

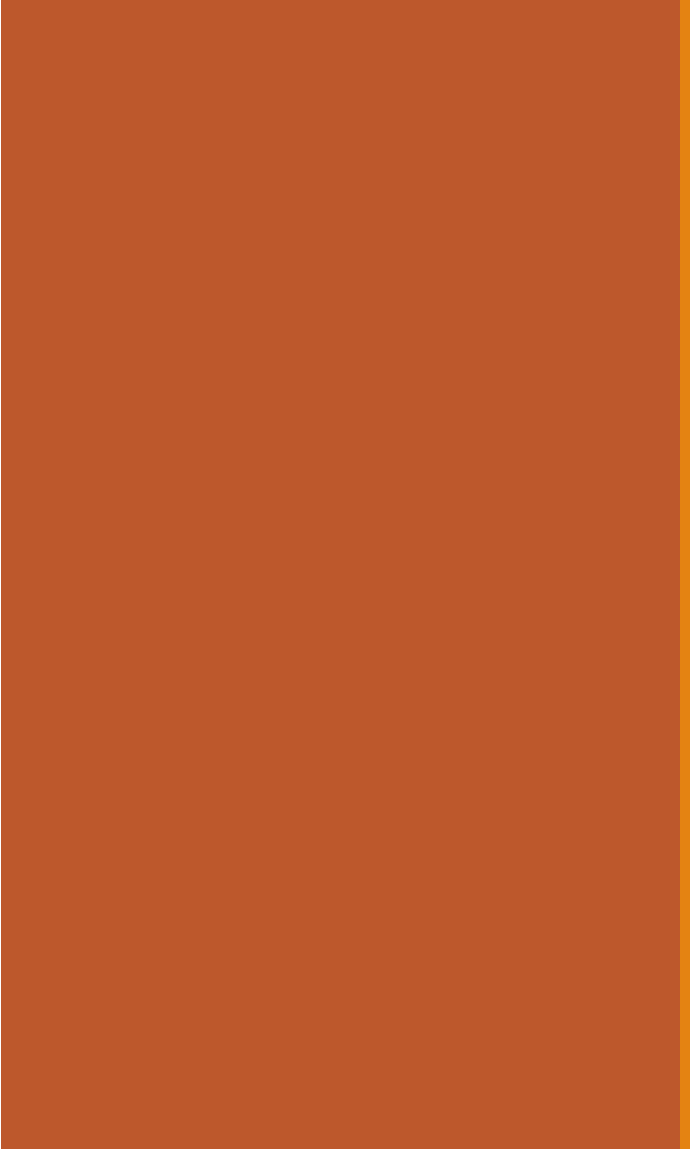
Table of Contents

2 Conditional Independence
12 Random Variables
18 PMF/CDF
26 Expectation
34 Exercises

o6: Random Variables

Jerry Cain
January 22, 2024

[Lecture Discussion on Ed](#)



Conditional Independence

Conditional Paradigm

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

*when in doubt, simply
assume that E is
really just the full
sample space S and
see if the formula
devolves into one
you learned during
Week 1 and 2.*

Axiom 1

$$0 \leq P(A|E) \leq 1$$

Corollary 1 (complement)

$$P(A|E) = 1 - P(A^c|E)$$

Transitivity

$$P(AB|E) = P(BA|E)$$

Chain Rule

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

Bayes' Theorem

$$P(A|BE) = \frac{P(B|AE)P(A|E)}{P(B|E)}$$



BAE 's theorem?

Conditional Independence

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

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



Conditional Independence

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

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:

each of these might be true, but only the third one conditions everything on E , so it's the only one that's equivalent to the above definition

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

E is the new sample space, so left and right side must both be conditioned on 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.



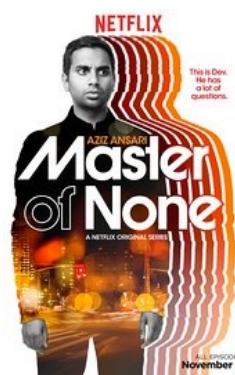
$$P(E) = 0.19$$

$$P(E|F) = 0.14$$



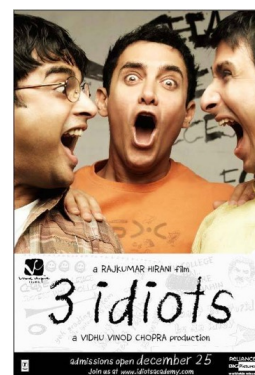
$$P(E) = 0.32$$

$$P(E|F) = 0.35$$



$$P(E) = 0.20$$

$$P(E|F) = 0.20$$



$$P(E) = 0.09$$

$$P(E|F) = 0.72$$



$$P(E) = 0.20$$

$$P(E|F) = 0.42$$

Lisa Yan, Chris Piech, Me **Independent!**, Winter 2024

Stanford University 7

Netflix and Condition (on many movies)

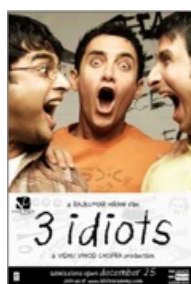
Watched:



E_1

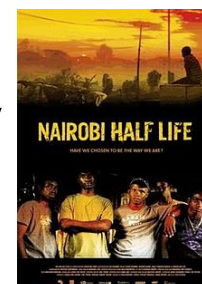


E_2



E_3

Will they
watch?



E_4

What if $E_1E_2E_3E_4$ are not independent? (e.g., all international emotional comedies)

$$P(E_4|E_1E_2E_3) = \frac{P(E_1E_2E_3E_4)}{P(E_1E_2E_3)} = \frac{\# \text{ people who have watched all 4}}{\# \text{ people who have watched those 3}}$$

We need to keep track of an exponential number of movie-watching statistics

Netflix and Condition (on many movies)



Assume: $E_1 E_2 E_3 E_4$ are conditionally independent given K
↪ simplifying assumption that works well in practice, even if it's not 100% accurate.

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

An easier probability to store and compute!

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 [[video of speech](#)],
“For fundamental contributions to artificial intelligence through the development of a calculus for probabilistic and causal reasoning”

Netflix and Condition

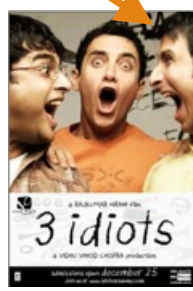
K : likes international emotional comedies



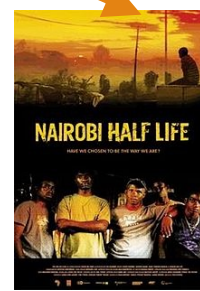
E_1



E_2



E_3



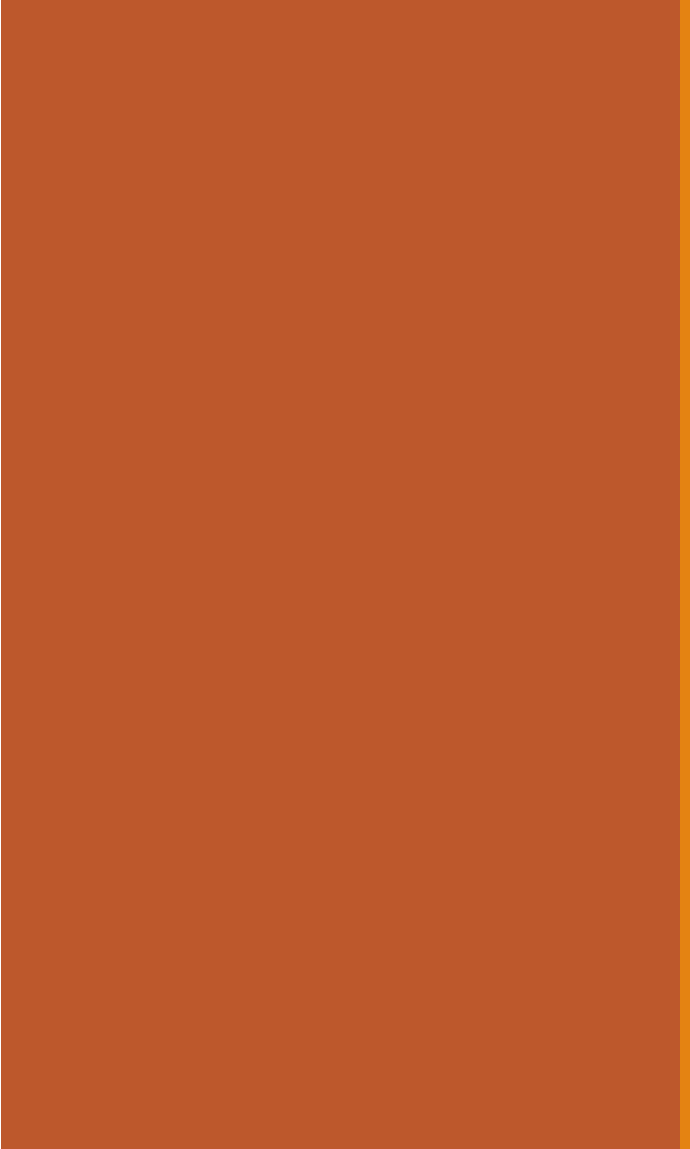
E_4

Challenge: How do we determine K ? Stay tuned in 6 weeks' time!

$E_1 E_2 E_3 E_4$ are
dependent

$E_1 E_2 E_3 E_4$ are
conditionally independent
given K

Dependent events can be conditionally independent.
(And vice versa: Independent events can be conditionally dependent.)



Random Variables

Random variables are like typed variables

type name value
`int` `a` = `5`;

`double` `b` = `4.2`;

`bit` `c` = `1`;

CS variables

A is the number of Pokemon we bring to our *future* battle.

$$A \in \{1, 2, \dots, 6\}$$



B is the amount of money we get *after* we win a battle.

$$B \in \mathbb{R}^+$$



C is 1 *if* we successfully beat the Elite Four. 0 otherwise.

$$C \in \{0, 1\}$$

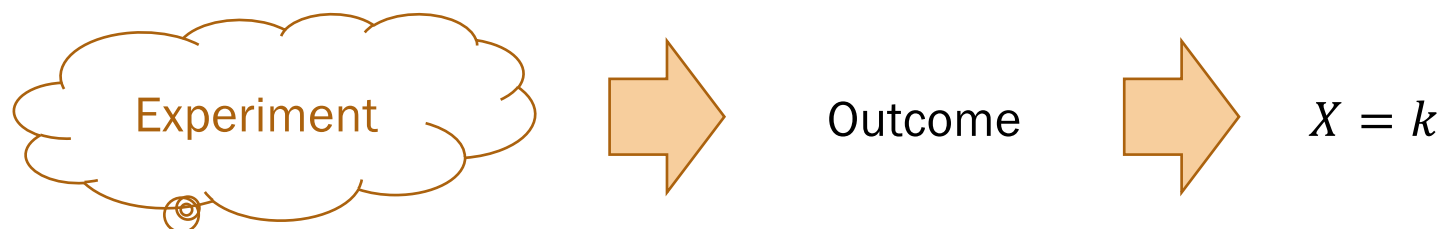


Random variables are like typed variables (with uncertainty)

Random
variables

Random Variable

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



Example:

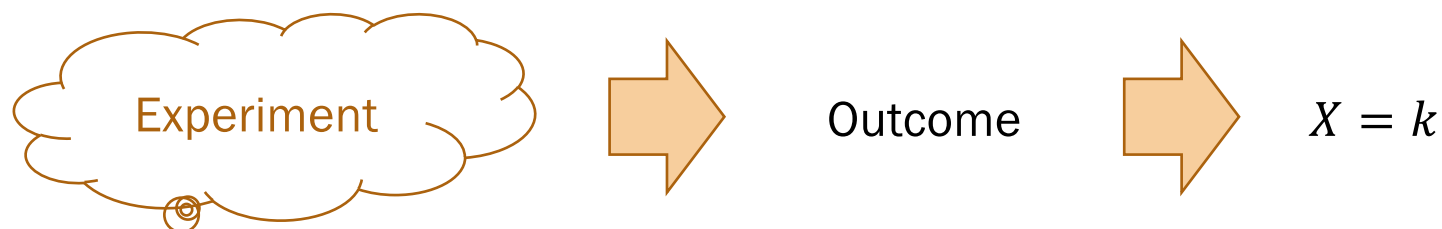
3 coins are flipped.
Let $X = \#$ of heads.
 X is a **random variable**.

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 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)? $X=0$
- (H,H,T)? $X=2$

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

$\{(H,H,T), (H,T,H), (T,H,H)\}$

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

$\frac{3}{8}$ $\leftarrow$ cardinality of this set
 $\leftarrow$ really 2^3

Random variables are **NOT** events!

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

- Random variables **≠** events.
- We can define an event to be a particular assignment of a random variable, or more generally, in terms of a random variable.

Example:

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

$X = 2$
event

$P(X = 2)$
probability
(number b/t 0 and 1)

Random variables are **NOT** events!

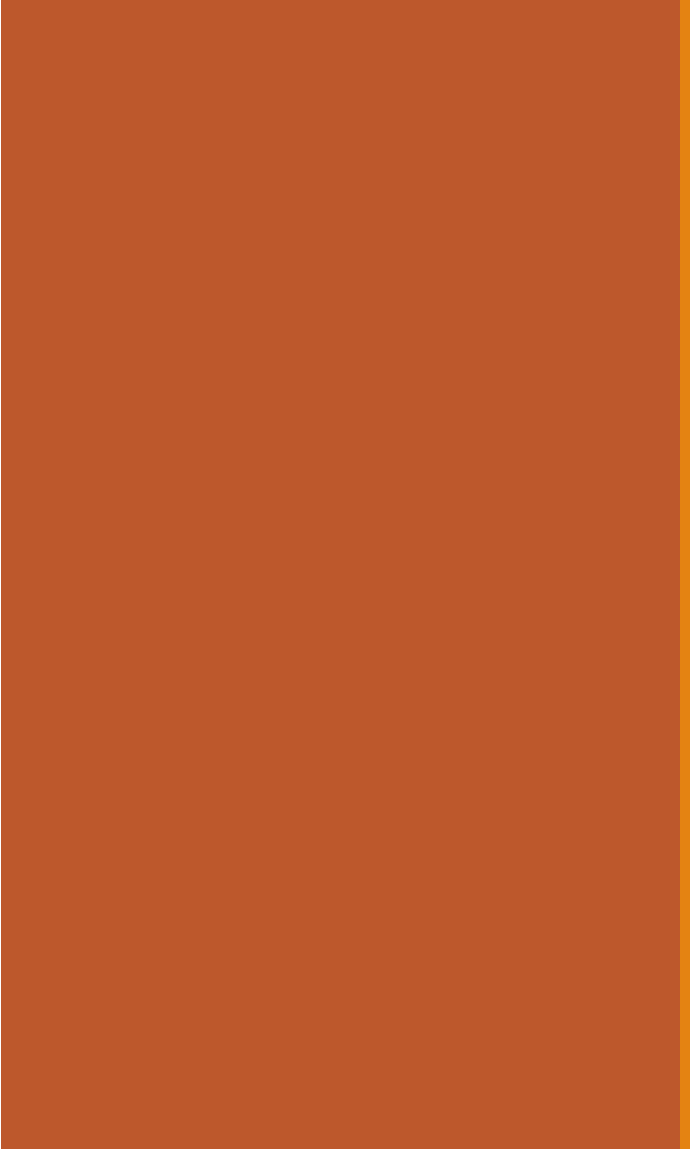
It is confusing that random variables and events use the same notation.

- Random variables **≠** events.
- We can define an event to be a particular assignment of a random variable, or more generally, in terms of a random variable.

	$X = x$	Set of outcomes	$P(X = k)$
Example: 3 coins are flipped. Let $X = \#$ of heads. X is a random variable.	$X = 0$	$\{(T, T, T)\}$	$1/8$
	$X = 1$	$\{(H, T, T), (T, H, T), (T, T, H)\}$	$3/8$
	$X = 2$	$\{(H, H, T), (H, T, H), (T, H, H)\}$	$3/8$
	$X = 3$	$\{(H, H, H)\}$	$1/8$
	$X \geq 4$	$\{\}$	0

← we just computed this a few slides back

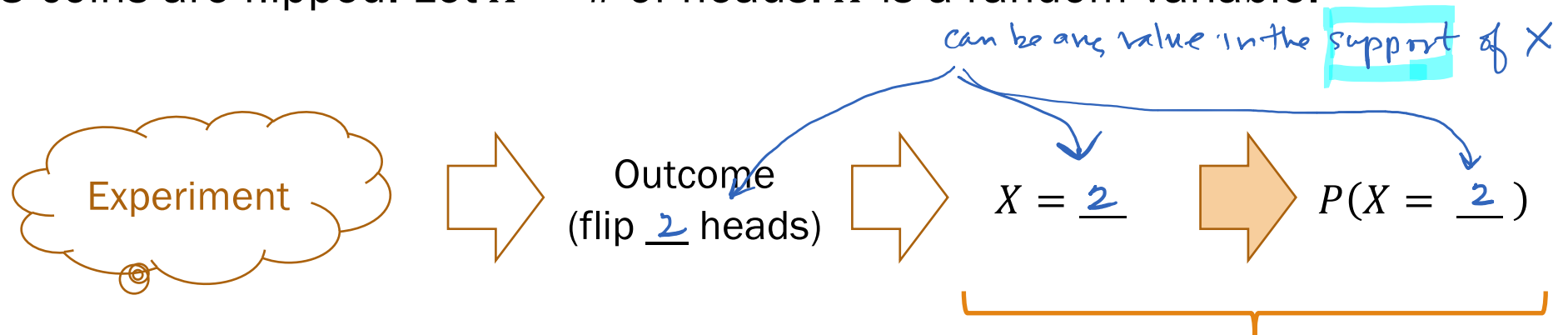
← not imperative to include this.



PMF/CDF

So far

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

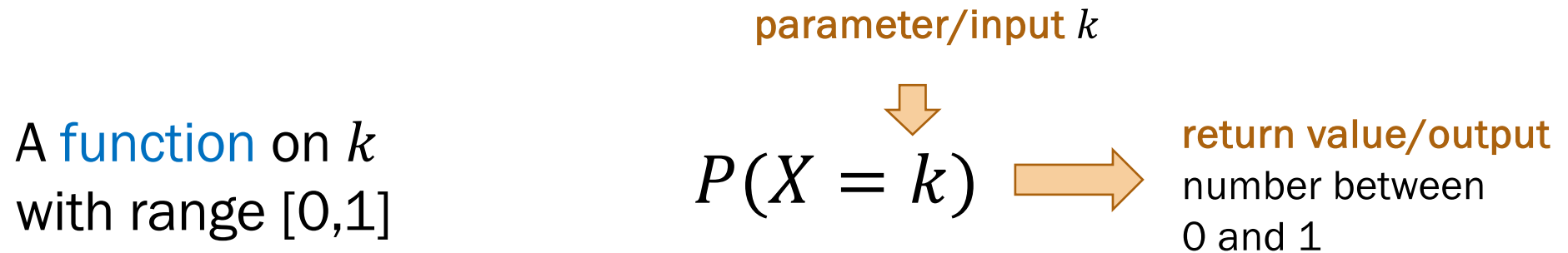


$X = x$	$P(X = k)$	Set of outcomes
$X = 0$	1/8	{(T, T, T)}
$X = 1$	3/8	{(H, T, T), (T, H, T), (T, T, H)}
$X = 2$	3/8	{(H, H, T), (H, T, H), (T, H, H)}
$X = 3$	1/8	{(H, H, H)}
$X \geq 4$	0	{ }

Can we get a "shorthand" for this last step?
Seems like it might be useful!

Probability Mass Function

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



What would be a *useful* function to define?

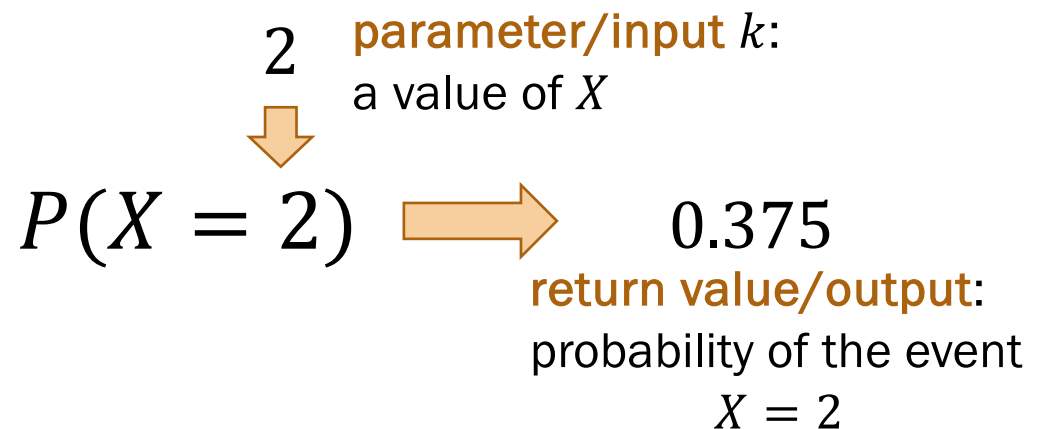
The probability of the event that a random variable X takes on the value k !

For **discrete random variables**, this is a **probability mass function**.

Probability Mass Function

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

A function on k
with range $[0,1]$



```
N = 3
P = 0.5
def prob_event_y_equals(k):
    n_ways = scipy.special.binom(N, k)
    p_way = np.power(P, k) * np.power(1 - P, N-k)
    return n_ways * p_way
```

probability mass function

seems like
 $p(X=k)$
 $= \binom{3}{k} 0.5^3$
for all reasonable
values of k .

Discrete RVs and Probability Mass Functions

A random variable X is **discrete** if it can take on countably many values.

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

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

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

- Probabilities must sum to 1:

$$\sum_{i=1}^{\infty} p(x_i) = 1$$

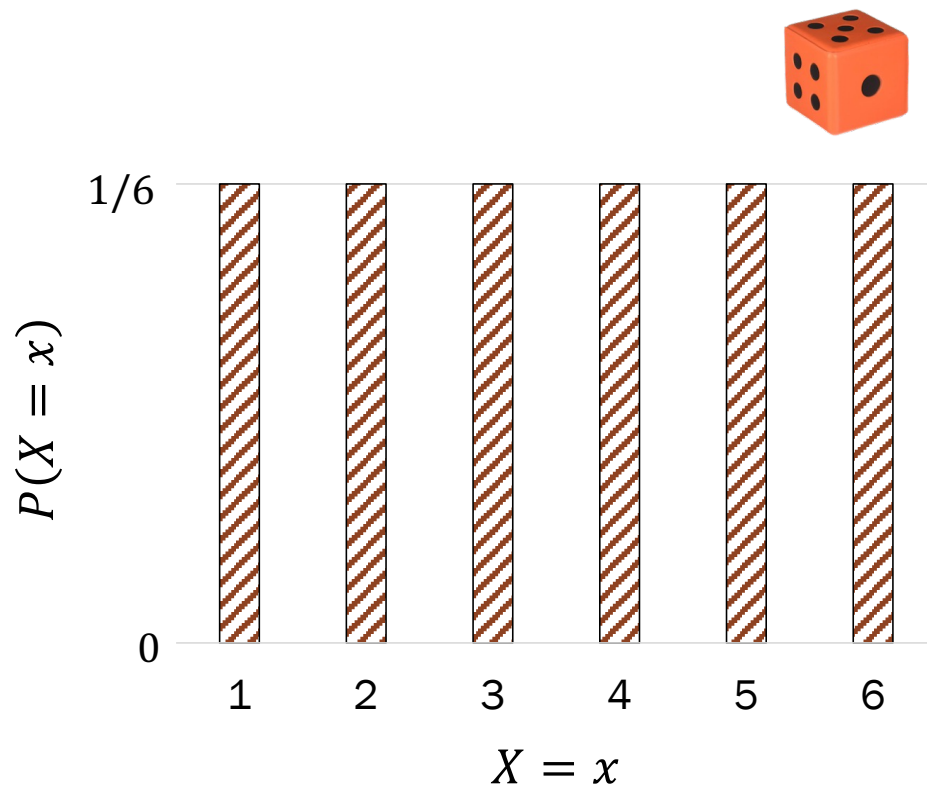
This last point is a good way to verify any PMF you create is valid

PMF for a single 6-sided die

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

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

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



Cumulative Distribution Functions

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

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

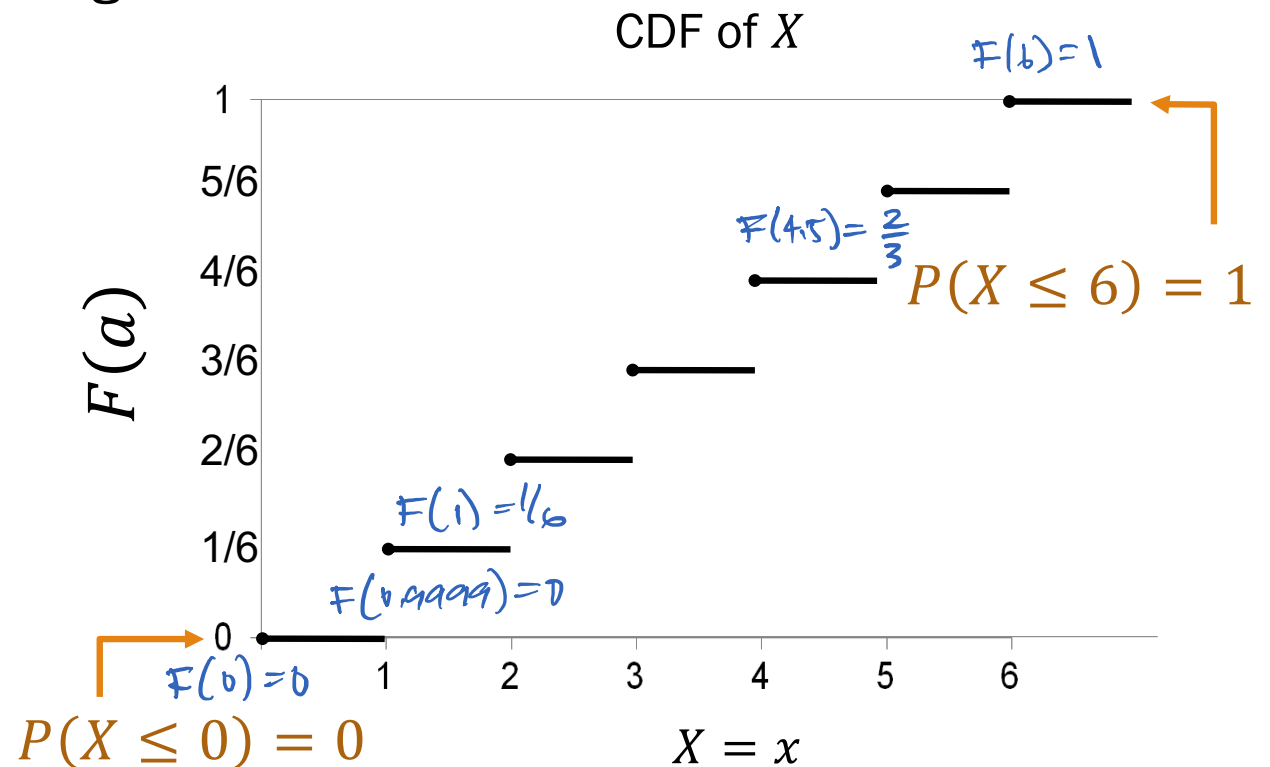
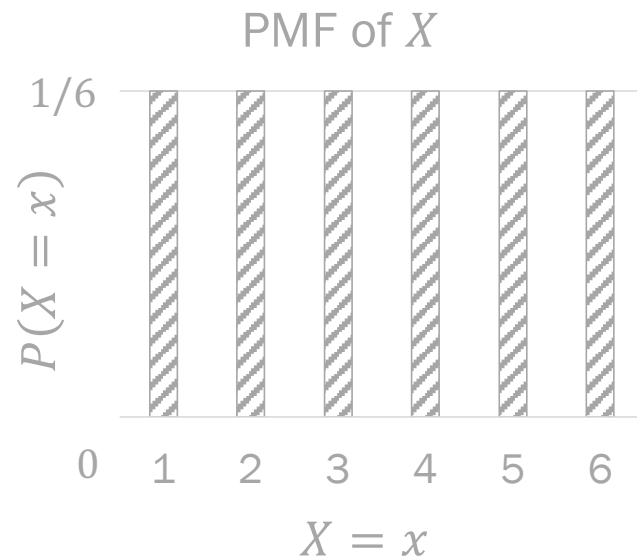
For a discrete RV X , the CDF is:

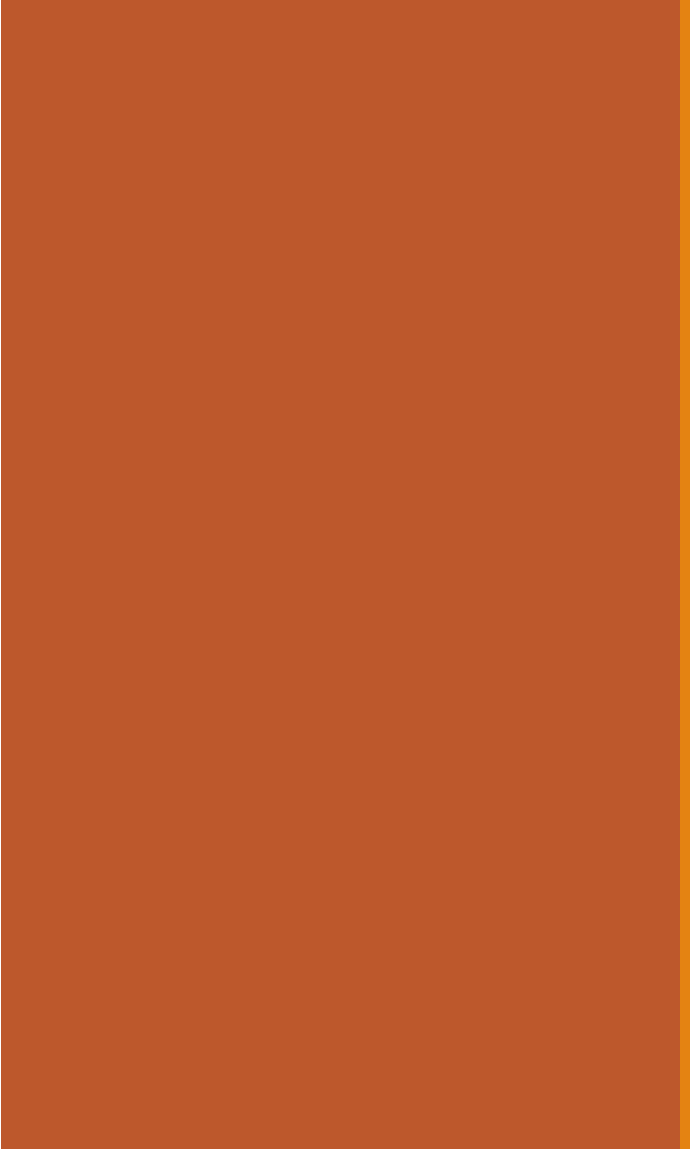
$$F(a) = P(X \leq a) = \sum_{\text{all } x \leq a} p(x)$$

CDFs as graphs

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

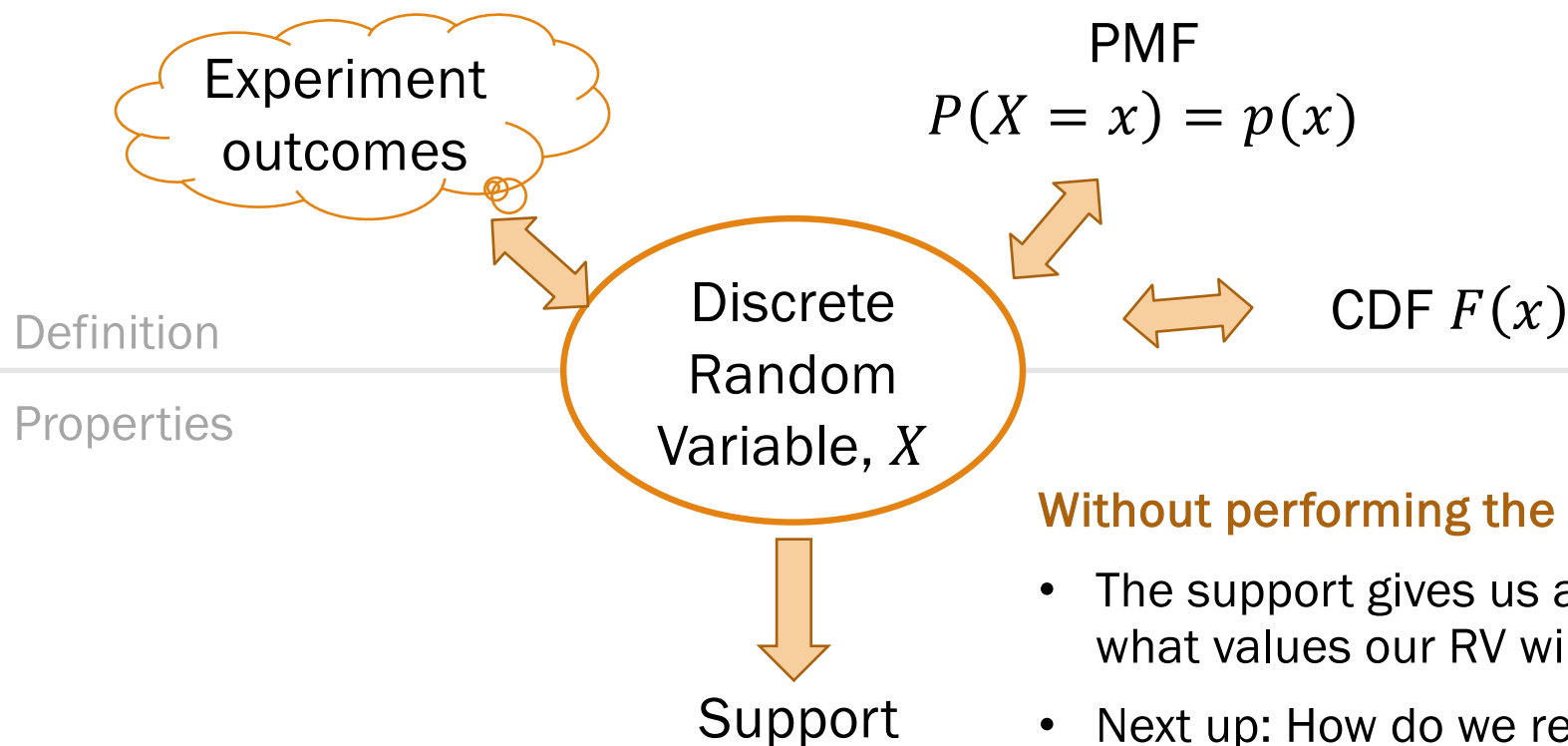
Let X be a random variable that represents the result of a single dice roll.





Expectation

Discrete random variables



Without performing the experiment:

- The support gives us a ballpark of what values our RV will take on
- Next up: How do we report the "average" value?

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

*I use these all the time, and
use them interchangeably.*

Expectation of a die roll

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



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

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

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. Unconscious statistician:

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

These properties let you avoid defining difficult PMFs.

Linearity of Expectation proof

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

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

Proof:

$$\begin{aligned} E[aX + b] &= \sum_x (ax + b)p(x) = \sum_x axp(x) + bp(x) \\ &= a \sum_x xp(x) + b \sum_x p(x) \\ &= a E[X] + b \cdot 1 \end{aligned}$$

Expectation of Sum intuition

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

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

we'll prove this in a few lectures

Intuition
for now:

X	Y	$X + Y$
3	6	9
2	4	6
6	12	18
10	20	30
-1	-2	-3
0	0	0
8	16	24

Average:

$$\frac{1}{n} \sum_{i=1}^n x_i + \frac{1}{n} \sum_{i=1}^n y_i = \frac{1}{n} \sum_{i=1}^n (x_i + y_i)$$
$$\frac{1}{7} (28) + \frac{1}{7} (56) = \frac{1}{7} (84)$$

LOTUS proof

$$E[g(X)] = \sum_x g(x)p(x) \quad \text{Expectation of } g(X)$$

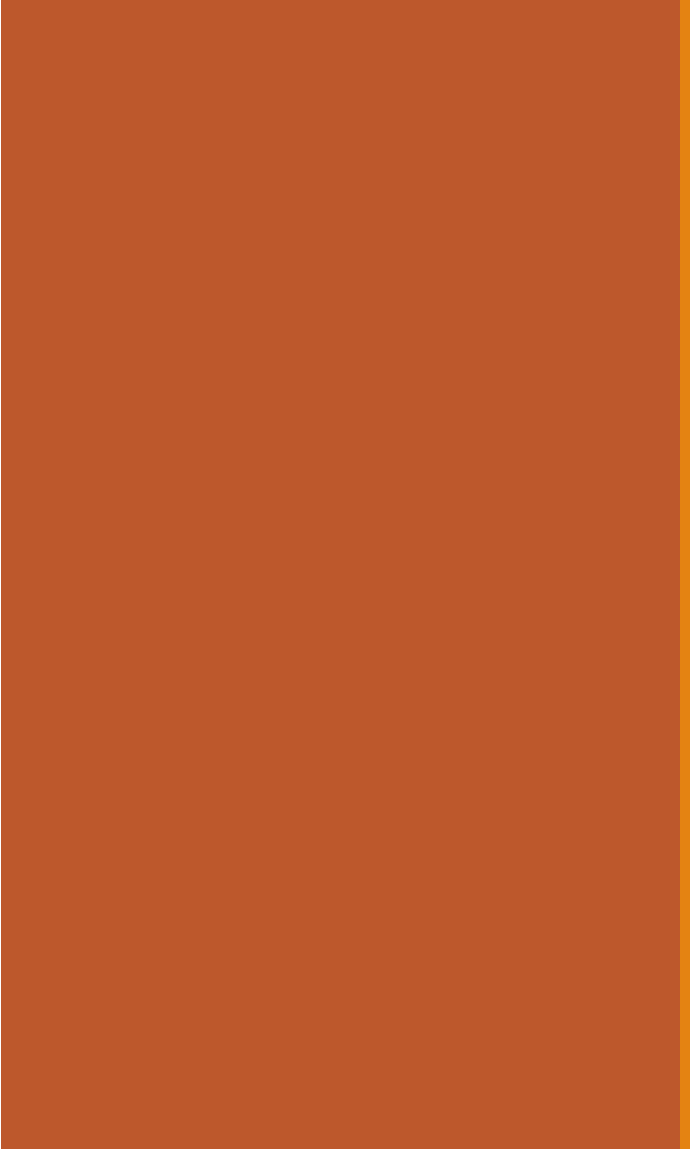
Let $Y = g(X)$, where g is a real-valued function.

$$\begin{aligned} E[g(X)] &= E[Y] = \sum_j y_j p(y_j) \\ &= \sum_j y_j \sum_{i: g(x_i) = y_j} p(x_i) \\ &= \sum_j \sum_{i: g(x_i) = y_j} y_j p(x_i) \\ &= \sum_j \sum_{i: g(x_i) = y_j} \boxed{g(x_i) p(x_i)} \\ &= \sum_i g(x_i) p(x_i) \end{aligned}$$

assume that $p(y_j) = \sum_{i: \text{such that } g(x_i) = y_j} p(x_i)$

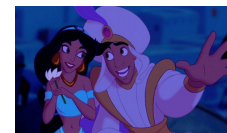
no dependence on j

For you to review
so that you can
sleep tonight!



Exercises

A Whole New World with Random Variables



Event-driven probability

- Relate only binary events
 - Either something happens (E)
 - or it doesn't happen (E^C)
- Can only report probability
- Lots of combinatorics



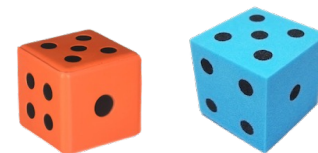
Random Variables

- Link multiple similar events together ($X = 1, X = 2, \dots, X = 6$)
- Can compute statistics: report the "average" outcome
- Once we have the PMF (for discrete RVs), we can do regular math



PMF for the sum of two dice

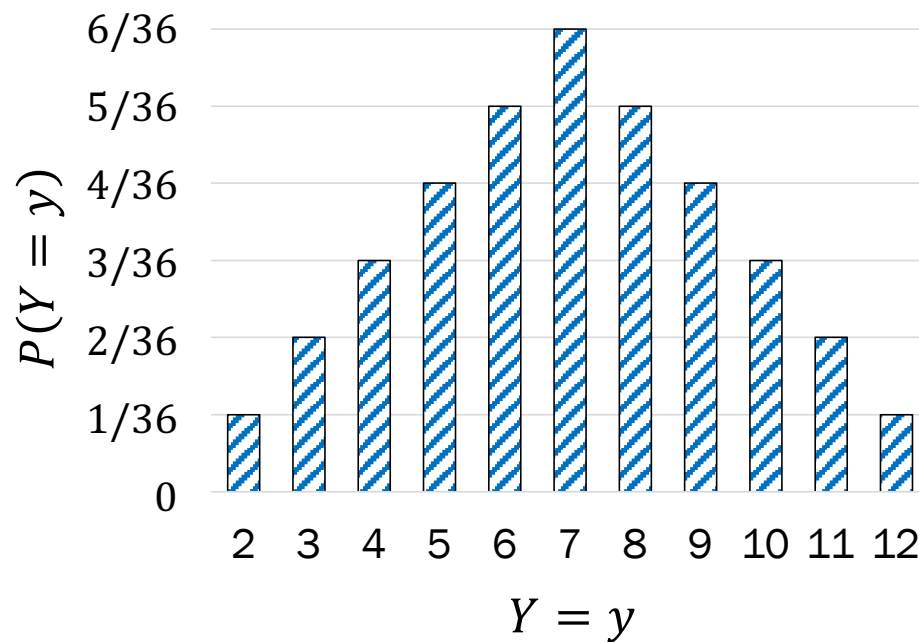
Let Y be a random variable that represents the sum of two independent dice rolls.



Support of Y : $\{2, 3, \dots, 11, 12\}$

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

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



Example random variable

Consider 5 flips of a coin which comes up heads with probability p . Each coin flip is an independent trial. **Let $Y = \#$ of heads on 5 flips.**

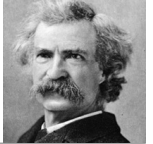
1. What is the **support** of Y ? In other words, what are the values that Y can take on with non-zero probability?
2. Define the event $Y = 2$. What is $P(Y = 2)$?
3. What is the PMF of Y ? In other words, what is $P(Y = k)$, for k in the support of Y ?



Example random variable

Consider 5 flips of a coin which comes up heads with probability p . Each coin flip is an independent trial. **Let $Y = \#$ of heads on 5 flips.**

1. What is the **support** of Y ? In other words, what are the values that Y can take on with non-zero probability? $\{0, 1, 2, 3, 4, 5\}$
2. Define the event $Y = 2$. What is $P(Y = 2)$? $P(Y = 2) = \binom{5}{2} p^2 (1 - p)^3$
3. What is the PMF of Y ? In other words, what is $P(Y = k)$, for k in the support of Y ? $P(Y = k) = \binom{5}{k} p^k (1 - p)^{5-k}$



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

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

2. Interpretation #2

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

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

What alumni relations usually reports

Average student perception of class size

Being a statistician unconsciously

$$E[g(X)] = \sum_x g(x)p(x) \quad \text{Expectation of } g(X)$$

Let X be a discrete random variable.

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

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

A. $\frac{1}{3} \cdot 1 + \frac{1}{3} \cdot 0 + \frac{1}{3} \cdot -1 = 0$

B. $E[Y] = E[0] = 0$

C. $\frac{1}{3} \cdot 0 + \frac{2}{3} \cdot 1 = \frac{2}{3}$

D. $\frac{1}{3} \cdot |-1| + \frac{1}{3} \cdot |0| + \frac{1}{3} |1| = \frac{2}{3}$

E. C and D



Being a statistician unconsciously

$$E[g(X)] = \sum_x g(x)p(x) \quad \text{Expectation of } g(X)$$

Let X be a discrete random variable.

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

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

A. $\frac{1}{3} \cdot 1 + \frac{1}{3} \cdot 0 + \frac{1}{3} \cdot -1 = 0$ ✗ $E[X]$

B. $E[Y] = E[0] = 0$ ✗ $E[E[X]]$

C. $\frac{1}{3} \cdot 0 + \frac{2}{3} \cdot 1 = \frac{2}{3}$

D. $\frac{1}{3} \cdot |-1| + \frac{1}{3} \cdot |0| + \frac{1}{3} \cdot |1| = \frac{2}{3}$

E. C and D

- 1. Find PMF of Y : $p_Y(0) = \frac{1}{3}, p_Y(1) = \frac{2}{3}$
- 2. Compute $E[Y]$

Use LOTUS by using PMF of X :

- 1. $P(X = x) \cdot |x|$
- 2. Sum up