Professor Hamlin
Day 13

Agenda

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

Admin

Midterm - Next Thursday
Check in grade a 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.

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

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
coinflip or dice roll

Outcome (of an experiment) - a particular result of experiment
Heads / Tails 1 or 2 ...

Sample Space (of an experiment) - the set of all possible outcomes
$S = \{Heads, Tails\}$
unique unordered elements
$S = \{1, 2, 3, 4, 5, 6\}$

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 outcome $\{$Cloudy, Sunny, Rainy$\}$ - sample space
Experiment: weather conditions
C: 40% S: 40% R: 20% - distribution

Random Variable:

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

a random variable W where $W \in S$ (sample space)

Say $S = \{$cloudy, rainy, 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 weather is cloudy is 20%

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

Probability that weather is sunny is 40%

Probability that weather is rainy is 40%

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

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

Syntax

random variable (for experiment are capitalized)

outcomes (are lowercase)

Also remember: $20\% = .2$

$100\% = 1$

Important facts

1) Probability of an outcome happening is positive or zero

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

$Pr[W = \text{"cloudy"}] = .4$
$Pr[W = \text{"rainy"}] = .2$ $\quad .4 + .2 + .4 = 1$
$Pr[W = \text{"sunny"}] = .4$ ~~.4~~ .3
$Pr[W = \text{"hail"}] = .1$

Uniform Distribution - all outcomes have equal probability

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

- X: RV over a uniform dist.
- x: one of the outcome
- $|S|$: size of sample space

Example: 1) 8-sided dice

$Pr[X = 4] = \frac{1}{|S|} = \frac{1}{8}$

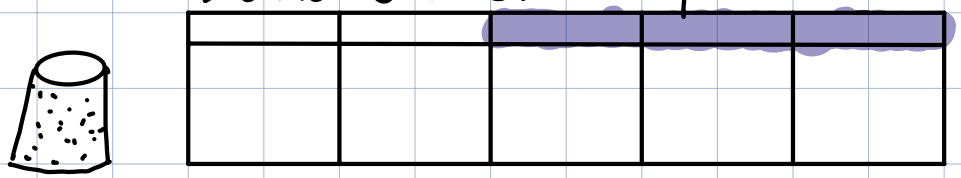
2) Picking a number between 0 and 111_2 (inclusive)

$Pr[X = 111_2] = \frac{1}{8}$

Event - subset of the sample space

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

Land on blue space



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

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

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

$|B|$: # of outcomes in event

$|S|$: # of outcomes in sample space

Exercise: Assume 8-sided die

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

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

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

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

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

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

Combining random variables (e.g. $X+Y$)

Let D be the outcome of 4-sided die

$$X = D_1 + D_2$$

D_2: 2nd 4 sided die

D_1: 1st 4 sided die

X: sum of two 4-sided die rolls

What is X distribution?

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

$S_X = \{2, 3, 4, 5, 6, 7, 8\}$

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

↑ 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

← win $2 half the time

~~←~~ $0 1/2 the time

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

← random variable

"expected value of Random Variable X"

add all of those values up

value of outcome

probability of outcome

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

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

$-1 \cdot .3 \quad + 100 \cdot .6 \quad + 4 \cdot .1$

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

Pr	Win
½	$2
½	$0

Pr	Win
½	$.9
½	$1.1

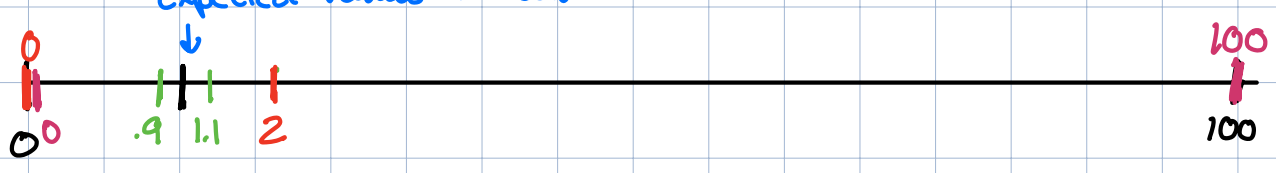
Pr	Win
1/100	$100
99/100	$0

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

$E[B] = .5 \cdot \$1.9 + .5 \cdot \$1.1$ [1]

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

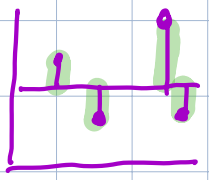
So expected values are the same? But the winnings vary a lot



B is pretty close to expected value

C is far from expected value

Variance



intuitively: measures how far outcomes range from expected value

Formally: $Var(X) = E[(X - E[X])^2]$

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

Ex

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

$E[X] = 1$

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

$1/2 + 1/2$

$= \boxed{1}$

Ex

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

$E[X] = 1$

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

$\boxed{99}$

Also have another formula that is equivalent

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

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

$(E[X])^2 = (1)^2$

$100 - 1 = \boxed{99}$

Standard Deviation: square root of variance

$$\sigma = \sqrt{Var(X)}$$

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

Exercise 1) Variance of lotto B $E[B] = 1$

Pr	
½	\$.9
½	\$1.1

$$E[X^2] = (.9)^2 \cdot .5 + (1.1)^2 \cdot .5$$
$$= 1.01$$
$$E[X^2] - (E[X])^2 = 1.01 - 1^2$$
$$= \boxed{.01}$$

2) Variance of 4 sided die

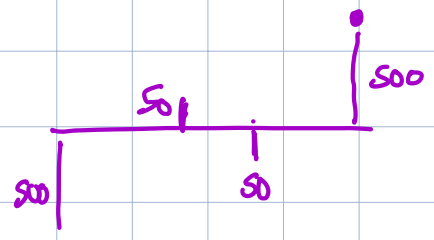
1	¼
2	¼
3	¼
4	¼

$$E[X] = 1 \cdot .25 + 2 \cdot .25 + 3 \cdot .25 + 4 \cdot .25$$
$$2.5$$
$$E[X^2] = 1^2 \cdot .25 + 2^2 \cdot .25 + 3^2 \cdot .25 + 4^2 \cdot .25$$
$$1/4 + 1 + 9/4 + 4$$
$$5 + 10/4$$
$$7.5$$

$$7.5 - (2.5)^2 = \boxed{1.25}$$

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



5/6, 1, 2

3 vs 4 think about it!