# CS145: Probability & Computing

## Lecture 2: Axioms of Probability, Conditioning and Bayes' Rule

IN DEO SPERAMUS

*Figure credits:*
*Bertsekas & Tsitsiklis, **Introduction to Probability**, 2008*
*Pitman, **Probability**, 1999*

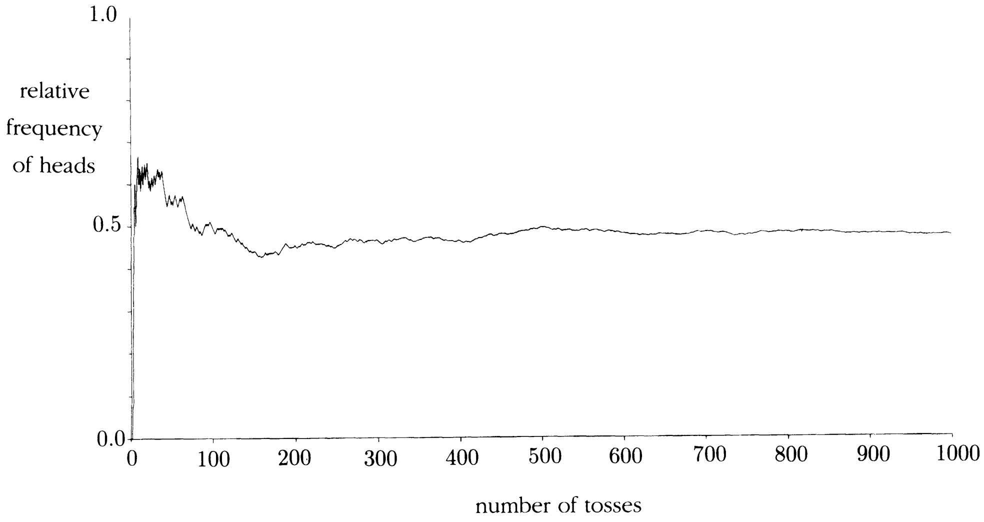
# Reasoning Under Uncertainty

- $1/1000$ of tourists who visit tropical country $X$ return with a dangerous virus $Y$.
- There is a test to check for the virus. The test has $5\%$ *false positive* rate and no *false negative* error.
- You returned from country $X$, took the test, and it was positive. Should you take the painful treatment for the virus?

https://infotainmentnews.net

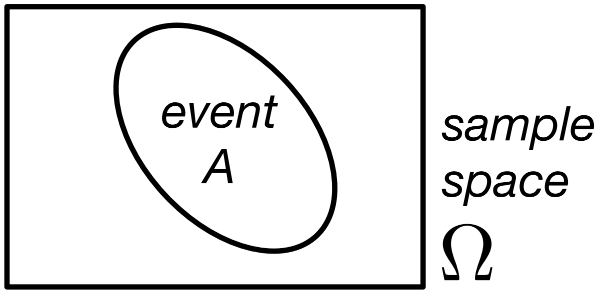
# CS145: Lecture 2 Outline

- ➢ Relative Frequencies and the Axioms of Probability
- ➢ Conditional Probabilities
- ➢ Bayes’ Rule



# The Discrete Uniform Law

*Formalizes the idea of "uniform random" sampling.*




- Let all outcomes be equally likely

- Then,

$$\mathbf{P}(A) = \frac{\text{number of elements of } A}{\text{total number of sample points}} = \frac{|A|}{|\Omega|}$$

- Computing probabilities $\equiv$ counting

- Defines fair coins, fair dice, well-shuffled decks

# Probabilities as Relative Frequencies

If an event A happens *m* times in *n* trials,
then *m/n* is the *relative frequency* of A in the *n* trials

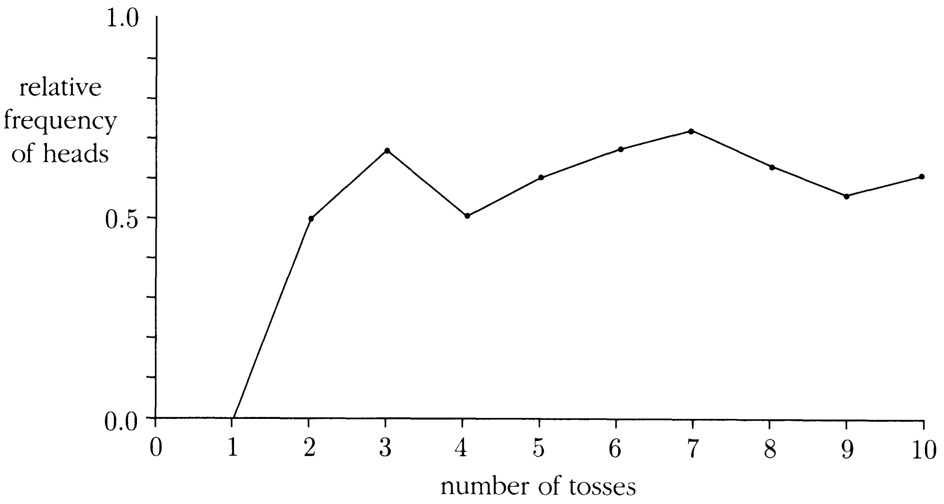
**Example:** Relative frequency of heads in coin toss sequence

*Experimental outcomes:*

$$t, h, h, t, h, h, h, t, t, h,$$

*Relative frequencies:*

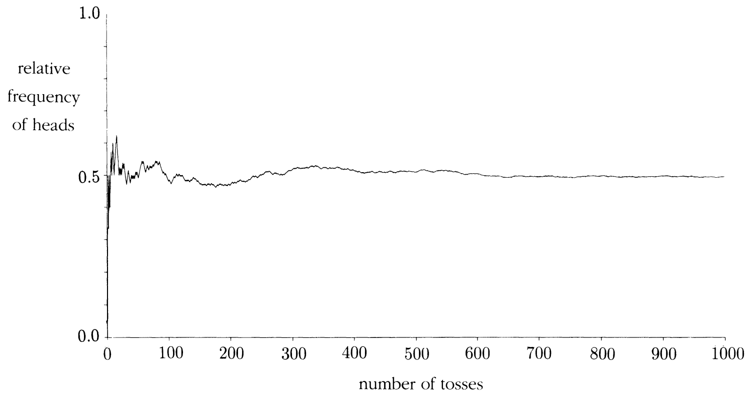
$$\frac{0}{1}, \frac{1}{2}, \frac{2}{3}, \frac{2}{4}, \frac{3}{5}, \frac{4}{6}, \frac{5}{7}, \frac{5}{8}, \frac{5}{9}, \frac{6}{10},$$

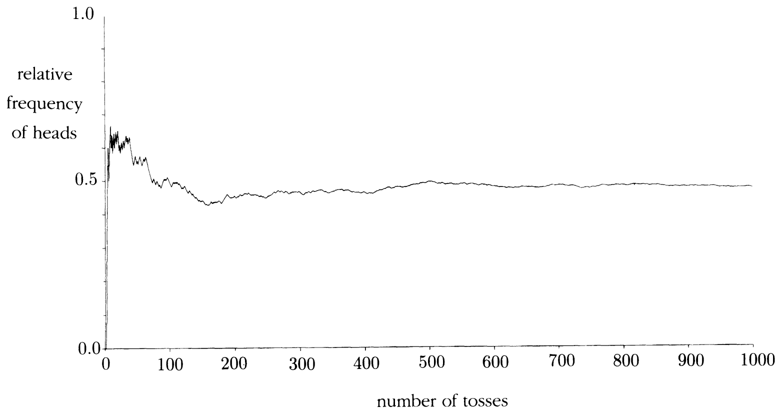


# Probabilities as Relative Frequencies

If an event A happens *m* times in *n* trials,
then *m*/*n* is the *relative frequency* of A in the *n* trials

We define the probability of $A$ as $P(A) = \lim_{n\to\infty} \frac{m(n)}{n}$




This definition is consistent (limit is defined and unique) due to the *Law of Large Numbers.*

# Non-Uniform Relative Frequencies

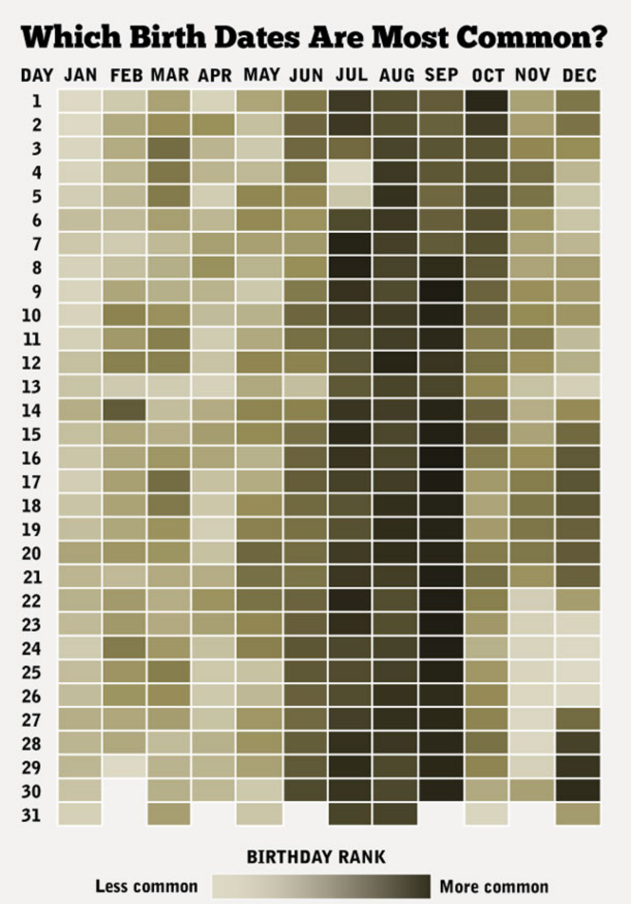
If an event A happens *m* times in *n* trials, then *m/n* is the *relative frequency* of A in the *n* trials

Let $m(n)$ be the number of occurrences of A in $n$ trials. We define the probability of $A$ as

$$P(A) = \lim_{n\to\infty} \frac{m(n)}{n}$$

This definition extends to outcomes (simple events) with different probabilities.

This definition extends to compound events (events with more than one outcome).

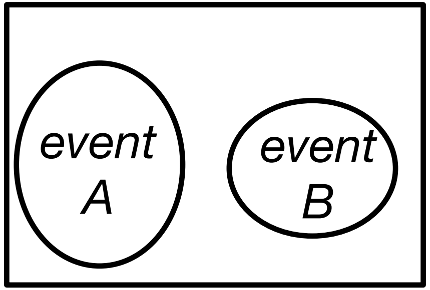



http://thedailyviz.com/2012/05/12/how-common-is-your-birthday/

# Axioms of Probability Functions

In the uniform random case we had:

$$\mathrm{P}(A) = \frac{\text{number of elements of } A}{\text{total number of sample points}} = \frac{|A|}{|\Omega|}$$




In general discrete sample spaces: any function $P : 2^\Omega \to [0,1]$ that satisfies:

**Axioms:**

1. **Nonnegativity:** $\mathrm{P}(A) \geq 0$
2. **Normalization:** $\mathrm{P}(\Omega) = 1$
3. **Additivity:** If $A \cap B = \varnothing$, then $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B)$

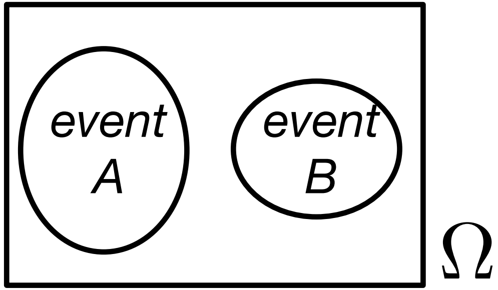
For infinite sample spaces we need:

3. **Additivity:** For a countable infinite set $C$, and disjoint events $\{A_i \mid i \in C\}$, $Pr(\cup_{i \in C} A_i) = \sum_{i \in C} Pr(A_i)$.

# Axioms of Probability Function

**Axioms:**

1. **Nonnegativity:** $\mathrm{P}(A) \geq 0$
2. **Normalization:** $\mathrm{P}(\Omega) = 1$
3. **Additivity:** If $A \cap B = \emptyset$, then $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B)$




---

- Valid probabilities - any function mapping subsets of $\Omega$ to [0,1] that satisfies these *axioms.*
- The *nonnegativity* and *additivity* axioms are fundamental to probability and uncertainty
- Unit *normalization* is just a convention, another options is probability between 0% and 100%
- The *additivity* axiom guarantees that the probabilities of any finite (countable infinite) set of disjoint events are additive (induction)

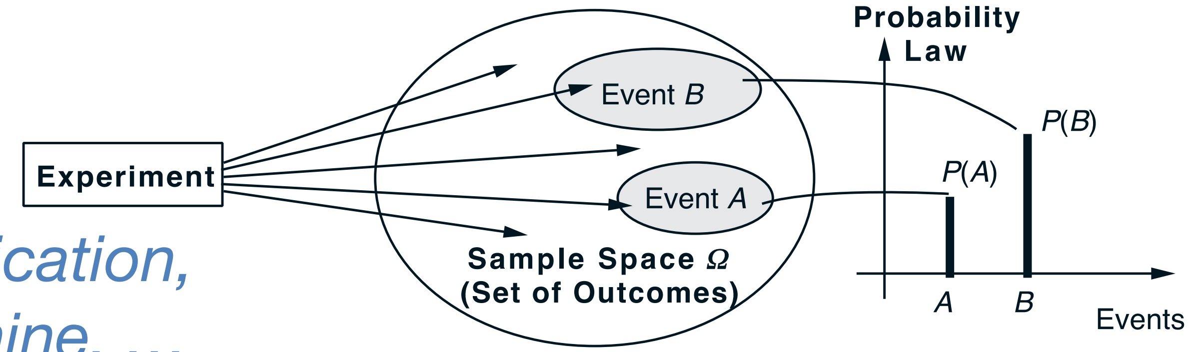
# Example:

We flip a fair coin until the first HEAD.

- $i$=number of flips till (included) the first head
- Sample space $\Omega = \{1, 2, 3, \dots\}$
- $P(\{i\}) = P(i) = 1/2^i$
- $P(\Omega) = \Pr(\cup_{i\geq 1}\{i\}) = \sum_{i\geq 1} P(i) = \sum_{i\geq 1} \frac{1}{2^i} = 1.$

# Defining a Probabilistic Model

*flip a coin,*
*roll a die,*
*take a medication,*
*test a machine, …*



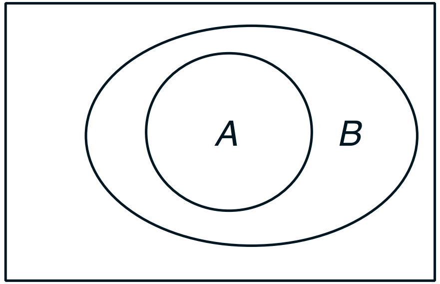

**Elements of a Probabilistic Model**

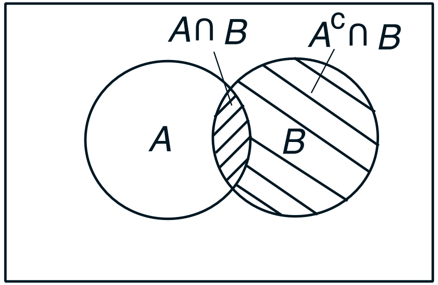
- The **sample space** $\Omega$, which is the set of all possible outcomes of an experiment.
- The **probability law**, which assigns to a set $A$ of possible outcomes (also called an **event**) a nonnegative number $\mathbf{P}(A)$ (called the **probability** of $A$) that encodes our knowledge or belief about the collective "likelihood" of the elements of $A$. The probability law must satisfy certain properties to be introduced shortly.

# Properties of Probability Laws

**Axioms:**

1. **Nonnegativity:** $\mathrm{P}(A) \geq 0$
2. **Normalization:** $\mathrm{P}(\Omega) = 1$
3. **Additivity:** If $A \cap B = \varnothing$, then $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B)$




**Some Properties of Probability Laws**

Consider a probability law, and let $A$, $B$, and $C$ be events.

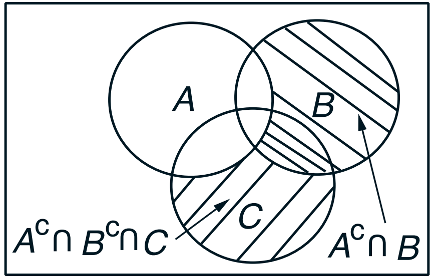
(a) If $A \subset B$, then $\mathbf{P}(A) \leq \mathbf{P}(B)$.

(b) $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$.

(c) $\mathbf{P}(A \cup B) \leq \mathbf{P}(A) + \mathbf{P}(B)$.

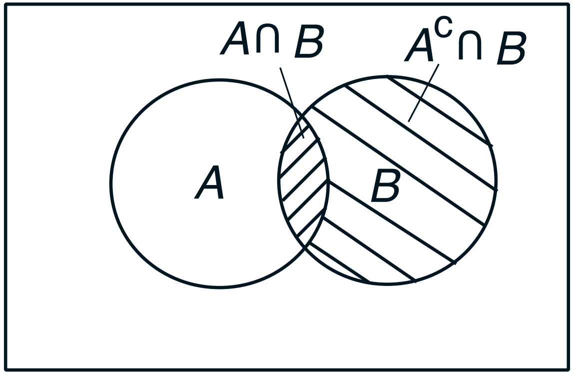
(d) $\mathbf{P}(A \cup B \cup C) = \mathbf{P}(A) + \mathbf{P}(A^c \cap B) + \mathbf{P}(A^c \cap B^c \cap C)$.

$(e) P(A^c) = 1 - P(A)$



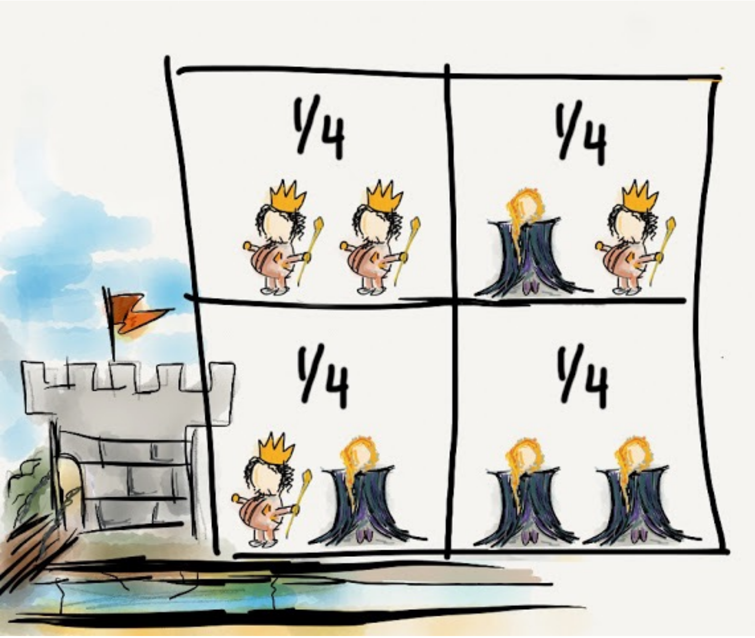
# CS145: Lecture 2 Outline

- Relative Frequencies and the Axioms of Probability
- Conditional Probabilities
- Bayes' Rule



# Example: The King's Sibling

The king comes from a family of two children.
Assume there is an equal probability of each birth being a boy or girl.
**What is the probability that the king has a sister (his sibling is a girl)?**




Event A: At least one child is a boy
Event B: Both children are boys
Event C: The king has a sister

$$P(A) = \frac{3}{4} \qquad P(B) = \frac{1}{4}$$

$$P(C|A) = 1 - P(B|A) = ?$$

***Figure credits:*** http://gaussiangeek.blogspot.com/2015/06/the-kings-sibling-how-well-do-you.html

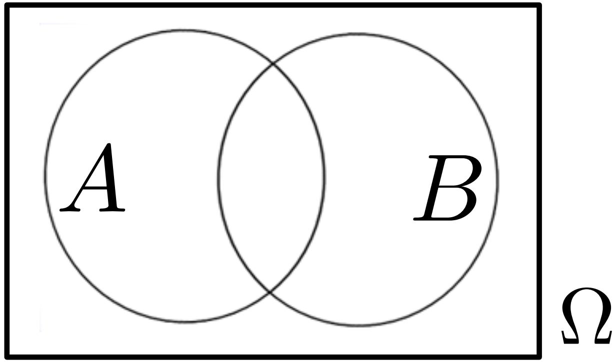
# Conditional Probability

- $\mathbf{P}(A \,|\, B) =$ probability of $A$,
  given that $B$ occurred
  - $B$ is our new universe

- **Definition:** Assuming $\mathbf{P}(B) \neq 0$,

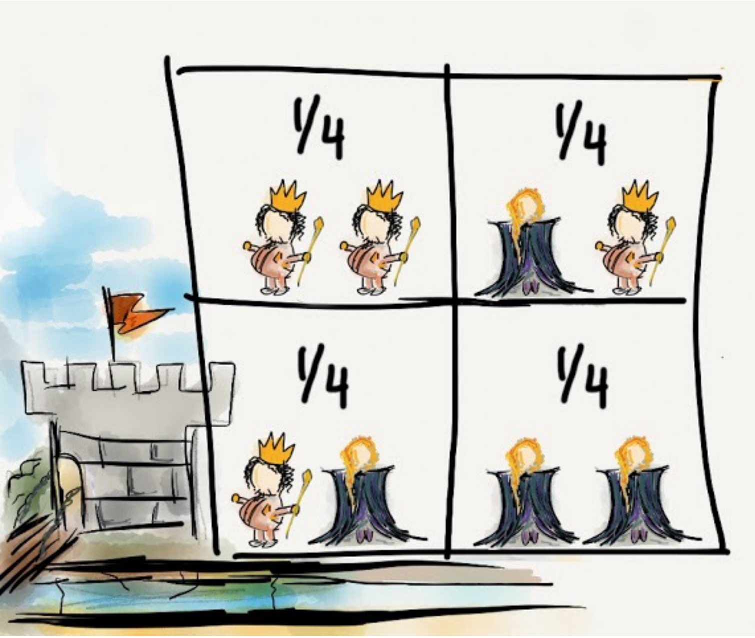
$$\mathbf{P}(A \mid B) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)}$$

$\mathbf{P}(A \mid B)$ undefined if $\mathbf{P}(B) = 0$



# Example: The King's Sibling

The king comes from a family of two children.
Assume there is an equal probability of each birth being a boy or girl.
**What is the probability that the king has a sister (his sibling is a girl)?**




Event A: At least one child is a boy
Event B: Both children are boys

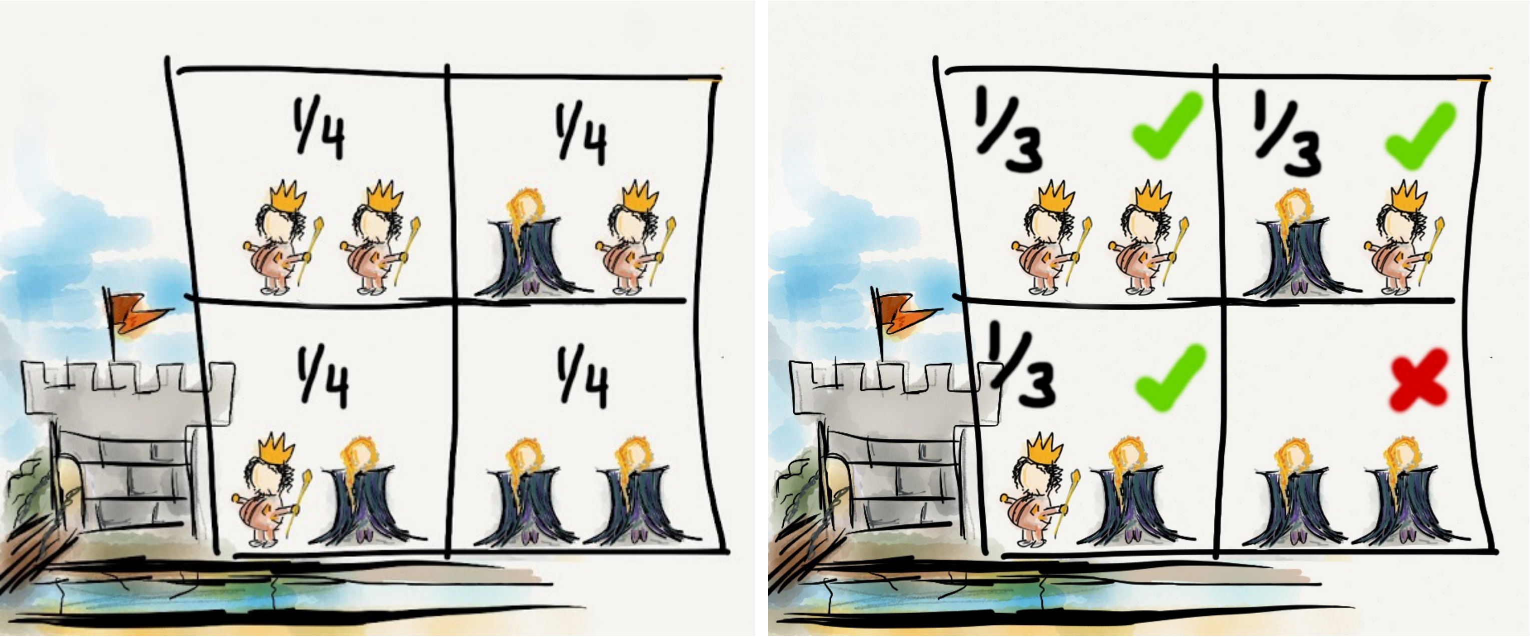
$$P(A) = \frac{3}{4} \qquad P(B) = \frac{1}{4}$$

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{1/4}{3/4} = \frac{1}{3}$$

***Figure credits:*** http://gaussiangeek.blogspot.com/2015/06/the-kings-sibling-how-well-do-you.html

# Example: The King's Sibling

The king comes from a family of two children.
Assume there is an equal probability of each birth being a boy or girl.
**What is the probability that the king has a sister (his sibling is a girl)?**

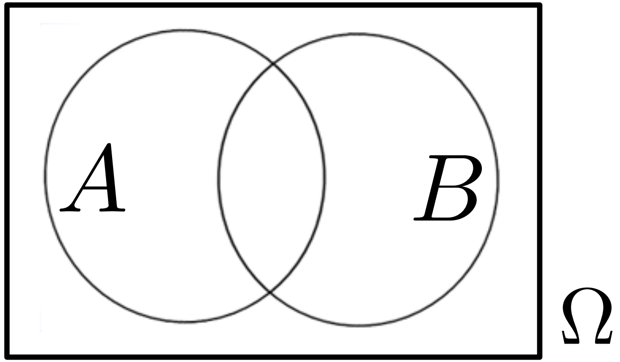



*With probability 2/3 the king has a sister!*

***Figure credits:*** *http://gaussiangeek.blogspot.com/2015/06/the-kings-sibling-how-well-do-you.html*

# Conditional Probability

- $\mathbf{P}(A \mid B) =$ probability of $A$,
  given that $B$ occurred
  - $B$ is our new universe




- **Definition:** Assuming $\mathbf{P}(B) \neq 0$,

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

$\mathbf{P}(A \mid B)$ undefined if $\mathbf{P}(B) = 0$

- Under discrete uniform law, where all outcomes equally likely:

$$\mathbf{P}(A \mid B) = \frac{\text{number of elements of } A \cap B}{\text{number of elements of } B} = \frac{|A \cap B|}{|B|}$$

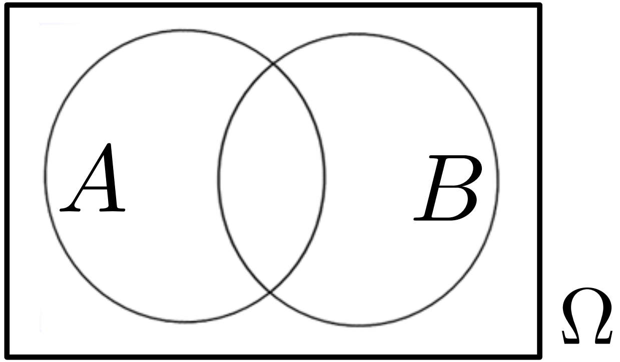
# Conditional Probability

- $\mathbf{P}(A \mid B) =$ probability of $A$,
  given that $B$ occurred
  - $B$ is our new universe

- **Definition:** Assuming $\mathbf{P}(B) \neq 0$,

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

  $\mathbf{P}(A \mid B)$ undefined if $\mathbf{P}(B) = 0$

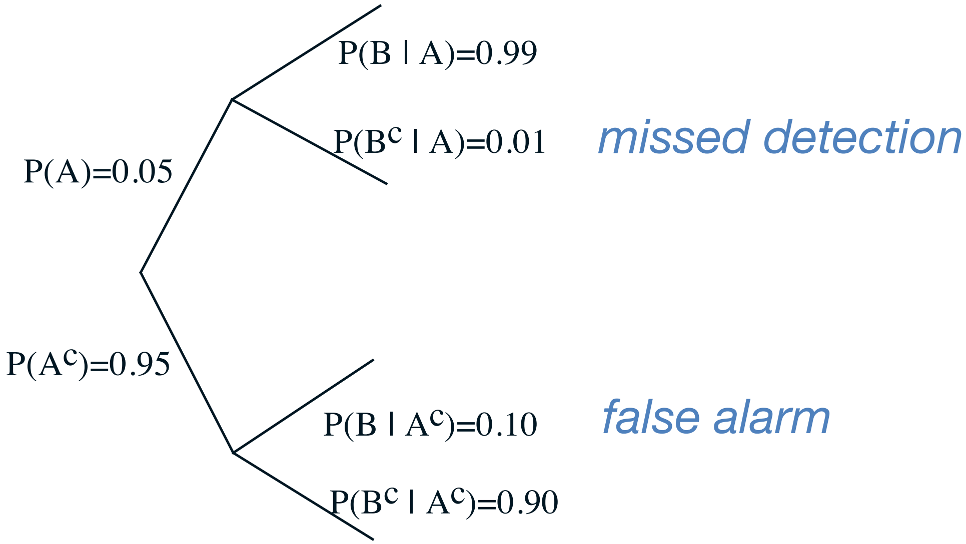



Conditioning on B defines a new probability model, with sample space B, and probability function P(A | B) for All A ⊑ B.

# Models Based on Conditional Probabilities

**Event A:** Airplane is flying above
**Event B:** Blip appears on radar

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




$$\mathbf{P}(A \cap B) =$$

$$\mathbf{P}(B) =$$

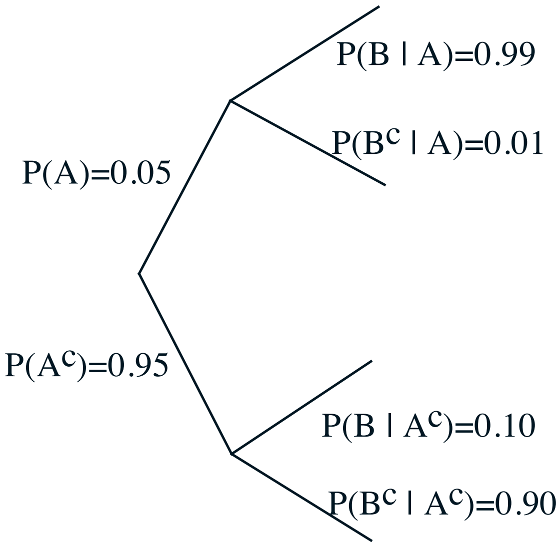
$$\mathbf{P}(A \mid B) =$$

# Models Based on Conditional Probabilities

**Event A:** Airplane is flying above

**Event B:** Blip appears on radar

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




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

$$\mathrm{P}(\mathrm{A} \cap \mathrm{B}) = P(B|A)P(A) = 0.99 \times 0.05$$

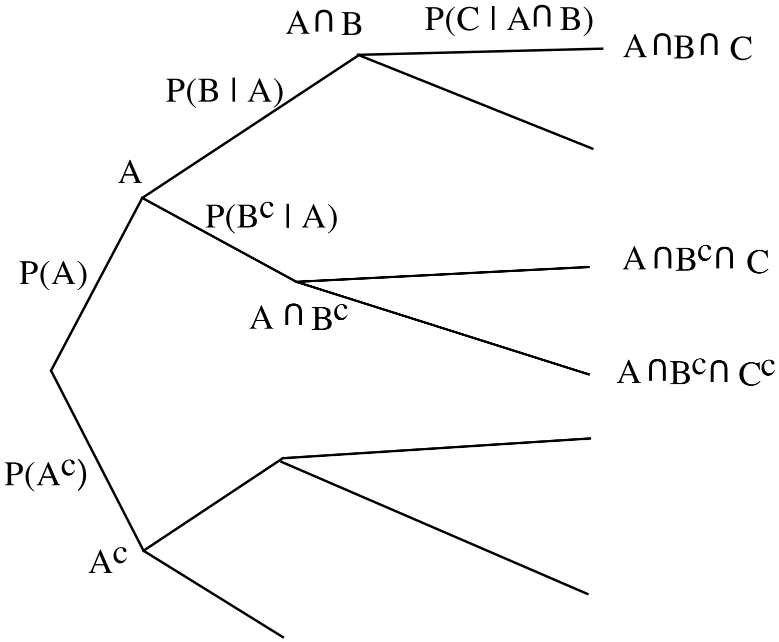
$$\mathrm{P}(\mathrm{B}) = \mathrm{P}(\mathrm{B} \cap \mathrm{A}) + \mathrm{P}(\mathrm{B} \cap \mathrm{A}^c)$$

$$\mathrm{P}(\mathrm{A} \mid \mathrm{B}) = \frac{P(B|A)P(A)}{\mathrm{P}(\mathrm{B} \cap \mathrm{A}) + \mathrm{P}(\mathrm{B} \cap \mathrm{A}^c)}$$

# Multiplication Rule

$$\mathbf{P}(A \cap B \cap C) = \mathbf{P}(A) \cdot \mathbf{P}(B \mid A) \cdot \mathbf{P}(C \mid A \cap B)$$

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



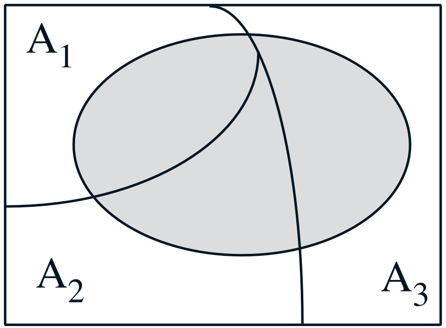

**Multiplication Rule**

Assuming that all of the conditioning events have positive probability, we have

$$\mathbf{P}\big(\cap_{i=1}^{n} A_i\big) = \mathbf{P}(A_1)\mathbf{P}(A_2 \mid A_1)\mathbf{P}(A_3 \mid A_1 \cap A_2) \cdots \mathbf{P}\big(A_n \mid \cap_{i=1}^{n-1} A_i\big).$$

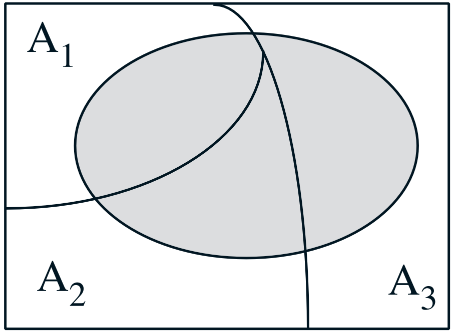
# CS145: Lecture 2 Outline

- Relative Frequencies and the Axioms of Probability
- Conditional Probabilities
- Bayes' Rule



# Total Probability Theorem

- Divide and conquer
- Partition of sample space into $A_1, A_2, A_3$
- Have $\mathbf{P}(B \mid A_i)$, for every $i$




$$
\begin{aligned}
P(A \mid B) &= \frac{P(A \cap B)}{P(B)} \\
P(A \cap B) &= P(B)P(A \mid B)
\end{aligned}
$$

$$
\begin{aligned}
B &= (B \cap A_1) \cup (B \cap A_2) \cup (B \cap A_3) \\
P(B) &= P(B \cap A_1) + P(B \cap A_2) + P(B \cap A_3) \\
P(B) &= P(A_1)P(B \mid A_1) + P(A_2)P(B \mid A_2) + P(A_3)P(B \mid A_3)
\end{aligned}
$$

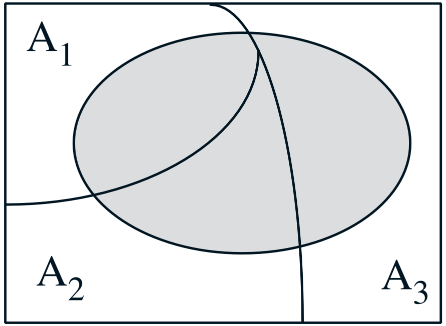
- One way of computing $\mathbf{P}(B)$:

$$
\begin{aligned}
\mathbf{P}(B) = \quad & \mathbf{P}(A_1)\mathbf{P}(B \mid A_1) \\
+ \; & \mathbf{P}(A_2)\mathbf{P}(B \mid A_2) \\
+ \; & \mathbf{P}(A_3)\mathbf{P}(B \mid A_3)
\end{aligned}
$$

# Bayes' Rule

- "Prior" probabilities $\mathbf{P}(A_i)$
  - initial "beliefs"
- We know $\mathbf{P}(B \mid A_i)$ for each $i$
- Wish to compute $\mathbf{P}(A_i \mid B)$
  - revise "beliefs", given that $B$ occurred

T. Bayes.




$$\mathbf{P}(A_i \mid B) = \frac{\mathbf{P}(A_i \cap B)}{\mathbf{P}(B)} = \frac{\mathbf{P}(A_i)\mathbf{P}(B \mid A_i)}{\mathbf{P}(B)} = \frac{\mathbf{P}(A_i)\mathbf{P}(B \mid A_i)}{\sum_j \mathbf{P}(A_j)\mathbf{P}(B \mid A_j)}$$

# Reasoning Under Uncertainty

- 1/1000 of tourists who visit tropical country $X$ return with a dangerous virus $Y$.
- There is a test to check for the virus. The test has 5% *false positive* rate and no *false negative* error.
- You returned from country $X$, took the test, and it was positive. Should you take the painful treatment for the virus?

$A$ - has the virus, $B$ - positive in the test.

$$
\begin{aligned}
P(A \mid B) &= \frac{P(A \cap B)}{P(B)} = \frac{P(A \cap B)}{P(A)P(B \mid A) + P(\bar{A})P(B \mid \bar{A})} \\
&= \frac{1/1000}{1/1000 + (999/1000)(5/100)} = \frac{20}{1019} \approx 2\%
\end{aligned}
$$

**Explanation:** Out of 1000 tourist, 1 will have the virus and another 50 will be false positive in the test.