CM 3.1:
Probability Review

In 1990, the following question appeared
in Parade Magazine’s “Ask Marilyn” column:

Suppose you’re on a game show, and you’re given the choice of three doors. Behind one door is a car, behind the others, goats. You pick a door, say #1, and the host, who knows what’s behind the doors, opens another door, say #3, which has a goat. He says to you, "Do you want to pick door #2?"

Is it to your advantage to switch your choice of doors?

Craig F. Whitaker
Columbia, Maryland

Yes; you should SWITCH

-- Marilyn vos Savant

http://marilynvossavant.com/game-show-problem/

What's the Best Strategy?

- If you always switch doors after Monty Hall reveals a goat, then your odds of winning are two-in-three, or 66.7 percent on average. If you keep your original choice, your chances of winning are just one-in-three, or 33 percent on average.
- That seems weird, because after Monty reveals a goat, there are two closed doors left, and it might seem as if there should be a 50-50 chance that the car is behind either door.

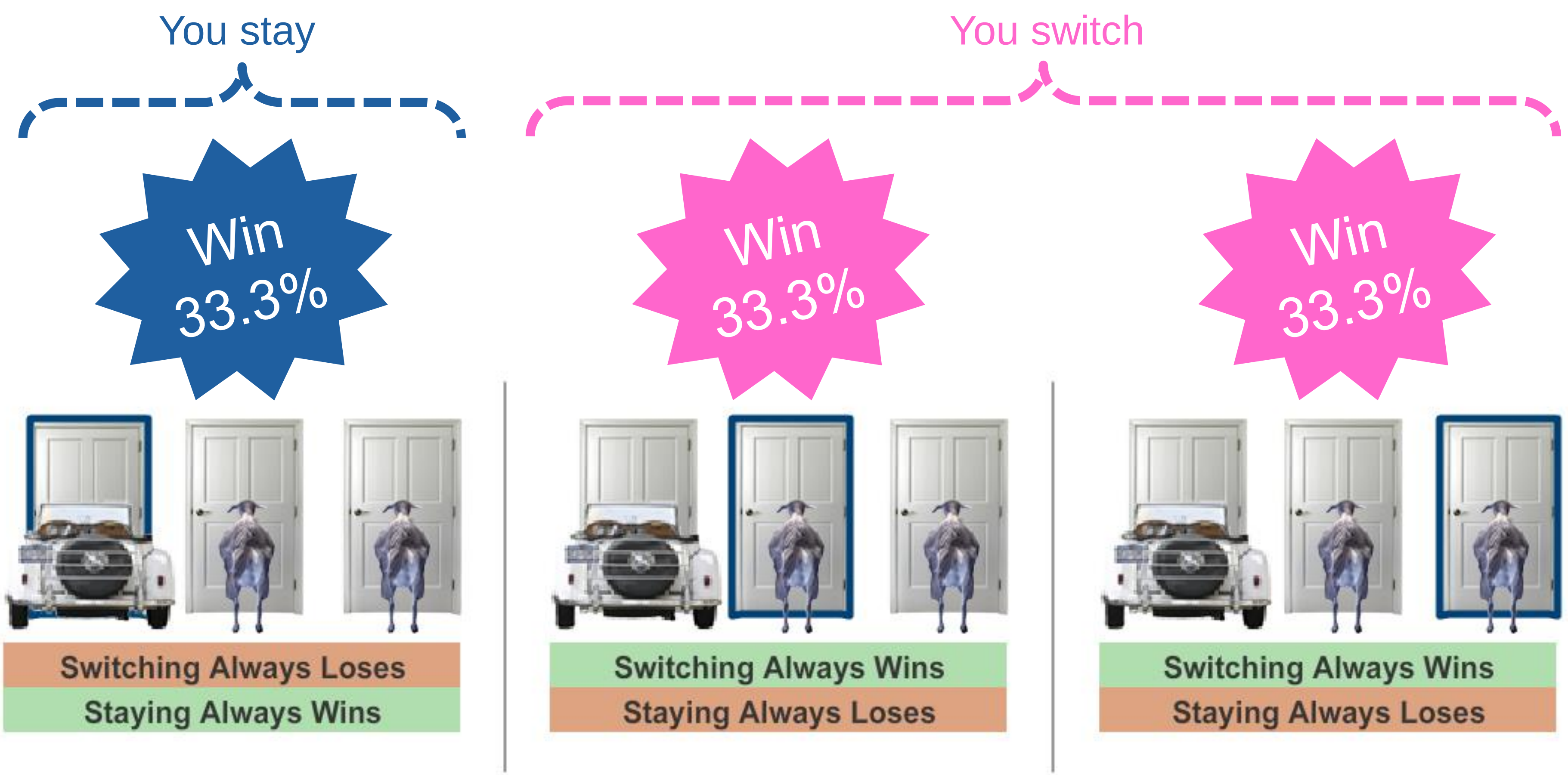
How to lose by switching

- To help explain, let's look at the situation from the other side, so we have as much information as Monty Hall does. The critical aspect of the problem is that Monty Hall always opens a door to reveal one of the goats.
- If you correctly chose the door with the car at the start, he can open either of the other doors to reveal a goat. If you accept his offer to switch doors, you will switch away from your winning choice and end up with a goat. So far, switching doesn't sound like a winning strategy.

How to win by switching

- But look what happens if your initial choice of door was hiding a goat. If you picked the middle door, Monty Hall opens the third door to show you a goat. In this case, if you switch doors, you switch to the door hiding the car.
- The same situation applies if you chose the third door initially. (Remember, Monty Hall knows where the car is and needs to open a door that will reveal a goat.) Again, switching from your initial choice to the other closed door means you trade a goat for a car.

- If your strategy is to always switch doors, you will lose only if your initial choice is the door with the car, which is a 33.3 percent chance. In the other two cases (66.7 percent of the time) you will switch to the car and walk away a winner.



“Our brains are just not

WIRED

to do probability problems well.”

-- Persi Diaconis
Professor of Statistics & Mathematics
Stanford University

http://www.nytimes.com/1991/07/21/us/behind-monty-hall-s-doors-puzzle-debate-and-answer.html

As a professional mathematician, I'm very concerned with the general public's lack of mathematical skills. Please help by confessing your error and in the future being more careful.
Robert Sachs, Ph.D.
George Mason University

You made a mistake, but look at the positive side. If all those Ph.D.'s were wrong, the country would be in some very serious trouble.
Everett Harman, Ph.D.
U.S. Army Research Institute

May I suggest that you obtain and refer to a standard textbook on probability before you try to answer a question of this type again?
Charles Reid, Ph.D.
University of Florida

I am sure you will receive many letters on this topic from high school and college students. Perhaps you should keep a few addresses for help with future columns.
W. Robert Smith, Ph.D.
Georgia State University

"I wrote her another letter telling her that after removing my foot from my mouth I'm now eating

humble pie.

I vowed as penance to answer all the people who wrote to castigate me. It's been an intense professional

embarrassment."

--Dr. Robert Sachs

http://www.nytimes.com/1991/07/21/us/behind-monty-hall-s-doors-puzzle-debate-and-answer.html

What is probability?

- A measure of chance that something will occur.
- Probability theory is the mathematical machinery necessary to answer questions about uncertain events that occur randomly.
- As scientists/engineers, we need to be more precise…

Random Experiment:

1. An experiment, trial, or observation that can be repeated numerous times under the same conditions.
2. The outcome of an individual random experiment must be independent and identically distributed.
3. It must in no way be affected by any previous outcome and cannot be predicted with certainty.

http://cnx.org/content/m13470/latest/

Random Experiment:

1. Experiment is **repeatable** (ideally)
2. Outcomes are **iid**
3. Outcomes are **uncertain**

http://cnx.org/content/m13470/latest/

What does iid mean?

independent
identically
distributed

- What it does not mean:
  - Independent ≠ uncorrelated
  - “identically distributed” ≠ equally likely

$$(\Omega, \mathcal{F}, P)$$

Probability Triple

The probability triple

a random experiment!!!

- From Wikipedia:

"a probability triple is a mathematical construct that models a real-world process (or "experiment") consisting of states that occur randomly."

- A probability triple has (predictably) 3 parts:

Ω	Sample space	The set of outcomes that we are sampling from.
$\mathcal{F}$	Set of events	What are the sane questions I can ask about this probability distribution?
$\mathbf{P}$	Probability measure	Function that maps elements from $\mathcal{F}$ to the interval [0, 1]

The probability triple

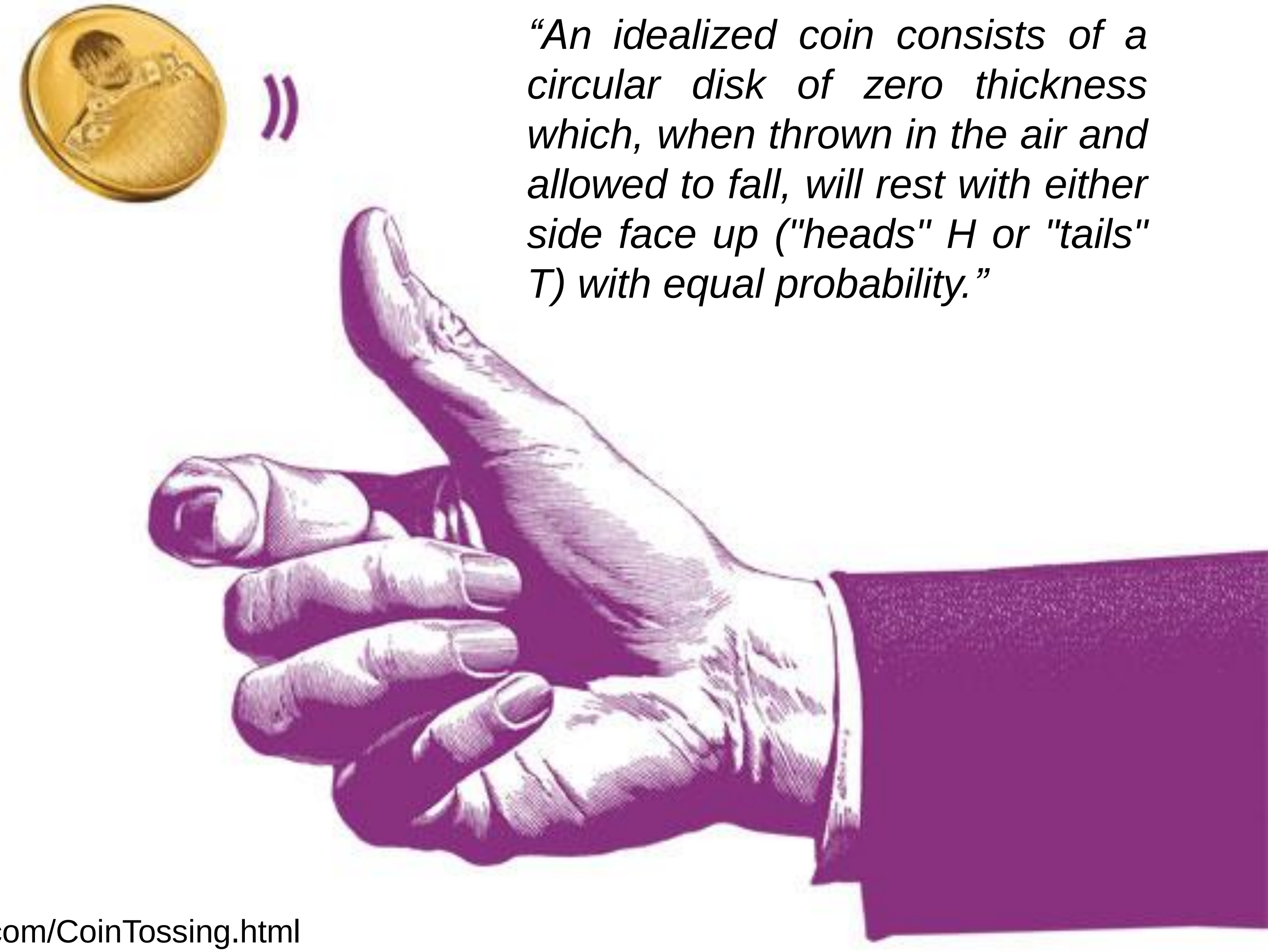
- Given a random experiment…

$\mathbf{\Omega}$	Sample space	What can possibly happen?
$\mathcal{F}$	Set of events	What are the sane questions I can ask about this probability distribution?
$\mathbf{P}$	Probability measure	Function that maps elements from $\mathcal{F}$ to the interval [0,1]

Sample space (Ω)

- The sample space is the set containing all possible outcomes from a random experiment
- “What can possibly happen (in the context of the random experiment?”
- Possible ≠ probable

Random experiment: toss a fair coin 2 times



“An idealized coin consists of a circular disk of zero thickness which, when thrown in the air and allowed to fall, will rest with either side face up ("heads" H or "tails" T) with equal probability.”

http://mathworld.wolfram.com/CoinTossing.html

Sample space (Ω)

“What can possibly happen?”

Sample space (Ω)

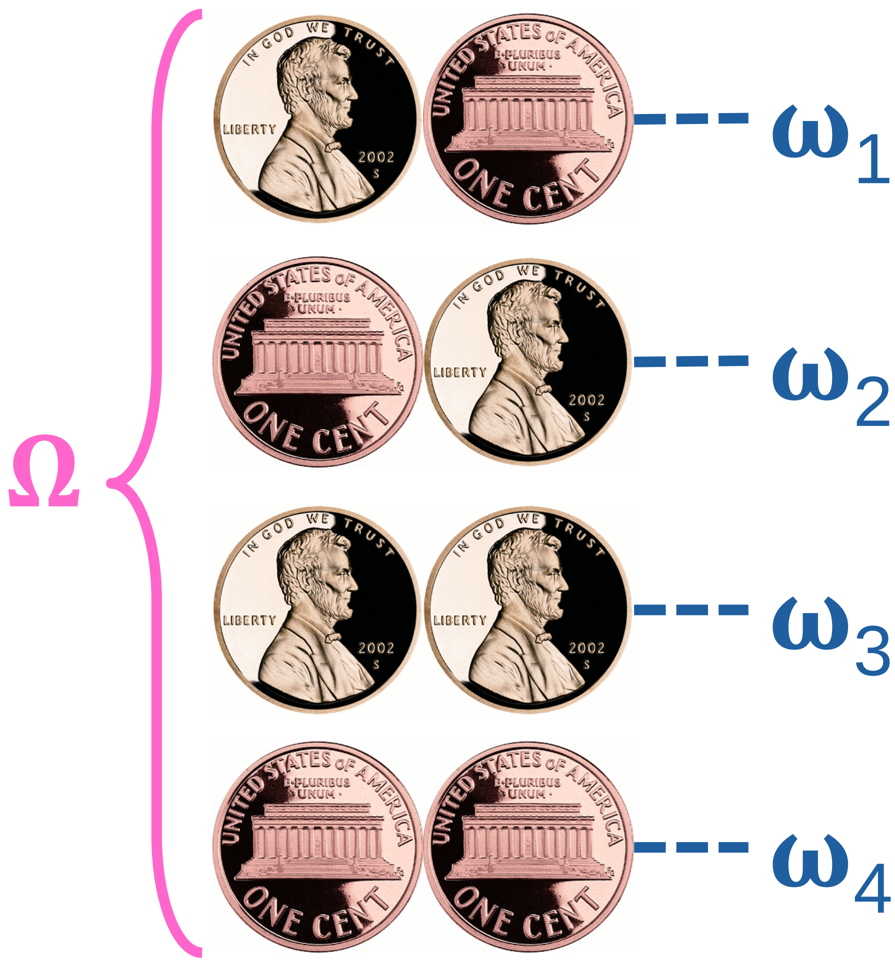
“What can possibly happen?”

“What can not possibly happen?”

$$(\Omega, \mathcal{F}, P)$$

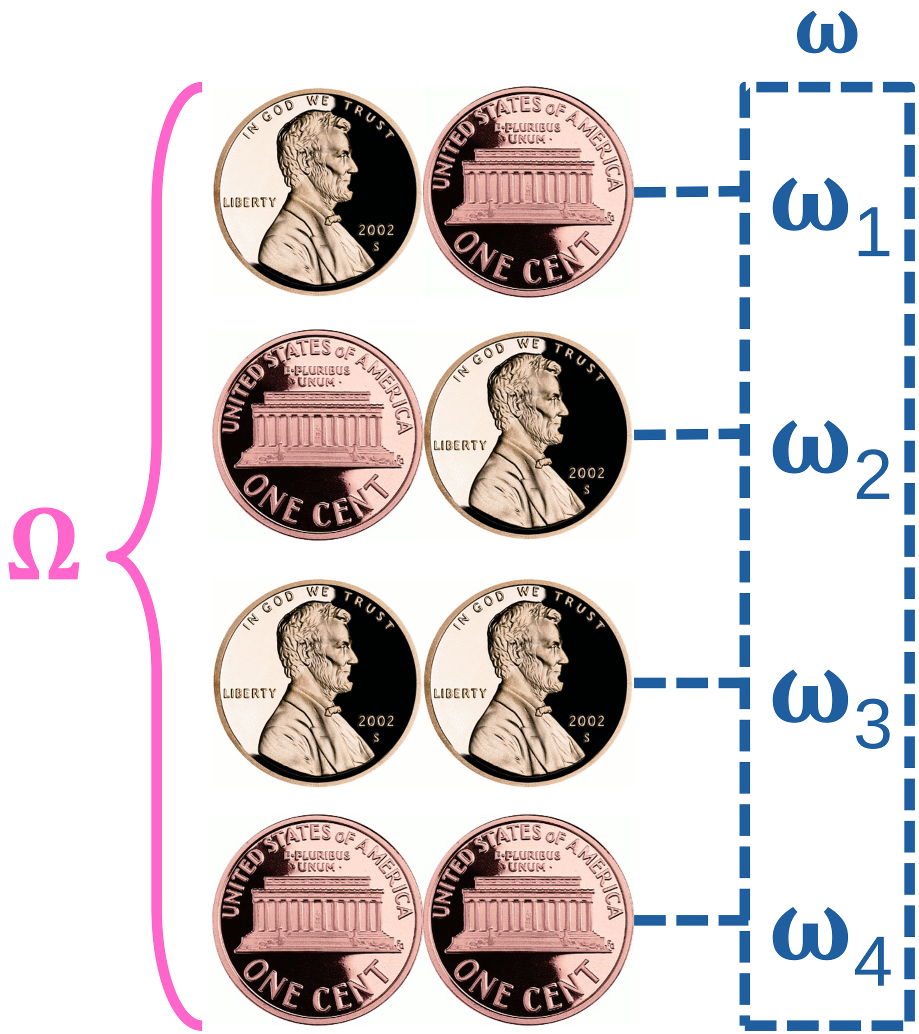
Probability Triple

Outcomes (ω)



One outcome is one element in the sample space, $\boldsymbol{\omega}_{1,\dots,n} \in \boldsymbol{\Omega}$

Outcomes (ω)

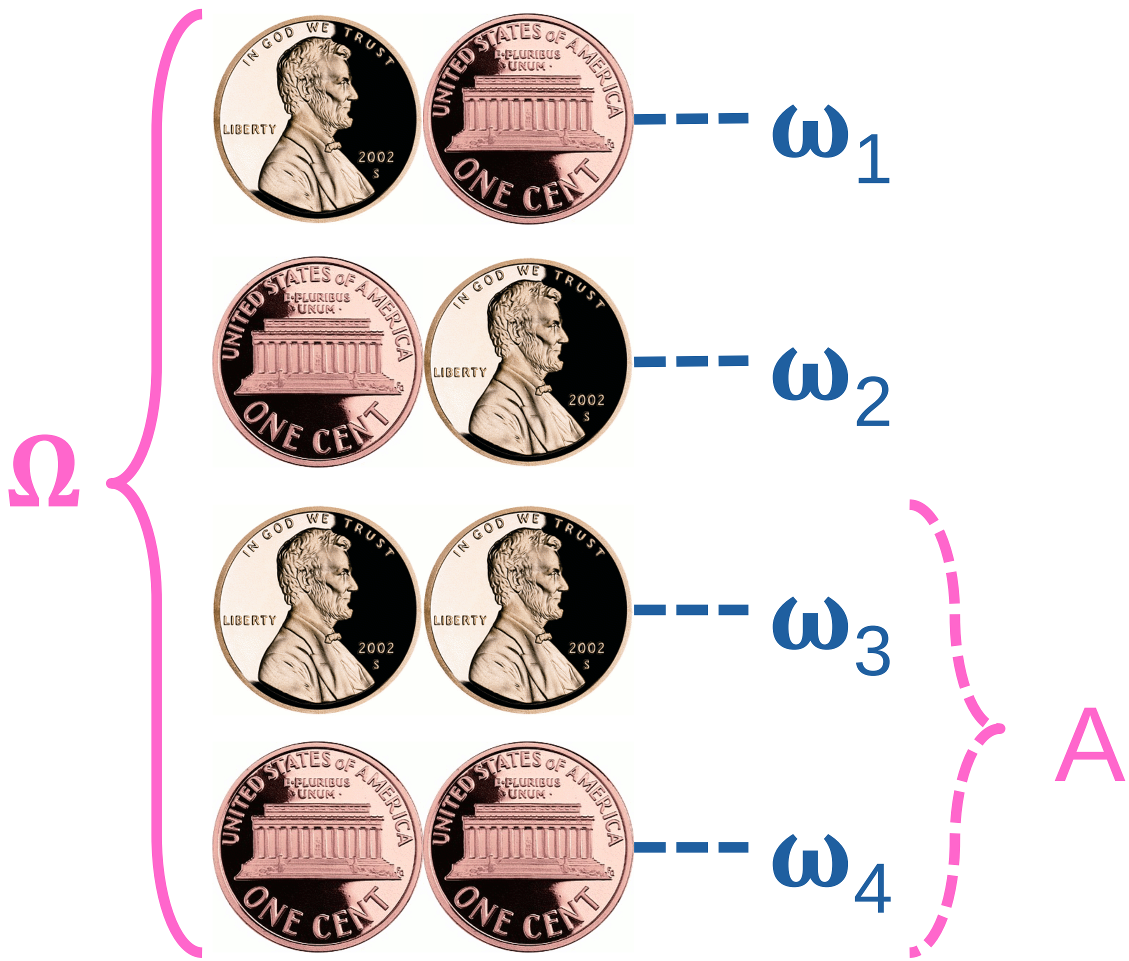


The set of all outcomes is denoted ω such that $\omega \in \Omega$

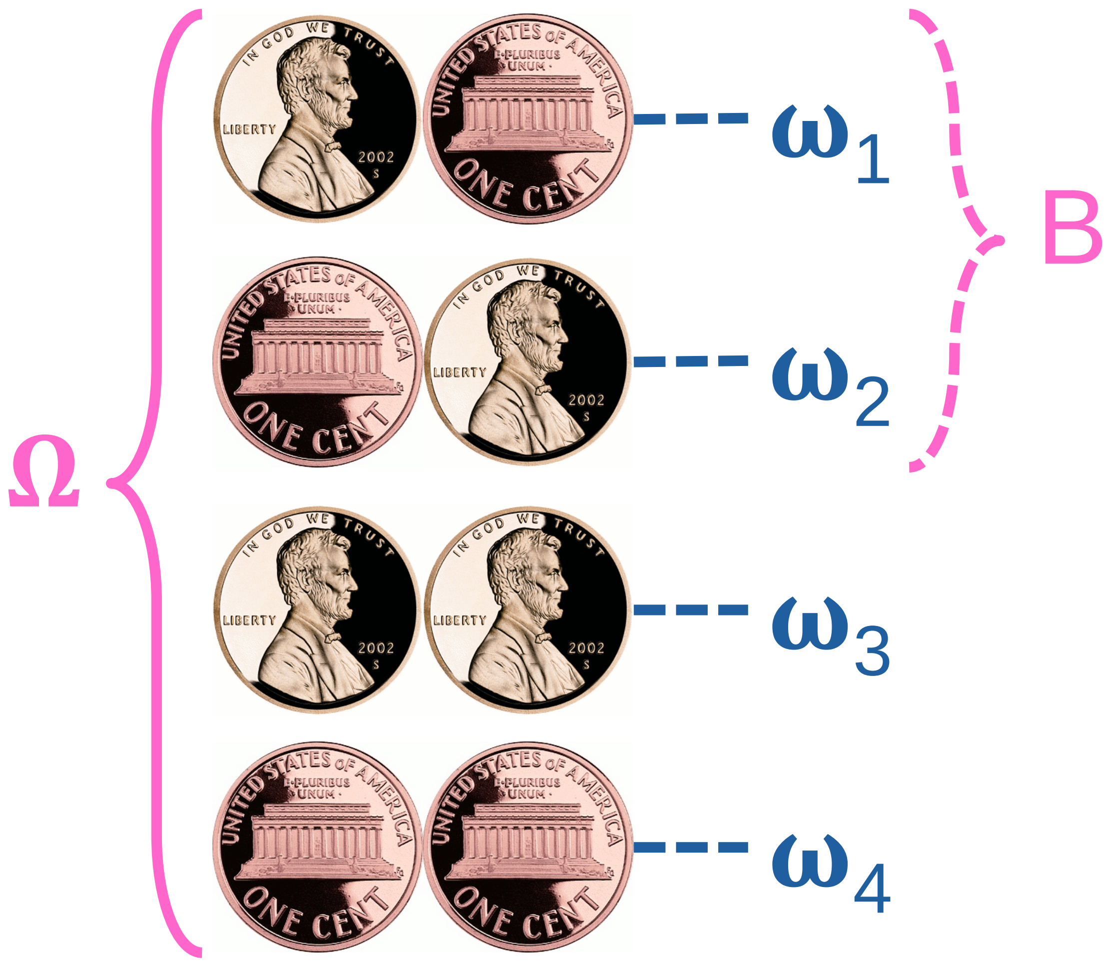
Events (A, B, etc.)

- An **event** is a set of outcomes to which a probability is assigned
  - Notation: capital letters (i.e., A)
  - The actual letter means nothing
- By definition, set of events is a subset of the sample space
  - $A \subseteq \mathbf{\Omega}$

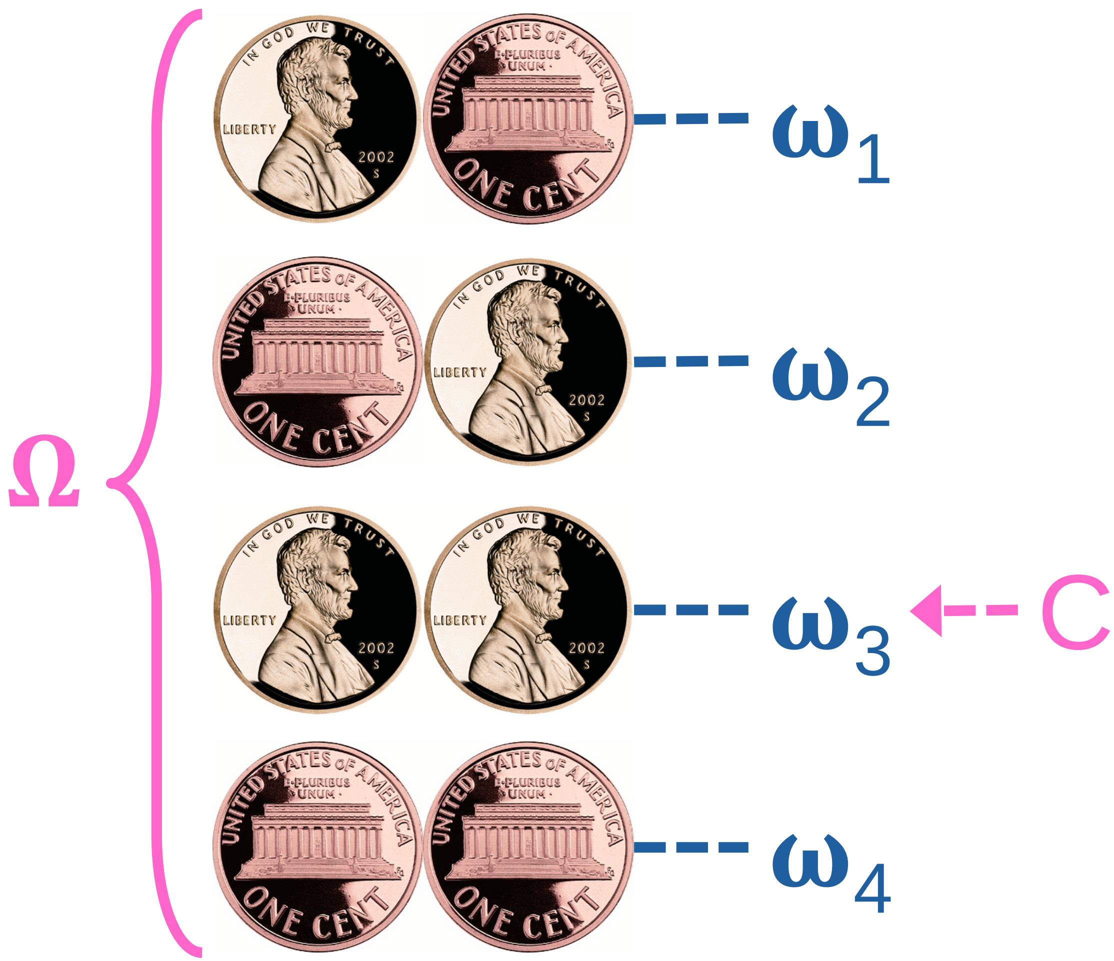
Event A: same result on both flips



Event B: exactly one head



Event C: two heads

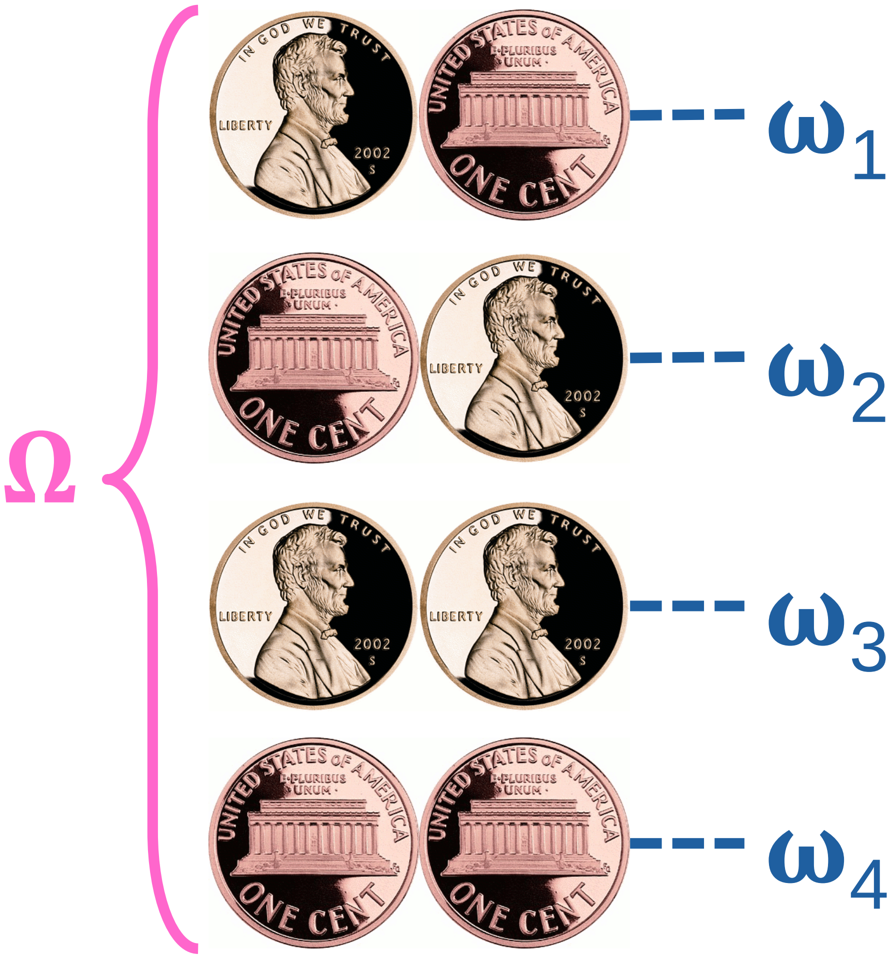


C is an elementary event

$$(\Omega, \mathcal{F}, P)$$

Probability Triple

Pictures are difficult to do math with…



Letters are also difficult to do math with…

1st flip f_1 2nd flip f_2

Ω

$\omega_1 = \{f_1 = H \cap f_2 = T\}$

$\omega_2 = \{f_1 = T \cap f_2 = H\}$

$\omega_3 = \{f_1 = H \cap f_2 = H\}$

$\omega_4 = \{f_1 = T \cap f_2 = T\}$

Letters are also difficult to do math with…

1^{st} flip f_1 | 2^{nd} flip f_2

Ω

$\omega_1 = \{f_1 = H \cap f_2 = T\}$

$\omega_2 = \{f_1 = T \cap f_2 = H\}$

$\omega_3 = \{f_1 = H \cap f_2 = H\}$

$\omega_4 = \{f_1 = T \cap f_2 = T\}$

You need a function that maps outcomes and events onto real numbers

RANDOM VARIABLES

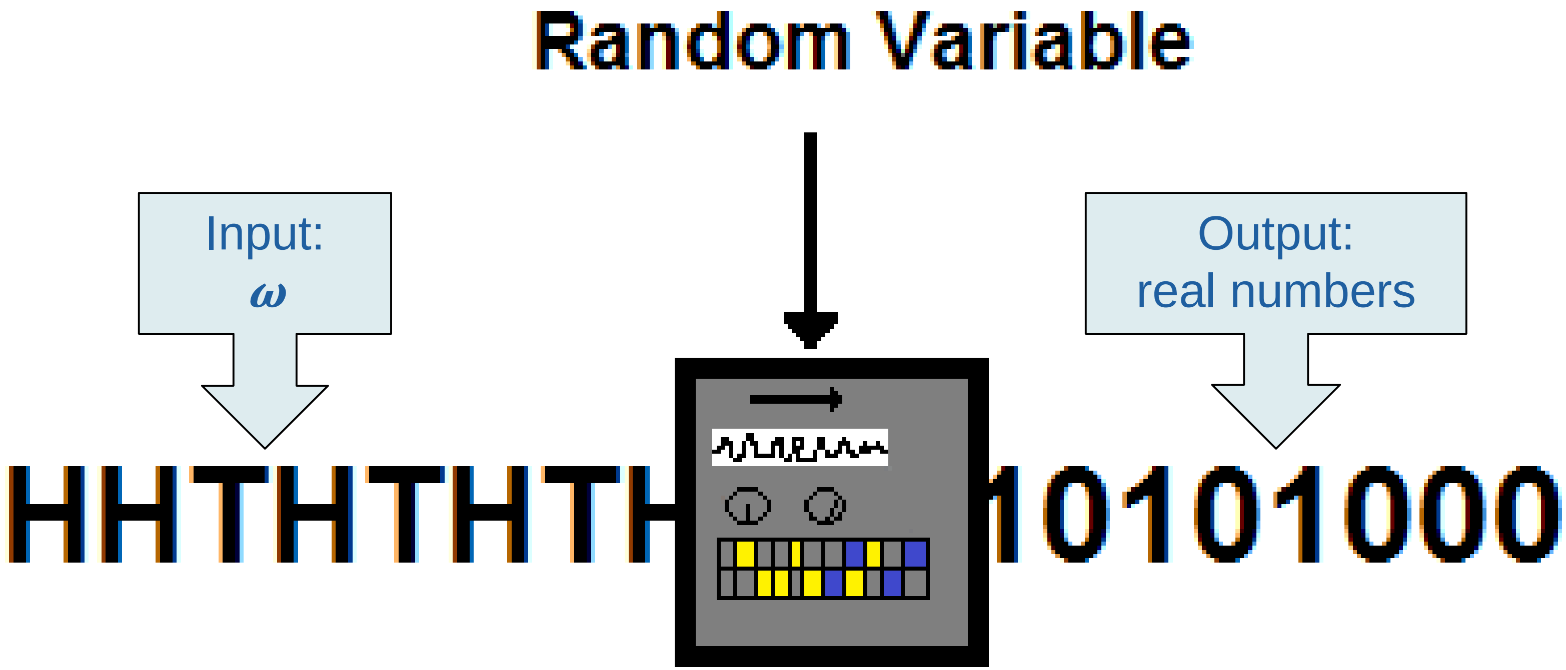
neither random nor variables

Random variable

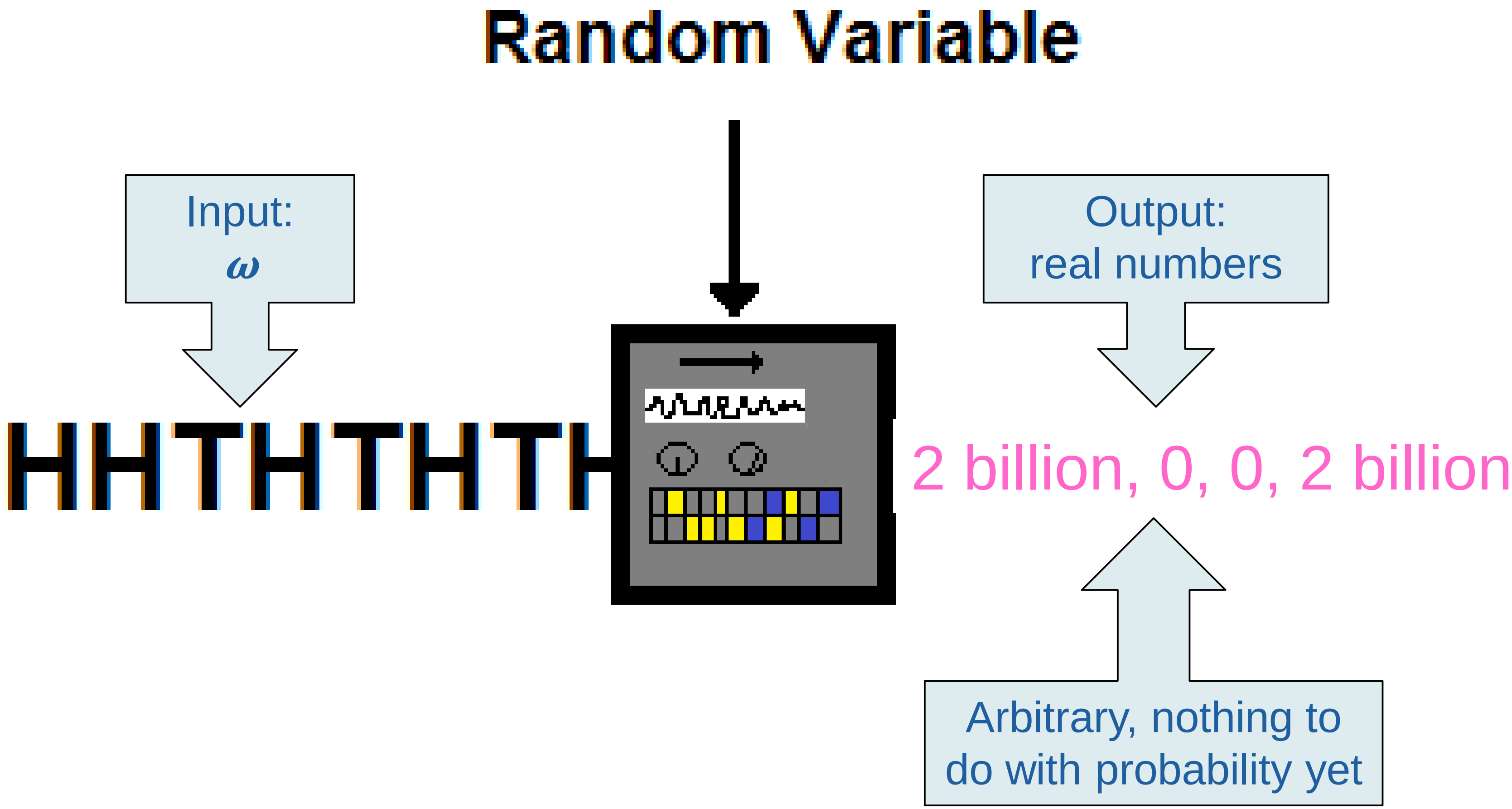
- A function that associates a real number with an event
  - **Input: Ω**

Ω	Sample space	The set of outcomes that we are sampling from.

  - **Output:** numeric sample space



http://www.countbayesie.com/blog/2015/2/20/random-variables-and-expectation



http://www.countbayesie.com/blog/2015/2/20/random-variables-and-expectation

$$X(\omega) = \begin{cases} 2, & \text{if } \omega \text{ is 2 heads} \\ 1, & \text{if } \omega \text{ is 1 head} \\ 0, & \text{if } \omega \text{ is 0 heads} \end{cases}$$

Ω

$X(\omega_1) = 1$

$X(\omega_2) = 1$

$X(\omega_3) = 2$

$X(\omega_4) = 0$

$$X(\omega) = \begin{cases} 2, & \text{if } \omega \text{ is 2 heads} \\ 1, & \text{if } \omega \text{ is 1 head} \\ 0, & \text{if } \omega \text{ is 0 heads} \end{cases}$$

Ω

$X(\omega_1) = 1$

$X(\omega_2) = 1$

$X(\omega_3) = 2$

$X(\omega_4) = 0$

These are x

Random variables: formalization

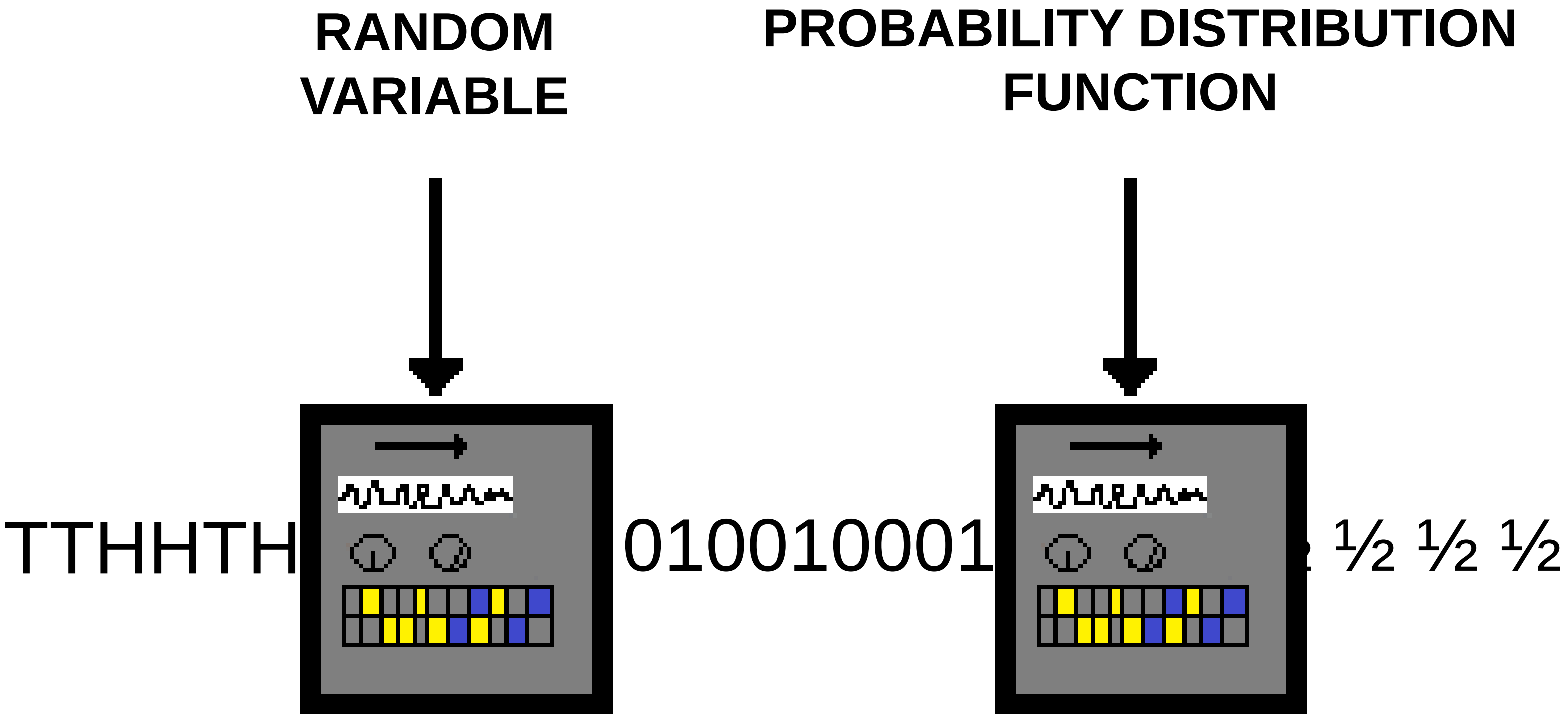
- Capital letters used to denote (X, Y, Z, etc.)
- Let Q = set of real numbers $\mathbb{R}$
- Random variable (rv): X is a function X: $\mathbf{\Omega} \rightarrow Q$
- Discrete rv: $Q \subseteq Z$ countable set, e.g., a subset of integers
- Continuous rv: $Q \subseteq \mathbb{R}$ is a subset of real numbers
- A function $f(X)$ of a rv is also an rv

Random variables: X, Y, Z, etc.

- Same letter, *but in lower case,* used to represent the outcomes or observed values

- **This is not a typo, it actually means something**: $X = x$ "the event that rv X takes on the value x"

- Why call it an rv?
  - Random because observed value depends on outcome of random experiment
  - Variable because different values are possible



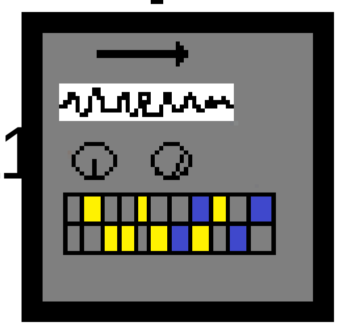
RANDOM VARIABLE
PROBABILITY DISTRIBUTION FUNCTION
TTHHTH
010010001
½ ½ ½

PROBABILITY DISTRIBUTION FUNCTION

Input: real numbers $X(\omega) = x$

Output: always between 0 and 1

01001000111



½ ½ ½ ½ ½ ½

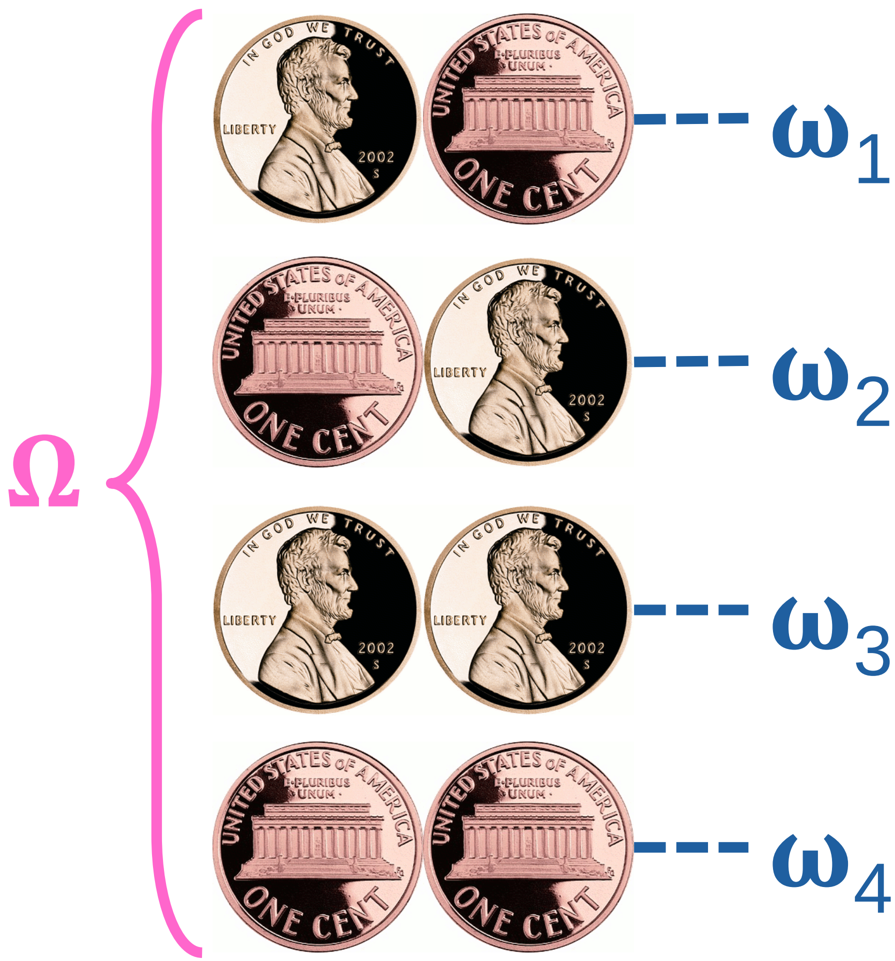
Output: always sum to 1

Simple probability

- If I flip a fair penny, the probability of heads is 50%
  - $p(\mathrm{H}) = p(\mathrm{T}) = .5$
- $\boldsymbol{\Omega} = \{(\mathrm{H}), (\mathrm{T})\}$
- $\boldsymbol{\omega} = \{(\mathrm{H}), (\mathrm{T})\}$
- $X(\boldsymbol{\omega}) = 1$ iff $\boldsymbol{\omega} = \mathrm{H}$; $X(\boldsymbol{\omega}) = 0$ iff $\boldsymbol{\omega} = \mathrm{T}$
- $\mathrm{P}(X(\boldsymbol{\omega}) = 1) = 0.5$
- But what if I flip two coins?

Equally likely (ordered) outcomes

1^{st} flip f_1 | 2^{nd} flip f_2



- If I flip two fair pennies, the probability of heads on each flip is still 50%
  - $p(\mathrm{H}) = p(\mathrm{T}) = .5$
- All *ordered* outcomes here are equally likely
- $\mathrm{P(H, T)} = \mathrm{P(H} \cap \mathrm{T)} = .5 \times .5 = ¼$
- $\mathrm{P(H, T)} = \mathrm{P(T, H)} = ¼$ so the probabilities are equal, but $\boldsymbol{\omega}_1 \neq \boldsymbol{\omega}_2$ in the sample space $\boldsymbol{\Omega}$

$$X(\omega) = \begin{cases} 2, & \text{if } \omega \text{ is 2 heads} \\ 1, & \text{if } \omega \text{ is 1 head} \\ 0, & \text{if } \omega \text{ is 0 heads} \end{cases}$$

X(***ω***) ignores order, but that doesn't mean that order does not matter when defining **ω∈Ω** or *P*(*X*(***ω***))

Ω

$X(\omega)$	probability
1	.25
1	.25
2	.25
0	.25

	$X(\omega) = x$	$P(X(\omega) = x)$
	1	.5
	2	.25
	0	.25

This notation is getting pretty cumbersome

$X = x$ $\quad$ $P(X = x)$

$X(\omega) = x$	$P(X(\omega) = x)$

1	.5

2	.25

0	.25

This notation is getting pretty cumbersome

x	$P(x)$
$X = x$	$P(X = x)$
$X(\omega) = x$	$P(X(\omega) = x)$
1	.5
2	.25
0	.25

equivalent

This notation is getting pretty cumbersome

	$P_X(x)$
x	$P(x)$
$X = x$	$P(X = x)$
$X(\omega) = x$	$P(X(\omega) = x)$

equivalent

1 .5

2 .25

0 .25

This notation is getting pretty cumbersome

These things are all different!

- Events: e.g., a fair coin flip lands on its tail
  - This is an elementary event b/c refers to 1 outcome
- A random variable: e.g., X
  - More clearly and less compactly: $X(\omega)$
- Observation of a random variable: e.g., $X = 5$
  - More clearly and less compactly: $X(\omega) = 5$
- A parameter: e.g., the probability p of heads is ½

Distinguish between what is random but attainable (actual data)
vs.
The unknown but ultimately important true state of nature (parameters).

Kolmogorov's three axioms

Probability is a function, $\mathbf{P}$, that satisfies these three conditions:

1. $P(A) \geq 0$, for all $A \subseteq \mathbf{\Omega}$

1. $P(\mathbf{\Omega}) = 1$

1. $P(A \cup B) = P(A) + P(B)$ if $A \cap B = 0$

Any function that satisfies these three axioms is a probability function.

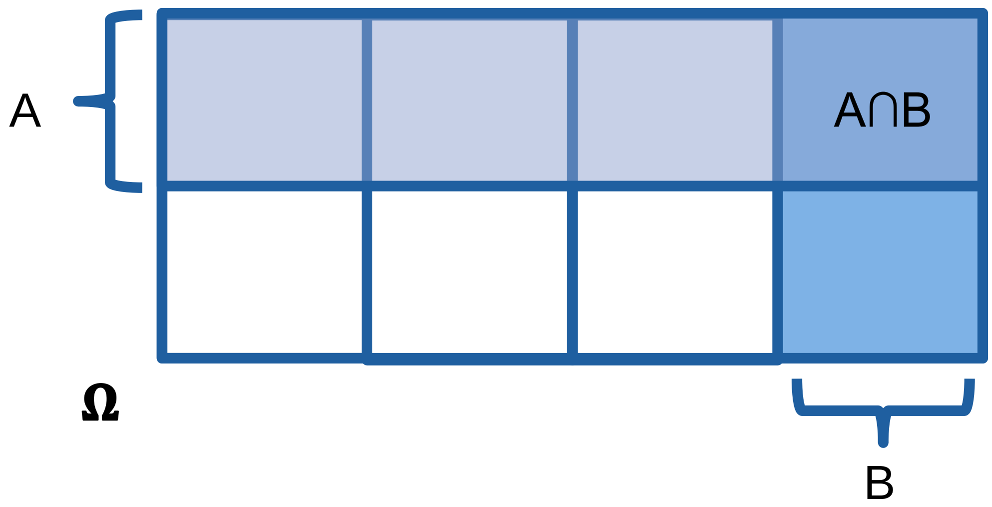
Basics

- $P(A') = 1 - P(A)$
- If $A \subset B$, then $P(A) \leq P(B)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

- Using set theory and our three axioms, you should be able to prove each of these.

Independence

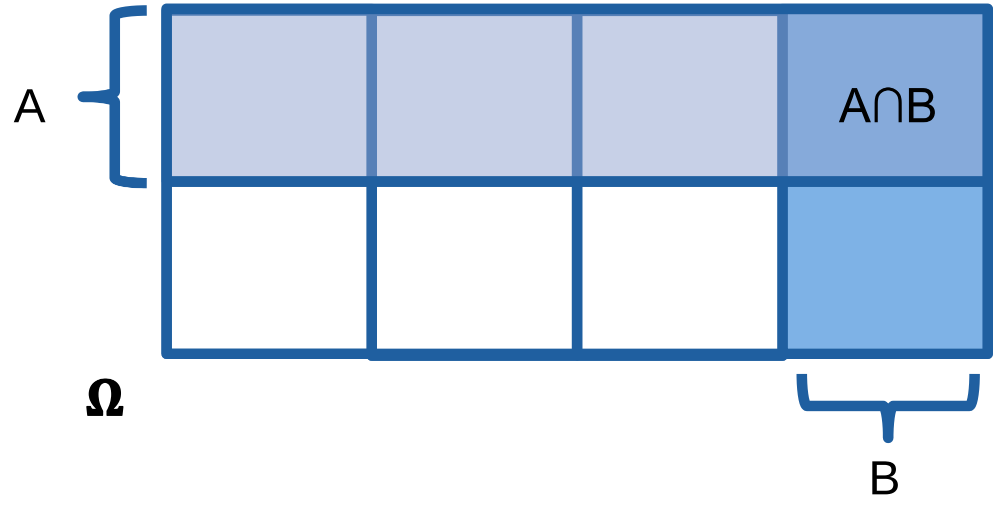
- If 2 events are independent, both of these must be true:
  - $P(A \cap B) = P(A) \times P(B)$
  - $P(B|A) = P(B)$
- Are A and B independent?



Independence

independent
identically
distributed

- Independence of events or rvs makes it much easier to write down the probability of joint events or the joint distribution. It allows you to write these as a simple product.
- Here, $P(A) = ½; P(B) = ¼$
- $P(A \cap B) = P(A) \times P(B) = ½ \times ¼ = 1/8$



When is an event independent of itself?
When is A independent of A?

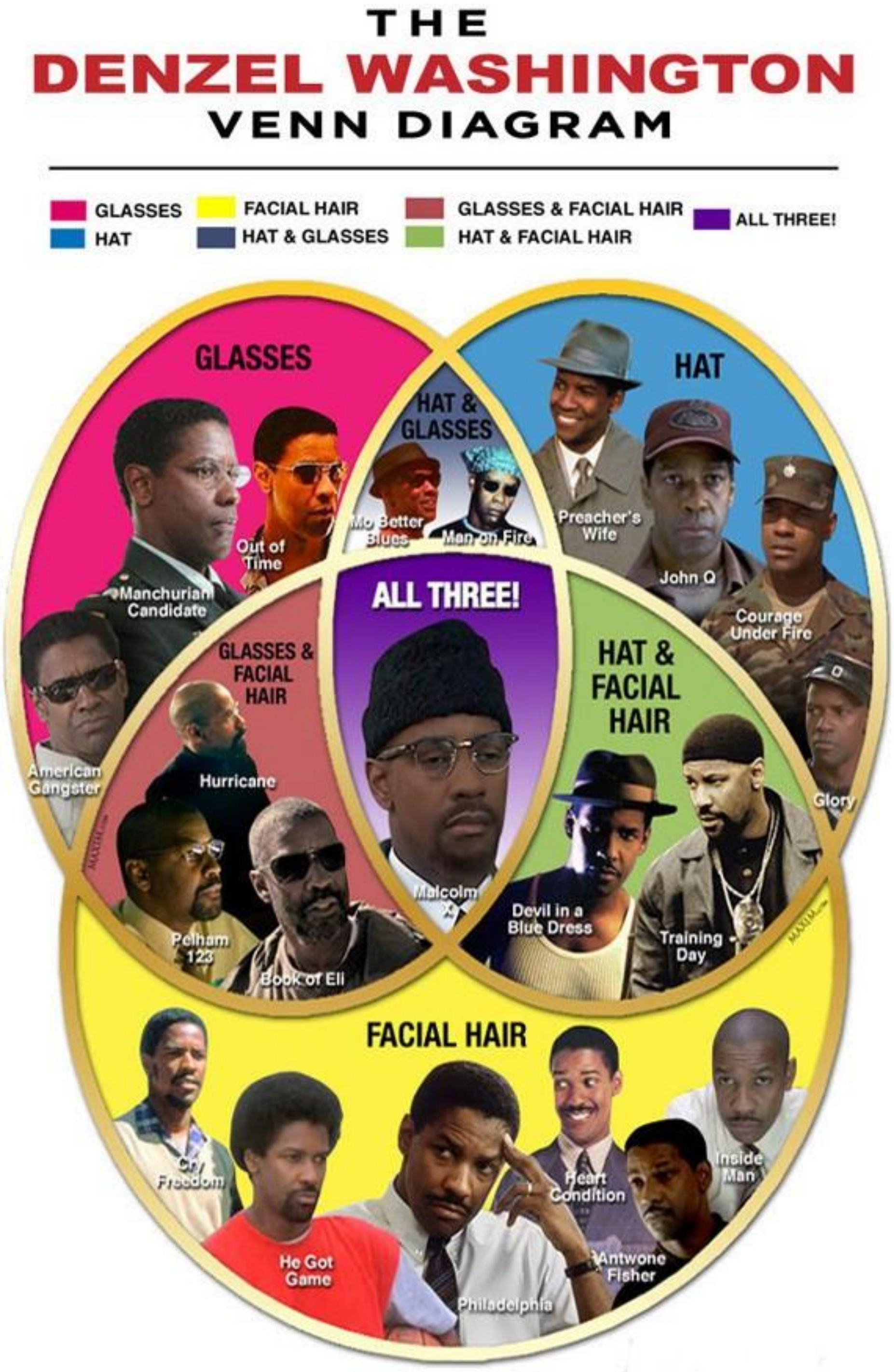
When is:

$P(A \cap A) = P(A) \times P(A)$

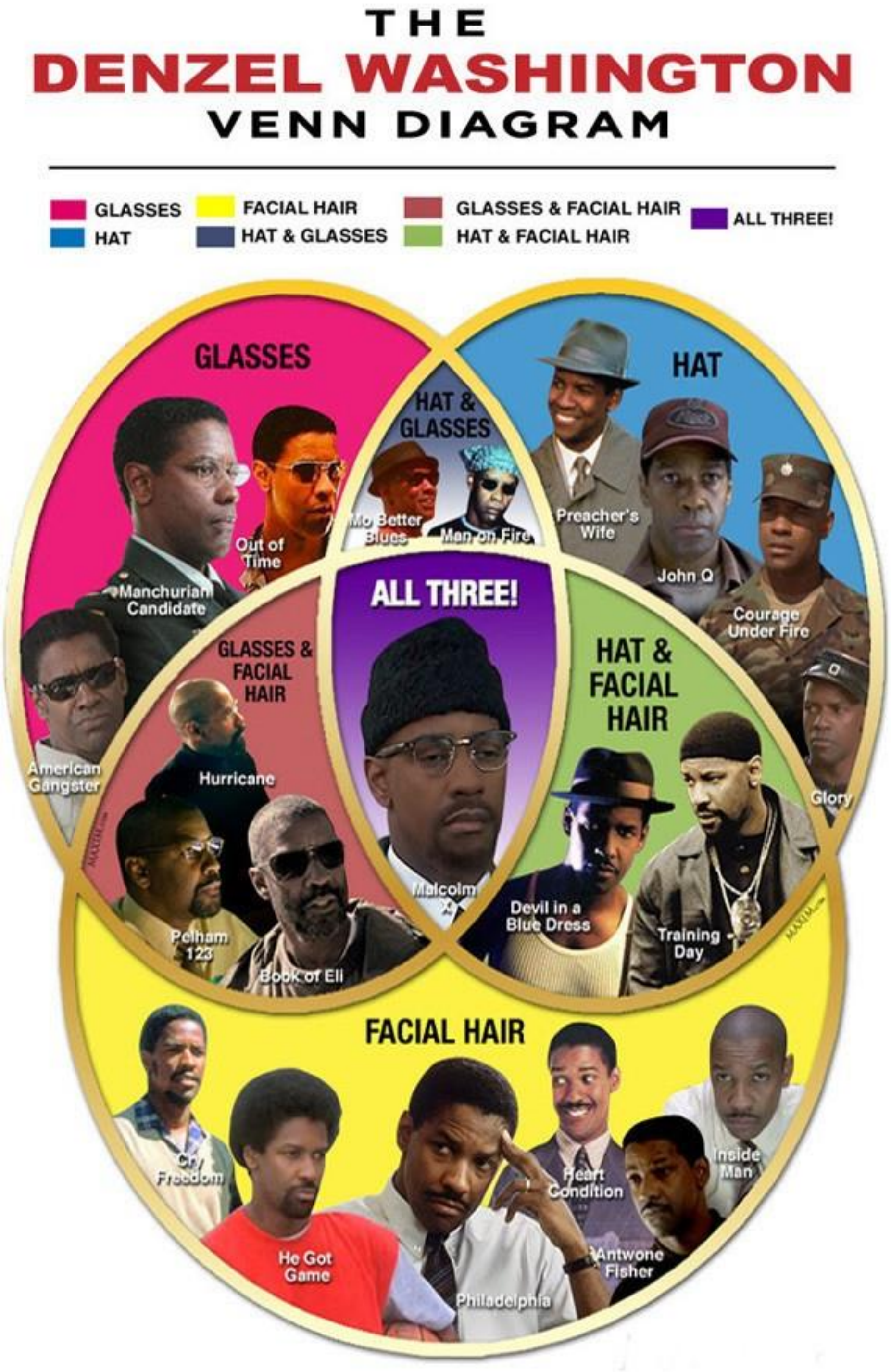
Hint: $P(A \cap A) = P(A)$

When is A independent of A?

- $P(A) = P(A \cap A) = P(A) \times P(A)$
- So only if $P(A) = 0$ or 1



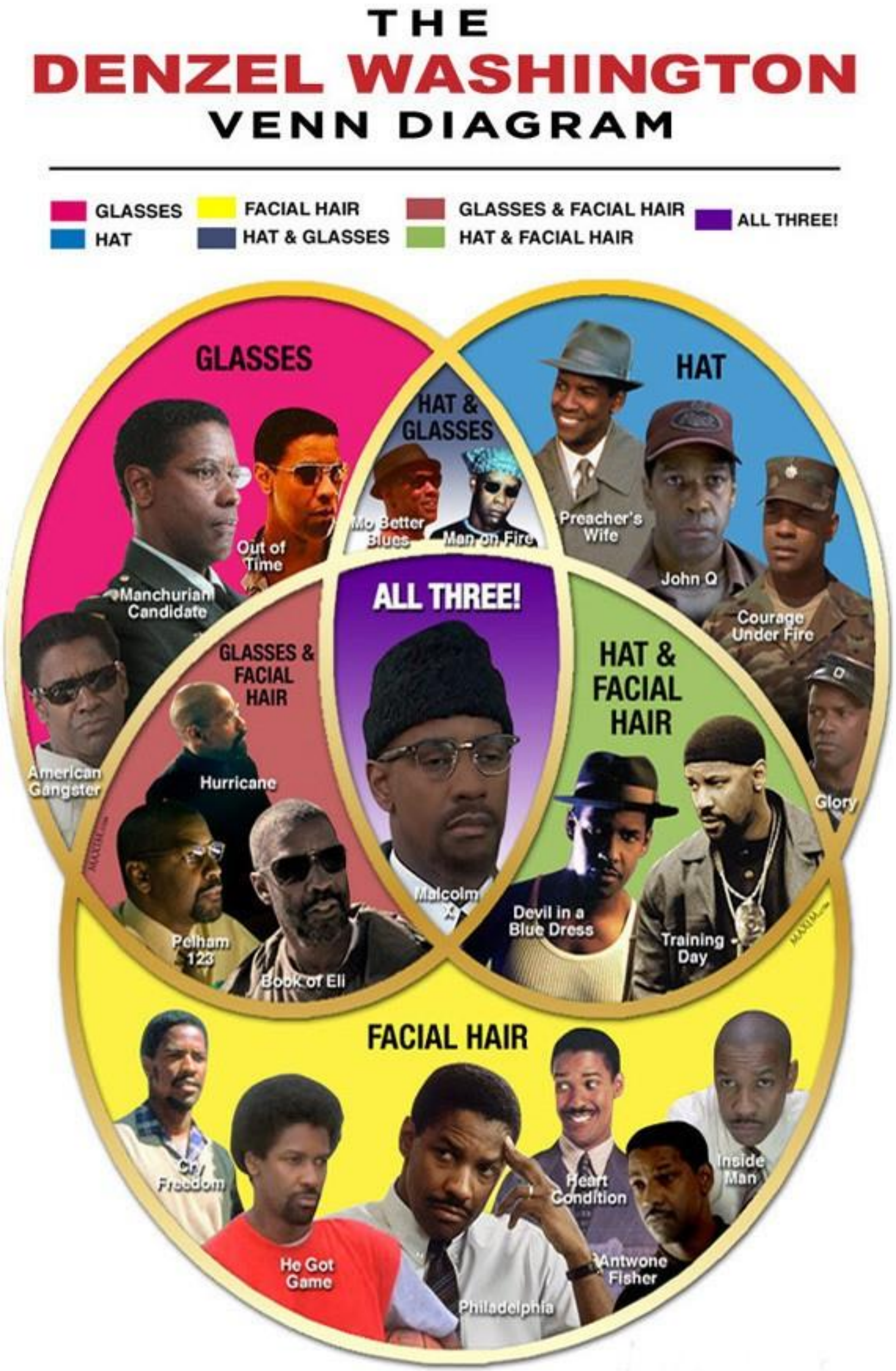
- $\boldsymbol{\Omega}$ = 21 Denzels
- A = Denzel wears glasses
- B = Denzel wears a hat
- Are the glasses and hat events independent?



- $\mathbf{\Omega}$ = 21 Denzels
- G = Denzel wears glasses
- H = Denzel wears a hat
- Are the glasses and hat events independent?
- $P(G) = 9/21 = 3/7$
- $P(H) = 9/21 = 3/7$
- $P(G \cap H) = 3/21 = .14$
- $P(G) \times P(H) = 9/49 = .18$
- Wearing glasses and hat together is (slightly) less likely than we'd expect if they were independent.

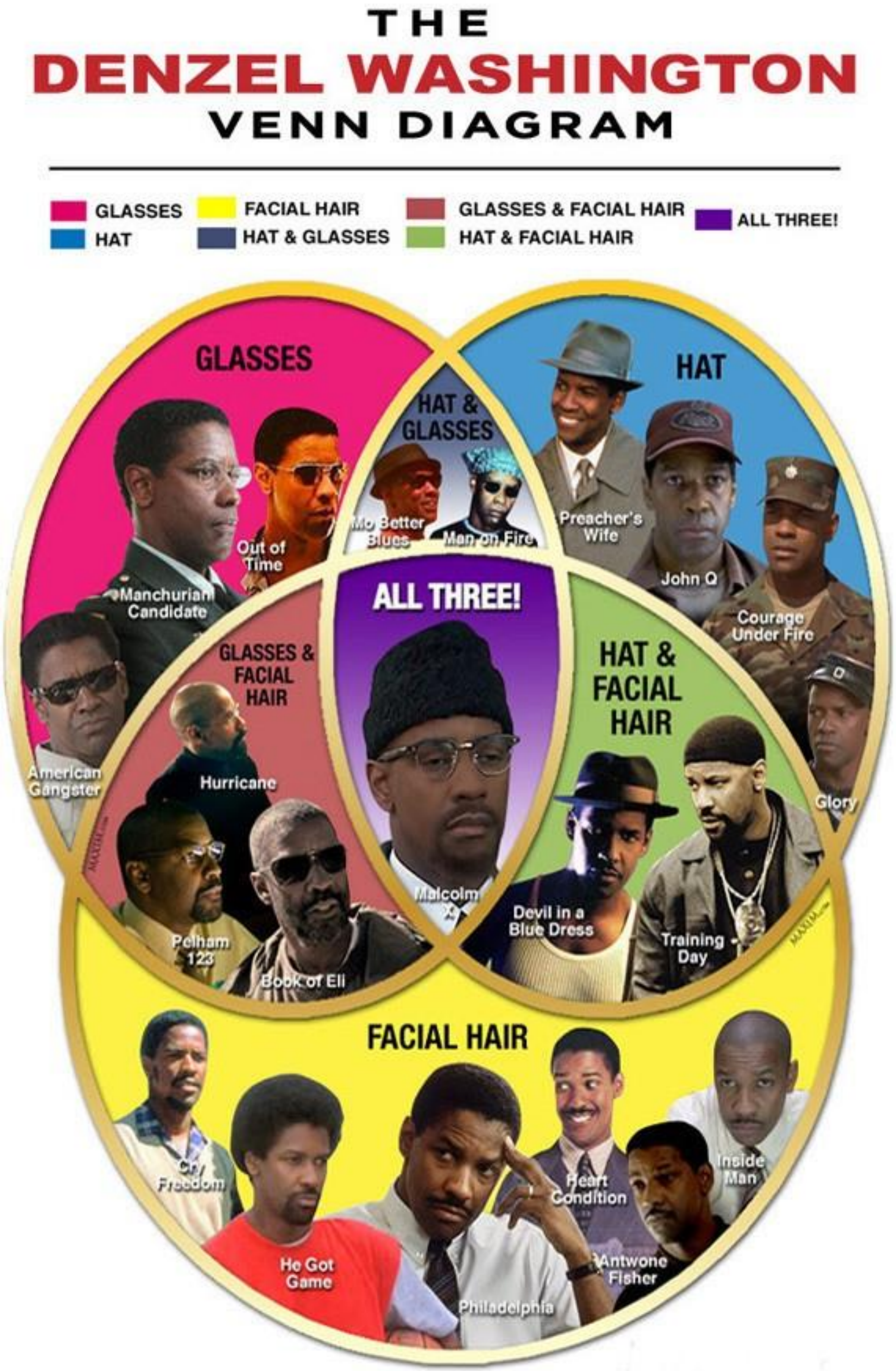
CONDITIONAL PROBABILITY

The probability of an event given that another event has occurred.

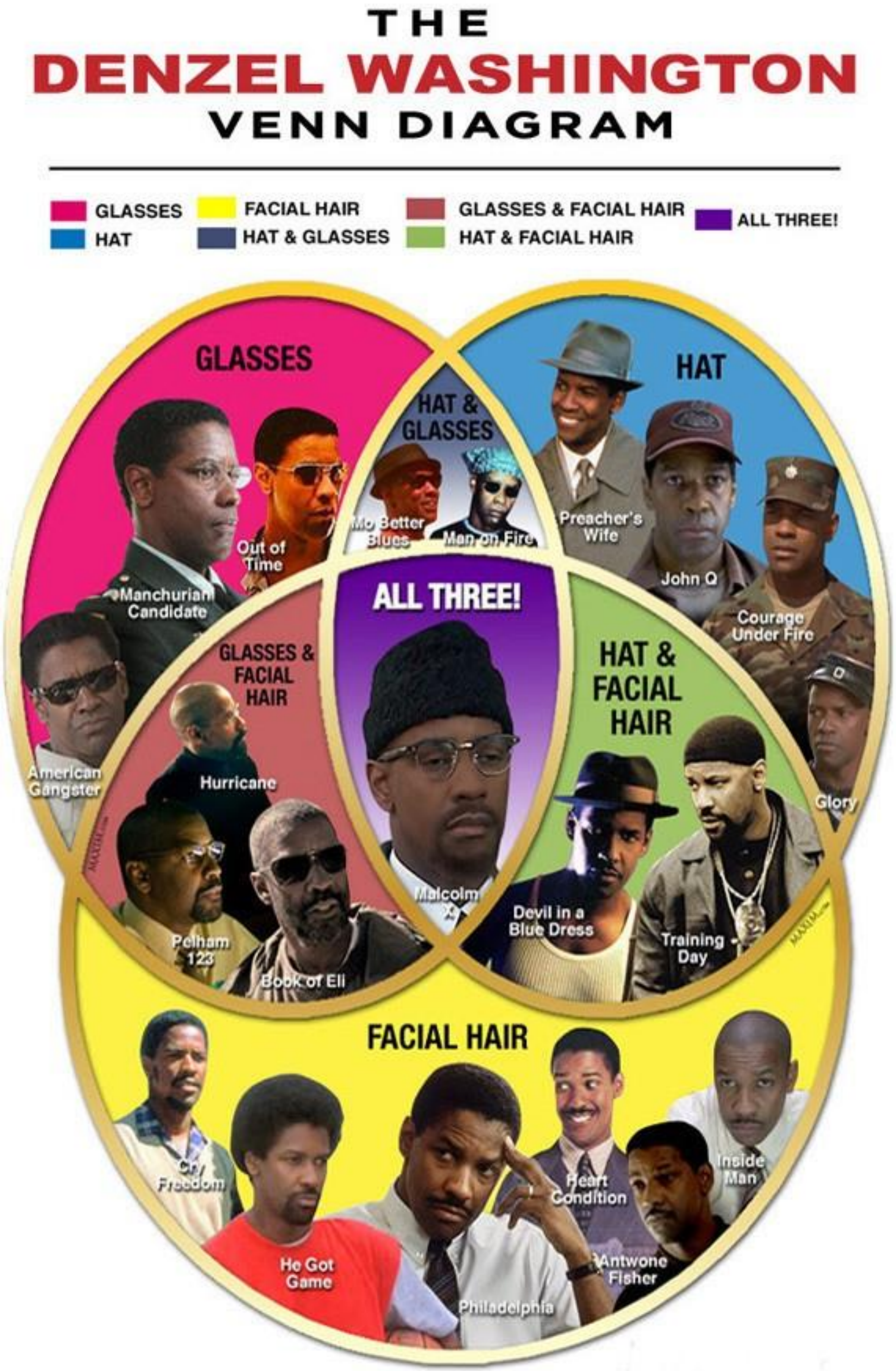


- We now know that glasses and hat events are not independent.
- New movie coming soon: you'll win $1000 if you correctly guess whether Denzel's character will have facial hair.
- What do you do?

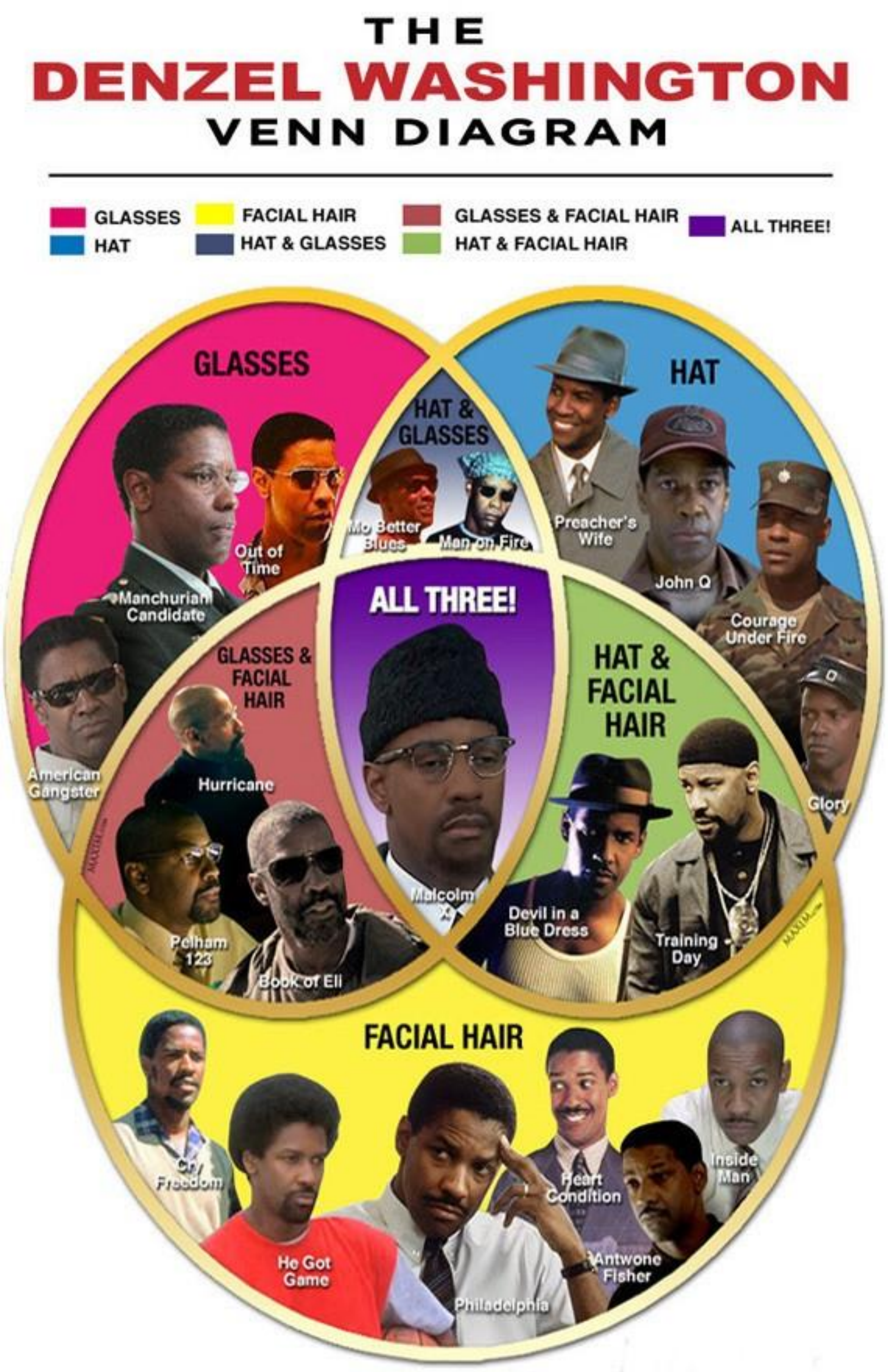
- Ω = 21 Denzels
- P(FH) = 12/21 = 57%



- $\mathbf{\Omega}$ = 21 Denzels
- P(FH) = 12/21 = 57%
- So Denzel has had facial hair in 57% of his movies- you'd be smart to guess that for a new Denzel movie, yes, he would have facial hair!



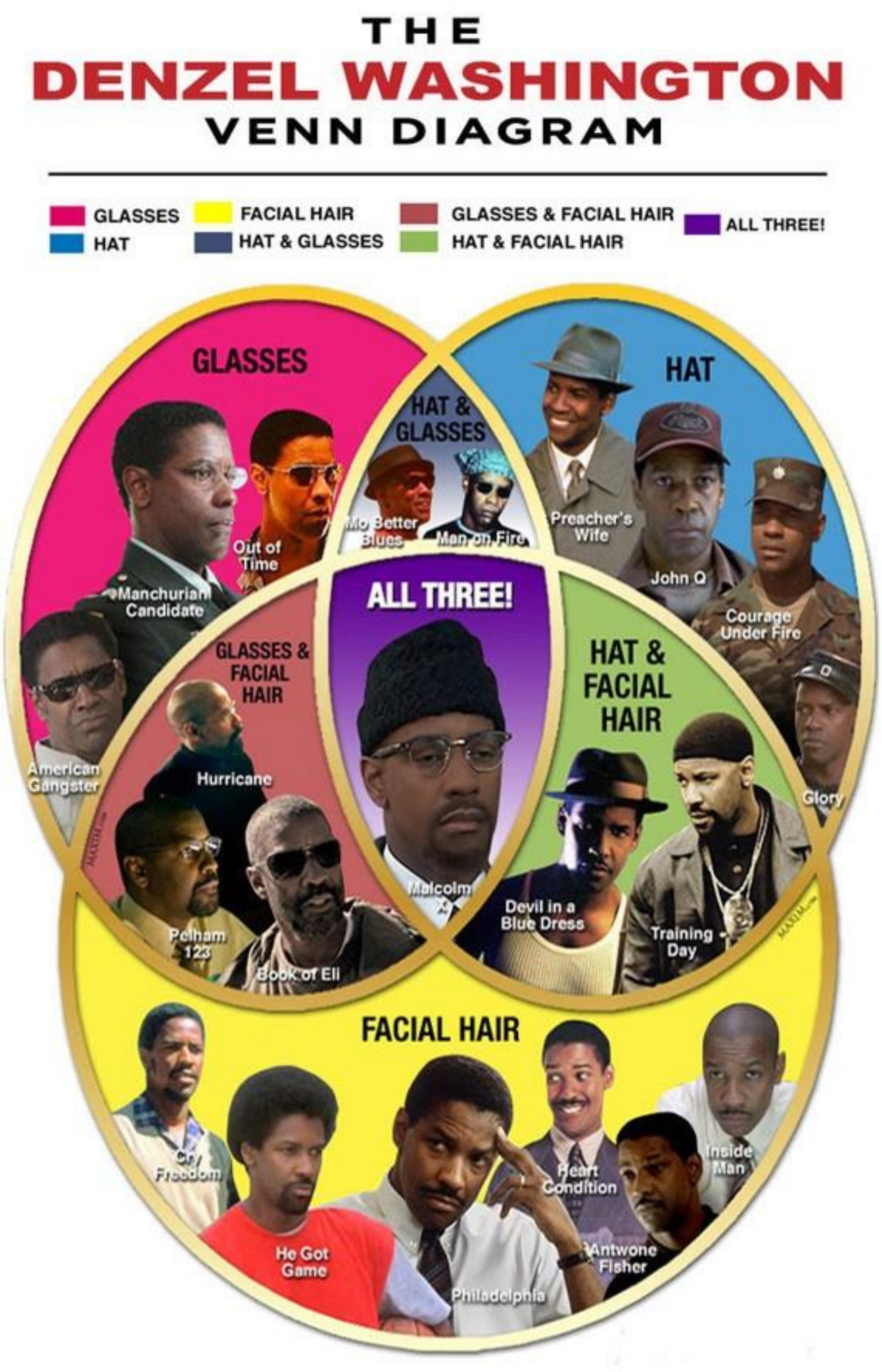
- **Ω** = 21 Denzels
- P(FH) = 12/21 = 57%
- BUT: just before you enter your answer, you spy the new movie poster- Denzel has a hat on!
- This is new information!
- What do you do?



- **Ω** = 21 Denzels
- P(FH) = 12/21 = 57%
- P(H) = 9/21 = 3/7
- P(FH | H) = 3/9 = 33.3%

Law of total probability

$$P(B) = P(A)P(B|A) + P(A')P(B|A')$$



- Law of total probability in English
  - If A_i, …A_j are mutually exclusive and jointly exhaustive, then the probability of B is the sum of the probabilities of B given A_i and conditioned on the probability of A_i within A.

$$P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i)$$

Bayes' rule

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Bayes' theorem

- Reverse the conditioning using the definition of conditional probability

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$= \frac{P(B|A) \times P(A)}{P(B|A) \times P(A) + P(B|A') \times P(A')}$$

Derive Bayes' law

$$= \frac{P(B|A) \times P(A)}{P(B|A) \times P(A) + P(B|A') \times P(A')}$$

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

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

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

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Bayes' rule

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

From the law of total probabilities we know that P(B) can also be stated this way:

$$P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i)$$

$$= \frac{P(B|A) \times P(A)}{P(B|A) \times P(A) + P(B|A') \times P(A')}$$

If you have HIV, the probability of testing positive on an HIV test is 99.9%.

If you don't have HIV, the probability of testing negative on the HIV test is 99.99%.

In Oregon, about 8 people per 100,000 have HIV.

Imagine you select a person at random and give them an HIV test. If the HIV test is positive, what is the probability that they have HIV?

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Your turn

- T = {test is positive}
- H = {has HIV}
- P(T|H) = 0.999
- P(T'|H') = 0.9999
- P(H) = 0.00008 ($8*10^{-5}$)
- We want to solve for P(H|T)
- P(T) = ?

$$P(H|T) = \frac{P(T|H)P(H)}{P(T)}$$

Your turn

- T = {test is positive}
- H = {has HIV}
- $P(T|H) = 0.999$
- $P(T'|H') = 0.9999$
- $P(H) = 0.00008$ ($8*10^{-5}$)
- We want to solve for $P(H|T)$
- $P(T) = P(T|H) \times P(H) + P(T|H') \times P(H')$
  $= 0.999 \times 0.00008 + (1 - 0.9999) \times (1 - 0.00008)$
  $= 0.000179912$

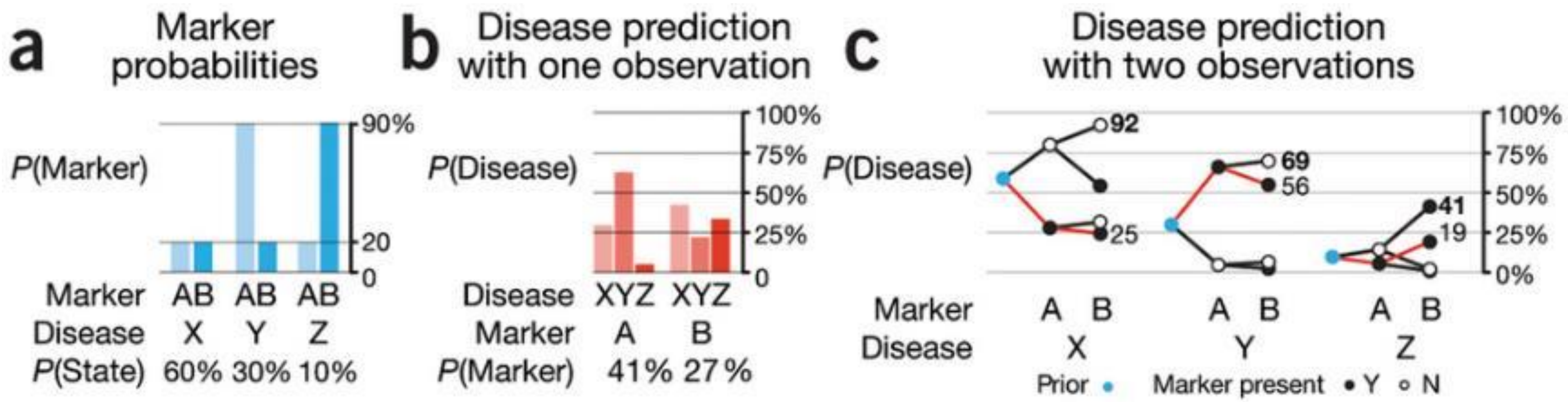
$$P(H|T) = \frac{P(T|H)P(H)}{P(T)}$$

Your turn

- If the HIV test is positive, what is the probability that they have HIV?

$$P(H|T) = \frac{.999 \times .00008}{0.000179912} = .44$$

Bayes' rule



http://www.nature.com/nmeth/journal/v12/n4/full/nmeth.3335.html

“Yes; you should switch.” –Marilyn vos Savant

- The first door has a 1/3 chance of winning, but the second door has a 2/3 chance.
- Here’s a good way to visualize what happened. Suppose there are a million doors, and you pick door #1. Then the host, who knows what’s behind the doors and will always avoid the one with the prize, opens them all except door #777,777.
- **You’d switch to that door pretty fast, wouldn’t you?**

Use Bayes to solve Monty

- A = {The car is behind door 1}
  - P(A) = 1/3; P(A') = 2/3
- B= {Monty shows you a goat behind door 2}
  - P(B|A) = 1/2; {He could choose either 2 or 3}
- Use law of total probability to solve for P(B):
  - P(B) = 1/3*1/2 + 1/3*0 + 1/3*1
    - Assume you chose door 1:
    - If the car is behind door 1, Monty will chose door 2 half the time. {1/2}
    - If the car is behind door 2, Monty will open door 3.{0}
    - If the car is behind door 3, Monty will open door 2. {1}

- P(A|B) =1/2*1/3 / (1/3*1/2) + (1/3*0) +(1/3*1)