Professor Hamlin
Day 13

Agenda

1) Probability
   - random variable
   - outcome
   - distribution
2) Expected Value
3) Variance

Admin

Midterm grades - next Thursday
Check in grade week after

Why study probability?

If I have a length 4 password - how likely can someone break?
Length 8?
Length 16?

My whole field (Cryptography) tries to make it low probability that a bad guy will break into things.

$\frac{1}{2^{80}}$ $\frac{1}{2^{256}}$

Other places probability shows up:

1. ChatGPT: probability of common next words - "The best cat is ...."
2. Self driving cars: probability car will change lanes
3. Games: spawn rate of monsters

Probability: how likely are future events to happen based on past events.

Definitions

Experiment - thing trying to model

coin flip rolling dice

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

Heads / Tails 1, 2, ... 6

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

→ unordered unique collection of items

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

$S = \{$Sunny, rainy, cloudy$\}$

Distribution (of an experiment) - set of probabilities of each outcome in distribution

e.g

Heads	Tails
50%	50%

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

Ex) $\{\boxed{\text{Red}}$, Green, Yellow$\}$ - sample space

outcome

traffic light color - experiment

R : 30% G : 40% Y : 30% - distribution

Random Variable:

- variable: $X+2$ where $X \in \mathbb{R}$

a <u>random</u> variable W where $W \in S$

sample space

Say $S = \{\text{cloudy}, \text{rainy}, \text{sunny}\}$ with distribution

cloudy	rainy	sunny
20%	40%	40%

W is a random variable representing the weather today.

$$\Pr[W = \text{"cloudy"}] = 20\%$$

Probability that W is cloudy is 20%

$$\Pr[W = \text{"sunny"}] = 40\%$$

Probabilty that W is sunny is 40%

Probability that W is rainy is 40%

$$\Pr[W = \text{"rainy"}] = 40\%$$

Convention: $\Pr[X = x_0] = 40\%$ ← prob of outcome

Syntax

random variable (for experiment are capitilized)

outcomes (are lowercase)

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

Also remember: $20\% = .2$

$100\% = 1$

Important facts (axioms)

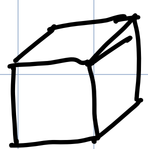
1) Probability of an outcome happening is positive or zero

2) Sum of probability of all outcomes in sample space is 1

$Pr[W = \text{"cloudy"}] = .4$
$Pr[W = \text{"rainy"}] = .2$
$Pr[W = \text{"sunny"}] = \cancel{.4}\ .3$
$Pr[W = \text{"Hail"}] = .1$

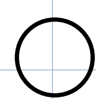
$.4 + .4 + .2 = 1$

Uniform Distribution - all outcomes have equal probabilty



6 sided dice

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



fair coin

H	T
.5	.5

Generally: $Pr[X = x] = \dfrac{1}{|S|}$

random variable over uniform dist.

one of outcomes

size of sample space

Example: 1) 8-sided dice

$Pr[X = 4] = \dfrac{1}{8}$

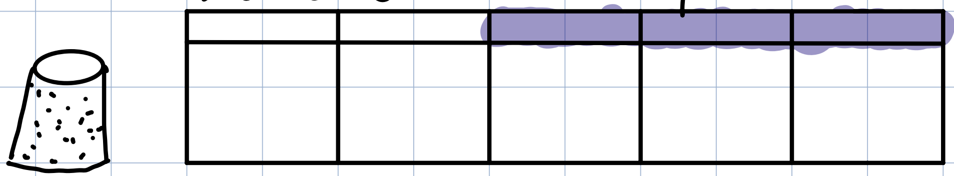
2) Pick a number between 0 and 111_2 inclusive

$Pr[X = 11_2] = 1/8$

Event - subset of the sample space

e.g. Roll an even number on 6-sided die

Land on blue space



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

# of outcomes in event

$$P[\text{Blue Space}] = \frac{|B|}{|S|} = \frac{3}{5}$$

event is $X=3 \vee X=4 \vee X=5$

# of outcomes in sample space

<u>Exercise</u>: Assume 8-sided die

1) Event: roll a 1 $\{1\}$ (One)

$Pr[\text{One}] = \frac{1}{8}$

2) Event: outcome is even (Even) $\{2, 4, 6, 8\}$

$Pr[\text{Even}] = \frac{4}{8}$

3) Event: roll a prime number (Prime) $\{2, 3, 5, 7\}$

$Pr[\text{Prime}] = \frac{4}{8}$

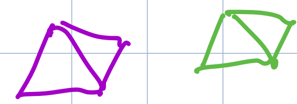
<u>Combining random variables</u> $X + Y$

Let D be the outcome of 4-sided die

sum of two 4-sided die rolls → $X = D_1 + D_2$ ← 2nd 4 sided die

1st 4 sided Die

What is X distribution?



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

Sample of $X = \{2,3,4,5,6,7,8\}$

$$S \times S = \{(1,1)\ (1,2)\ (1,3)\ (1,4)$$
$$(2,1)\ (2,2)\ (2,3)\ (2,4)$$
$$(3,1)\ (3,2)\ (3,3)\ (3,4)$$
$$(4,1)\ (4,2)\ (4,3)\ (4,4)\}$$

Cartesian product

$Pr[X=2] = 1/16$ $Pr[X=3] = 2/16$

$Pr[X=5] = 4/16$ $Pr[X=6] = 3/16$

Expected Value:

- should you play the lotto? -

expected value is an "average" outcome of random variable.

Ex)

Pr	Winnings
.5	\$2
.5	\$0

1/2 I win \$2

1/2 \$0

Expected winnings $= \$2 \cdot .5 + \$0 \cdot .5$
$= \$1$

Intuitively: multiply every outcome by probability and add it up

→ only works for outcomes that are #

Formally: $E[X] = \sum_{x \in S} x \cdot Pr[X = x]$

"expected value of Random Variable X"

add all of those values up

value of outcome

probability of outcome

Ex) $S = \{-1, 100, 4\}$

$$E[X] = -1 \cdot Pr[X=-1] + 100 \cdot Pr[X=100] + 4 \cdot Pr[X=4]$$

Exercise Given probabilities (left) and value (right) which lotto would you prefer to play? Calculate $E[X]$

Pr	Win
1/2	\$2
1/2	\$0

$E[A] = 2 \cdot .5 + 0 \cdot .5$ $\boxed{1}$

Pr	Win
1/2	\$.9
1/2	\$1.1

$E[B] = .9 \cdot .5 + 1.1 \cdot .5$ $\boxed{1}$

Pr	Win
1/100	\$100
99/100	\$0

$E[C] = 100 \cdot 1/100 + 99/100 \cdot 0$ $\boxed{1}$

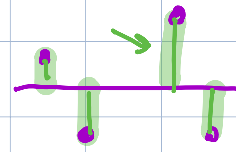
So expected values are the same? But the winnings vary a lot

B is pretty close to expected value

C is far from expected value

Variance

intuitively: measures how far outcomes range from expected value



Formally: $\mathrm{Var}(X) = E[(X - E[X])^2]$

$$= \sum_{x \in S} (x - E[X])^2 \cdot \Pr[X = x]$$

Ex)

Pr	Winnings	$x - E[x]$
1/2	\$2	\$1
1/2	\$0	−\$1

$E[X] = 1$

$\mathrm{Var}(X) = (1)^2 \cdot 1/2 + (-1)^2 \cdot 1/2$

$\boxed{1}$

Ex)

Pr	Winnings	$x - E[x]$
1/100	\$100	99
99/100	\$0	−1

$E[X] = 1$

$\mathrm{Var}(X) = (99)^2 \cdot 1/100 + (-1)^2 \cdot 99/100$

$\boxed{99}$

Also have another formula that is equivalent

$$\mathrm{Var}(X) = E[X^2] - (E[X])^2$$

Ex) $E[X^2] = (100)^2 \cdot 1/100 + (0)^2 \cdot 99/100 = 100$

$(E[X])^2 = 1^2 = 1$ $\quad 100 - 1 = \boxed{99}$

Standard Deviation : square root of variance (stdev)

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

Why use it? Kinda like radius & diameter on circle, two ways to describe same thing

Exercise Variance of lotto B $\quad E[x]=1$

Pr	
1/2	\$.9
1/2	\$1.1

$E[x^2] = (.9)^2 \cdot .5 + (1.1)^2 \cdot .5$
$= 1.01$

$E[x^2] - (E[x])^2 = 1.01 - 1^2 = .01$

2) Variance of 4-sided die

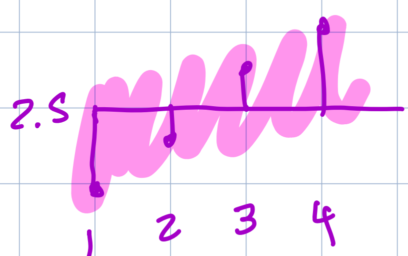
1	.25
2	25%
3	.25
4	.25

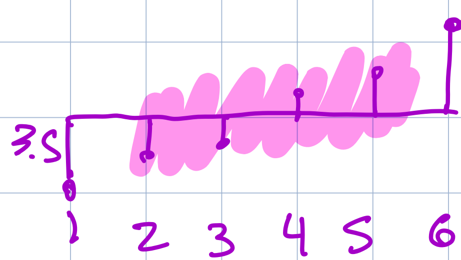
$E[x] = 1 \cdot 1/4 + 2 \cdot 1/4 + 3 \cdot 1/4 + 4 \cdot 1/4$
$2\,1/2$

$E[x^2] = 1^2 \cdot 1/4 + 2^2 \cdot 1/4 + 3^2 \cdot 1/4 \cdot 4^2 \cdot 1/4$
$1/4 + 1 + 9/4 + 4$
$7\,1/2$

$7.5 - (2.5)^2 = 1.25$

3) Variance of 6-sided die bigger or smaller than 4 sided die?





Bigger

Exercise: Order the experiments from smallest to largest variance

1) X = outcomes of 100 sided die
2) Y = outcomes of 1000 sided die
3) Z = height of students, chosen uniformly, in meters
4) A = height of students, chosen uniformly, in miles
5) B = Always 1
6) C = Always 2

$B/C, A, Z, X, Y$