Probability and Random Processes

ECS 315

Asst. Prof. Dr. Prapun Suksompong

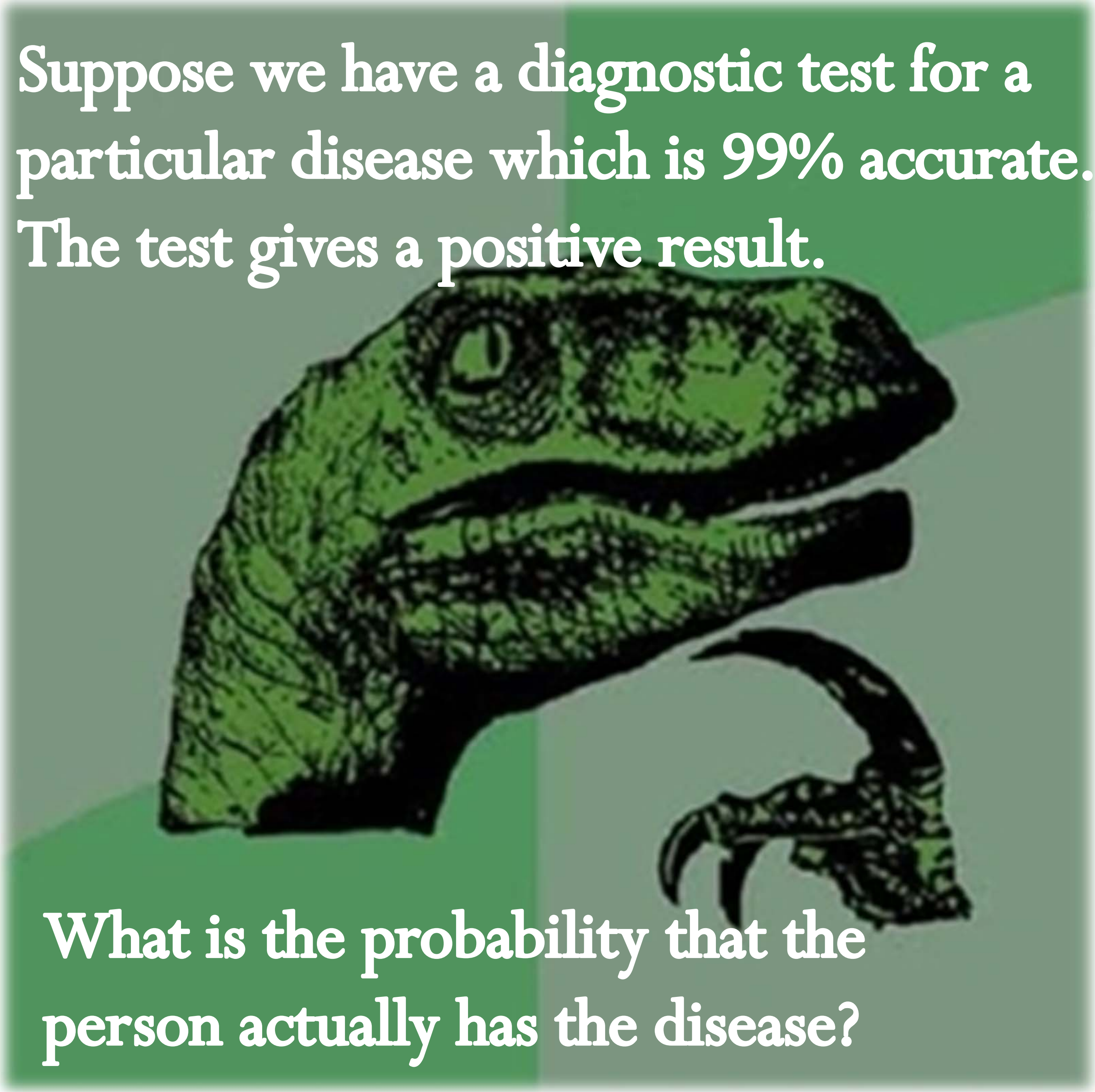
prapun@siit.tu.ac.th

6.1 Conditional Probability

Office Hours:

BKD, 6th floor of Sirindhralai building

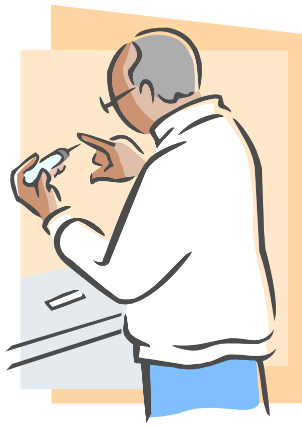
Tuesday	9:00-10:00
Wednesday	14:20-15:20
Thursday	9:00-10:00


Suppose we have a diagnostic test for a particular disease which is 99% accurate.
The test gives a positive result.
What is the probability that the person actually has the disease?

News: September 2015

Disease Testing

- Suppose we have a diagnostic test for a particular **disease** which is 99% accurate.
- A person is picked at random and tested for the disease.
- The test gives a **positive result**.
- Q1: What is the probability that the person actually has the disease?
- Natural answer: 99% because the test gets it right 99% of the times.



99% accurate test?

- Two kinds of error
- If you use this test on many persons **with** the disease, the test will indicate correctly that those persons have disease 99% of the time.
  - **False negative** rate = 1% = 0.01 | 1 → 0
- If you use this test on many persons **without** the disease, the test will indicate correctly that those persons do not have disease 99% of the time.
  - **False positive** rate = 1% = 0.01 | 0 → 1

Disease Testing: The Question

- Suppose we have a diagnostic test for a particular **disease** which is 99% accurate.
- A person is picked at random and tested for the disease.
- The test gives a **positive result**.
- Q1: What is the probability that the person actually has the disease?
- Natural answer: 99% because the test gets it right 99% of the times.
- Q2: Can the answer be 1% or 2%?
- Q3: Can the answer be 50%?

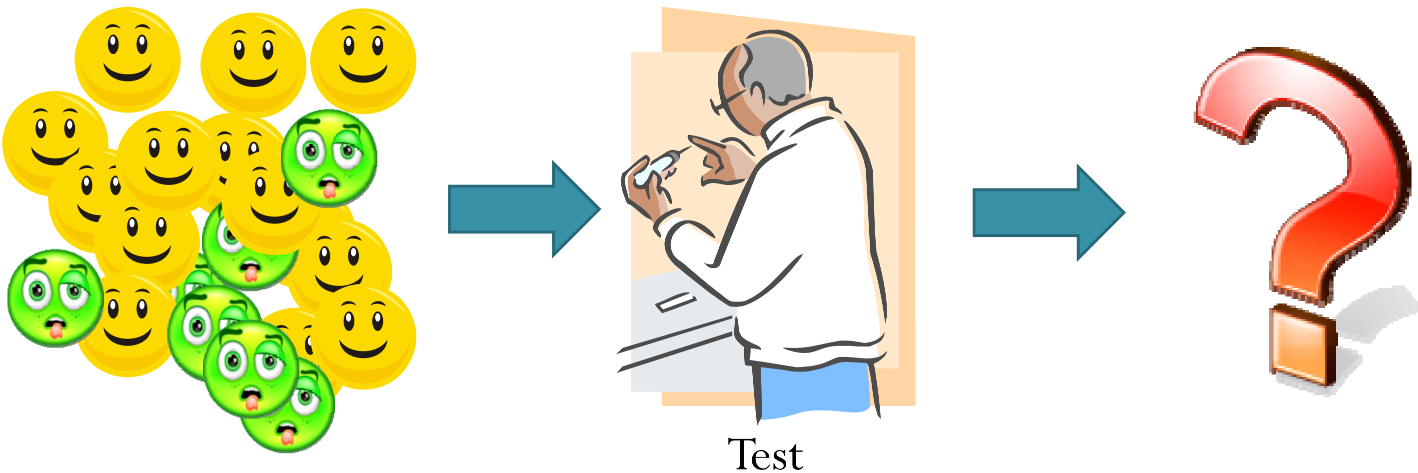
Disease Testing: The Answer

Q1: What is the probability that the person actually has the disease?

A1: The answer actually depends on how **common** or how **rare** the disease is!

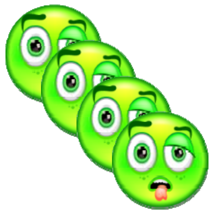
Why?

- Let's assume **rare disease**.
  - The disease affects about 1 person in 10,000.
- Try an experiment with **10^6 people**.
- Approximately **100 people** will have the disease.
- What would the (99%-accurate) test say?



10^6 people

Results of the test

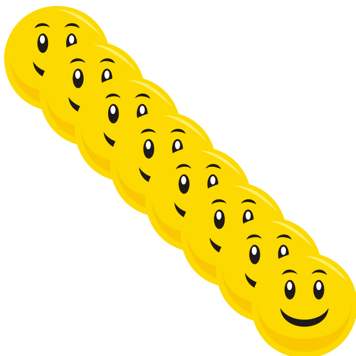


100 people w/ disease

approximately

99 of them will test positive

1 of them will test negative

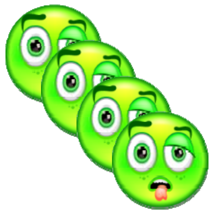


999,900 people w/o disease

989,901 of them will test negative

9,999 of them will test positive

Results of the test

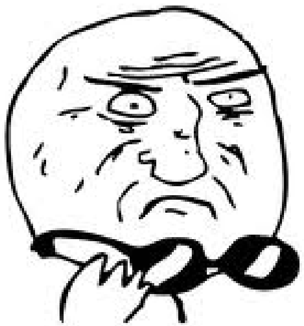


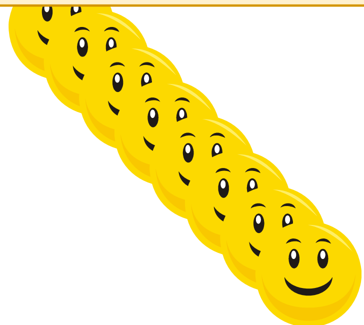
100 people w/ disease

99 of them will test positive

1 of them will test negative

Of those who test positive, only $\frac{99}{99+9,999} \approx 1\%$ actually have the disease!



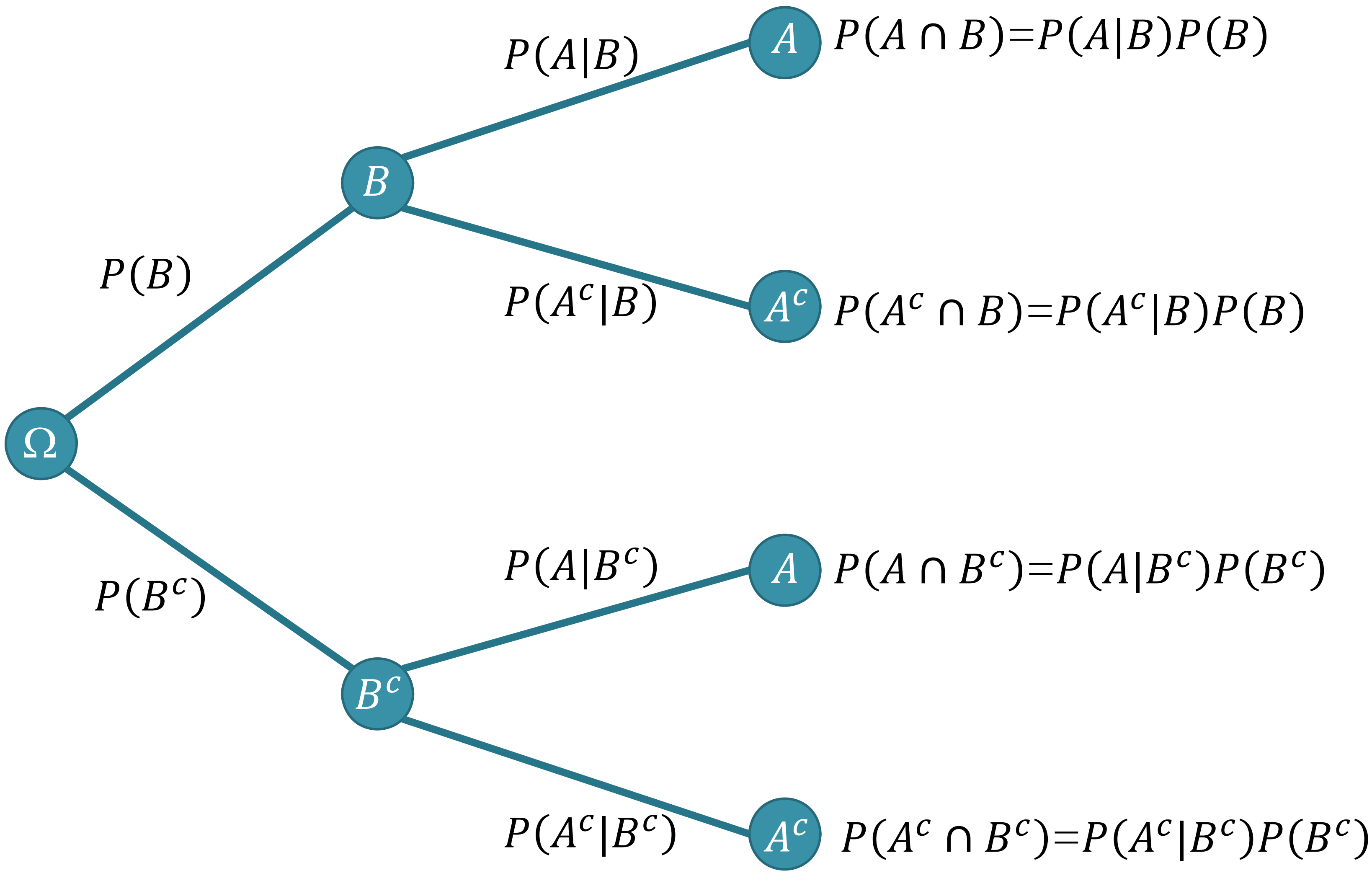


999,900 people w/o disease

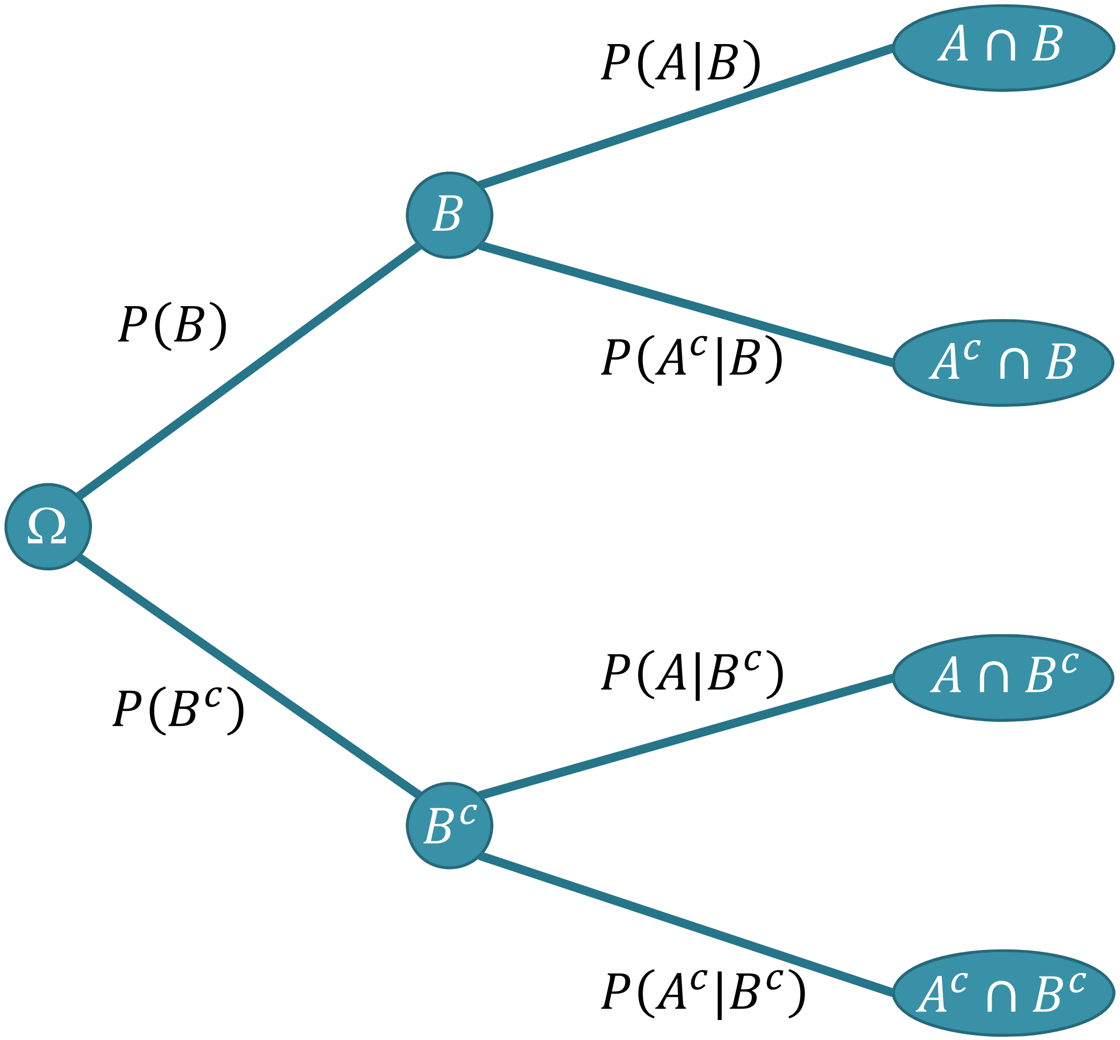
989,901 of them will test negative

9,999 of them will test positive

Tree Diagram and Conditional Probability



Tree Diagram and Conditional Probability



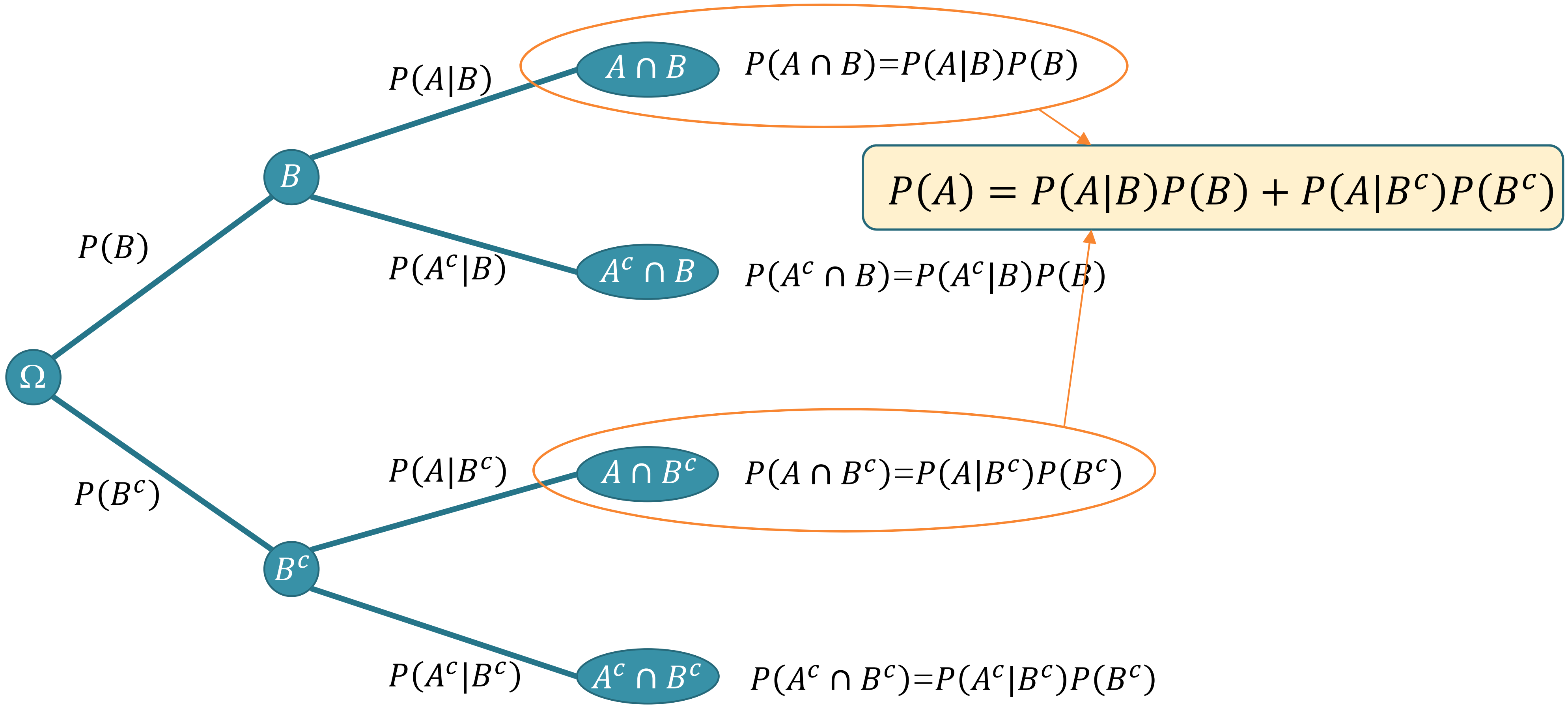
$P(A \cap B) = P(A|B)P(B)$

$P(A^c \cap B) = P(A^c|B)P(B)$

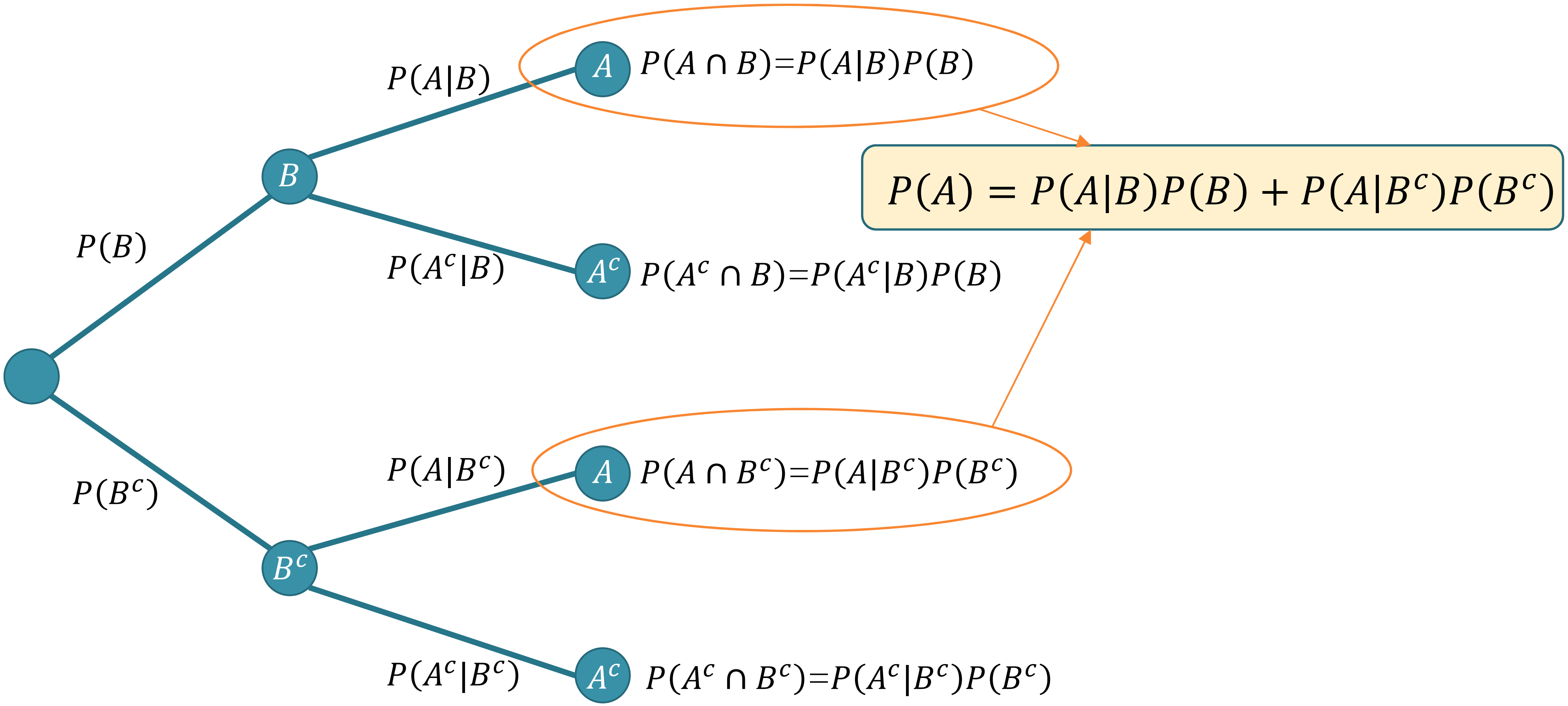
$P(A \cap B^c) = P(A|B^c)P(B^c)$

$P(A^c \cap B^c) = P(A^c|B^c)P(B^c)$

Tree Diagram and Total Probability Theorem

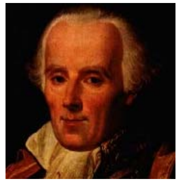


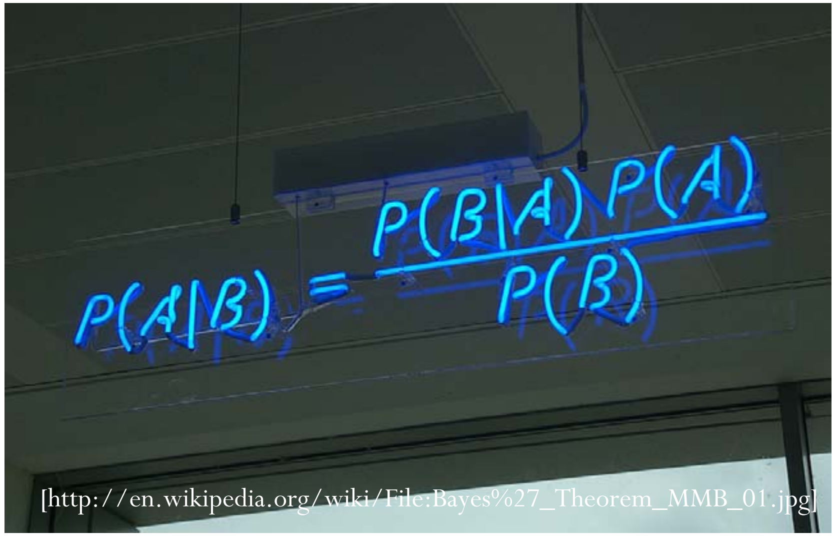
Tree Diagram and Total Probability Theorem



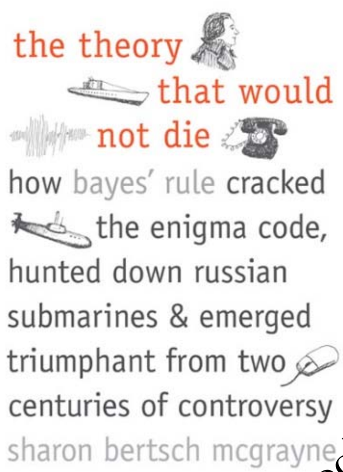
Bayes' Theorem: History

- Named after the **Thomas Bayes** (1701–61)
  - Father of mathematical decision making
- Bayes studied how to compute a distribution for the probability parameter of a binomial distribution in **1740s**. ≈ 250-year-old!
- His friend Richard **Price** edited and presented this work in 1763, after Bayes's death, as "An Essay towards solving a Problem in the Doctrine of Chances".
- **Laplace** independently rediscovered and extended Bayes' results in 1774.
  - Over the next forty years he developed it into the form we use today.

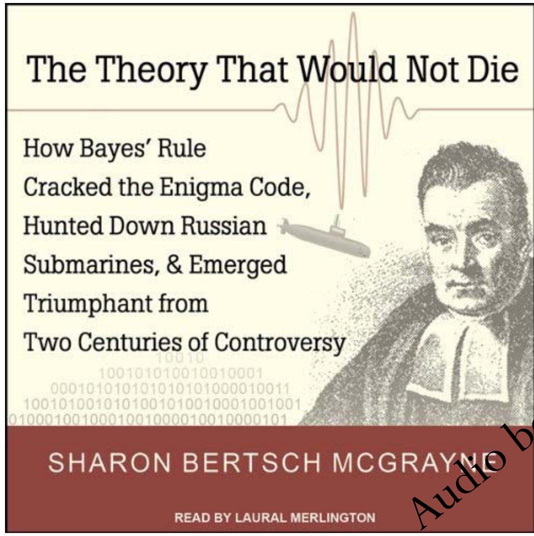




[http://en.wikipedia.org/wiki/File:Bayes%27_Theorem_MMB_01.jpg]



Book



Audio book

[http://www.bbc.com/news/uk-wales-24007735]

Bayes' Theorem: Scientific Battle

- An example of "science gone awry".
- The **scientific battle** over Bayes' theorem (Bayesian analysis) is lasted for **150 years**.
  - Respected statisticians rendered it **professionally taboo**
    - while practitioners relied on it to solve problems
  - Similar case: Geologists accumulated the evidence for Continental Drift in 1912 and then spent 50 years arguing that continents cannot move.
- Sometime during the 1740s, Bayes made this discovery but then mysteriously abandoned it.
  - Bayes' theorem began life amid an inflammatory religious controversy in England in the 1740s: can we make rational conclusions about God based on evidence about the world around us?
- Laplace gave it its modern mathematical form and scientific application and then moved on to other methods.

Bayes' Theorem

Using the concept of **conditional probability** and **Bayes' Theorem**, we can show that

> the probability that a (randomly selected) person will have the disease (defined as event D) *given that* the test result (for that person) is positive (defined as event T_P)

is given by

$$P(D|T_P) = \frac{\boldsymbol{P(T_P|D)}P(D)}{\boldsymbol{P(T_P|D)}P(D) + \boldsymbol{P(T_P|D^c)}P(D^c)}$$

$$= \frac{\boldsymbol{P(T_P|D)}P(D)}{\boldsymbol{P(T_P|D)}P(D) + \left(1 - \boldsymbol{P(T_P^c|D^c)}\right)\left(1 - P(D)\right)}$$

Positive Predictive Values (PPV)

		Reality	
		Have disease	No disease
Test outcome	+	Sensitivity (True Positive) $P(T_P \mid D)$	False Positive (Type I Error) $P(T_P \mid D^c)$
	-	False Negative (Type II Error) $P(T_P^c \mid D)$	Specificity (True Negative) $P(T_P^c \mid D^c)$

$P(D)$ = prevalence

PPV:

$$P(D|T_P) = \frac{P(T_P|D)P(D)}{P(T_P|D)P(D) + P(T_P|D^c)P(D^c)}$$

$$= \frac{P(T_P|D)P(D)}{P(T_P|D)P(D) + \left(1 - P(T_P^c|D^c)\right)\left(1 - P(D)\right)}$$

In our example,

		Reality	
		Have disease	No disease
Test outcome	+	Sensitivity (True Positive) $P(T_P \mid D) = 1 - p_{TE} = 0.99$	False Positive (Type I Error) $P(T_P \mid D^c) = p_{TE} = 0.01$
	-	False Negative (Type II Error) $P(T_P^c \mid D) = p_{TE} = 0.01$	Specificity (True Negative) $P(T_P^c \mid D^c) = 1 - p_{TE} = 0.99$

PPV:

$$P(D \mid T_P) = \frac{P(T_P \mid D)P(D)}{P(T_P \mid D)P(D) + P(T_P \mid D^c)P(D^c)}$$

$$= \frac{P(T_P \mid D)P(D)}{P(T_P \mid D)P(D) + \left(1 - P(T_P^c \mid D^c)\right)\left(1 - P(D)\right)}$$

$$= \frac{(1 - p_{TE})p_D}{(1 - p_{TE})p_D + p_{TE}(1 - p_D)}$$

$$P(D) \equiv p_D$$

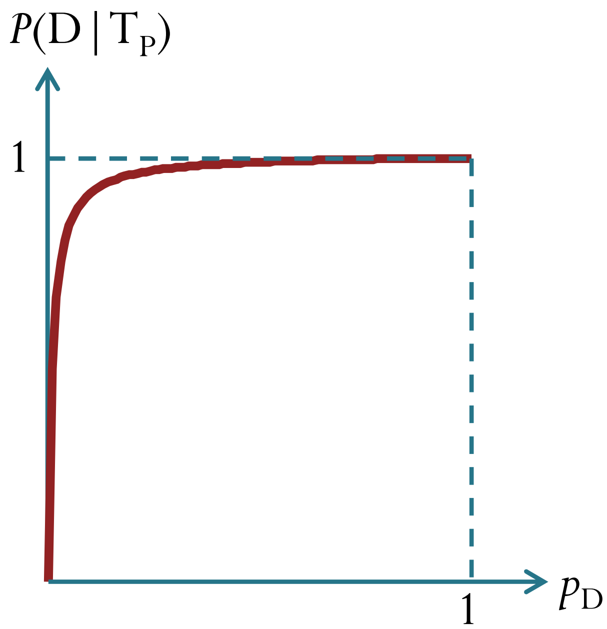
In our example,

When different value of p_D is assumed,

We get different value of $P(D \mid T_P)$.

<u>Conclusion</u>: ***Any*** value (between 0 and 1)

can be obtained by varying the value of p_D.

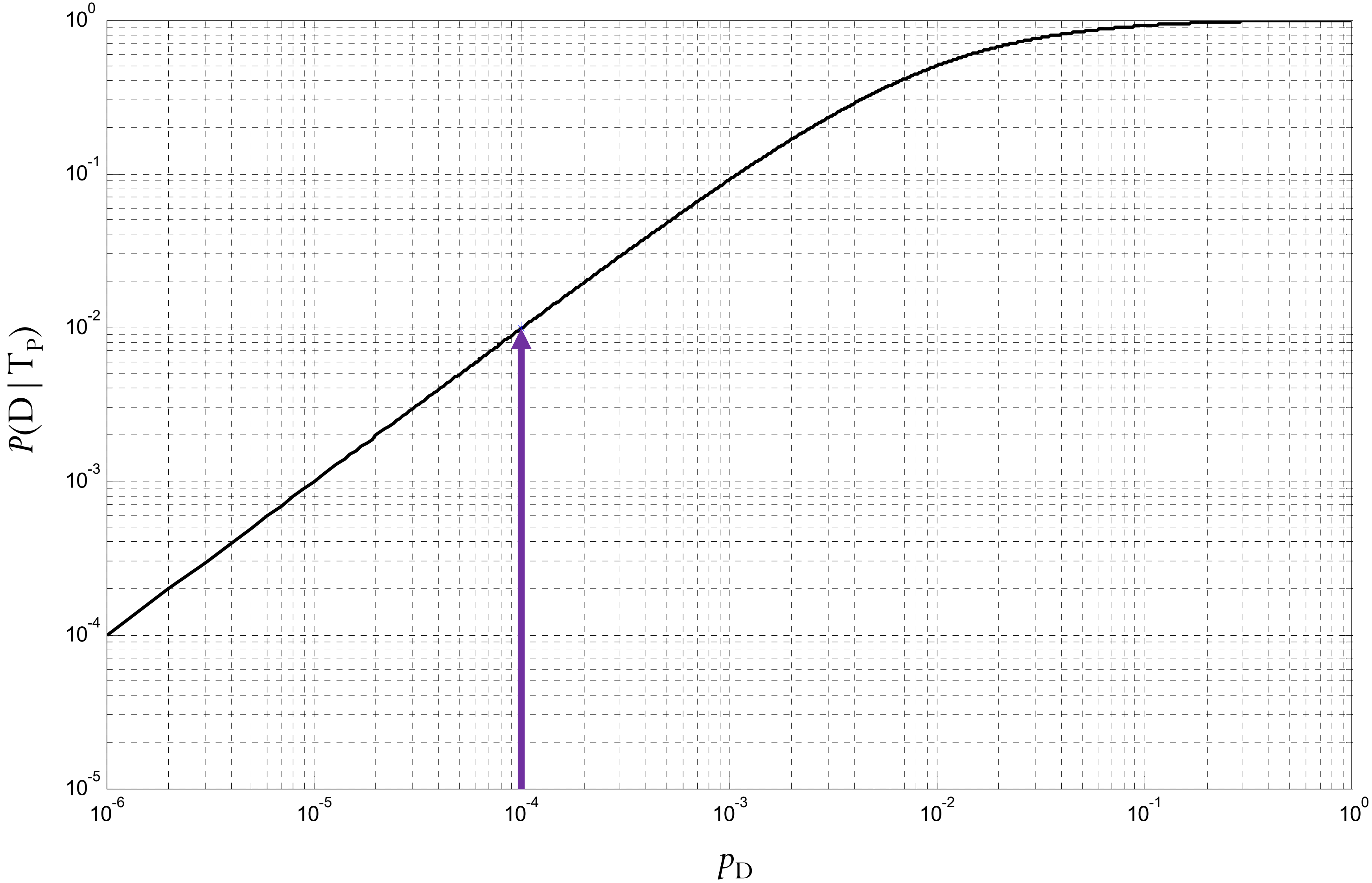


PPV: $$P(D|T_P) = \frac{P(T_P|D)P(D)}{P(T_P|D)P(D) + P(T_P|D^c)P(D^c)}$$

$$= \frac{P(T_P|D)P(D)}{P(T_P|D)P(D) + \left(1 - P(T_P^c|D^c)\right)\left(1 - P(D)\right)}$$

$$= \frac{(1 - p_{TE})p_D}{(1 - p_{TE})p_D + p_{TE}(1 - p_D)}$$

In log scale…



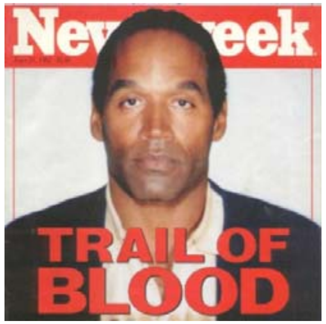
Wrap-up

- Q1: What is the probability that the person actually has the disease?
- A1: The answer actually depends on how common or how rare the disease is! (The answer depends on the value of P_D.)
- Q2: Can the answer be 1% or 2%?
- A2: Yes.
- Q3: Can the answer be 50%?
- A3: Yes.

Prosecutor's fallacy

- Criminal trial for murder
  - "one of the biggest media events of 1994–95"
  - "the most publicized criminal trial in American history"
    (การพิจารณาคดีในศาล)
  - Often characterized as "the trial of the century"

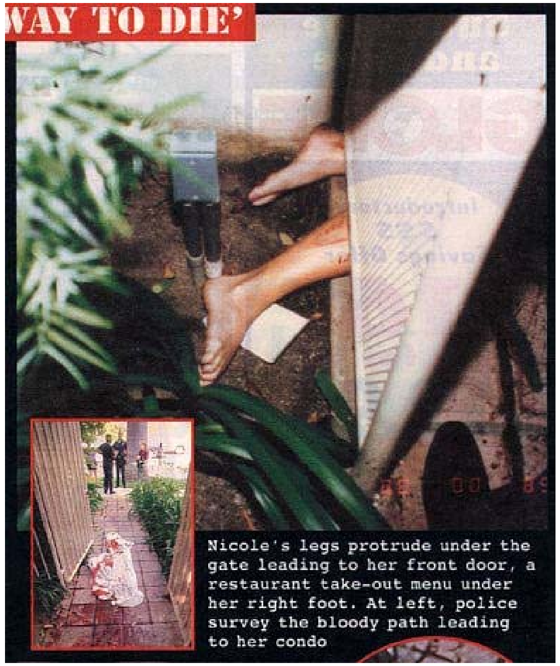
- O. J. Simpson
  - At the time a well-known celebrity famous both as a **TV actor** and as a retired **professional football star**.
- Defense lawyer (ทนาย): Alan Dershowitz
  - Renowned attorney and **Harvard Law School professor**


[Mlodinow, 2008, p. 119-121],[Tijms, 1007, Ex 8.7]

The murder of Nicole

- **Nicole** Brown was murdered at her home in Los Angeles on the night of June 12, 1994.
  - So was her friend Ronald Goldman.
- The **prime suspect** (ผู้ต้องสงสัย) was her (ex-) **husband** O.J. Simpson.
  - (They were divorced in 1992.)



Prosecutor* = a government official who conducts criminal prosecutions on behalf of the state (เป็นฝ่ายผู้ฟ้องร้อง/โจทก์)
(พนักงานอัยการ)

Prosecutors' argument

- Prosecutors* spent the first ten days of the trial entering **evidence** of Simpson's history of **physically abusing** her and claimed that this alone was a good reason to suspect him of her murder.
- As they put it,
  "a slap is a prelude to homicide." (ฆาตกรรม)

Counterargument

- The **defense attorneys** (ทนายฝ่ายจำเลย) argued
  - that the prosecution* had spent two weeks trying to **mislead** the jury
  - and that the **evidence** that O. J. had battered (ทุบตี) Nicole on previous occasions **meant nothing**.
- **Dershowitz's reasoning**:
  - 4 million women are battered annually by husbands and boyfriends in the US.
  - In 1992, a total of 1,432, or 1 in 2,500, were killed by their (ex)husbands or boyfriends.
  - Therefore, few men who slap or beat their domestic partners go on to murder them.
- True? …Yes…Convincing?

The verdict:

Not guilty for the two murders!

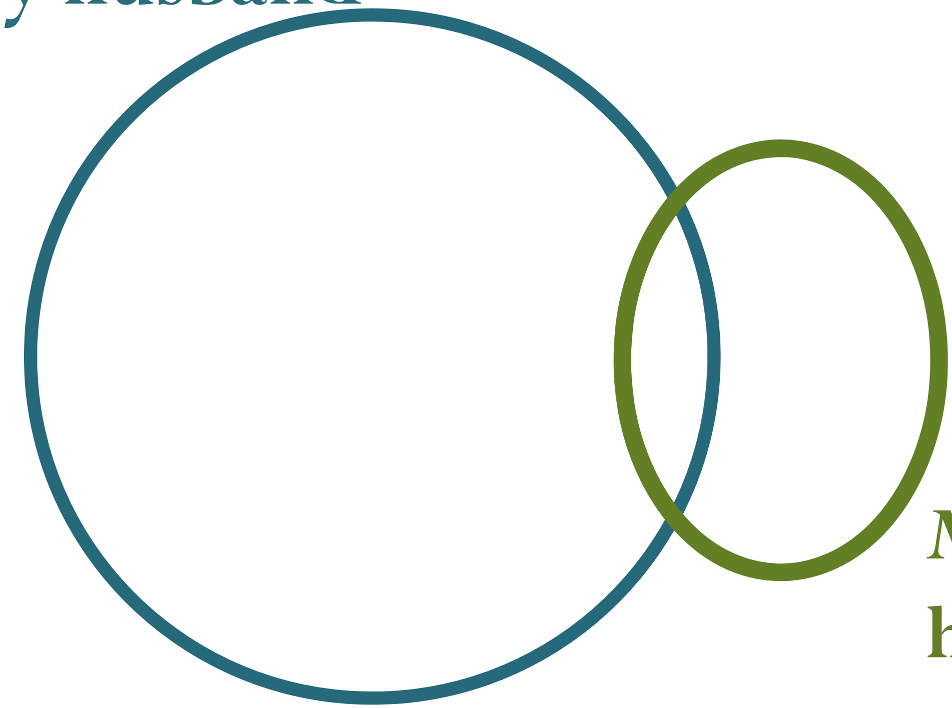
The verdict was seen live on TV by more than half of the U.S. population, making it one of the most watched events in American TV history.

The Truth: Another number…

- It is important to make use of the crucial fact that Nicole Brown was murdered.
- The relevant number is not the probability that a man who batters his wife will go on to kill her (1 in 2,500) but rather the probability that a battered wife who was murdered was murdered by her abuser.

This event has happened and should be used in probability evaluation

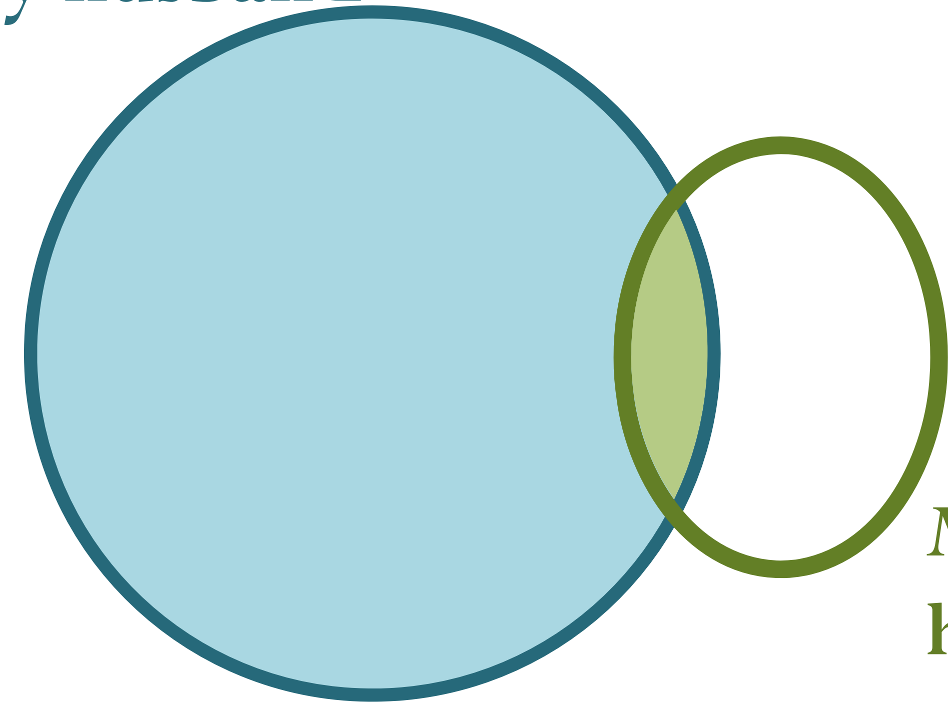
A Simplified Diagram



A Simplified Diagram

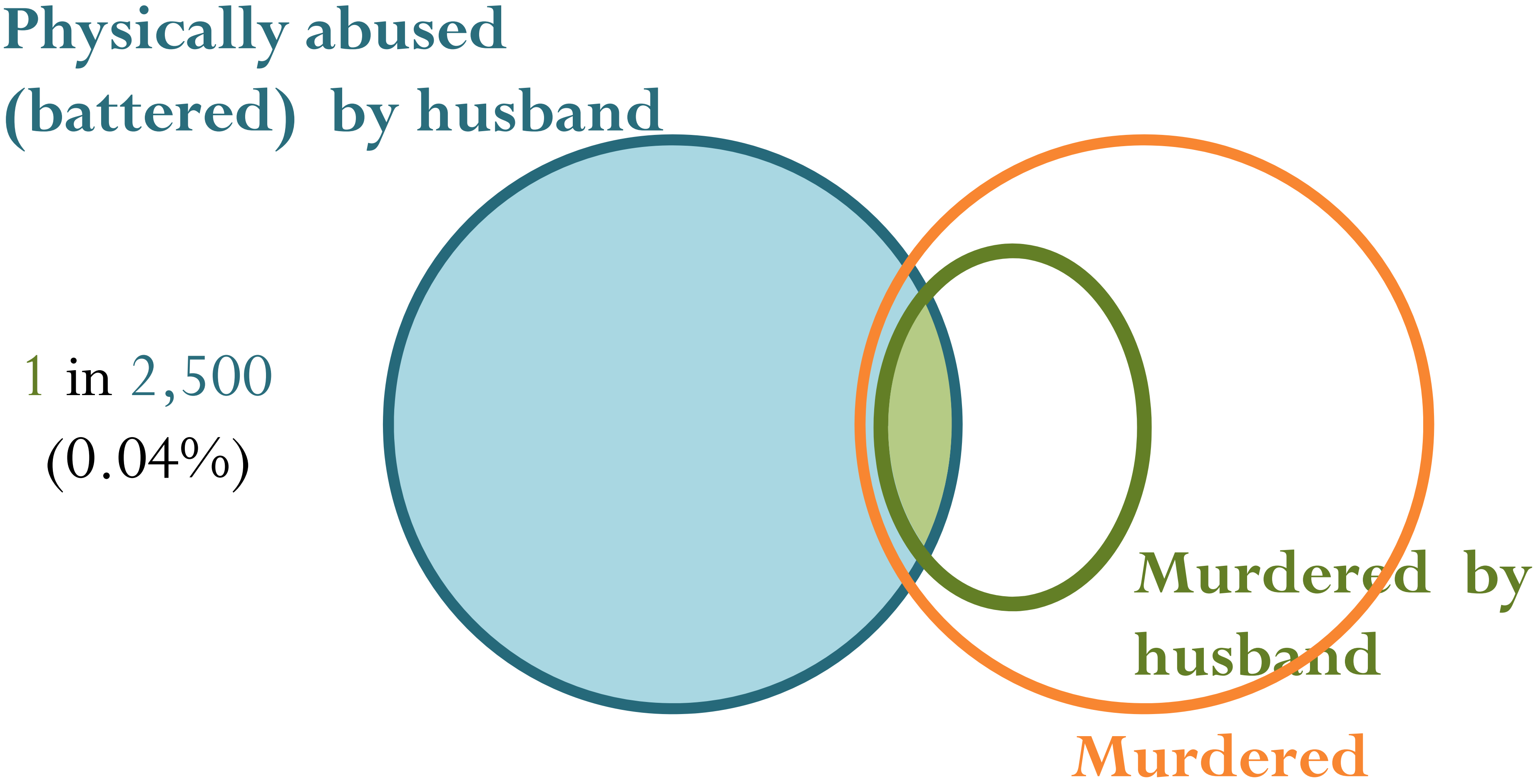
Physically abused (battered) by husband

1 in 2,500
(0.04%)



Murdered by husband

A Simplified Diagram



The Truth: Another number…

- It is important to make use of the crucial fact that Nicole Brown was murdered.
- The relevant number is not the probability that a man who batters his wife will go on to kill her (1 in 2,500) but rather the probability that a battered wife who was murdered was murdered by her abuser.

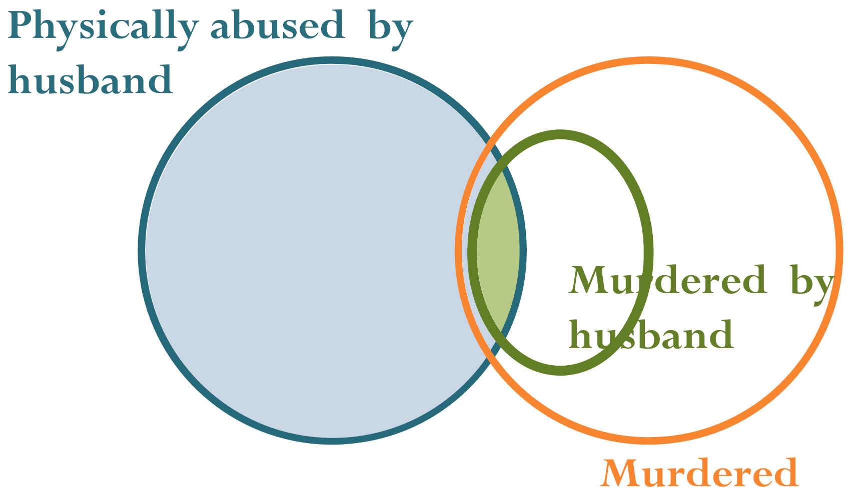
This event has happened and should be used in probability evaluation

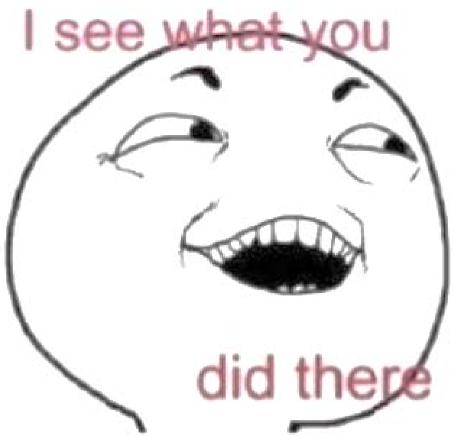
- According to the Uniform Crime Reports for the United States and Its Possessions in 1993, the probability Dershowitz (or the prosecution) should have reported was this one: of all the battered women murdered in the United States in 1993, some 90 percent were killed by their abuser.
- That statistic was **not mentioned at the trial**.

Probability Comparison

The orange event is ignored.

1 in 2,500
(0.04%)





90%