Lecture 6

Probability

H T

Example: When you toss a coin, there are only two possible outcomes, heads and tails.

$P(H) = P(T) = \frac{1}{2}$

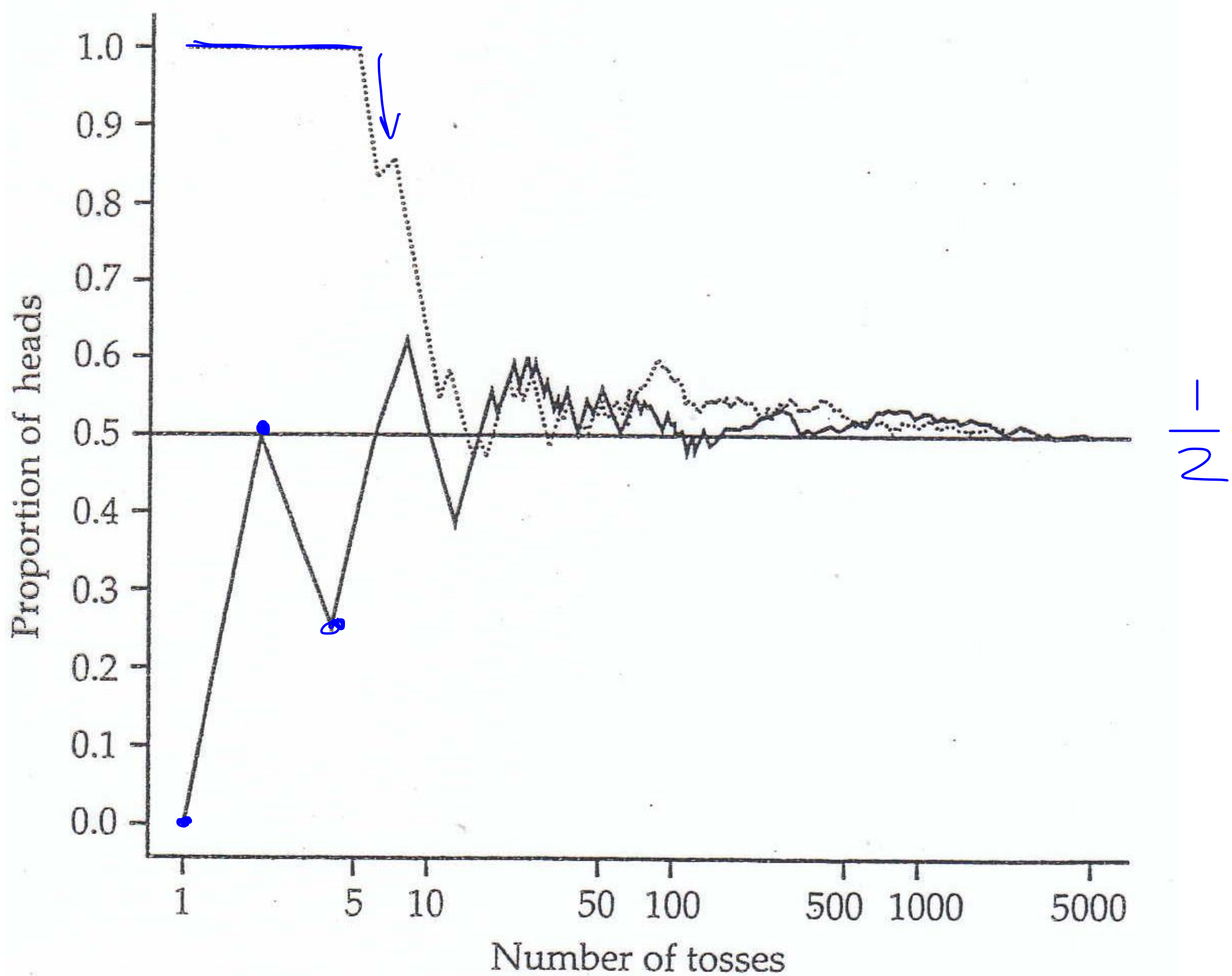
What if we toss a coin two times?

HH, TT, TH, HT

$P(\text{all heads}) = P(HH)$

$= \frac{1}{4}$

Figure below shows the results of tossing a coin 5000 times twice.
For each number of tosses from 1 to 5000, we have plotted the proportion of those tosses that gave a head.



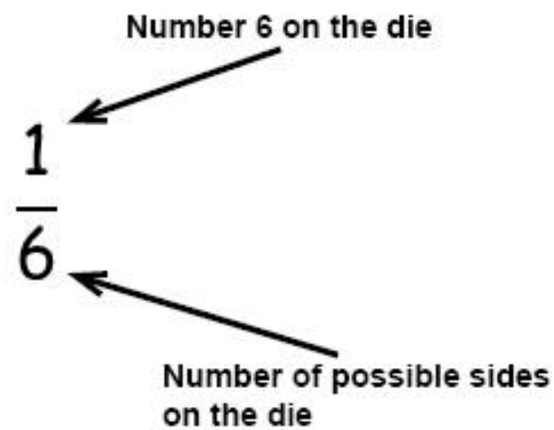
Trial A (solid line) begins tail, head, tail, tail. Trial B starts with five straight heads, so the proportion of heads is 1 until the sixth toss.

Caution: Probability describes only what happens in the long run
(so **probability of a head is 0.5**).

Randomness

Definition: We call a phenomenon **random** if individual outcomes are uncertain but there is nonetheless a regular distribution of outcomes in a large number of repetitions.

The **probability** of any outcome of a random phenomenon is the proportion of times the outcome would occur in a very long series of repetitions.



$$\frac{1000}{6000}$$

History of many tosses of a coin:

- The French naturalist Count Buffon (1707-1788) tossed a coin 4,040 times.

  Result:

  2,048 heads, or proportion 2,048/4,040 = 0.5069 for heads.

- Around 1900, the English statistician Karl Pearson heroically tossed a coin 24,000 times.

  Result:

  12,012 heads, a proportion of 0.5005.

- While imprisoned by the Germans during World War II, the South African statistician John Kerrich tossed a coin 10,000 times.

  Result:

  5,067 heads, proportion of heads = 0.5067.

Probability Models

A description of a random phenomenon in the language of mathematics is called a **probability model**.

For example, to build a probability model for a coin tossing example we have to

- List all possible outcomes

- Assign probabilities to those outcomes

Definition: The **sample space** ***S*** of a random phenomenon is the set of all possible outcomes.

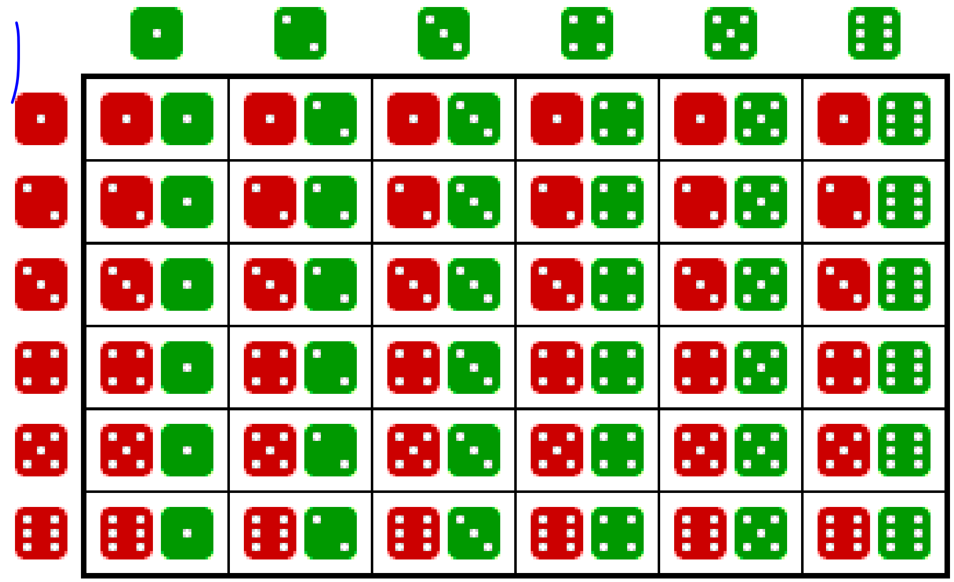
Examples:

(1) Toss a coin (we call a coin **fair** if the probability of heads is 0.5)

$S = \{H, T\}$

(2) Roll two dice

$S = \{(1,1), (1,2), \ldots, (1,6), \ldots, (6,1), \ldots, (6,6)\}$



\# of outcomes $= 36$

$6 \times 6 = 36$

(3) Random digits

$S = \{0, 1, 2, \ldots, 9\}$

$P: \frac{1}{10}, \frac{1}{10}, \frac{1}{10}, \ldots, \frac{1}{10}$

probability model

(4) Toss a coin 3 times

$S = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$

<u>Notation</u>: $n(S) = |S| =$ number of elements in S.

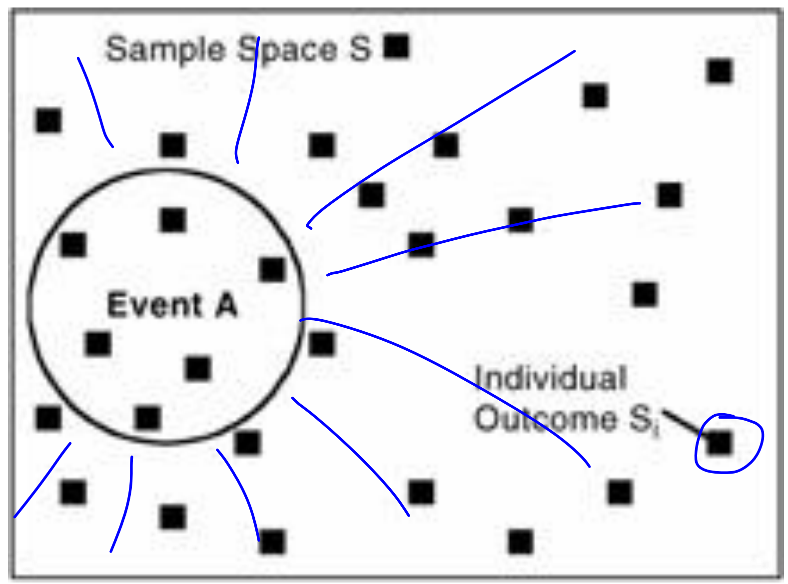
$|S| = 8 \leftarrow$ we toss a coin 3 times

<u>Definition</u>: An **event** is an outcome or a set of outcomes of a random phenomenon. That is, an event is a **subset** of the sample space.

Toss a coin 3 times

$A = \{\text{exactly 2 heads}\}$

$= \{HHT, HTH, THH\}$



$|S| = 8$

$|A| = 3$

$P(A) = \frac{3}{8}$

Probability Rules:

1. $0 \leq P(A) \leq 1$, for any event A.

sure event

2. If S is the sample space in a probability model, then $P(S) = 1$.

mutually exclusive

3. Two events A and B are **disjoint** if they have no outcomes in common and so can never occur together. If A and B are disjoint,

$$\boxed{P(A \text{ or } B) = P(A) + P(B)}$$

This is the **addition rule for disjoint events**.

4. The **compliment** of any event A is the event that A does not occur, written as A^c or $\bar{A}$. The **complement rule** states that

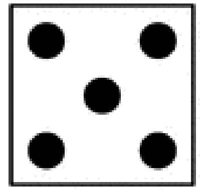
$$P(A^c) = 1 - P(A)$$

$S = \{A, A^c\}$, A, A^c are disjoint

$1 = P(S) = P(A) + P(A^c)$

So, $P(A^c) = 1 - P(A)$

Example: Roll a die.

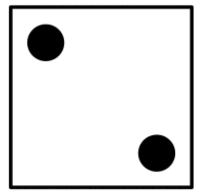


$S = \{1, 2, 3, 4, 5, 6\}$

A = {number on the upper face is odd}

$A = \{1, 3, 5\}$

$P(A) = \frac{3}{6} = \frac{1}{2}$



B = {number on the upper face is even)

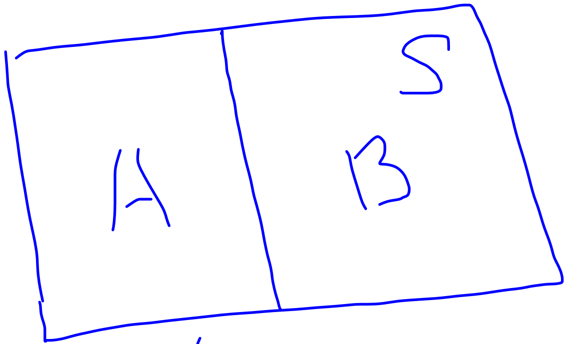
$B = \{2, 4, 6\}$

$P(B) = \frac{3}{6} = \frac{1}{2}$

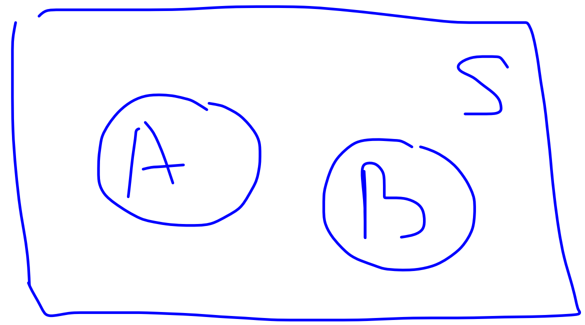
$P(A \text{ or } B) = P(A) + P(B) = \frac{1}{2} + \frac{1}{2} = 1$

A, B are disjoint, $\{A \text{ or } B\} = S$

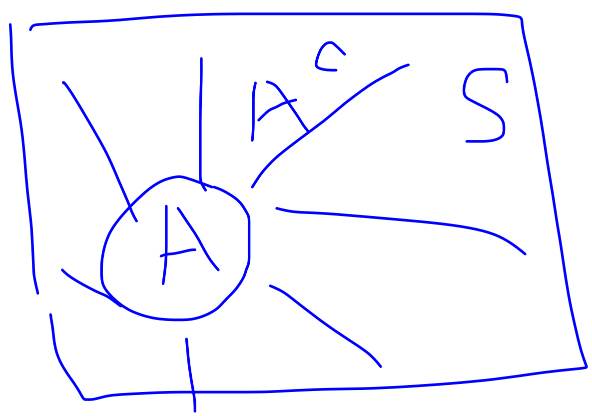
Venn Diagrams



for the prev. ex.



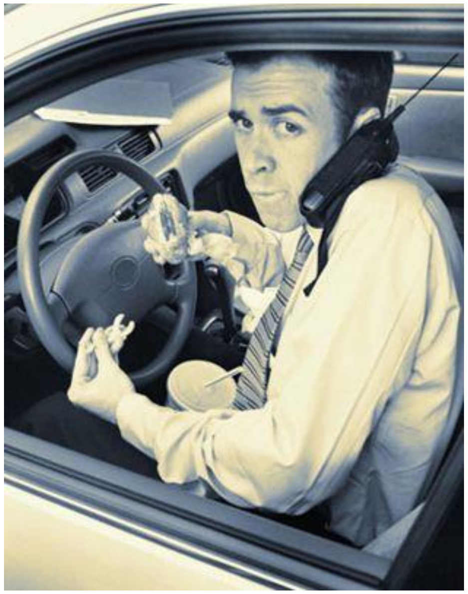
A, B are disjoint



$P(A \text{ or } A^c) = P(S) = 1$

Example:

> Some countries are considering laws that will ban the use of cell phones while driving because they believe that the ban will reduce phone-related car accidents.



One study classified these types of accidents by the day of the week when they occurred.

For this example, we use the values from this study as our probability model. Here are the probabilities:

Day	Sun	Mon	Tue	Wed	Thu	Fri	Sat
Probability	0.03	0.19	0.18	0.23	0.19	0.16	0.02

Day	Sun	Mon	Tue	Wed	Thu	Fri	Sat
Probability	0.03	0.19	0.18	0.23	0.19	0.16	0.02

$\sum p = 1$

sum of all probabilities

$P(\text{accident on weekend})$

$= P(\text{Sat or Sun})$

$= P(\text{Sat}) + P(\text{Sun})$

$= 0.02 + 0.03 = 0.05$

5% of accidents happen on weekends

$P(\text{weekdays}) = P(\text{Mon}) + P(\text{Tue}) + \ldots + P(\text{Fri})$

$= 1 - P(\text{weekend}) = 0.95$

How to assign probabilities?

<u>Finite Sample Space</u>:

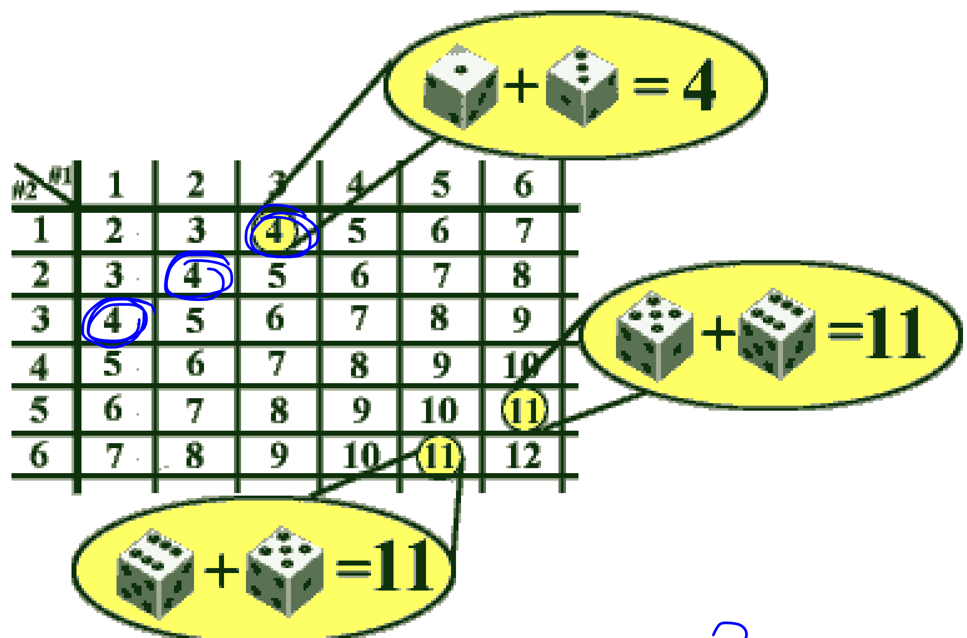
- Assign a probability to each individual outcome. These probabilities must be numbers between 0 and 1 and must have sum 1.

- The probability of any event is the sum of probabilities of the outcomes making up the event.

<u>Example</u>: We toss two dice.

$A = \{\text{sum} = 4\}$

$= \{(1,3), (2,2), (3,1)\}$



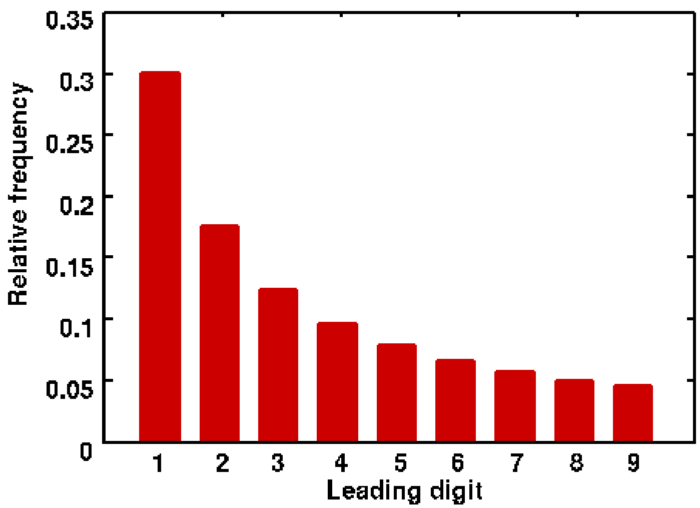
$P(A) = P(1,3) + P(2,2) + P(3,1)$

$= \frac{1}{36} + \frac{1}{36} + \frac{1}{36} = \frac{3}{36}$

$B = \{\text{sum} = 11\} = \{(5,6), (6,5)\}$; $P(B) = \frac{2}{36}$

Example (Benford's Law): Faked numbers in tax returns, payment records, invoices, expense account claims, and many other settings often display patterns that aren't present in legitimate records. It is a striking fact that the first digits of numbers in legitimate records often follow a distribution known as **Benford's law**. Here is it:

First digit	1	2	3	4	5	6	7	8	9
Probability	0.301	0.176	0.125	0.097	0.079	0.067	0.058	0.051	0.046



$A = \{1^{st}$ digit is $1\}$

$B = \{1^{st}$ digit is 6 or greater$\}$

$C = \{1^{st}$ digit is odd$\}$

$X = 1^{st}$ digit, then

$A = \{X = 1\}$ $\quad P(A) = 0.301$

$B = \{X \geq 6\}$ $\quad P(B) = P(6) + P(7) + P(8) + P(9)$

First digit	1	2	3	4	5	6	7	8	9
Probability	0.301	0.176	0.125	0.097	0.079	0.067	0.058	0.051	0.046

$P(B) = 0.067 + 0.058 + 0.051 + 0.046$

$= 0.222$

$C = \{X \text{ is odd}\} = \{1, 3, 5, 7, 9\}$

$P(C) = P(1) + P(3) + P(5) + P(7) + P(9) = 0.609$

$P(B \text{ or } C) = P(1) + P(3) + P(5) + P(6) + \ldots + P(9)$

$\{B \text{ or } C\} = \{1, 3, 5, 6, 7, 8, 9\} \quad = 0.727$

$P(B) + P(C) = 0.222 + 0.609$

because B and C are not disjoint

Equally Likely Outcomes

<u>Example</u>: Roll a die.

x:	1	2	3	4	5	6
p:	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

If a random phenomenon has k possible outcomes, all equally likely, then each individual outcome has probability $\mathbf{1/k}$.

The probability of any event A is

$$P(A) = \frac{|A|}{|S|} = \frac{|A|}{k}$$

<u>Example</u>: Roll a die.

$A = \{\text{odd}\} = \{1, 3, 5\}$

$$P(A) = \frac{|A|}{|S|} = \frac{3}{6} = \frac{1}{2}$$

Example (Benford's Law):

First digit	1	2	3	4	5	6	7	8	9
Probability	0.301	0.176	0.125	0.097	0.079	0.067	0.058	0.051	0.046

$$\begin{array}{ccccc} 1 & 2 & 3 & \cdots & 9 \\ 1/9 & 1/9 & 1/9 & \cdots & 1/9 \end{array}$$

$\leftarrow$ If we use random digits to create fake numbers

$$B = \{X \geq 6\}$$

$$P(B) = P(6) + P(7) + P(8) + P(9)$$

$$= \frac{1}{9} + \frac{1}{9} + \frac{1}{9} + \frac{1}{9} = \frac{4}{9} = 0.444$$

$\times$

0.222

Independence

Example: Toss a coin twice.

$S = \{HH, HT, TH, TT\}$

A = {first toss is a head}

$A = \{HH, HT\}$ $P(A) = \frac{2}{4} = \frac{1}{2}$

B = {second toss is a head}

$B = \{HH, TH\}$ $P(B) = \frac{2}{4} = \frac{1}{2}$

A and B are not disjoint

A, B are independent, because tosses are independent

$P(A \text{ and } B) = P(HH) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = P(A)P(B)$

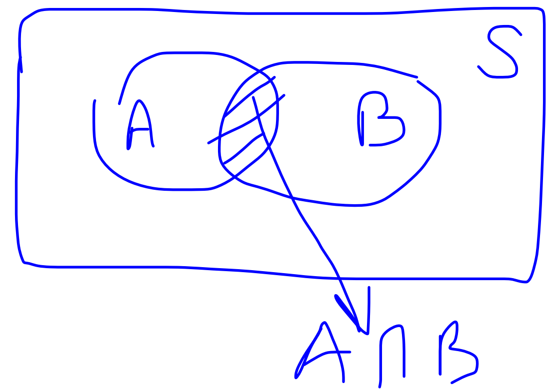
The Multiplication Rule for Independent Events: Two events A and B are **independent** if knowing that one occurs does not change the probability that the other occurs. If A and B are independent,

$$\boxed{P(A \text{ and } B) = P(A)P(B)}$$

Independence $\not\Rightarrow$ disjoint
(by prev. ex)

Definition: The **intersection** of any collection of events is the event that all of the events occur.

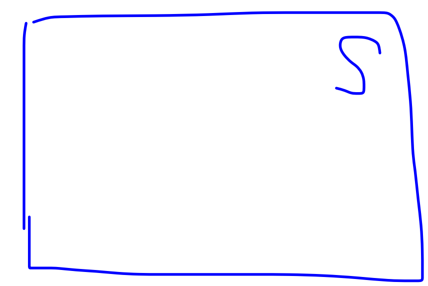
A and B $= A \cap B$



$A = \{1, 2, 3\}$

$B = \{2, 3, 5\}$

$A \cap B = \{2, 3\}$



$S^c = \emptyset$

Example: Roll a die.

$A = \{\text{odd}\} = \{1, 3, 5\}$

$B = \{\text{even}\} = \{2, 4, 6\}$

A and B $= \emptyset \rightarrow$ empty set

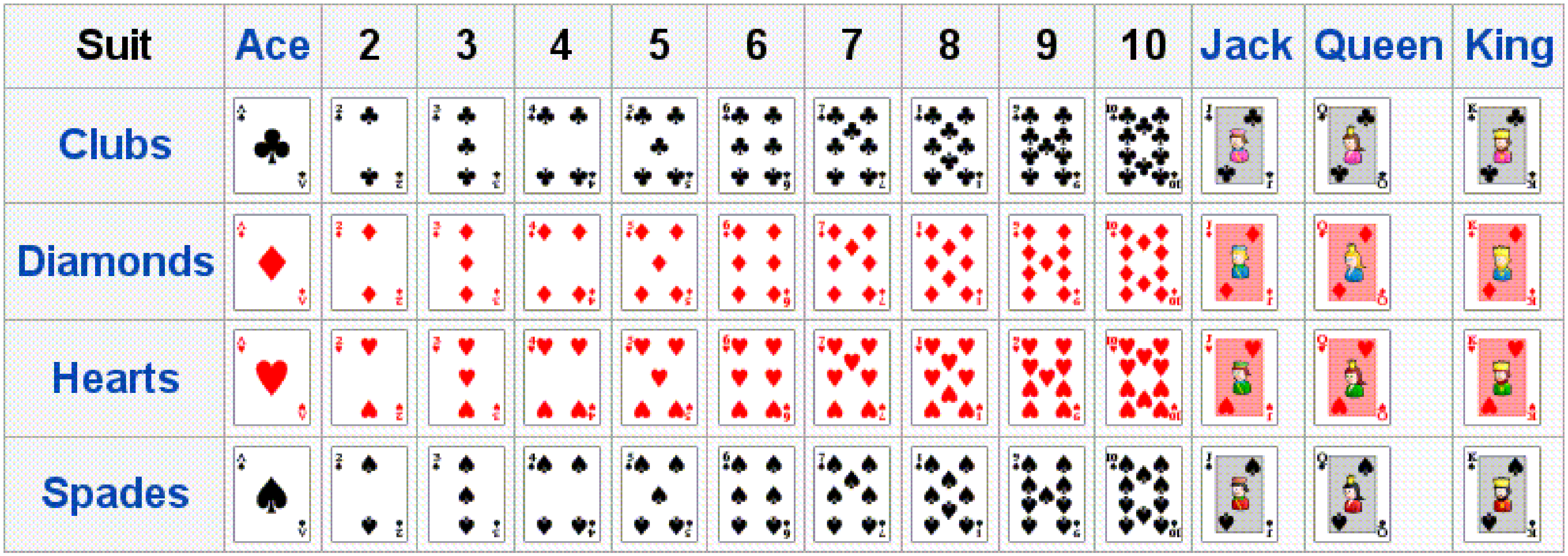
A, B are disjoint

$P(\emptyset) = 0 = P(A \cap B) \neq P(A)P(B) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$

$\Rightarrow A, B$ are dependent

Example: The colors of successive cards dealt from the same deck are not independent. A standard 52-deck contains 26 red and 26 black cards.

Example set of 52 poker playing cards



P(first card dealt is red) = $\frac{26}{52} = \frac{1}{2}$

P(second card dealt is also red) = $\frac{25}{51}$

Example:

IQ Test

If you take an IQ test or other test twice in succession, the two test scores are not independent.

The learning that occurs on the first attempt influences your second attempt. If you learn a lot, then your second test score might be a lot higher than your first test score.

This phenomenon is called a *carry-over effect*.

General Probability Rules

So far we learned five rules:

1. $0 \leq P(A) \leq 1$ for any event A.

2. $P(S) = 1$. $P(\emptyset) = 0$

3. **Addition rule:** If A and B are **disjoint**, then

$$P(A \text{ or } B) = P(A) + P(B)$$

4. **Compliment rule:** for any event

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

5. **Multiplication rule**: If A and B are independent, then

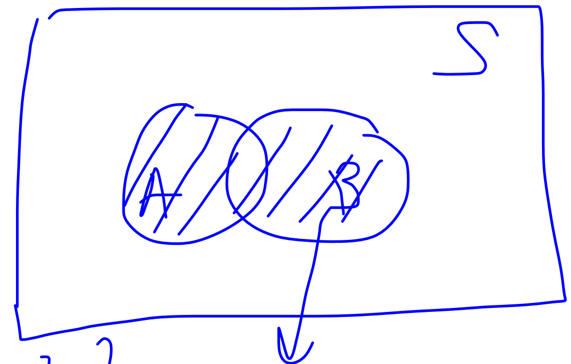
$$P(A \text{ and } B) = P(A)P(B)$$

Definition: The **union** of any collections of events is the event that at least one of the collections occurs.

A or B = $A \cup B$

$A = \{1, 2, 3\}$, $B = \{3, 4, 5\}$

$A \cup B = \{1, 2, 3, 4, 5\}$, $A \cap B = \{3\}$

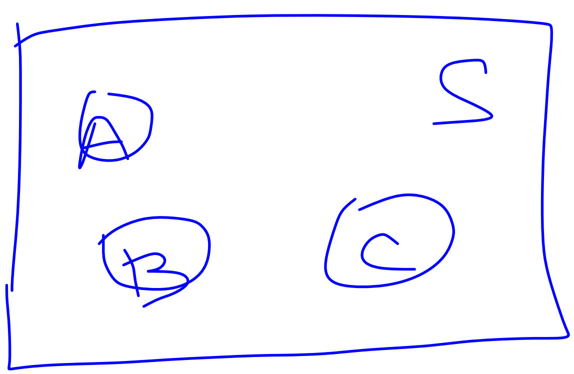


$A \cup B$

Addition Rule for Disjoint Events: If events *A*, *B*, and *C* are disjoint in the sense that no two have any outcomes in common, then

$$P(one\ or\ more\ of\ A, B, C) = P(A) + P(B) + P(C)$$

This rule extends to any number of disjoint events.

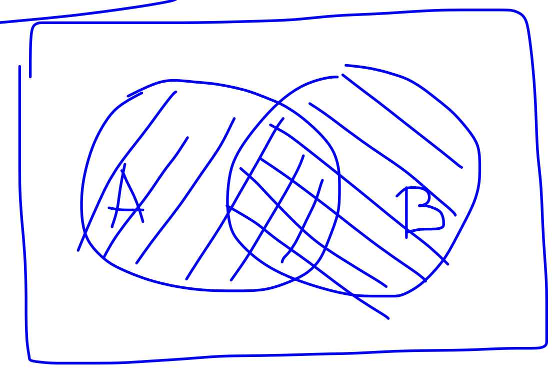


$P(A \cup B \cup C) = P(A) + P(B) + P(C)$

A or B or C

<u>General Addition Rule for Unions of Two Events</u>: For any two events A and B,

$$\boxed{P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)}$$



<u>Note</u>: If A, B are disjoint,

then $A \cap B = \emptyset$

$$P(A \cap B) = P(\emptyset) = 0$$

$$\boxed{P(A \text{ or } B) = P(A) + P(B)}$$

<u>Example</u>: Deborah and Matthew are awaiting word on whether they have been made partners of their law firm.

$P(A)$

$\|$

Deborah guesses that her probability of making partner is 0.7,

and that Matthew's is 0.5. $= P(B)$

She also guesses that the probability that both she and Matthew are made partners is 0.3. $= P(A \text{ and } B)$

$P(\text{at least one is promoted}) = P(A \cup B)$ or

$A = \{\text{Deborah is promoted}\}$

$B = \{\text{Matthew is promoted}\}$

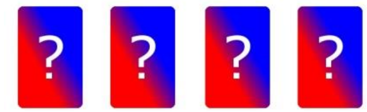
$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$

$= 0.7 + 0.5 - 0.3 = 0.9$

$P(\text{neither is promoted}) = 1 - 0.9 = 0.1$

Conditional Probability

<u>Example</u>: There are 52 cards in the deck. Four cards are dealt, but we didn't look at them.



P(next card is ace) = $\frac{4}{52} = \frac{1}{13}$

Suppose there is one ace among first four cards, then

P($\underbrace{\text{next card is ace}}_{B}$, $\underbrace{\text{given one of four is ace}}_{A}$) = $\frac{3}{48} = \frac{1}{16}$

P(B|A) = P(B given A) gives the probability of one event under the condition that another event has occurred. Such probabilities are called **conditional probabilities**.

Example: Students of the University of Toronto received 10,000 grades last semester. Table below breaks down these grades by which school of the university taught the course. The schools are Liberal Arts, Engineering and Physical Sciences, and Health and Human Services.

	A	B	Below B	**Total**
Liberal Arts	2,142	1,890	2,268	**6,300**
Engineering and Physical Sciences	368	432	800	**1,600**
Health and Human Services	882	630	588	**2,100**
Total	**3,392**	**2,952**	**3,656**	**10,000**

Consider the two events:

A = the grades comes from an engineering and physical sciences course

B = the grade is below B

	A	B	Below B	Total
Liberal Arts	2,142	1,890	2,268	6,300
Engineering and Physical Sciences	368	432	800	1,600
Health and Human Services	882	630	588	2,100
Total	3,392	2,952	3,656	10,000

A = the grades comes from an engineering and physical sciences course
B = the grade is below B

$P(A) = \frac{1600}{10,000} = 0.16 \rightarrow$ marginal probability

$P(B) = \frac{3656}{10000} = 0.3656$

$P(A \text{ and } B) = \frac{800}{10000} = 0.08 \rightarrow$ joint probability

$P(B \mid A) = \frac{800}{1600} = 0.5$ - conditional probability

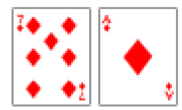
$P(A \text{ and } B) = P(A)\,P(B \mid A)$

$0.08 \qquad\qquad 0.16 \cdot 0.5$

Multiplication Rule: The probability that both of two events A and B happen together can be found by

$$\boxed{P(A \text{ and } B) = P(A)P(B|A)}$$

Example: You are at the poker table.

You want to draw two diamonds in a row. 

You can see 11 cards on the table. Of these, 4 are diamonds.

$$P(\underbrace{1^{st} \text{ card is } \diamondsuit}_{A} \text{ and } \underbrace{2^{nd} \text{ card is } \diamondsuit}_{B}) = ?$$

$$P(A \text{ and } B) = P(A)P(B|A)$$

$$P(A) = \frac{9}{41} \; ; \; P(B|A) = \frac{8}{40}$$

$$P(A \text{ and } B) = \frac{9}{41} \cdot \frac{8}{40} = 0.044$$

Definition: When $P(A) > 0$, the **conditional probability** of B given A is

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

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

Note: Be sure to keep in mind the distinct roles in $P(B|A)$ of the event B whose probability we are computing and the event A that represents the information we are given.

Example: What is the conditional probability that a grade at the UofT is an A, given that it comes from a liberal arts course?

	A	B	Below B	Total
Liberal Arts	2,142	1,890	2,268	6,300
Engineering and Physical Sciences	368	432	800	1,600
Health and Human Services	882	630	588	2,100
Total	3,392	2,952	3,656	10,000

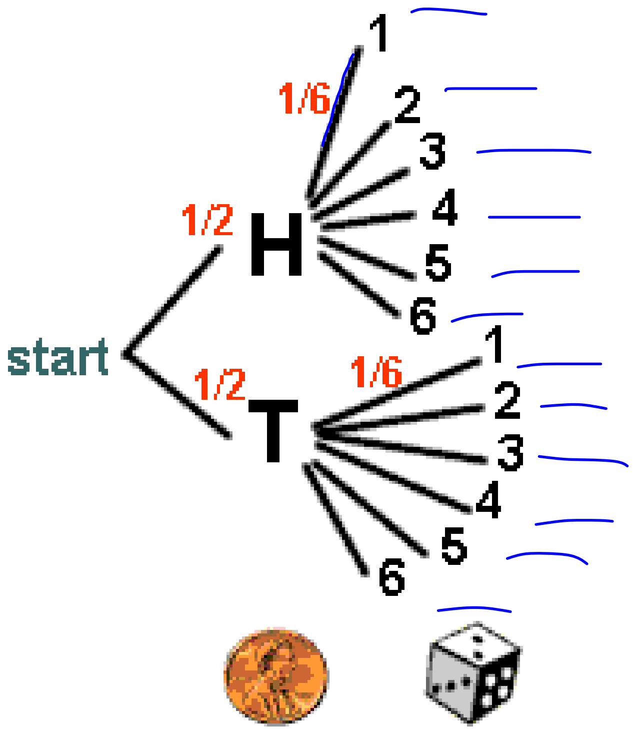
$$P(\text{grade is A} \mid \text{liberal arts})$$

$$= \frac{P(\text{grade is A and liberal arts})}{P(\text{liberal arts})} = \frac{2142/10000}{6300/10000} = \frac{2142}{6300} = 0.34$$

Tree Diagrams

- Tree diagrams are a way to draw out different possible outcomes.
- A tree diagram helps us figure out the sample space for an event.
- We can use a tree diagram any time we have a situation where there is more than one possible outcome.

Example: Show the sample space for tossing one coin and rolling one die.

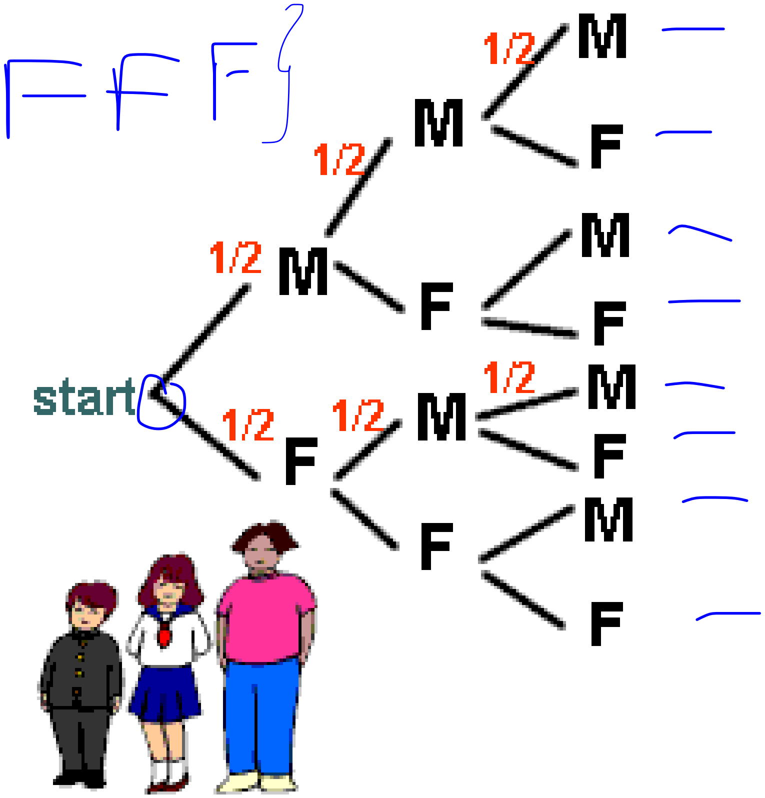


$$S = \{(H,1), (H,2), \ldots, (H,6), (T,1), (T,2), \ldots, (T,6)\}$$

$$|S| = 2 \cdot 6 = 12$$

Example: A family has three children. How many outcomes in the sample space that indicates the sex of the children?

$S = \{MMM, MMF, \ldots, FFF\}$



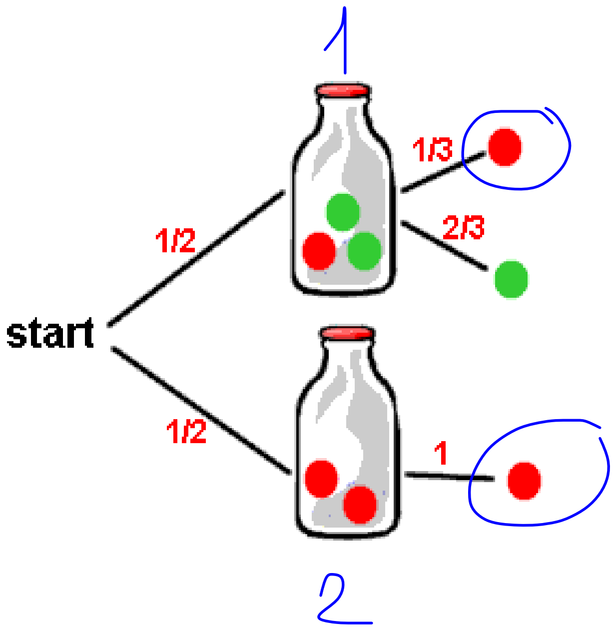
$|S| = 8$

$P(MMM) = \frac{1}{8}$

$= P(M)P(M)P(M) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2}$

$= \frac{1}{8}$

<u>Example:</u> There are two identical bottles. One bottle contains 2 green balls and 1 red ball. The other contains 2 red balls. A bottle is selected at random and a single ball is drawn. What is the probability that the ball is red?



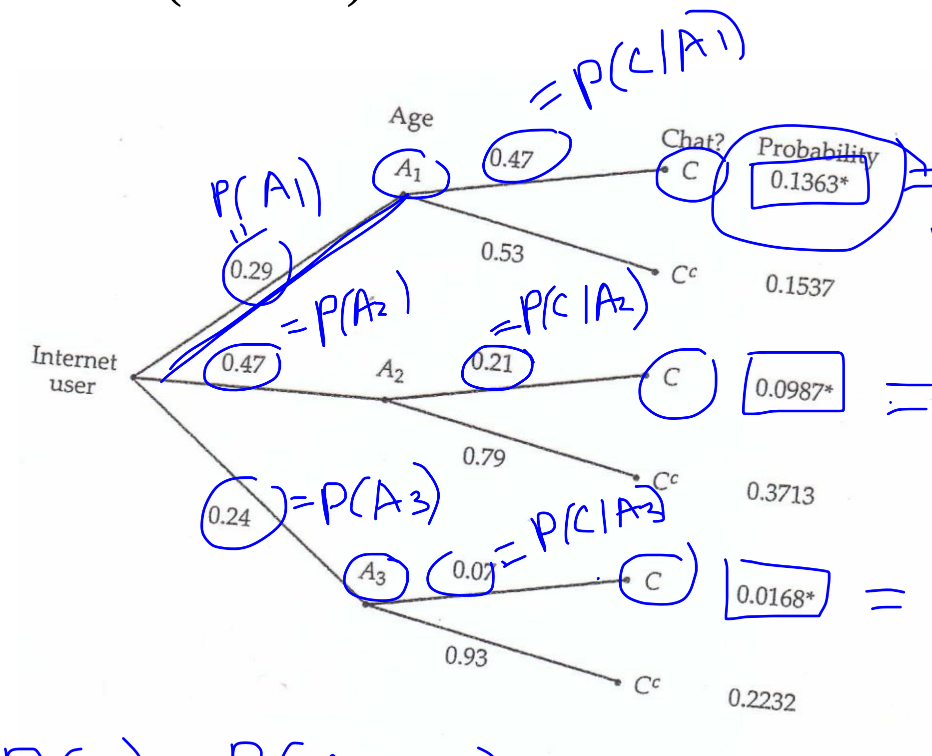
$$P(\text{red ball})$$

$$= P(\text{bottle 1 and red ball}) + P(\text{bottle 2 and red ball})$$

$$= P(\text{bottle 1})\,P(\text{red ball} \mid \text{bottle 1}) + P(\text{bottle 2})\,P(\text{red ball} \mid \text{bottle 2})$$

$$= \frac{1}{2}\cdot\frac{1}{3} + \frac{1}{2}\cdot 1 = \frac{1}{6} + \frac{1}{2} = \frac{2}{3}$$

Example: Online chat rooms are dominated by the young. Teens are the biggest users. If we look only at adult Internet users (aged 18 and over), 47% of the 18 to 29 age group chat, as do 21% of the 30 to 49 age group and just 7% of those 50 and over. To learn what percent of all Internet users participate in chat, we also need the age breakdown of users. Here is it: 29% of adult Internet users are 18 to 29 years old (event A_1), another 47% are 30 to 49 (event A_2), and the remaining 24% are 50 and over (event A_3). What is the probability that a randomly chosen user of the Internet participates in chat rooms (event C)?



$= 0.29 \cdot 0.47$
$\boxed{P(A_1)\,P(C \mid A_1)}$

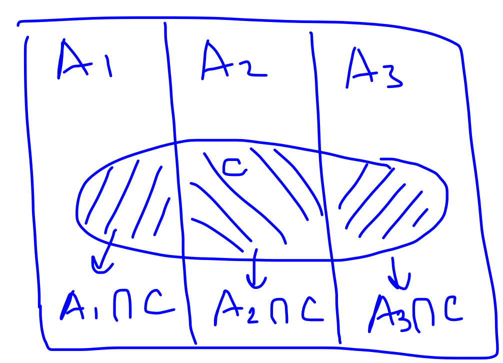
$= 0.47 \cdot 0.21$
$P(A_2)\,P(C \mid A_2)$

$= 0.24 \cdot 0.07$
$P(A_3)\,P(C \mid A_3)$

$P(C) = P(A_1 \cap C) + P(A_2 \cap C) + P(A_3 \cap C)$

$= P(A_1)P(C \mid A_1) + P(A_2)P(C \mid A_2) + P(A_3)P(C \mid A_3)$

$= 0.1363 + 0.0987 + 0.0168 = 0.2518$



Bayes's Rule: Suppose that $A_1, A_2, \ldots, A_k$ are disjoint events whose probabilities are not 0 and add to exactly 1. That is, any outcome is in exactly one of these events. Then if C is any other event whose probability is not 0 or 1,

$$P(A_i|C) = \frac{P(C|A_i)P(A_i)}{P(C|A_1)P(A_1) + \cdots + P(C|A_k)P(A_k)}$$

$$= \frac{P(A_i \text{ and } C)}{P(A_1 \text{ and } C) + \ldots + P(A_k \text{ and } C)} = \frac{P(A_i \text{ and } C)}{P(C)}$$

Example: What % of adult chat room participants are aged 18 to 29?

$$P(A_1|C) = \frac{P(C|A_1)P(A_1)}{P(C|A_1)P(A_1) + P(C|A_2)P(A_2) + P(C|A_3)P(A_3)}$$

$$= \frac{0.1363}{0.1363 + 0.0587 + 0.0168}$$

$$= \frac{0.1363}{0.2518} = 0.5413$$

$$P(A_2|C) = \frac{P(A_2|C)P(A_2)}{P(C|A_1)P(A_1) + P(C|A_2)P(A_2) + P(C|A_3)P(A_3)}$$

$$= \frac{0.0987}{0.2518}$$

Definition: Two events A and B that both have positive probability are **independent** if $P(B|A) = P(B)$.

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