CS1800 day 13

Admin:

- exam1 & hw4 graded to you next week
- hw5 released next friday
  enjoy the break from hw :)

Content:

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

objective: predict the outcome of a coin flip
prize: fleeting satisfaction of having been correct once
approach?

a similar? problem:

objective: predict the outcome of a coin flip
prize: world peace, universal happiness and calorie free cake (that tastes just as good) for all
approach? same as above?

Why study probability?

Probability allows us to build simple, effective models of the world from relevant data, (otherwise we must model complex details of how something really works)

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, what are popular movies which the user hasn't yet seen?

Self driving cars:

- among all the times I've been in this position on the road, how often does this car turn right even 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

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

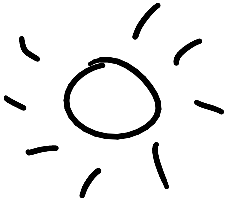
$S = \{1, 2, 3, 4, 5, 6\}$

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: Notations



$P(\text{SUN}) = 40\%$

ICONS | TEXT A BIT AMBIGUOUS

LET W BE RANDOM VARIABLE REPRESENTING WEATHER TODAY

ITS SAMPLE SPACE IS

$$S_W = \{ \underset{\text{CLOUDY}}{w_0}, \underset{\text{RAINY}}{w_1}, \underset{\text{SUNNY}}{w_2} \}$$

$$P(W = w_0) = 40\%$$

Probability: Notations

LET W BE RANDOM VARIABLE REPRESENTING WEATHER TODAY

ITS SAMPLE SPACE IS

$$S_W = \{ 0, 1, 2 \}$$

CLOUDY RAINY SUNNY

$$P(W = 0) = 40\%$$

SOME FOLKS PREFER O INSTEAD OF W_0

(PREFER W_0 IF MORE THAN 1 RANDOM VARIABLE)

CONVENTION

(A REALLY HELPFUL ONE)

$$P(W = w_0) = 40\%$$

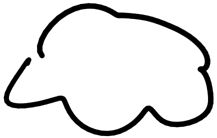
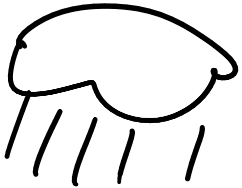
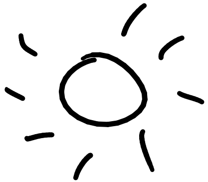
RANDOM VARIABLES (EXPERIMENTS W/ UNKNOWN OUTCOME) ARE CAPITALIZED

OUTCOMES ARE LOWERCASE

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%

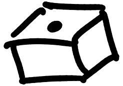
Uniform Distribution

Uniform distribution assigns equal prob to all outcomes in sample space

Fair coin

H	T
50%	50%

Fair die

	$\boxed{2}$	$\boxed{3}$	$\boxed{4}$	$\boxed{5}$	$\boxed{6}$
1/6	1/6	1/6	1/6	1/6	1/6

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

S is sample space

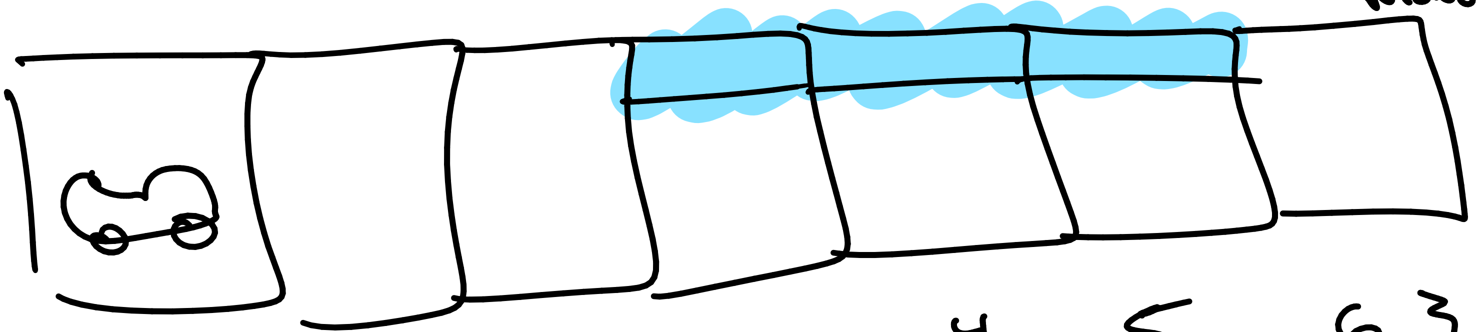
$|S|$ is # elements in S

EVENT

Subset of Sample Space

Event B = "Land on Blue Prop"

Assume
one die Monopoly



Sample Space $S = \{1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6\}$

$B = \{3 \quad 4 \quad 5\}$

COMPUTING PROB FROM UNIFORM DISTRIBUTION

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

IF X
IS UNIFORM

IN CLASS ACTIVITY:

GIVEN A FAIR DIE $\rightarrow$ 6 SIDED COMPUTE PROB OF EACH EVENT

X = ROLL A 1	Y = ROLL AN EVEN #	Z = ROLL A PRIME #
$E = \{1\}$	$E = \{2\ 4\ 6\}$	$E = \{2, 3, 5\}$
$P(X) = \frac{\lvert E \rvert}{\lvert S \rvert} = \frac{1}{6}$	$P(Y) = \frac{\lvert E \rvert}{\lvert S \rvert} = \frac{3}{6}$	$P(Z) = \frac{\lvert E \rvert}{\lvert S \rvert} = \frac{3}{6}$

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	1	4	7	12	5	6	6
~~PROB~~	1/41	4/41	7/41	12/41	5/41	6/41	6/41

WHAT IS DISTRIBUTION OF $X = D_1 + D_2$

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

SAMPLE SPACE OF D_i

$S \times S = \{$

(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)

$\}$

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

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

$$P(X=2) = P(X=8) = 1/16$$

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

$\$1 = 2 \cdot 1/2 + 0 \cdot 1/2$

W	P(W)
\$2	3/4
\$0	1/4

EXPECTED VALUE: COMPUTATION

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

SUPPOSE X HAS SAMPLE SPACE $\{-1, 100, 4\}$

"EXPECTED VALUE OF X"

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

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

COMPUTE INNER TERM FOR ALL $x \in S$ AND ADD THEM TOGETHER

$$E[2x] = \sum_{x \in S} 2x \cdot P(x)$$

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

$$E[D] = \sum_{x \in S} x \cdot P(x) = 2 \cdot 1/2 + 0 \cdot 1/2 = 1$$

"Steady Lotto"

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

$$E[S] = \sum_{s \in S} s \cdot P(s) = .9 \cdot 1/2 + 1.1 \cdot 1/2 = 1$$

"Shoot for Moon Lotto"

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

$$E[M] = \sum_{m \in S} m \cdot P(m) = 1000 \cdot \frac{1}{1000} + 0 \cdot \frac{999}{1000} = 1$$

"Double Lotto"

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

"Steady Lotto"

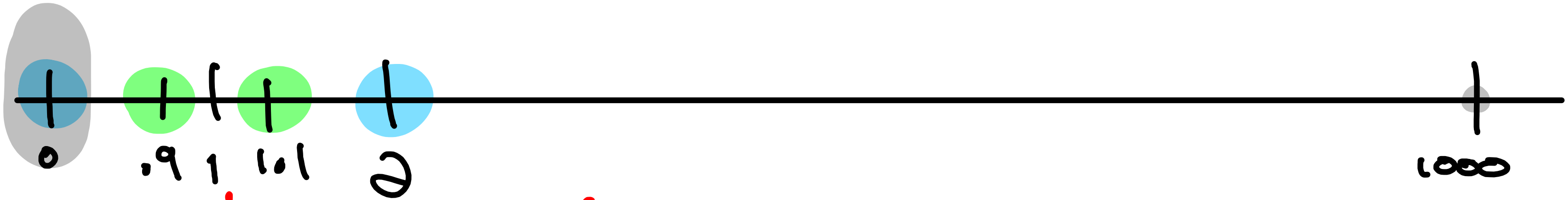
P(S)	S
1/2	\$.9
1/2	\$1.1

"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

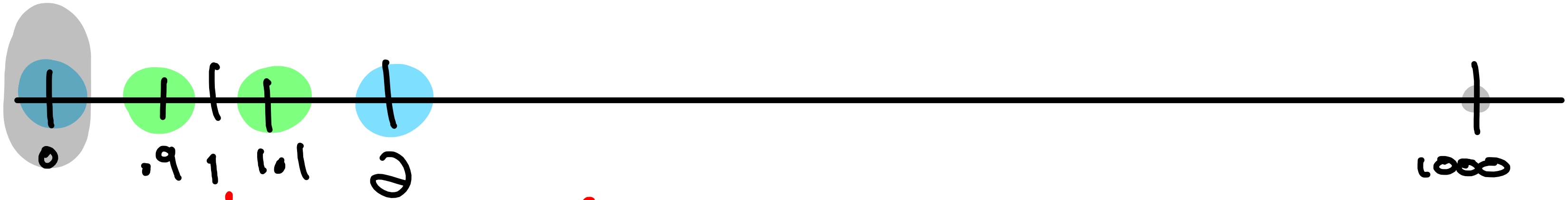
Variance of a random variable:

Intuition: variance is a measurement of typical distance is to its own 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)



EXPECTED VAL OF ALL LOTTOS IS 1

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

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

Quantification:

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

$$VAR(D) = \sum_{x \in S} \left(D - E[D]\right)^2 \cdot P(D=d)$$

$$= 1^2 \cdot 1/2 + (-1)^2 \cdot 1/2$$

$$= 1$$

"DOUBLE LOTTO"

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

$E[D] = 1$

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?)

Quantification (2 of 2):

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

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

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

$$= 2 - 1^2 = 1$$

"DOUBLE LOTTO"

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

$E[D] = 1$

$E[D^2] = 4 \cdot 1/2 + 0 \cdot 1/2 = 2$

WHY GIVE TWO EQUATIONS FOR SAME THING?

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

INTUITIVE:
TYPICAL DISTANCE
TO EXP VAL
SQUARED

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

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?

Sometimes easier to use one vs the other. Consider that radius & diamter have similar redundancy.

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, clarify them by rewriting to a true statement.

- "good deal" outcomes must always be higher than the other lottos
- every "good deal" outcome is further from the "good deal" expected value than all other lotto outcomes are to their own expected values

"STEADY LOTTO"

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

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

$$= 1.01 - 1^2 = .01$$

$$E[s] = \sum_{x \in S} s \cdot P(s) = .9 \cdot 1/2 + 1.1 \cdot 1/2 = 1$$

$$E[s^2] = \sum_{x \in S} s^2 \cdot P(s) = (.9)^2 \cdot 1/2 + (1.1)^2 \cdot 1/2 = 1.01$$

"SHOOT FOR MOON LOTTO"

$E[M] = 1$

$P(M)$	M	$M - E[M]$
$1/1000$	\$1000	999
$999/1000$	\$0	-1

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

$$= \sum_{m \in S} \left(m - E[M]\right)^2 \cdot P(m)$$

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

$$\simeq 999$$

Variance (intuition building):

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

X = outcome of a 100 sided die
Y = outcome of a 1000 sided die
Z = height of student, uniformly chosen, from this room (measured in meters)
A = height of student, uniformly chosen, from this room (measured in miles)
B = 1 with probability of 100%
C = 2 with probability of 100%