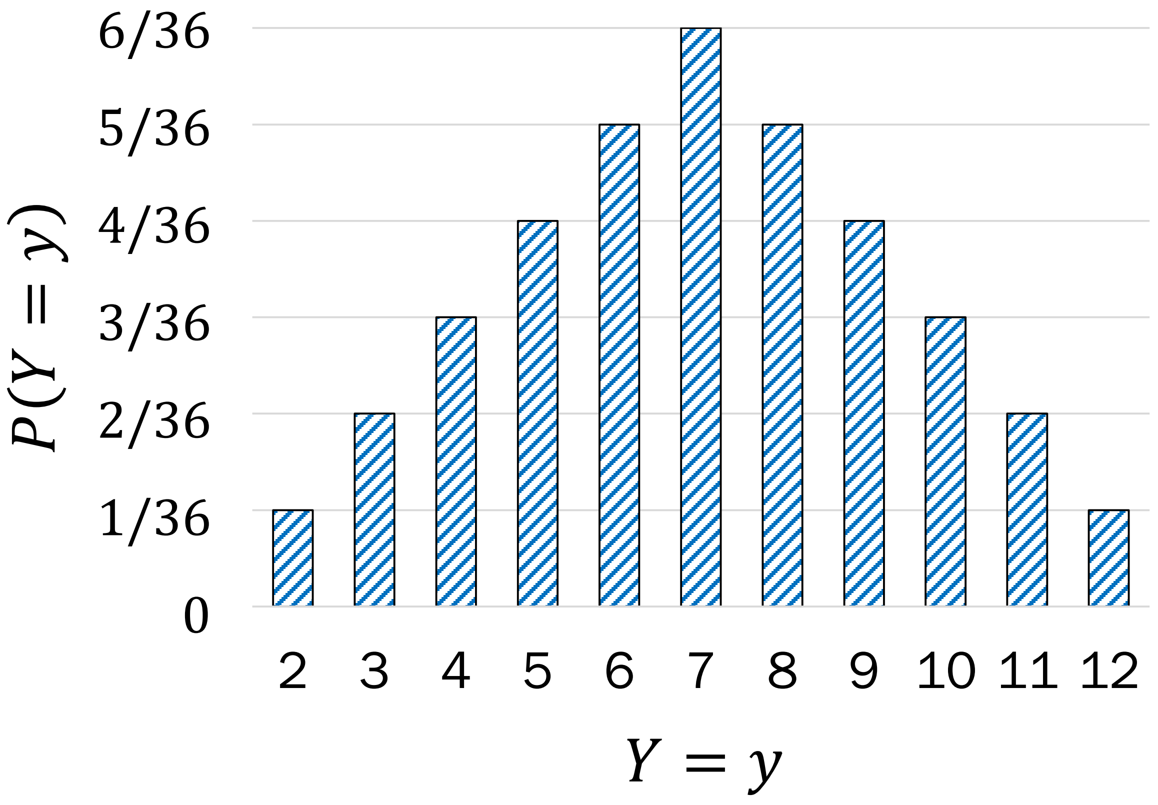


Range of Y: $\{2, 3, \ldots, 11, 12\}$

$$p(y) = \begin{cases} \frac{y-1}{36} & y \in \mathbb{Z}, 2 \le y \le 6 \\ \frac{13-y}{36} & y \in \mathbb{Z}, 7 \le y \le 12 \\ 0 & \text{otherwise} \end{cases}$$

Check:

$$\sum_{y=2}^{12} p(y) = 1$$



Announcements

Problem Set 1

Due:	an hour ago
On-time grades:	next Friday
Solutions:	next Friday

Problem Set 2

Out:	today
Due:	Monday 1/27
Covers:	through today

Cumulative Distribution Functions

For a random variable X, the cumulative distribution function (CDF) is defined as

$$F(a) = F_X(a) = P(X \leq a), \text{where } -\infty < a < \infty$$

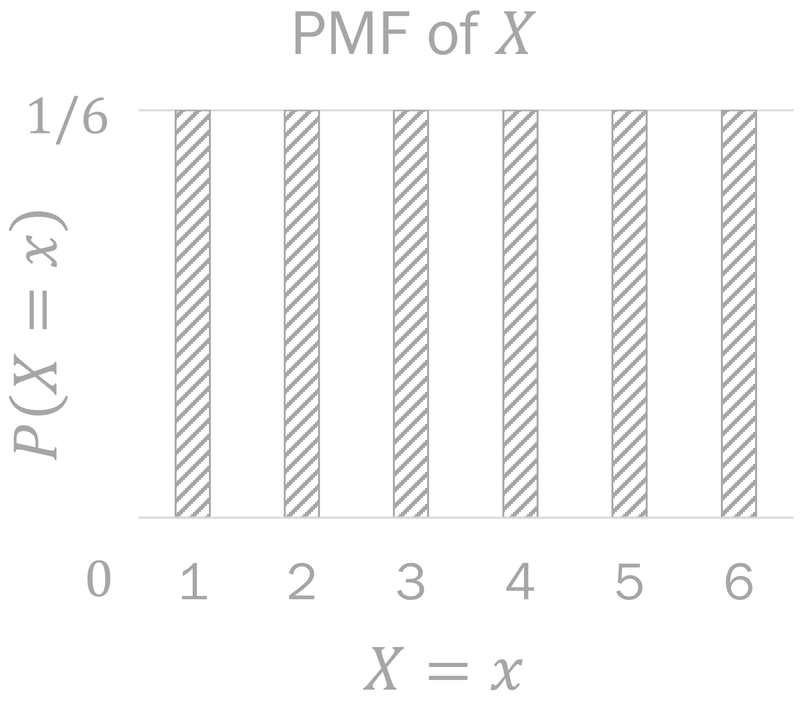
For a discrete RV X, the CDF is:

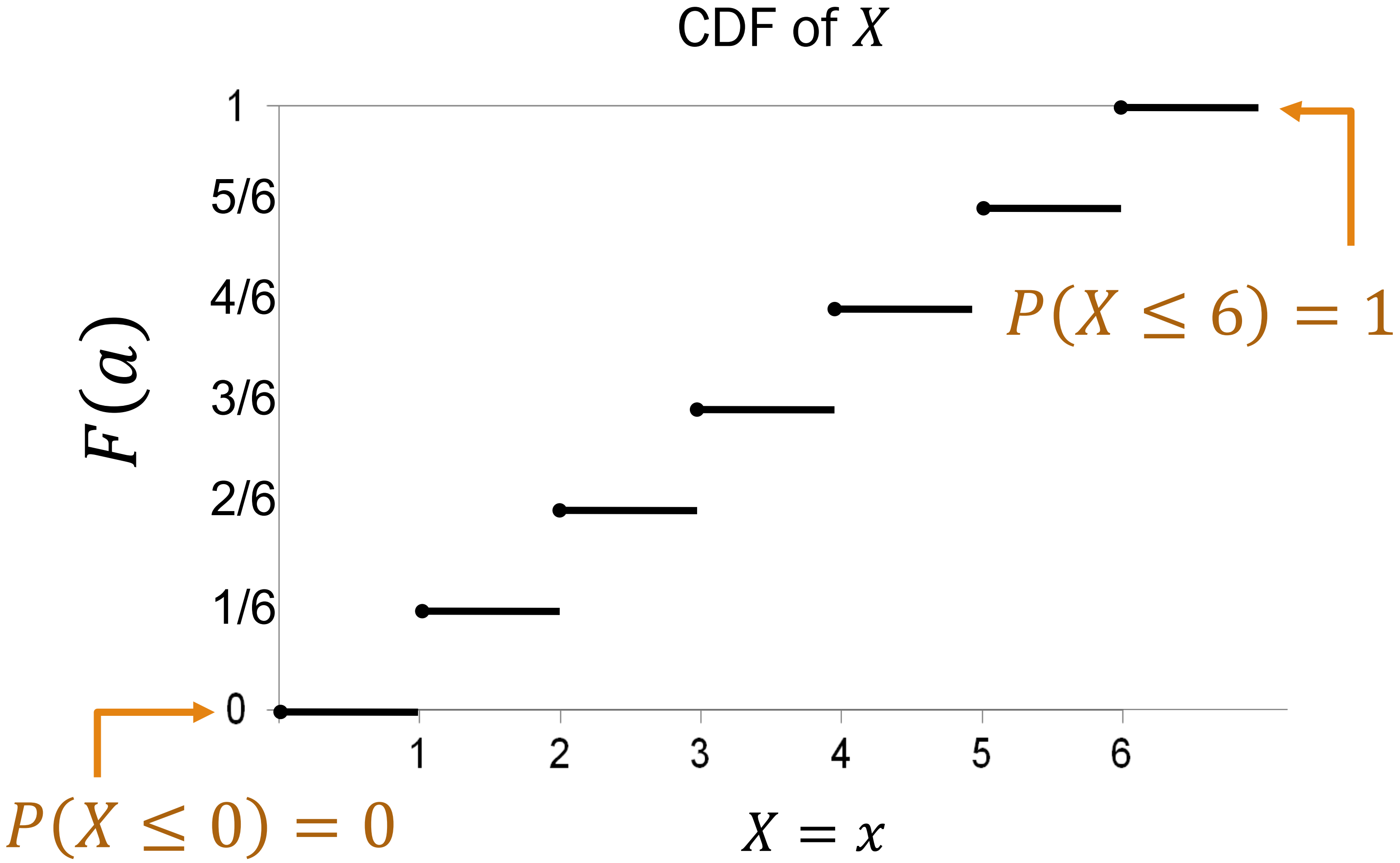
$$F(a) = P(X \leq a) = \sum_{\text{all } x \leq a} p(x)$$

CDFs as graphs

CDF (cumulative distribution function) $F(a) = P(X \le a)$

Let X be a random variable that represents the result of a single dice roll.





Today's plan

Conditional Independence

Random Variables

PMFs and CDFs

Expectation

Expectation

The expectation of a discrete random variable X is defined as:

$$E[X] = \sum_{x:p(x)>0} p(x) \cdot x$$

- Note: sum over all values of $X = x$ that have non-zero probability.
- Other names: **mean**, expected value, **weighted average**,
  center of mass, first moment

Expectation of a die roll

$$E[X] = \sum_{x:p(x)>0} p(x) \cdot x$$ Expectation of X

What is the expected value of a 6-sided die roll?

1. Define random variables

X = RV for value of roll

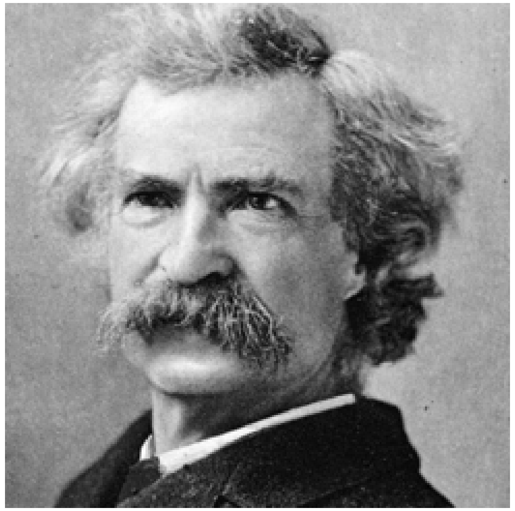
$$P(X = x) = \begin{cases} 1/6 & x \in \{1, \dots, 6\} \\ 0 & \text{otherwise} \end{cases}$$

2. Solve

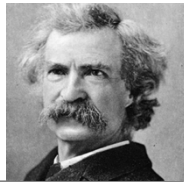
$$E[X] = 1\left(\frac{1}{6}\right) + 2\left(\frac{1}{6}\right) + 3\left(\frac{1}{6}\right) + 4\left(\frac{1}{6}\right) + 5\left(\frac{1}{6}\right) + 6\left(\frac{1}{6}\right) = \frac{7}{2}$$

Lying with statistics

"There are three kinds of lies:
lies, damned lies, and statistics"
–popularized by Mark Twain, 1906



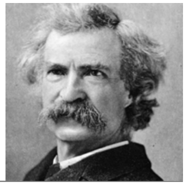
Lying with statistics



A school has 3 classes with 5, 10, and 150 students.

What is the average class size?

Lying with statistics



A school has 3 classes with 5, 10, and 150 students.
What is the average class size?

1. Interpretation #1
   - Randomly choose a class with equal probability.
   - X = size of chosen class

$$E[X] = 5\left(\frac{1}{3}\right) + 10\left(\frac{1}{3}\right) + 150\left(\frac{1}{3}\right)$$
$$= \frac{165}{3} = 55$$

2. Interpretation #2
   - Randomly choose a student with equal probability.
   - Y = size of chosen class

$$E[Y]$$
$$= 5\left(\frac{5}{165}\right) + 10\left(\frac{10}{165}\right) + 150\left(\frac{150}{165}\right)$$
$$= \frac{22635}{165} \approx 137$$

Important properties of expectation

1. Linearity:

$$E[aX + b] = aE[X] + b$$

- Let X = 6-sided dice roll, $Y = 2X - 1$.
- $E[X] = 3.5$
- $E[Y] = 6$

2. Expectation of a sum = sum of expectation:

$$E[X + Y] = E[X] + E[Y]$$

Sum of two dice rolls:

- Let X = roll of die 1
  Y = roll of die 2
- $E[X + Y] = 3.5 + 3.5 = 7$

3. Law of the unconscious statistician (LOTUS):

$$E[g(X)] = \sum_x g(x)p(x)$$

Being a statistician unconsciously

$E[g(X)] = \sum_x g(x)p(x)$ Expectation of $g(X)$

Let X be a discrete random variable.

- $P(X = x) = \frac{1}{3}$ for $x \in \{-1, 0, 1\}$

Let $Y = |X|$. What is $E[Y]$?