CS1800 day 12 (we'll start @ 13:37, want to ensure everyone hears announcements)

SECTION 4: REMIND ME TO SHARE SCREEN PLEASE

Admin:

- exam1 & hw4 graded to you next week
    - grade estimate pushed to canvas 1 week after exam1 released (likely sooner)
    - talk about how to make this course more productive / fun for you? grab an OH appointment with me (monday afternoons)
- hw5 released next friday
    enjoy the break from hw :)

Content:

- Probability definitions (random variable, outcome, distribution)
- Computing prob of event from equal prob outcomes
- Expected Value
- Variance

a problem:

objective: predict the outcome of a coin flip

reward for correct: fleeting satisfaction of having been correct once

how might we approach this problem?

a (similar?) problem:

objective: predict the outcome of a coin flip

reward for correct: world peace, universal happiness and calorie free cake (that tastes just as good!)

how might we approach this problem? (how is it different from the previous problem)

Why study probability?

Probability allows us to build simple, effective models to make predictions in complex situations, (avoiding modelling how something really works with all its complexities!)

ChatGPT: a glorified "next-word" prediction

- how common is "dog" if the preceding N words were:
  "the quick brown fox jumps over the lazy ..."

Netflix reccomendation:

- among all people who like similar movies as you, what are popular movies which they've rated highly which you haven't seen?

Self driving cars:

- among all the times I've been in a similar position on the road, how often does this car turn right without signalling?

Probability: intro definitions

"Experiment" - the thing we're trying to model

COIN FLIP

DIE ROLL

Outcome (of an experiment) - a particular result of the experiment

HEADS

3

Sample space (of an experiment) - the set of all possible outcomes

$S = \{ \text{HEADS}, \text{TAILS} \}$

$S = \{1, 2, 3, 4, 5, 6\}$

Distribution (of an experiment) - the probablility of each outcome

HEADS	TAILS
50%	50%

1	2	3	4	5	6
1/6	1/6	1/6	1/6	1/6	1/6

Probability: Notation

The weather tomorrow is going to be:

SUN

50%

CLOUD

30%

FOG

20%

Probability: Notation

The weather tomorrow is going to be:

$w_0 =$ SUN

$P(W = w_0) = 50\%$

$w_1 =$ CLOUD

$P(W = w_1) = 30\%$

$w_2 =$ FOG

$P(W = w_2) = 20\%$

capital W is a random variable, it represents an undecided experiment (no particular outcome yet assigned)

each lowercase w_i is a particular outcome, the result of the experiment

convention: capitals for Random Variables, lowercase-same-letter-with-index for outcomes

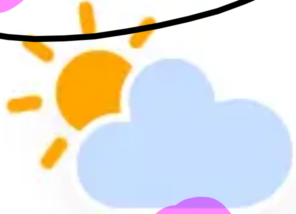
Probability: Notation (another common convention)

The weather tomorrow is going to be:

0 = SUN

$P(W=0) = 50\%$

1 = CLOUD



$P(W=1) = 30\%$

2 = FOG

$P(W=2) = 20\%$

capital W is a random variable, it represents an undecided experiment (no particular outcome yet assigned)

each lowercase w_i is a particular outcome, the result of the experiment

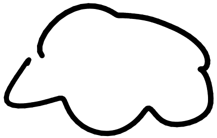
convention: capitals for Random Variables, some natural number for outcomes

(I don't like this: unclear association of outcomes to random variables with multiple random variables...)

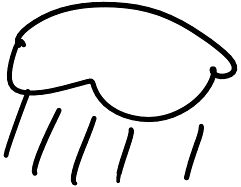
Probability: axioms (necessary conventions, kind of like a definition)

Axiom 1. Probability is positive

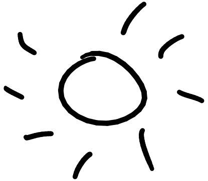
"Axiom 2 & 3-ish". The sum of the probability of all outcomes in the sample space is 1



P(CLOUDY) = 40%



P(RAINY) = 20%



P(SUN) = 40%

Uniform Distribution

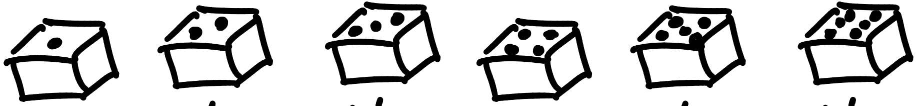
Assigns equal probability to all outcomes in the sample space

FAIR COIN

H　　T

50%　　50%

FAIR DIE



1/6　1/6　1/6　1/6　1/6　1/6

$$P(x) = 1/|S| \leftarrow$$

S is sample space

|S| is # elements in S

Event

Experiment:

a player rolls two six-sided die and moves this many spaces. if they start from "just visiting", where do they land?

$$S = \{2, 3, 4, \ldots, 10, 11, 12\}$$

Event: a subset of the sample space

Event "lands on an orange property"

$$E = \{6, 8, 9\}$$

$$E \subseteq S$$

START

Computing event probabilities (from a uniform distribution of outcomes)

$$P(\text{EVENT}) = \frac{\#\text{ ELEMENTS IN EVENT}}{\#\text{ ELEMENTS IN SAMPLE SPACE}}$$

EXAMPLE: APPLY TO "ORANGE PROPERTIES" EXAMPLE (STATE ANY ASSUMPTIONS)

⚠ 2 DIE SUM IS NOT UNIFORM

$$P(E) = \frac{|E|}{|S|} = \frac{3}{11}$$

In Class Activity

GIVEN A <u>FAIR</u> 6 SIDED DIE COMPUTE PROB OF EACH EVENT

X = ROLL A 1

$X = \{1\}$

$$P(X) = \frac{|X|}{|S|} = \frac{1}{6}$$

Y = ROLL AN EVEN #

$Y = \{2, 4, 6\}$

$$P(Y) = \frac{|Y|}{|S|} = \frac{3}{6}$$

Z = ROLL A PRIME #

$Z = \{2, 3, 5\}$

$$P(Z) = \frac{|Z|}{|S|} = \frac{3}{6}$$

$\cancel{P(Y\ Y_{even})}$

RANDOM VARIABLE

LET D_i BE OUTCOME OF FAIR DIE

4-SIDED ROLL

RANDOM VARIABLES

1ST 4-SIDED DIE

2ND 4-SIDE DIE

$$X = D_1 + D_2$$

SUM OF TWO 4-SIDED DIE ROLLS

Simulate 2 four sided die:

https://www.gigacalculator.com/randomizers/random-dice-roller.php

OUTCOME COUNTER (HISTOGRAM)

OUTCOME	2	3	4	5	6	7	8
# TIMES OBSERVED	0	4	12	5	10	4	1
PROB (SEE NEXT)	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{3}{16}$	$\frac{4}{16}$	$\frac{3}{16}$	$\frac{2}{16}$	$\frac{1}{16}$

WHAT IS DISTRIBUTION OF $X = D_1 + D_2$

$S = \{1\ 2\ 3\ 4\}$

SAMPLE SPACE OF D_i

EACH OF THESE IN $S \times S$ IS EQUALLY LIKELY

$$S \times S = \left\{ \begin{matrix} (1,1) & (1,2) & (1,3) & (1,4) \\ (2,1) & (2,2) & (2,3) & (2,4) \\ (3,1) & (3,2) & (3,3) & (3,4) \\ (4,1) & (4,2) & (4,3) & (4,4) \end{matrix} \right\}$$

$P(X=2) = \frac{1}{16}$

$P(X=3) = \frac{2}{16}$

$P(X=4) = \frac{3}{16}$

$P(X=5) = \frac{4}{16}$

$P(X=6) = 3/16$

$P(X=7) = 2/16$

$P(X=8) = 1/16$

Expected Value

Expected value is an "average" outcome of a random variable

$S = \{2, 0\}$

"Double Lotto"

$P(w)$	w
$1/2$	\$2
$1/2$	\$0

Half time win \$2

Half time 'win' \$0

EXPECTED VALUE

EXPECTED VALUE IS AN "AVERAGE" OUTCOME OF A RANDOM VARIABLE

"DOUBLE LOTTO"

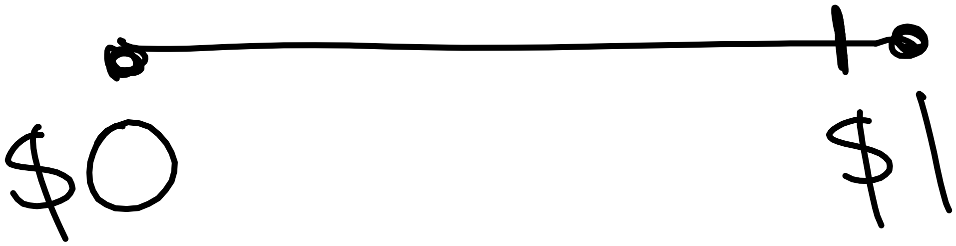
$S = \{2, 0\}$

$P(w)$	w
$1/2$	\$2
$1/2$	\$0

$E[x] = \$1 = 2 \cdot 1/2 + 0 \cdot 1/2$

$P(x)$	X
$1/1000$	\$0
$\frac{999}{1000}$	\$1

$$\frac{0+1}{2}$$



Expected Value: Computation

Intuition: multiply every outcome by its corresponding probability, add up all results

Suppose X has sample space $\{-1, 100, 4\}$

$$E[X] = -1 \cdot P(X=-1) + 100 \cdot P(X=100) + 4 \cdot P(X=4)$$

$$= \sum_{x \in S} x \cdot P(x)$$

Inner term

Compute inner term for all $x \in S$ and add them together

$$E[X] = \sum_{x \in S} x \cdot P(x)$$

In Class Activity:

The following three distributions describe the winnings (right column) and their associated probs (left).

Compute the expected value of each of the following lottery tickets.
How are the tickets similar, how are they different? Which would you prefer to have?

"Double Lotto"

P(D)	D
1/2	\$2
1/2	\$0

"Steady Lotto"

P(S)	S
1/2	\$.9
1/2	\$1.1

"Shoot for Moon Lotto"

P(M)	M
1/1000	\$1000
999/1000	\$0

"Double Lotto"

$P(D)$	D
$1/2$	\$2
$1/2$	\$0

$$E[D] = 1/2 \cdot 2 + 1/2 \cdot 0$$

$$= 1$$

"Steady Lotto"

$P(S)$	S
$1/2$	\$.9
$1/2$	\$1.1

$$E[S] = 1/2 \cdot .9 + 1/2 \cdot 1.1$$

$$= 1$$

"Shoot for Moon Lotto"

$P(M)$	M
$1/1000$	\$1000
$999/1000$	\$0

$$E[M] = \frac{1}{1000} \cdot 1000 + \frac{999}{1000} \cdot 0$$

$$= 1$$

"Double Lotto"

$P(O)$	O
1/2	\$2
1/2	\$0

"Steady Lotto"

$P(S)$	S
1/2	\$.9
1/2	\$1.1

"Shoot For Moon Lotto"

$P(M)$	M
1/1000	\$1000
999/1000	\$0

Comparing Distributions
(Larger shaded area → More Prob)

Expected val of all lottos is 1

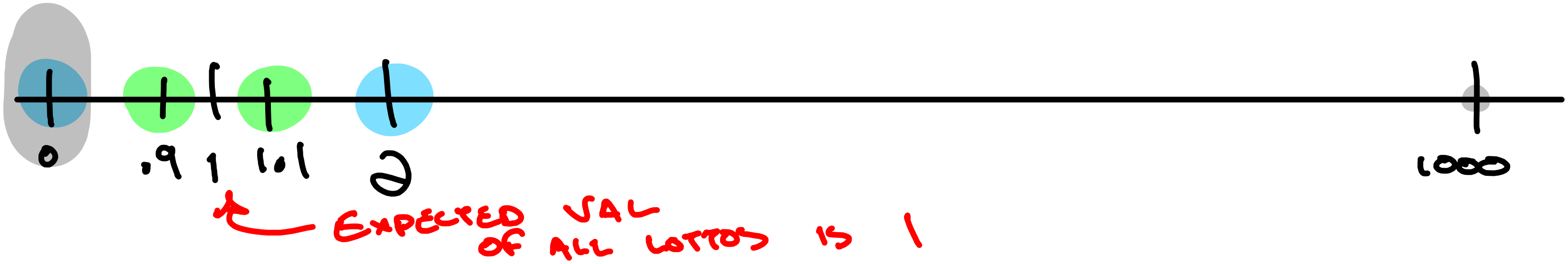
Variance of a random variable:

Intuition: variance measures how close, on average, outcomes of a RV are to their expected value
(how much "varying" do the outcomes do?)

"Steady Lotto" is typically very close to its expected value (small variance)

"Double Lotto" isn't super close or super far from expected value (medium variance)

"Shoot for moon lotto" is typically far from its expected value (large variance)



Variance of a random variable: computing (1 of 2)

variance: how close is a typical outcome to its expected value?
(how much "varying" do the outcomes do?)

$$E[D] = 1$$

Quantification:

$$VAR(X) = E\left[\left(X - E[X]\right)^2\right]$$

"DOUBLE LOTTO"

$P(O)$	O	$D - E[D]$	$(D - E[D])^2$
1/2	\$2	1	1
1/2	\$0	-1	1

$$VAR(D) = E\left[\left(D - E[D]\right)^2\right]$$

$$= 1/2 \cdot 1 + 1/2 \cdot 1 = 1$$

Variance of a random variable: computing (2 of 2)

Intuition: variance is a measurement of typical distance is to its own expected value
(how much "varying" do the outcomes do?)

$E[D] = 1$

Quantification (2 of 2):

"DOUBLE LOTTO"

$P(D)$	D	D^2
1/2	\$2	4
1/2	\$0	0

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

$$= E[X^2] - E[X]^2$$

$$VAR(D) = E[D^2] - E[D]^2$$

$$= \frac{1}{2}\cdot 4 + \frac{1}{2}\cdot 0 - 1^2 = 2 + 0 - 1 = 1$$

WHY GIVE TWO EQUATIONS FOR SAME THING?

$$\mathrm{VAR}(X) = E\left[\left(X - E[X]\right)^2\right]$$

INTUITIVE:
TYPICAL DISTANCE
TO EXP VAL
SQUARED

$$= E\left[X^2\right] - E\left[X\right]^2$$

OFTEN EASIER TO
COMPUTE

Standard Deviation:

The square root of variance (intuition is the very same)

$$\sigma = \sqrt{VAR(x)}$$

STANDARD DEVIATION

FOR THIS REASON WE ALSO USE σ^2 AS NOTATION FOR $VAR(x)$

Why have two measurements of the same thing?

In Class Activity: Variance (building intuition)

Order the following experiments from smallest to largest variance (or maybe two hvae equivilent variance?)

X = outcome of a 100 sided die STD DEV = 25
Y = outcome of a 1000 sided die STD DEV = 250
Z = height of student, uniformly chosen, from this room (measured in meters) STD DEV = 1
A = height of student, uniformly chosen, from this room (measured in miles) STD DEV = 1/1600
B = outcome is always 1928421984, with probability 100%
C = outcome is always 239832974, with probability 100%

0 = VAR(B) = VAR(C) A Z X Y

SMALL VAR ⟷ LARGE VAR

In Class Activity:

Compute the variance of the remaining two lottos. Validate that your quantification is consistent with the intuitions we've previously developed

"STEADY LOTTO"

$P(s)$	s
1/2	\$.9
1/2	\$1.1

"SHOOT FOR MOON LOTTO"

$P(m)$	m
1/1000	\$1000
999/1000	\$0

Suppose there is one more lotto:

"Good deal lotto":

- has a larger expected value than all others
- has a larger variance than all others

Tell if the following statements are true or false. If false, provide a particular "good deal" lotto distribution (e.g. table as shown) which has the two properties immediately above while violating the statement below.

- "good deal" outcomes are, on average, further from the "good deal" expected value than other lotto outcomes are to their own expected values
- every "good deal" outcome is larger than all other lotto outcomes

"Steady Lotto"

$E[S]=1$

$P(S)$	S	$S-E[S]$	$(S-E[S])^2$	S^2
$1/2$	\$.9	$.9-1=-.1$	.01	$.9^2$
$1/2$	\$1.1	$1.1-1=.1$	.01	1.1^2

$$VAR(S)=E\left[(S-E[S])^2\right]$$

$$=\tfrac{1}{2}\,.01+\tfrac{1}{2}\cdot .01$$

$$=.01$$

$$VAR(S)=E[S^2]-E[S]^2$$

$$=\tfrac{1}{2}\cdot .9^2+\tfrac{1}{2}\cdot 1.1^2-1^2$$

$$=.01$$

"SHOOT FOR MOON LOTTO"

$P(M)$	M	M^2
$1/1000$	\$1000	1000^2
$999/1000$	\$0	0^2

$E[M] = 1$

$$VAR(M) = E[M^2] - E[M]^2$$

$$= \frac{1}{1000} 1000^2 + \frac{999}{1000} 0^2 - 1^2$$

$$= 1000 + 0 - 1 = 999$$