some pictures are taken from the internet. credits belong to original authors.
**Basic language and notation**

**Experiment** $\rightarrow$ **outcome** $\rightarrow$ **number**

**Example 1**: Roll a die

**Sample space** $\Omega$: The set of all the outcomes (or sample points, elements).

*Visualize*: randomly sample an outcome from the sample space.
Basic language and notation

**Experiment → outcome → number**

**Example 1**: Roll a die

**Sample space** $\Omega$: The set of all the outcomes.

**Event** $A$:
1. A **statement** about the outcome, e.g., bigger than 4.
2. A **subset** of sample space, e.g., $\{5, 6\}$. 
**Basic language and notation**

**Experiment** → **outcome** → **number**

**Example 1**: Roll a die

Assume the die is fair so that all the outcomes are **equally likely**.

**Probability**: defined on event:

\[
P(A) = \frac{|A|}{|\Omega|} = \frac{2}{6} = \frac{1}{3}.
\]

$|A|$ counts the size of $A$, i.e., the number of elements in $A$. 
Basic language and notation

**Experiment** $\rightarrow$ **outcome** $\rightarrow$ **number**

**Example 1**: Roll a die

![Dice Image]

**Random variable**: Let $X$ be the number:

$$P(X > 4) = \frac{1}{3}.$$

An event is a **math statement** about the random variable. We can either use events or use random variables. In Parts 2 and 3, we will focus on random variables.
Basic language and notation

**Experiment** → **outcome** → **number**

**Example 1**: Roll a die

![Diagram of a die with sample points and event A]

**Conditional probability**: Let $B$ be the event that the number is 6. Given that $A$ happens, what is the probability of $B$?

$$P(B|A) = \frac{1}{2}.$$  

As if we randomly sample a number from $A$.  
As if $A$ is the sample space.
Basic language and notation

**Experiment → outcome → number**

**Example 1:** Roll a die

Random variable

\[ P(X = 6 | X > 4) = \frac{1}{2}. \]
**Basic language and notation**

**Example 1:** Roll a die

![Sample Space and Event A](image.png)

**Complement**

Statement: Not $A$

Subset: $A^c = \{1, 2, 3, 4\}$. 
Example 1: Roll a die

\[ A = \{1,2,3\} \]
\[ B = \{3,4,5\} \]
\[ A \cup B = \{1,2,3,4,5\} \]

Venn diagram

**Union**

Statement: \( A \) or \( B \).

Subset: \( A \cup B \).
**Example 1:** Roll a die

\[
\begin{align*}
A &= \{1,2,3,4\} \\
B &= \{3,4,5,6\} \\
A \cap B &= \{3,4\}
\end{align*}
\]

**Intersection**

Statement: \( A \) and \( B \).

Subset: \( A \cap B \).
Population proportion

**Experiment** $\rightarrow$ **outcome** $\rightarrow$ **number**

**Example 2**: Sample a random person from a population of 100 people, 50 males and 50 females. 30 males are taller than 6 ft, 10 females are taller than 6 ft.

*The sample space $\Omega$ is the population.*

<table>
<thead>
<tr>
<th></th>
<th>taller than 6 ft</th>
<th>shorter than 6 ft</th>
</tr>
</thead>
<tbody>
<tr>
<td>male</td>
<td>30</td>
<td>50</td>
</tr>
<tr>
<td>female</td>
<td>10</td>
<td>50</td>
</tr>
</tbody>
</table>
Experiment $\rightarrow$ outcome $\rightarrow$ number

**Example 2:** Let $A$ be the event that the person is male. Let $B$ be the event that the person is taller than 6 feet (or simply the person is tall). $A$ is the sub-population of males, and $B$ is the sup-population of tall people.

<table>
<thead>
<tr>
<th></th>
<th>male</th>
<th>female</th>
</tr>
</thead>
<tbody>
<tr>
<td>taller than 6 ft</td>
<td>30</td>
<td>10</td>
</tr>
<tr>
<td>shorter than 6 ft</td>
<td>50</td>
<td>50</td>
</tr>
</tbody>
</table>
Population proportion

**Experiment → outcome → number**

**Example 2:** A male, B tall.

<table>
<thead>
<tr>
<th></th>
<th>male</th>
<th>female</th>
</tr>
</thead>
<tbody>
<tr>
<td>taller than 6 ft</td>
<td>30</td>
<td>10</td>
</tr>
<tr>
<td>shorter than 6 ft</td>
<td>50</td>
<td>50</td>
</tr>
</tbody>
</table>

\[
P(A) = \frac{|A|}{|\Omega|} = \frac{50}{100} = 50\%.
\]

\[
P(B) = \frac{|B|}{|\Omega|} = \frac{30 + 10}{100} = 40\%.
\]

**Probability = population proportion.**
Experiment → outcome → number

Example 2: $A$ male, $B$ tall.

<table>
<thead>
<tr>
<th></th>
<th>taller than 6 ft</th>
<th>shorter than 6 ft</th>
</tr>
</thead>
<tbody>
<tr>
<td>male</td>
<td>50</td>
<td>30</td>
</tr>
<tr>
<td>female</td>
<td>10</td>
<td>50</td>
</tr>
</tbody>
</table>

$$P(A|B) = \frac{|A \cap B|}{|B|} = \frac{30}{40} = 75\%.$$  

Among tall people, what is the proportion of males?

$$P(B|A) = \frac{|A \cap B|}{|A|} = \frac{30}{50} = 60\%.$$  

Among males, what is the proportion of tall people?

Conditional probability = proportion within sub-population.
Link between event and random variable.

**Example 2**: A male, B tall.

Let $\omega \in \Omega$ be a person. Let $X(\omega)$ be the gender of $\omega$, so that $X(\omega) = 1$ if $\omega$ is male, and $X(\omega) = 0$ if $\omega$ is female. Let $Y(\omega)$ be the height of $\omega$. Then

$$A = \{\omega : X(\omega) = 1\}, \quad B = \{\omega : Y(\omega) > 6\}.$$  

$$P(A) = P(\{\omega : X(\omega) = 1\}) = P(X = 1).$$  

$$P(B) = P(\{\omega : Y(\omega) > 6\}) = P(Y > 6).$$  

$$P(B|A) = P(Y > 6|X = 1), \quad P(A|B) = P(X = 1|Y > 6).$$
**Equally likely scenario**

A real population of people, under purely random sampling

An imagined population of equally likely possibilities

\[
P(A) = \frac{|A|}{|\Omega|}.
\]

Axiom 0.
Can always translate a problem into equally likely setting.
Equally likely scenario

\[ P(A|B) = \frac{|A \cap B|}{|B|} = \frac{|A \cap B|/|\Omega|}{|B|/|\Omega|} = \frac{P(A \cap B)}{P(B)}. \]

As if \( B \) is the sample space.
Axiom 4
Or definition of conditional probability.
(1) \( X \) is uniform random number in \([0, 1]\).

(2) \((X, Y)\) are two independent random numbers in \([0, 1]\).

(3) \((X, Y, Z)\) are three independent random numbers in \([0, 1]\).

\(\Omega = [0, 1] \) or \([0, 1]^2 \) or \([0, 1]^3 = \) set of points.

Population of points (uncountably infinitely many).
Random point in a region

Example 3: throwing point into region

$X$ and $Y$ are independent uniform random numbers in $[0, 1]$. 
$(X, Y)$ is a random point in $\Omega = [0, 1]^2$.

$A = \{(x, y) : x^2 + y^2 \leq 1\}$.

$$P(A) = \frac{|A|}{|\Omega|} = \frac{\pi}{4}.$$ 

$|A|$ is the size of $A$, e.g., area (length, volume).
Example 3: throwing point into region

\[ X \text{ and } Y \text{ are independent uniform random numbers in } [0, 1]. \]
\[ (X, Y) \text{ is a random point in } \Omega = [0, 1]^2. \]
\[ A = \{(x, y) : x^2 + y^2 \leq 1\}. \]

\[ P(X^2 + Y^2 \leq 1) = \frac{\pi}{4}. \]

\[ P(X^2 + Y^2 = 1) = 0. \]

Capital letters for random variables.
Discretization $\rightarrow$ finite population of small squares.
Area = number of small squares $\times$ area of each small square.

**Inner measure**: fill inside by small squares $\rightarrow$ upper limit.

**Outer measure**: cover outside by small squares $\rightarrow$ lower limit.

**Measurable**: inner measure $=$ outer measure.

The collection of all measurable sets, $\sigma$-algebra.

**Integral**: area under curve.
Axioms

**Probability as measure**, i.e., count, length, area, volume ...

Axiom 0: $P(A) = \frac{|A|}{|\Omega|}$ in equally likely scenario.

**Axiom 1**: $P(\Omega) = 1$.

**Axiom 2**: $P(A) \geq 0$.

**Axiom 3**: Additivity: If $A \cap B = \emptyset$ (empty), then

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

Axiom 4: $P(A|B) = \frac{P(A \cap B)}{P(B)}$, assuming $P(B) > 0$. 
Long run frequency

Throw $n$ points into $\Omega$. $m$ of them fall into $A$.

$$P(A) = \frac{|A|}{|\Omega|} \approx \frac{m}{n}.$$ 

As $n \to \infty$, $\frac{m}{n} \to P(A)$ in probability. $P(A)$ can be interpreted as **long run frequency**.
Long run frequency

Throw \( n \) points into \( \Omega \). \( m \) of them fall into \( A \). Among all equally likely possibilities, 99.999% are like below, where \( m/n \) is close to \( P(A) \).

0.00000001% are like below, where \( m/n \) are far from \( P(A) \).

Can prove \( P(\left| \frac{m}{n} - P(A) \right| > \epsilon) \rightarrow 0 \) for any fixed \( \epsilon > 0 \).
Long run frequency

**Example 3: \( \pi \)**

Throw \( n \) points into \( \Omega \). \( m \) of them fall into \( A \).

\[
P(A) = \frac{|A|}{|\Omega|} = \frac{\pi}{4} \approx \frac{m}{n}.
\]

**Monte Carlo method:**

\[
\hat{\pi} = \frac{4m}{n}.
\]

As \( n \to \infty \), \( \frac{m}{n} \to P(A) \) in probability.

\( P(A) \) can be interpreted as **long run frequency**.
Monte Carlo

Deterministic method

Go over all the $n = 100 = 10^2$ square cells, count inner or outer measure, i.e., how many $(m)$ fall into $A$.
3-dimensional? $n = 10^3$ cubic cells.
4-dimensional? $n = 10^4$ cells.
10000-dimensional? $n = 10^{10000}$ cells.

**Monte Carlo**: sample $n = 1000$ points in the hyper-cube. Count how many $(m)$ fall into $A$. 
Example 3: $\pi$, buffon needle

Lazzarini threw $n = 3408$ times.

$$P(A) \approx \frac{m}{n}.$$ 

Monte Carlo method:

$$\hat{\pi} = \frac{355}{113}$$

Too accurate. $m$ is random.

For fixed $n$, $m$ is random. $m/n$ fluctuates around $P(A)$.

As $n \to \infty$, $\frac{m}{n} \to P(A)$ in probability, law of large number. $P(A)$ can be interpreted as long run frequency, how often $A$ happens in the long run.
Example 3: throwing point into region

$X$ and $Y$ are independent uniform random numbers in $[0, 1]$. 
$(X, Y)$ is a random point in $\Omega = [0, 1]^2$. 
$A = \{(x, y) : x < 1/2\}$.

$$P(A) = P(X < 1/2) = \frac{|A|}{|\Omega|} = \frac{1}{2}.$$
Example 3: throwing point into region

\(X\) and \(Y\) are independent uniform random numbers in \([0, 1]\). 
\((X, Y)\) is a random point in \(\Omega = [0, 1]^2\). 
\(B = \{(x, y) : x + y < 1\}\). 

\[ P(B) = P(X + Y < 1) = \frac{|B|}{|\Omega|} = 1/2. \]
Example 3: throwing point into region

\[
P(A|B) = \frac{|A \cap B|}{|B|} = \frac{1/2 - 1/8}{1/2} = \frac{3}{4}.
\]

\[
P(X < 1/2|X + Y < 1).
\]

(1) As if randomly throw a point into \(B\), as if \(B\) is the sample space. Then what is the probability the point falls into \(A\)?
Example 3: throwing point into region

\[ P(A|B) = \frac{|A \cap B|}{|B|} = \frac{1/2 - 1/8}{1/2} = \frac{3}{4}. \]

\[ P(X < 1/2|X + Y < 1). \]

(2) Consider throwing a lot of points into \( \Omega \).
How often \( A \) happens? How often \( B \) happens?
When \( B \) happens, how often \( A \) happens?
Among all the points in \( B \), what is the fraction belongs to \( A \)?
Experiment $\rightarrow$ outcome $\rightarrow$ number

Example 4: Coin flipping

(4.1) Flip a coin $\rightarrow$ head or tail $\rightarrow$ 1 or 0
(4.2) Flip a coin twice $\rightarrow$ (head, head), or (head, tail), or (tail, head) or (tail, tail) $\rightarrow$ 11 or 10 or 01 or 00

List:
- HH
- HT
- TH
- TT

Table:

<table>
<thead>
<tr>
<th></th>
<th>H</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
<td>HH</td>
<td>HT</td>
</tr>
<tr>
<td>T</td>
<td>TH</td>
<td>TT</td>
</tr>
</tbody>
</table>

Tree Diagram:

The sample space is \{HH, HT, TH, TT\}
Experiment $\rightarrow$ outcome $\rightarrow$ number

Example 4: Coin flipping

(4.3) Flip a coin $n$ times $\rightarrow 2^n$ binary sequences.

Sample space $\Omega$: all $2^n$ sequences.
Each $\omega \in \Omega$ is a sequence.

**Visualize:** randomly pick a sequence from $2^n$ sequences.

$Z_i(\omega) = 1$ if $i$-th flip is head; $Z_i(\omega) = 0$ if $i$-th flip is tail.
Example 4: Coin flipping

$Z_i(\omega) = 1$ if $i$-th flip is head; $Z_i(\omega) = 0$ if $i$-th flip is tail.

HHHH, THHH, HTHT, TTHT,
HHHT, HHTT, THHT, THTT,
HHTH, TTHH, HTTH, HTTT,
HTHH, THTH, TTHH, TTTT

Flip a fair coin 4 times independently, let $A$ be the event that there are 2 heads.

Visualize: randomly pick a sequence from 16 sequences.

$$P(A) = \frac{|A|}{|\Omega|} = \frac{6}{2^4} = \frac{3}{8}.$$
Example 4: Coin flipping

$Z_i(\omega) = 1$ if $i$-th flip is head; $Z_i(\omega) = 0$ if $i$-th flip is tail.

Let $X(\omega)$ be the number of heads in the sequence $\omega$.

$$X(\omega) = Z_1(\omega) + Z_2(\omega) + Z_3(\omega) + Z_4(\omega).$$

$$P(A_k) = P(\{\omega : X(\omega) = k\}) = P(X = k) = p_k.$$  

$$(p_k, k = 0, 1, 2, 3, 4) = (1, 4, 6, 4, 1)/16.$$
**Example 4: Coin flipping**

| $A_2| = 6.
| $A_2| = \binom{4}{2} = \frac{4 \times 3}{2}.

4 positions, choose 2 of them to be heads, and the rest are tails.

- HHHH, THHH, HTHT, TTHT,
- HHHT, HHTT, THHT, THTT,
- HTHH, TTHH, HTTH, HTTT,
- HTHH, THTH, TTTH, TTTT
Ordered pair: roll a die twice

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(1,1)</td>
<td>(1,2)</td>
<td>(1,3)</td>
<td>(1,4)</td>
<td>(1,5)</td>
<td>(1,6)</td>
</tr>
<tr>
<td>2</td>
<td>(2,1)</td>
<td>(2,2)</td>
<td>(2,3)</td>
<td>(2,4)</td>
<td>(2,5)</td>
<td>(2,6)</td>
</tr>
<tr>
<td>3</td>
<td>(3,1)</td>
<td>(3,2)</td>
<td>(3,3)</td>
<td>(3,4)</td>
<td>(3,5)</td>
<td>(3,6)</td>
</tr>
<tr>
<td>4</td>
<td>(4,1)</td>
<td>(4,2)</td>
<td>(4,3)</td>
<td>(4,4)</td>
<td>(4,5)</td>
<td>(4,6)</td>
</tr>
<tr>
<td>5</td>
<td>(5,1)</td>
<td>(5,2)</td>
<td>(5,3)</td>
<td>(5,4)</td>
<td>(5,5)</td>
<td>(5,6)</td>
</tr>
<tr>
<td>6</td>
<td>(6,1)</td>
<td>(6,2)</td>
<td>(6,3)</td>
<td>(6,4)</td>
<td>(6,5)</td>
<td>(6,6)</td>
</tr>
</tbody>
</table>

Experiment 1 has $n_1$ outcomes. For each outcome of experiment 1, experiment 2 has $n_2$ outcomes. The number of all possible pairs is $n_1 \times n_2$. 
**Multiplication**

Ordered pair: flip a coin and roll a die

![Diagram showing multiplication of coin and dice outcomes]

- **Coin:** Head, Tail
- **Dice:** 1, 2, 3, 4, 5, 6
- **Outcomes:**
  - (H, 1)
  - (H, 2)
  - (H, 3)
  - (H, 4)
  - (H, 5)
  - (H, 6)
  - (T, 1)
  - (T, 2)
  - (T, 3)
  - (T, 4)
  - (T, 5)
  - (T, 6)
Multiplication

Ordered triplet
$n$ different cards. Choose $k$ of them. Order matters. Number of different sequences:

$$P_{n,k} = n(n - 1)\ldots(n - k + 1). \quad P_{4,2} = 4 \times 3 = 12.$$  

$$P_{n,n} = n!.$$  

How many different ways to permute things.
Combination

\( n \) different balls. Choose \( k \) of them. Order does NOT matters.

Number of different combinations:

\[
\binom{n}{k} = \frac{P_{n,k}}{k!} = \frac{n(n-1)(n-k+1)}{k!} = \frac{n!}{k!(n-k)!}.
\]

\[
\binom{4}{2} = \frac{4 \times 3}{2} = 6.
\]
Each combination corresponds to $k!$ permutations.

\[
\binom{n}{k} = \frac{P_{n,k}}{k!} = \frac{n(n-1)...(n-k+1)}{k!} = \frac{n!}{k!(n-k)!}.
\]

\[
\binom{4}{2} = \frac{4 \times 3}{2} = 6.
\]
Example 4: Coin flipping

In general, flip a fair coin $n$ times independently,

$$ P(A_k) = P(\{\omega : X(\omega) = k\}) = P(X = k) = \frac{\binom{n}{k}}{2^n}. $$
Random walk

Either go forward or backward by flipping a fair coin. Walk $n$ steps.

Number of heads $X = k$, then random walk ends up at $Y = m = k - (n - k) = 2k - n$, $k = (m + n)/2$.

$$P(Y = m) = P(X = k) = \binom{n}{k} \frac{1}{2^n} = \frac{\binom{n}{(m+n)/2}}{2^n}.$$
Either go forward or backward by flipping a fair coin.

Number of heads $X = k$, then random walk ends up at $Y = m = k - (n - k) = 2k - n$, $k = (m + n)/2$.

$$P(Y = m) = P(X = k) = \binom{n}{k} \binom{2n}{2k} = \binom{n}{m+n}/2.$$
Random walk

Example 4: Coin flipping

Pascal triangle

H H H H 4 heads
H H H T 3 heads
H T H H 3 heads
H H T H 3 heads
T H H H 3 heads
H H T T 2 heads
H T H T 2 heads
H T T H 2 heads
H T T T 2 heads
T H H T 2 heads
T H T H 2 heads
T H T T 2 heads
T T H H 2 heads
T T H T 1 heads
T T T H 1 heads
T T T T 1 heads
T T T T 0 heads
Example 4: Coin flipping

All $2^n$ paths are equally likely.

Number of paths that end up in $k$-th bin $= \binom{n}{k}$.

$X$: destination. $P(X = k) = \binom{n}{k} / 2^n$.

How often the balls fall into $k$-th bin.
Transition probability

Either go forward or backward

\[ X_t = Z_1 + Z_2 + \ldots + Z_t. \]

\( Z_k = 1 \) or \(-1\) with probability \(1/2\) each.

\[ X_{t+1} = X_t + Z_{t+1}. \]

\[ P(X_{t+1} = x + 1 | X_t = x) = P(X_{t+1} = x - 1 | X_t = x) = 1/2. \]
Example 5: Random walk over three states

With probability $1/2$, stay. With probability $1/4$, go to either states.

$$K_{ij} = P(X_{t+1} = j|X_t = i).$$

**Markov** property: past history before $X_t$ does not matter.
Example 5: Random walk over three states

With probability 1/2, stay. With probability 1/4, go to either of the other two states.

\[ K_{ij} = P(X_{t+1} = j \mid X_t = i). \]

Imagine 1 million people migrating. At each step, for each state, half of the people stay, 1/4 go to each of the other two states.
Example 5: Random walk over three states

With probability 1/2, stay. With probability 1/4, go to either of the other two states.

\[ K_{ij} = P(X_{t+1} = j | X_t = i). \]

\[
K = \begin{bmatrix}
0.5 & 0.25 & 0.25 \\
0.25 & 0.5 & 0.25 \\
0.25 & 0.25 & 0.375
\end{bmatrix}
\]
Marginal probability

Example 5: Random walk over three states

With probability $1/2$, stay. With probability $1/4$, go to either of the other two states.

$$K_{ij} = P(X_{t+1} = j | X_t = i).$$

$$p_i^{(t)} = P(X_t = i).$$

Imagine 1 million people migrating. $p_i^{(t)}$ is the number of people (in million) in state $i$ at time $t$.

$$p^{(t)} = (p_1^{(t)}, p_2^{(t)}, p_3^{(t)}).$$
Example 5: Random walk over three states

Imagine 1 million people migrating. $p_i^{(t)}$ is the number of people (in million) in state $i$ at time $t$.

$$p_i^{(t)} = P(X_t = i).$$

$$p^{(t)} = (p_1^{(t)}, p_2^{(t)}, p_3^{(t)}).$$
Example 5: Random walk over three states

\[ K_{ij} = P(X_{t+1} = j \mid X_t = i) \].

\[ p_i^{(t)} = P(X_t = i) \].

\[ p_j^{(t+1)} = \sum_i p_i^{(t)} K_{ij} \].

Number of people in state \( j \) at time \( t + 1 \) = sum number of people in state \( i \) at time \( t \) \times fraction of those in \( i \) who go to \( j \).
Example 5: Random walk over three states

\[ p^{(0)} = (1, 0, 0) \]

\[ p^{(1)} = \left( \frac{1}{2}, \frac{1}{4}, \frac{1}{4} \right) \]

\[ p^{(2)} = \left( \frac{3}{8}, \frac{5}{16}, \frac{5}{16} \right) \]

\[ \pi = \left( \frac{1}{3}, \frac{1}{3}, \frac{1}{3} \right) \]

Stationary distribution, arrow of time.
Example 5: Random walk over three states

\[ p_{j}^{(t+1)} = \sum_{i} p_{i}^{(t)} K_{ij}. \]

\[ p^{(t+1)} = p^{(t)} K. \]

\[ p^{(t)} = p^{(0)} K^{t} \rightarrow \pi. \]
Example 5: Random walk

\[ p_j^{(t+1)} = \sum_i p_i^{(t)} K_{ij}. \]

\[ p_i^{(t)} \rightarrow \pi_i. \]

\[ \pi_j = \sum_i \pi_i K_{ij}. \]

\(\pi_i\): proportion of people who are in page \(i\).

Popularity of \(i\) depends on the popularities of pages linked to \(i\).
Chain rule

\[ P(A|B) = \frac{P(A \cap B)}{P(B)} . \]

(1) As if we randomly throw a point into \( B \).
(2) When \( B \) happens, how often \( A \) happens.

Chain rule:

\[ P(A \cap B) = P(B)P(A|B). \]

\( B \) happens 1/2 times. When \( B \) happens, \( A \) happens 3/4 times. How often \( A \) and \( B \) happen together?
Chain rule and rule of total probability

Chain rule:

\[ P(A \cap B) = P(B)P(A|B). \]

\[ P(X_{t+1} = j \cap X_t = i) = P(X_t = i)P(X_{t+1} = j|X_t = i) \]
\[ = p_i^{(t)} K_{ij}. \]

Rule of total probability:

\[ P(X_{t+1} = j) = \sum_i P(X_{t+1} = j \cap X_t = i). \]

\[ p_j^{(t+1)} = \sum_i p_i^{(t)} K_{ij}. \]

Add up probabilities of alternative chains of events.
Independence

Conditional:

\[ P(A|B) = \frac{P(A \cap B)}{P(B)}. \]

Chain rule:

\[ P(A \cap B) = P(B)P(A|B). \]

**Independence**

\[ P(A|B) = P(A). \]

\[ P(A \cap B) = P(A)P(B). \]

*A and B have nothing to do with each other.*
Independence

Definition 1:
\[ P(A|B) = P(A). \]

Definition 2:
\[ P(A \cap B) = P(A)P(B). \]
Conditional independence

**Markov chain:** $C \rightarrow B \rightarrow A,$


$$P(X_{t+1} = j|X_t = i, X_{t-1}, \ldots, X_0) = P(X_{t+1} = j|X_t = i).$$

Future is independent of the past given present.
Immediate cause (parent), remote cause (grandparent).

**Shared cause:** $C \leftarrow B \rightarrow A,$

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

Children given parent.
Example 6: Rare disease example
1% of population has a rare disease.
A random person goes through a test.
If the person has disease, 90% chance test positive.
If the person does not have disease, 90% chance test negative.
If tested positive, what is the chance he or she has disease?
\[ P(D) = 1\% . \]
\[ P( + | D ) = 90\% , \ P( - | N ) = 90\% . \]
\[ P(D|+) = ? \]
Example 6: Rare disease example

\[ P(D) = 1\% . \]
\[ P(+|D) = 90\%, \quad P(-|N) = 90\%. \]
\[ P(D|+) = ? \]

\[ P(D|+) = \frac{9}{9 + 99} = \frac{1}{12}. \]

\[ P(\text{alarm} | \text{fire}) \ vs \ P(\text{fire} | \text{alarm}). \]
Example 6: Rare disease example

\[ P(D) = 1\% . \]
\[ P(\text{+}|D) = 90\% , \ P(\text{−}|\neg N) = 90\% . \]

\[ P(D \cap \text{+}) = P(D)P(\text{+}|D) = 1\% \times 90\% . \]
\[ P(N \cap \text{+}) = P(N)P(\text{+}|N) = 99\% \times 10\% . \]
\[ P(\text{+}) = P(D \cap \text{+}) + P(N \cap \text{+}) = 1\% \times 90\% + 99\% \times 10\% . \]
\[ P(D|\text{+}) = \frac{P(D \cap \text{+})}{P(\text{+})} = \frac{9}{9+99} = \frac{1}{12} . \]
Chain rule, rule of total probability, Bayes rule

**General formula**

$m$ causes: $C_1, \ldots, C_i, \ldots, C_m$.

$n$ effects: $E_1, \ldots, E_j, \ldots, E_n$.

**Given:**

Prior: $P(C_i), i = 1, \ldots, m$.

Conditional: $P(E_j|C_i), j = 1, \ldots, n$.
Chain rule, rule of total probability, Bayes rule

Prior: \( P(C_i), i = 1, \ldots, m. \)
Conditional: \( P(E_j|C_i), j = 1, \ldots, n. \)

Posterior: \( P(C_i|E). \)

\[
P(C_i \cap E) = P(C_i)P(E|C_i).
\]

\[
P(E) = \sum_i P(C_i \cap E) = \sum_i P(C_i)P(E|C_i).
\]

\[
P(C_i|E) = \frac{P(C_i \cap E)}{P(E)} = \frac{P(C_i)P(E|C_i)}{\sum_{i'} P(C_{i'})P(E|C_{i'})}.
\]
Chain rule, rule of total probability, Bayes rule

\[ X = \text{cause} \in \{1, \ldots, i, \ldots, m\}. \]
\[ Y = \text{effect} \in \{1, \ldots, j, \ldots, n\}. \]

\[
P(X = i | Y = j) = \frac{P(X = i \cap Y = j)}{P(Y = j)}
\]
\[
= \frac{P(X = i)}{\sum_{i'} P(X = i') P(Y = j | X = i')}. 
\]
Chain rule, rule of total probability, Bayes rule

\[
P(X = i \mid Y = j) = \frac{P(X = i \cap Y = j)}{P(Y = j)}
\]

\[
= \frac{P(X = i)P(Y = j \mid X = i)}{\sum_{i'} P(X = i')P(Y = j \mid X = i')}
\]

\[
p(x \mid y) = \frac{p(x, y)}{p(y)} = \frac{p(x)p(y \mid x)}{\sum_{x'} p(x')p(y \mid x')}
\]
Reasoning

Conditional probability is regular probability, (1) cause → effect, physical, $P(E|C)$ is given, $C$ defines experiment.

(2) effect → cause, mental, $P(C|E) = P(C \cap E)/P(E)$, as if $E$ is the sample space.

**Bayes network**, directed acyclic graph, graphic model

**Conditional independence:**
(1) Sibling nodes are independent given parent node.
(2) Child node is independent of grandparents given parent.
Take home message

As long as you can count
Count the number of people (equally likely possibilities)
Count the number of points (or repetitions)
Count the number of sequences (of coin flipping)

Two things
(1) Intuition, visualization and motivation
(2) Precise notation and formula

Accomplished
Most of the important concepts via intuitive examples

Next
Systematic and more in-depth treatments