COMP3055
Machine Learning

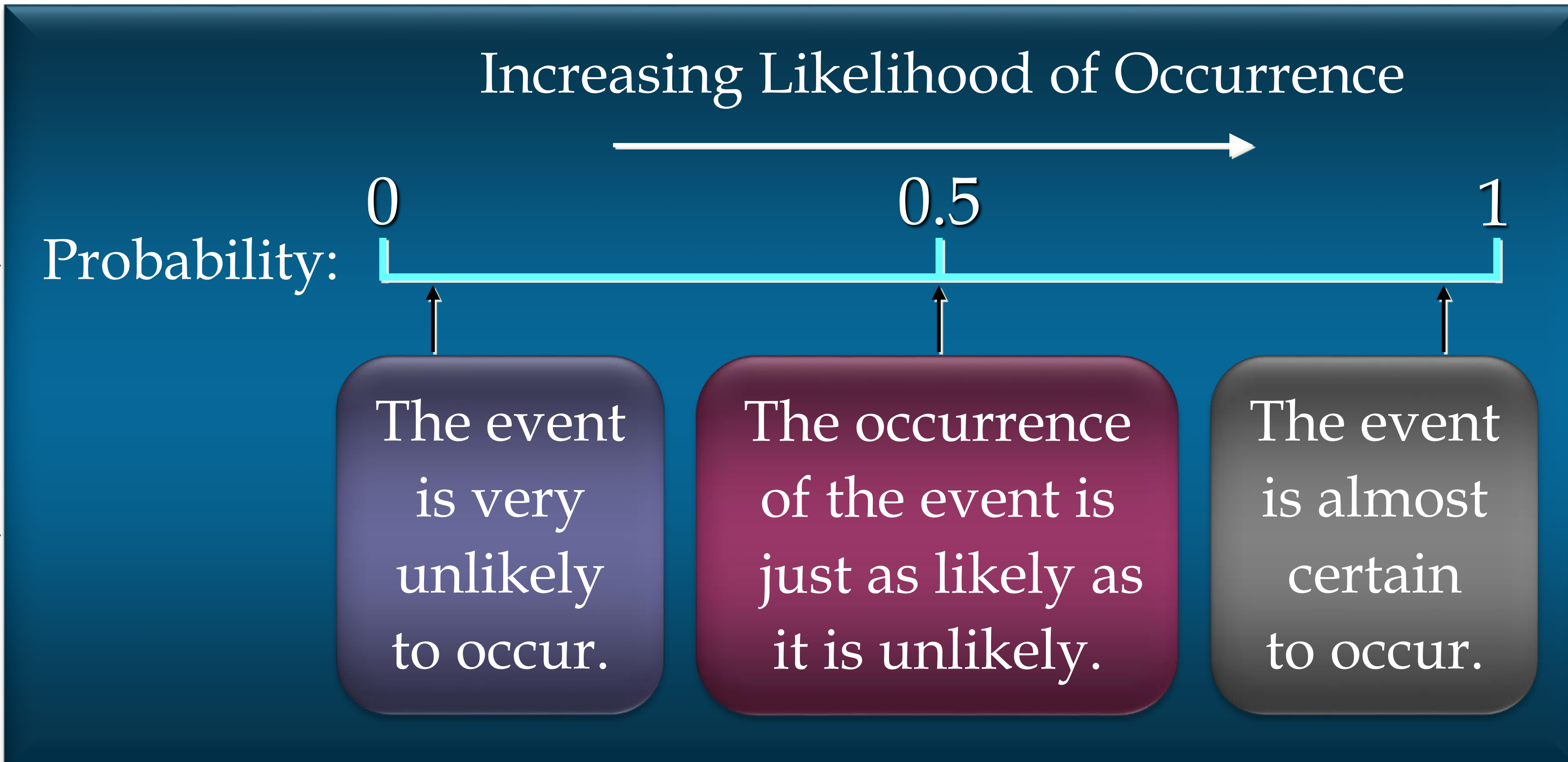
Topic 4 – Probability Theory

Ying Weng
2024 Autumn

Probability

- Probability is a numerical measure of the likelihood that an event will occur.
- Probability values are always assigned on a scale from 0 to 1.
- A probability near zero indicates an event is quite unlikely to occur.
- A probability near one indicates an event is almost certain to occur.

Probability as a Numerical Measure

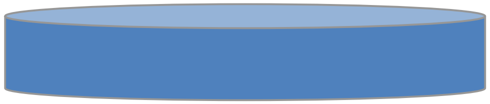


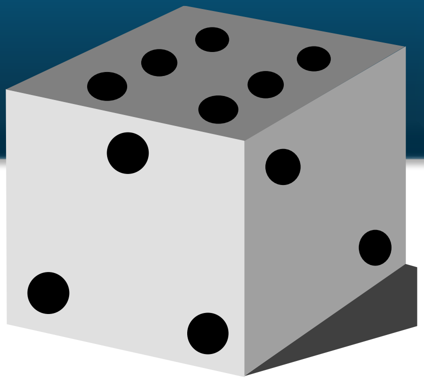
Sample Space

- The set of all possible outcomes of an experiment.
- A sample space can be finite or infinite, & discrete or continuous.

An Experiment and Its Sample Space

Experiment	Experiment Outcomes
Flip a coin	Head, tail
Inspect a product	Defective, non-defective
Conduct a sales call	Purchase, no purchase
Roll a die	1, 2, 3, 4, 5, 6
Play a football game	Win, lose, tie

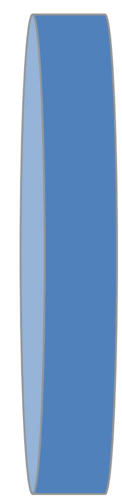




Experiment

- Something capable of replication under stable conditions.

- Example: Flipping a coin



Example: discrete, finite sample space

- Experiment: Flipping a coin once.
- Sample space: {**H**ead, **T**ail}.
- This sample space is **discrete**
- This sample space is also finite. There are just two elements: that’s a **finite** number.

Example: continuous sample space

- Experiment: Burning a light bulb until it burns out.
  Suppose there is a theoretical maximum number of hours that a bulb can burn and that is 10,000 hours.
- Sample space: The set of all real numbers between 0 & 10,000.
- Between any two numbers you can pick in the sample space, there is another number.
- For example, the bulb could burn for 99.777 hours or 99.778 hours. But it could also burn for 99.7775, which is in between.
- This sample space is **continuous**.
- It is also **infinite**, since there are an infinite number of possibilities.
- ***All continuous sample spaces are infinite.***

Two Basic Properties of Probability

- 1. $0 \leq Pr(E) \leq 1$
  for every subset of the sample space S

- 2. $Pr(S) = 1$

Assigning Probabilities

- **Basic Requirements for Assigning Probabilities**

1. The probability assigned to each experimental outcome must be between 0 and 1, inclusively.

$$0 \le P(E_i) \le 1 \text{ for all } i$$

where:

E_i **is the** i**th experimental outcome**

and $P(E_i)$ **is its probability**

Assigning Probabilities

- Basic Requirements for Assigning Probabilities

2. The sum of the probabilities for all experimental outcomes must equal to 1.

$$P(E_1) + P(E_2) + \ldots + P(E_n) = 1$$

where:

n **is the number of experimental outcomes**

Assigning Probabilities

- **Classical Method**

  Assigning probabilities based on the assumption of equally likely outcomes

- **Relative Frequency Method**

  Assigning probabilities based on experimental or historical data

- **Subjective Method**

  Assigning probabilities based on judgment

Classical Method

- **Example: Roll a die**
- If an experiment has n possible outcomes, the classical method would assign a probability of $1/n$ to each outcome.

- Sample Space: $S = \{1, 2, 3, 4, 5, 6\}$
- Probabilities: Each sample point has a 1/6 chance of occurring

Relative Frequency Method

- **Example: Lucas Tool Rental**
- Lucas Tool Rental would like to assign probabilities to the number of car polishers it rents each day. Office records show the following frequencies of daily rentals for the last 40 days.

Number of Polishers Rented	Number of Days
0	4
1	6
2	18
3	10
4	2

Relative Frequency Method

- **Example: Lucas Tool Rental**

Each probability assignment is given by dividing the frequency (number of days) by the total frequency (total number of days).

Number of Polishers Rented	Number of Days	Probability
0	4	.10
1	6	.15
2	18	.45
3	10	.25
4	2	.05
	40	1.00

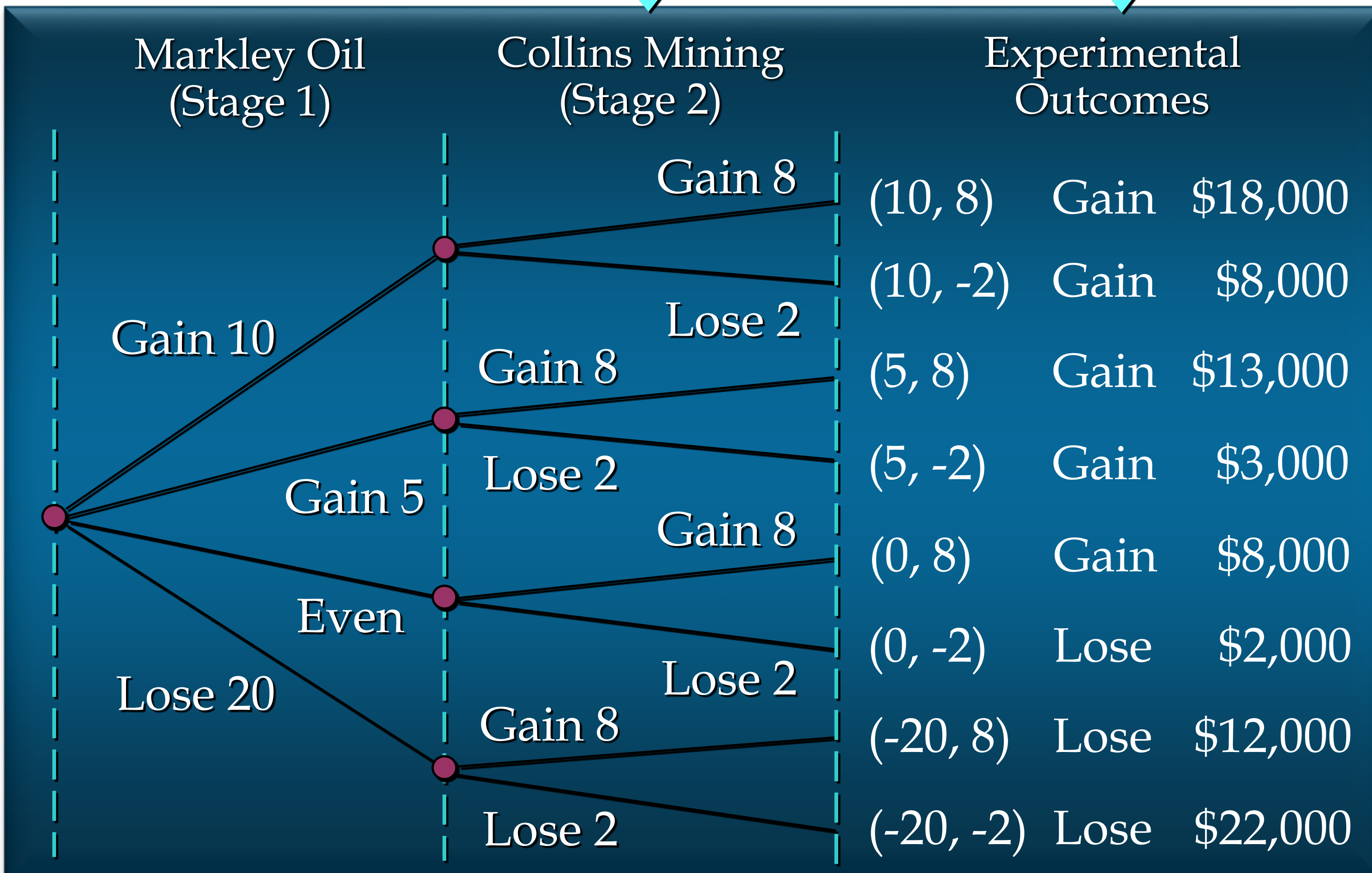
4/40

Subjective Method

- **When economic conditions and a company's circumstances change rapidly it might be inappropriate to assign probabilities based solely on historical data.**
- **We can use any data available as well as our experience and intuition, but ultimately a probability value should express our degree of belief that the experimental outcome will occur.**
- **The best probability estimates often are obtained by combining the estimates from the classical or relative frequency approach with the subjective estimate.**

Tree Diagram

- **Example: Bradley Investments**



Subjective Method

- **Example: Bradley Investments**
- An analyst made the following probability estimates.

Experimental Outcome	Net Gain or Loss	Probability
(10, 8)	$18,000 Gain	.20
(10, –2)	$8,000 Gain	.08
(5, 8)	$13,000 Gain	.16
(5, –2)	$3,000 Gain	.26
(0, 8)	$8,000 Gain	.10
(0, –2)	$2,000 Loss	.12
(–20, 8)	$12,000 Loss	.02
(–20, –2)	$22,000 Loss	.06

Counting Rules

- We'll look at:

  Basic multiplication rule

Multiplication Rule Example

In general,

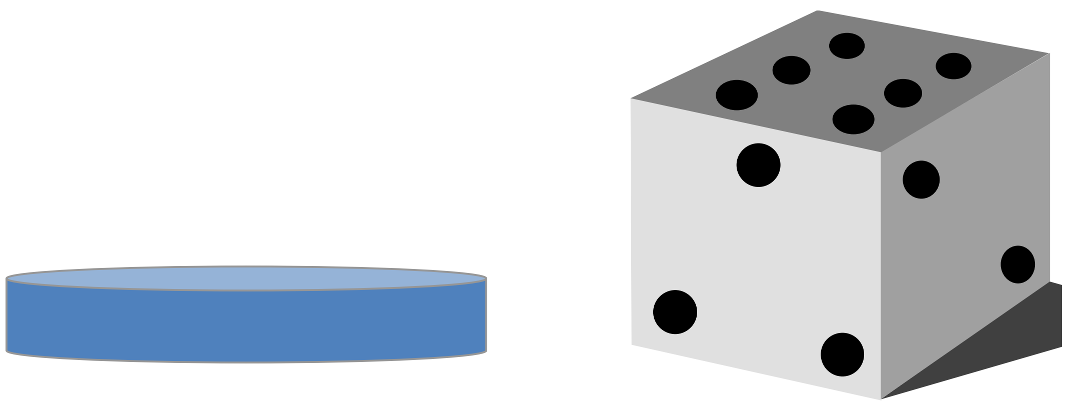
- If we have an experiment with k parts (such as 3 flips)
- and each part has n possible outcomes (such as heads & tails)
- then the total number of possible outcomes for the experiment is
  - n^k
- This is the simplest multiplication rule

As a variation, suppose that we have
an experiment with 2 parts,
the 1st part has m possibilities,
& the 2nd part has n possibilities.

- How many possibilities are there for the experiment?

In particular, we might have a coin & a die.

- So one possible outcome could be H6 (a head on the coin & a 1 on the die).
- How many possible outcomes does the experiment have?



We have 2 x 6 = 12 possibilities.

- Return to our more general question about the 2-part experiment with m & n possibilities for each part.

- We now see that the total number of outcomes for the experiment is

- ***mn***

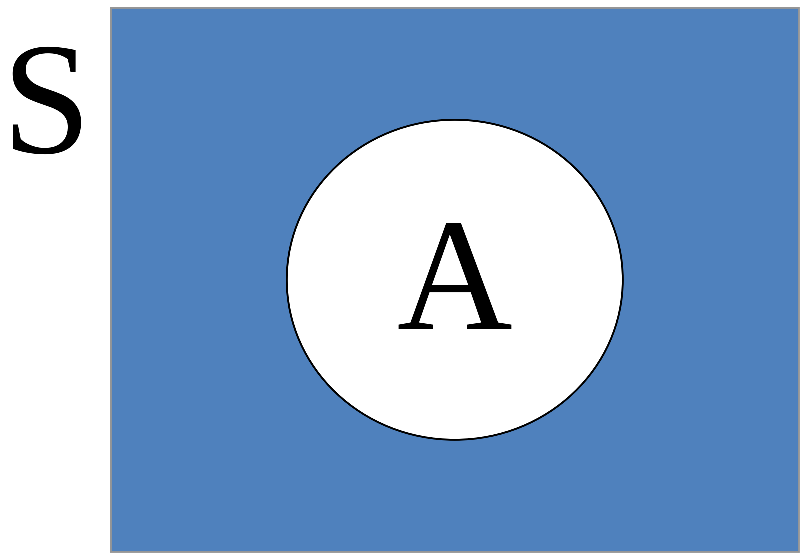
If we had a 3-part experiment,
the 1st part has m possibilities,
the 2nd part has n possibilities,
& the 3rd part has p possibilities,
how many possible outcomes would the experiment have?

- $m\ n\ p$

Complements, Unions, & Intersections

- Suppose A & B are events.

The ***complement*** of A is everything in the sample space S that is NOT in A.



- If the rectangular box is S, and the white circle is A, then everything in the box that's outside the circle is A^c , which is the complement of A.

Theorem

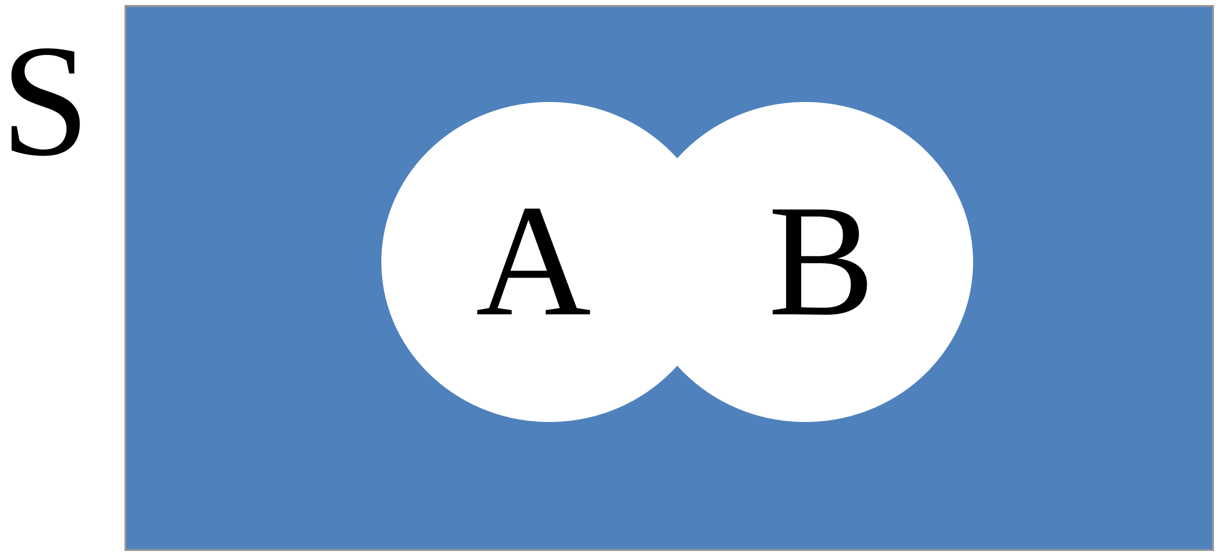
- $Pr(A^c) = 1 - Pr(A)$

- Example:

- If A is the event that a randomly selected student is male, and the probability of A is 0.6, what is A^c and what is its probability?

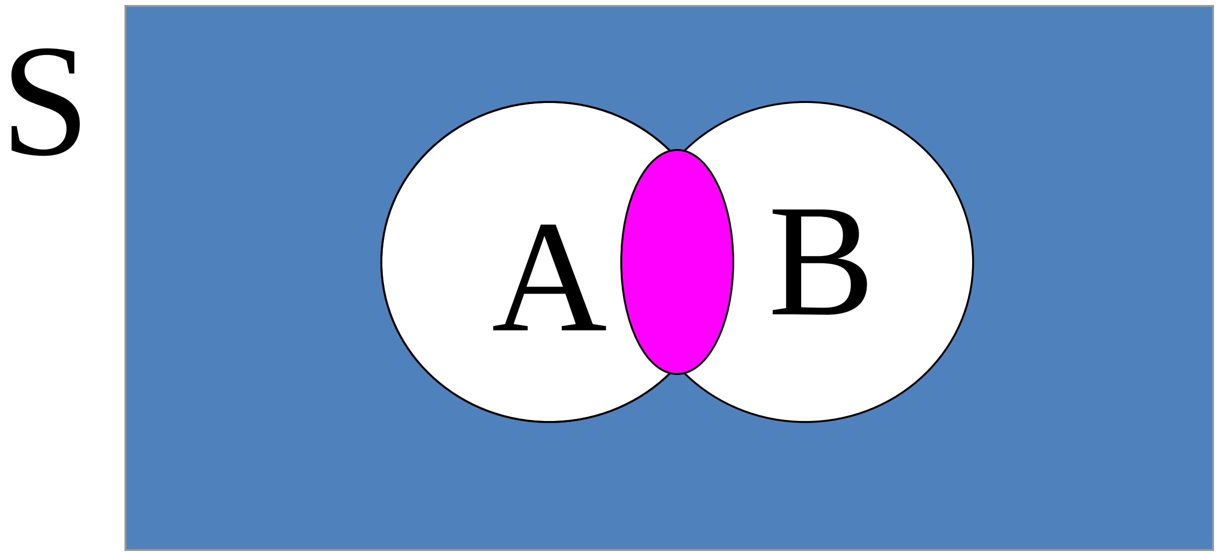
- A^c is the event that a randomly selected student is female, and its probability is 0.4.

The ***union*** of A & B (denoted A U B)
is everything in the sample space that is in either A or B or both.



- The union of A & B is the whole white area.

The *intersection* of A & B (denoted A∩B) is everything in the sample space that is in both A & B.



• The intersection of A & B is the pink overlapping area.

Example

- A family is planning to have 2 children.
- Suppose boys (B) & girls (G) are equally likely.

- What is the sample space S?

    - S = {BB, GG, BG, GB}

Example continued

- If E is the event that both children are the same sex, what does E look like & what is its probability?

- E = {BB, GG}

- Since boys & girls are equally likely, each of the four outcomes in the sample space S = {BB, GG, BG, GB} is equally likely & has a probability of 1/4.

- So Pr(E) = 2/4 = 1/2 = 0.5

Example continued:

Recall that $E = \{BB, GG\}$ & $Pr(E)=0.5$

- What is the complement of E and what is its probability?

- $E^c = \{BG, GB\}$

- $Pr\ (E^c) = 1 - Pr(E) = 1 - 0.5 = 0.5$

Example continued

- If F is the event that at least one of the children is a girl, what does F look like & what is its probability?

- F = {BG, GB, GG}

- Pr(F) = 3/4 = 0.75

Recall: $E = \{BB, GG\}$ & $Pr(E)=0.5$
$F = \{BG, GB, GG\}$ & $Pr(F) = 0.75$

- What is $E \cap F$?
  - $\{GG\}$
- What is its probability?
  - $1/4 = 0.25$

Recall: E = {BB, GG} & Pr(E)=0.5

F = {BG, GB, GG} & Pr(F) = 0.75

- What is the EUF?
  - {BB, GG, BG, GB} = S
- What is the probability of EUF?
  - 1
- If you add the separate probabilities of E & F together, do you get Pr(EUF)? Let's try it.
- Pr(E) + Pr(F) = 0.5 + 0.75 = 1.25 $\neq$ 1 = Pr (EUF)
- Why doesn't it work?
- We counted GG (the intersection of E & F) twice.

A formula for Pr(EUF)

- $Pr(E \cup F) = Pr(E) + Pr(F) - Pr(E \cap F)$

- If E & F do not overlap, then the intersection is the empty set, & the probability of the intersection is zero.

- When there is no overlap,
$Pr(E \cup F) = Pr(E) + Pr(F)$.

Conditional Probability of A given B
Pr(A|B)

- $Pr(A|B) = Pr(A \cap B) / Pr(B)$

Example

Suppose there are 10,000 students at a university.
2,000 are seniors (S). 3,500 are female (F).
800 are seniors & female.

- Determine the probability that a randomly selected student is (1) a senior, (2) a female, (3) a senior & female.
- 1. $Pr(S) = 2{,}000/10{,}000 = 0.2$
- 2. $Pr(F) = 3{,}500/10{,}000 = 0.35$
- 3. $Pr(S \cap F) = 800/10{,}000 = 0.08$

Use the definition of conditional probability
$Pr(A|B) = Pr(A \cap B) / Pr(B)$
& the previously calculated information
$Pr(S) = 0.2$; $Pr(F) = 0.35$; $Pr(S \cap F) = 0.08$
to answer the questions below.

- 1. If a randomly selected student is female, what is the probability that she is a senior?
  - $Pr(S|F) = Pr(S \cap F) / Pr(F)$
    - $= 0.08 / 0.35 = 0.228$
- 2. If a randomly selected student is a senior, what is the probability the student is female?
  - $Pr(F|S) = Pr(F \cap S) / Pr(S)$
    - $= 0.08 / 0.2 = 0.4$
- Notice that $S \cap F = F \cap S$, so the numerators are the same, but the denominators are different.

Joint Probability Distributions & Marginal Distributions

- **Example:**

  Suppose a firm's employees in three departments :
  10% are male & in dept. 1,
  30% are male & in dept. 2,
  20% are male & in dept. 3,
  15% are female & in dept. 1,
  20% are female & in dept. 2,
  5% are female & in dept. 3

  Then the ***joint probability distribution*** of gender & dept. is as in the table below:

	D_1	D_2	D_3
M	0.10	0.30	0.20
F	0.15	0.20	0.05

Example continued: What is the probability that a randomly selected employee is ***male***?

	D_1	D_2	D_3
M	0.10	0.30	0.20
F	0.15	0.20	0.05

Example continued: What is the probability that a randomly selected employee is ***male***?

	D_1	D_2	D_3	
M	0.10	0.30	0.20	0.60
F	0.15	0.20	0.05	

Example continued: What is the probability that a randomly selected employee is ***female***?

	D_1	D_2	D_3	
M	0.10	0.30	0.20	0.60
F	0.15	0.20	0.05	

Example continued: What is the probability that a randomly selected employee is ***female***?

	D_1	D_2	D_3	
M	0.10	0.30	0.20	0.60
F	0.15	0.20	0.05	0.40

Example continued: What is the probability that a randomly selected employee is in ***dept. 1***?

	D_1	D_2	D_3	
M	0.10	0.30	0.20	0.60
F	0.15	0.20	0.05	0.40

Example continued: What is the probability that a randomly selected employee is in ***dept. 1***?

	D_1	D_2	D_3	
M	0.10	0.30	0.20	0.60
F	0.15	0.20	0.05	0.40
	0.25			

Example continued: What is the probability that a randomly selected employee is in ***dept. 2***?

	D_1	D_2	D_3	
M	0.10	0.30	0.20	0.60
F	0.15	0.20	0.05	0.40
	0.25			

Example continued: What is the probability that a randomly selected employee is in ***dept. 2***?

	D_1	D_2	D_3	
M	0.10	0.30	0.20	0.60
F	0.15	0.20	0.05	0.40
	0.25	0.50		

Example continued: What is the probability that a randomly selected employee is in ***dept. 3***?

	D_1	D_2	D_3	
M	0.10	0.30	0.20	0.60
F	0.15	0.20	0.05	0.40
	0.25	0.50		

Example continued: What is the probability that a randomly selected employee is in ***dept. 3***?

	D_1	D_2	D_3	
M	0.10	0.30	0.20	0.60
F	0.15	0.20	0.05	0.40
	0.25	0.50	0.25	

Example continued: The ***marginal distribution*** of gender is in first & last columns (or left & right ***margins*** of the table) & gives the probability of each possibility for gender.

	D_1	D_2	D_3	
M	0.10	0.30	0.20	0.60
F	0.15	0.20	0.05	0.40
	0.25	0.50	0.25	

Example continued: The ***marginal distribution*** of department is in first & last rows (or top & bottom ***margins*** of the table) & gives the probability of each possibility for dept.

	D_1	D_2	D_3	
M	0.10	0.30	0.20	0.60
F	0.15	0.20	0.05	0.40
	0.25	0.50	0.25	

Notice that when you add the numbers in the last column or the last row, you must get ***one***, because you're adding all the probabilities for all the possibilities.

	D_1	D_2	D_3	
M	0.10	0.30	0.20	0.60
F	0.15	0.20	0.05	0.40
	0.25	0.50	0.25	1.00

Independence

- Two events are independent, if knowing that one event happened doesn't give you any information on whether the other happened.

- Example:
- A: It rained a lot in Beijing, China last year.
- B: You did well in your courses last year.
- These two events are independent

One of these events occurring tells you nothing about whether the other occurred.

So in terms of probability, two events A & B are independent if and only if

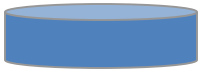
- * $\Pr(A|B) = \Pr(A)$

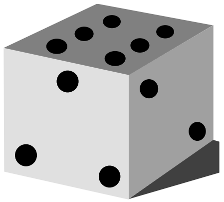
- Using the definition of conditional probability, this statement is equivalent to
  - $\Pr(A \cap B) / \Pr(B) = \Pr(A)$.
- Multiplying both sides by $\Pr(B)$, we have
  - * $\Pr(A \cap B) = \Pr(A)\Pr(B)$.
- Dividing both sides by $\Pr(A)$, we have
  - $\Pr(A \cap B) / \Pr(A) = \Pr(B)$,
  - which is equivalent to
  - * $\Pr(B|A) = \Pr(B)$.
- This makes sense. If knowing about B tells us nothing about A, then knowing about A tells us nothing about B.

We now have 3 equivalent statements for 2 independent events A & B

- $\Pr(A|B) = \Pr(A)$
- $\Pr(B|A) = \Pr(B)$
- $\Pr(A \cap B) = \Pr(A)\,\Pr(B)$.
- The last equation says that you can calculate the probability that both of two independent events occurred by multiplying the separate probabilities.

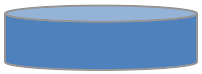
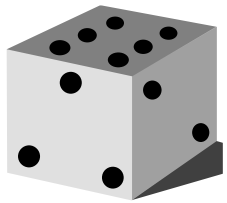
Example: flip a fair coin & roll a fair die





- A: You get a H on the coin.
- B: You get a 6 on the die.
- Recall that we counted 12 possible outcomes for this experiment.
- Since the coin & the die are fair, each outcome is equally likely, & the probability of getting a H & a 6 is 1/12.

Example continued

- The probability of a H on the coin is 1/2
- The probability of a 6 on the die is 1/6.
- So $Pr(H)\ Pr(6) = (1/2)(1/6)$
- $= 1/12$
- $= Pr(H \cap 6)$,

& we can see that these 2 events are independent of each other.

Mutually Exclusive

- Two events are mutually exclusive if you know that one occurred, then you know that the other could not have occurred.
- Example: You selected a student at random.
- A: The student is a male.
- B: The student is a female.
- These 2 events are mutually exclusive, because you know that if A occurred, B did not.

Mutually exclusive events are NOT independent!

- Remember that for independent events, knowing that one event occurred tells you nothing about whether the other occurred.

- For mutually exclusive events, knowing that one event occurred tells you that the other definitely did not occur!

The Birthday Problem

The Birthday Problem

- What is the probability that in a group of k people at least two people have the same birthday?

- (We are going to ignore leap day, which complicates the analysis, but doesn't have much effect on the answer.)

For our group of k people, let p = Pr(at least 2 people have the same birthday).

- At least 2 people having the same birthday is the complement (opposite) of no 2 people having the same birthday, or everyone having different birthdays.
- It’s easier to calculate the probability of different birthdays.
- So we can do that & then subtract the answer from one to get the probability we want.

p = 1- Pr(all different birthdays)

$$= 1 - \frac{(\#\text{ of ways k people can pick different bdays})}{(\#\text{ of ways k people can pick any bdays})}$$

p = 1- Pr(all different birthdays)

$$= 1 - \frac{(\text{\# of ways k people can pick different bdays})}{(\text{\# of ways k people can pick any bdays})}$$

$$= 1 - \frac{365}{}$$

p = 1- Pr(all different birthdays)

$$= 1 - \frac{(\#\text{ of ways k people can pick different bdays})}{(\#\text{ of ways k people can pick any bdays})}$$

$$= 1 - \frac{365 \bullet 364}{}$$

p = 1- Pr(all different birthdays)

$$= 1 - \frac{(\#\text{ of ways k people can pick different bdays})}{(\#\text{ of ways k people can pick any bdays})}$$

$$= 1 - \frac{365 \bullet 364 \bullet 363}{}$$

p = 1- Pr(all different birthdays)

$$= 1 - \frac{(\#\text{ of ways k people can pick different bdays})}{(\#\text{ of ways k people can pick any bdays})}$$

$$= 1 - \frac{365 \bullet 364 \bullet 363 \bullet 362 \bullet 361 \bullet \ldots (\text{until you have k \#s})}{}$$

p = 1- Pr(all different birthdays)

$$= 1 - \frac{(\text{\# of ways k people can pick different bdays})}{(\text{\# of ways k people can pick any bdays})}$$

$$= 1 - \frac{365 \bullet 364 \bullet 363 \bullet 362 \bullet 361 \bullet \ldots (\text{until you have k \#s})}{365}$$

p = 1- Pr(all different birthdays)

$$= 1 - \frac{(\#\text{ of ways k people can pick different bdays})}{(\#\text{ of ways k people can pick any bdays})}$$

$$= 1 - \frac{365 \bullet 364 \bullet 363 \bullet 362 \bullet 361 \bullet \ldots (\text{until you have k \#s})}{365 \bullet 365}$$

p = 1- Pr(all different birthdays)

$$= 1 - \frac{(\#\text{ of ways k people can pick different bdays})}{(\#\text{ of ways k people can pick any bdays})}$$

$$= 1 - \frac{365 \bullet 364 \bullet 363 \bullet 362 \bullet 361 \bullet \ldots (\text{until you have k \#s})}{365 \bullet 365 \bullet 365}$$

p = 1- Pr(all different birthdays)

$$= 1 - \frac{(\#\text{ of ways k people can pick different bdays})}{(\#\text{ of ways k people can pick any bdays})}$$

$$= 1 - \frac{365 \bullet 364 \bullet 363 \bullet 362 \bullet 361 \bullet \ldots (\text{until you have k \#s})}{365 \bullet 365 \bullet 365 \bullet 365 \bullet 365 \bullet \ldots (\text{until you have k \#s})}$$

- This is very messy, but you can calculate the answer for any number k.
- I have the answers computed for some sample values.

Birthday Problem Probabilities

k	p
5	0.027
10	0.117
15	0.253
20	0.411
22	0.476
23	0.507
25	0.569
30	0.706
40	0.891
50	0.970
100	0.9999997

Any Questions?