STATS 100A: BASICS & EXAMPLES

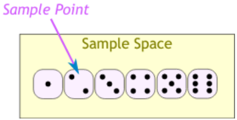
Ying Nian Wu

Department of Statistics
University of California, Los Angeles

Some pictures are taken from the internet.
Credits belong to original authors.

Sample space

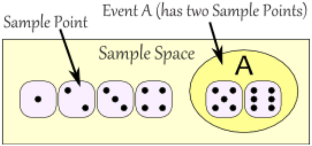
Experiment $\rightarrow$ outcome $\rightarrow$ number
Example 1: Roll a die



Sample space Ω: The set of all the outcomes (or sample points, elements).
Randomly sample an outcome from the sample space.

Event

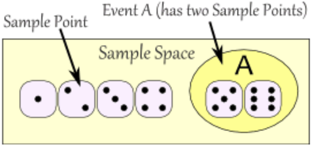
Experiment $\rightarrow$ **outcome** $\rightarrow$ **number**
Example 1: Roll a die



Sample space Ω: 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., $\{5, 6\}$.

Counting equally likely possibilities

Experiment $\rightarrow$ **outcome** $\rightarrow$ **number**
Example 1: Roll a die

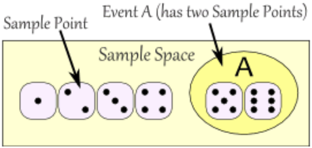


Assume the die is fair so that all the outcomes are **equally likely**.
Probability: defined on event:

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

$|A|$ counts the size of A, i.e., the number of elements in A.

Experiment $\rightarrow$ **outcome** $\rightarrow$ **number**
Example 1: Roll a die

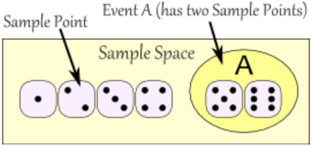


Random variable: Let X be the number:

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

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.

Experiment $\rightarrow$ outcome $\rightarrow$ number
Example 1: Roll a die



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.

Conditional probability

100A

Ying Nian Wu

Basics

Population

Region

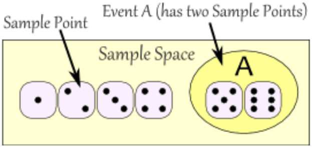
Coin

Markov

Reasoning

Experiment → outcome → number

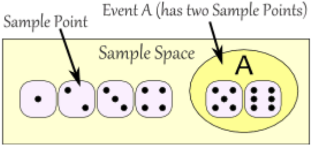
Example 1: Roll a die



Random variable

$$P(X = 6|X > 4) = \frac{1}{2}.$$

Example 1: Roll a die



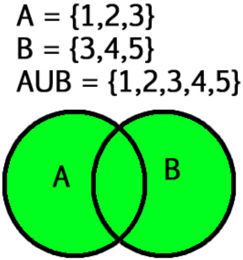
Complement

Statement: Not A

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

Relations

Example 1: Roll a die



Venn diagram

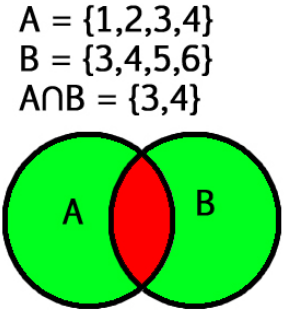
Union

Statement: A or B.

Subset: $A \cup B$.

Relations

Example 1: Roll a die



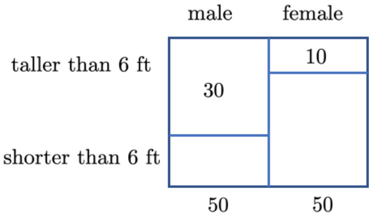
Intersection
Statement: A and B.
Subset: $A \cap B$.

Sample space is population

Experiment → outcome → 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.

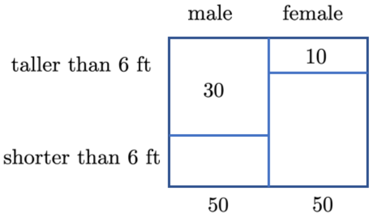
The sample space Ω is the population.



Events as sub-populations

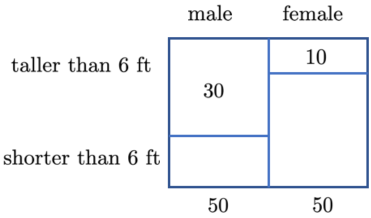
Experiment $\to$ outcome $\to$ number

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.



Probability is population proportion

Experiment → outcome → number
Example 2: A male, B tall.



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

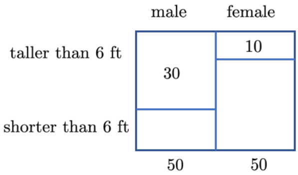
$$P(B) = \frac{|B|}{|\Omega|} = \frac{30+10}{100} = 40\%.$$

Probability = population proportion.

Conditional probability is proportion of sub-population

Experiment $\rightarrow$ outcome $\rightarrow$ number

Example 2: A male, B tall.



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

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

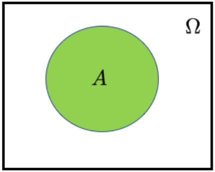
$$A = \{\omega : X(\omega) = 1\},\ B = \{\omega : Y(\omega) > 6\}.$$

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

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

$$P(B|A) = P(Y > 6|X = 1),\ P(A|B) = P(X = 1|Y > 6).$$

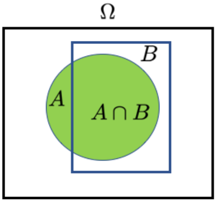
Equally likely scenario
A real population of people, under purely random sampling
or imagined population of equally likely possibilities



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

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

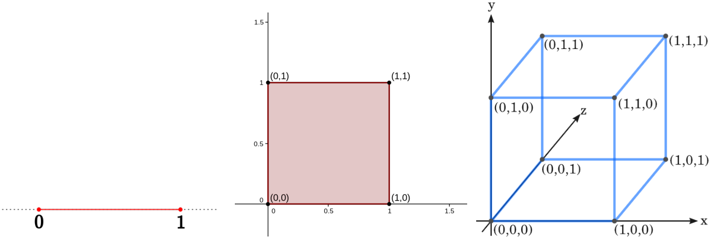
Equally likely scenario



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

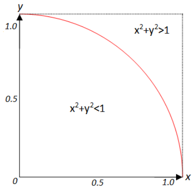
Physical: sample from B. B defines condition.
Mental: know that B happened, as if sample from B.
Axiom 4 or definition of conditional probability.

Sample space is region



(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.
Region = population of points (uncountably infinite).

Random point in a region
Example 3: throwing point into region

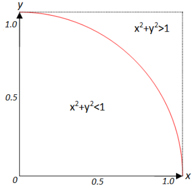


X and Y are independent uniform random numbers in [0, 1].
(X, Y) is a random point in $\Omega = [0, 1]^2$.
$A = \{(x, y) : x^2 + y^2 \leq 1\}$.

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

$|A|$ is the size of A, e.g., area (length, volume).

Example 3: throwing point into region



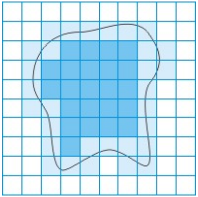
X and Y are independent uniform random numbers in [0, 1].
(X, Y) is a random point in $\Omega = [0, 1]^2$.
$A = \{(x, y) : x^2 + y^2 \leq 1\}$.

$$P(X^2 + Y^2 \leq 1) = \pi/4.$$

$$P(X^2 + Y^2 = 1) = 0.$$

Capital letters for random variables.

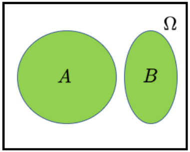
Measuring by counting



Discretization $\rightarrow$ finite population of tiny squares.
Area $=$ number of tiny squares $\times$ area of each tiny square.
Inner measure: fill inside by tiny squares $\rightarrow$ upper limit.
Outer measure: cover outside by tiny squares $\rightarrow$ lower limit.
Measurable: inner measure $=$ outer measure.
The collection of all measurable sets, σ-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: $P(A) \geq 0$.

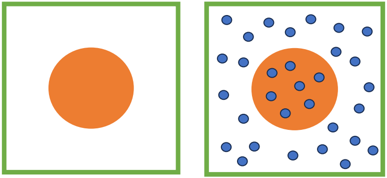


Axiom 3: Additivity: If $A \cap B = \phi$ (empty), then

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

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

Counting repetitions



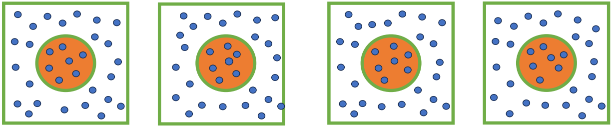
Throw n points into Ω. m of them fall into A.

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

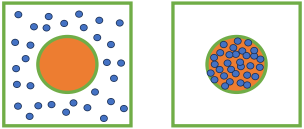
As $n \to \infty$, $\frac{m}{n} \to P(A)$ in probability.
$P(A)$ can be interpreted as **long run frequency**.

Fluctuations

Repeat random sampling n times independently.
Throw n points into Ω. m of them fall into A.
Among all equally likely possibilities, 99.999% are like below, where m/n is close to $P(A)$.

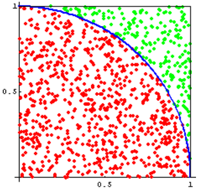


.00000001% are like below, where m/n are far from $P(A)$.



Can prove $P(|\frac{m}{n} - P(A)| > \epsilon) \to 0$ for any fixed $\epsilon > 0$.

Example 3: π



Throw n points into Ω. m of them fall into A.

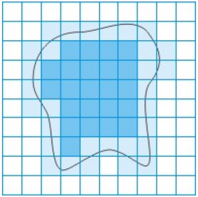
$$P(A) = \frac{|A|}{|\Omega|} = \frac{\pi}{4} \approx \frac{m}{n}.$$

Monte Carlo method:

$$\hat{\pi} = \frac{4m}{n}.$$

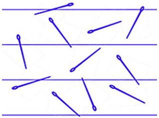
As $n \to \infty$, $\frac{m}{n} \to P(A)$ in probability.
$P(A)$ can be interpreted as **long run frequency**.

Deterministic method



Go over all the $n = 100 = 10^2$ tiny squares, count inner or outer measure, i.e., how many (m) fall into A.
3-dimensional? $n = 10^3$ tiny cubes.
4-dimensional? $n = 10^4$ tiny cells.
10000-dimensional? $n = 10^{10000}$ tiny cells.
Monte Carlo: sample $n = 1000$ points in the hyper-cube. Count how many (m) fall into A.

Example 3: π, buffon needle



Lazzarini threw $n = 3408$ times.

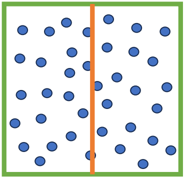
$$P(A) \approx \frac{m}{n}.$$

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.

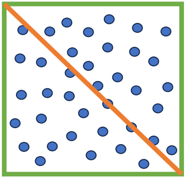
Example 3: throwing point into region



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

$$P(A) = P(X < 1/2) = \frac{|A|}{|\Omega|} = 1/2.$$

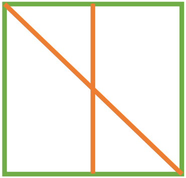
Example 3: throwing point into region



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

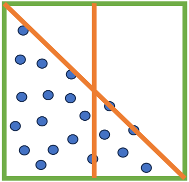


$$P(A|B) = \frac{|A \cap B|}{|B|} = \frac{1/2 - 1/8}{1/2} = 3/4.$$

$$P(X < 1/2 | X + Y < 1).$$

(1) 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



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

$$P(X < 1/2 | X + Y < 1).$$

(2) Consider throwing a lot of points into Ω.
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?

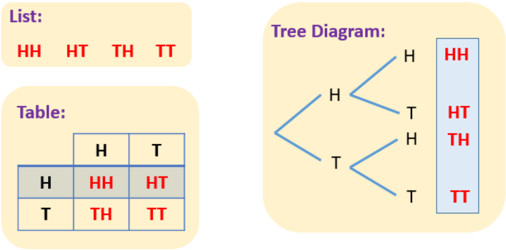
Coin flipping

Experiment $\to$ outcome $\to$ number

Example 4: Coin flipping

(4.1) Flip a coin $\to$ head or tail $\to$ 1 or 0

(4.2) Flip a coin twice $\to$ (head, head), or (head, tail), or (tail, head) or (tail, tail) $\to$ 11 or 10 or 01 or 00

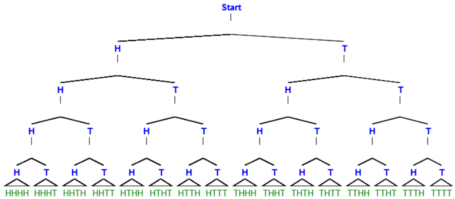


The sample space is {HH, HT, TH, TT}

Experiment $\rightarrow$ outcome $\rightarrow$ number

Example 4: Coin flipping

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



Sample space Ω: all 2^n sequences.

Each $\omega \in \Omega$ is a sequence.

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, TTTH, TTTT

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

Randomly pick a sequence from 16 sequences.

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

$$A = \{\omega : Z_1(\omega) + Z_2(\omega) + Z_3(\omega) + Z_4(\omega) = 2\}.$$

Number of heads

Example 4: Coin flipping

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

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
T	H	H	T	2 heads
T	H	T	H	2 heads
T	T	H	H	2 heads
H	T	T	T	1 heads
T	H	T	T	1 heads
T	T	H	T	1 heads
T	T	T	H	1 heads
T	T	T	T	0 heads

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

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

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

$$(p_k, k = 0, 1, 2, 3, 4) = (1, 4, 6, 4, 1)/16.$$

100A
Ying Nian Wu
Basics
Population
Region
Coin
Markov
Reasoning

Example 4: Coin flipping

$$\begin{array}{llll} \text{HHHH}, & \text{THHH}, & \text{HTHT}, & \text{TTHT}, \\ \text{HHHT}, & \text{HHTT}, & \text{THHT}, & \text{THTT}, \\ \text{HHTH}, & \text{TTHH}, & \text{HTTH}, & \text{HTTT}, \\ \text{HTHH}, & \text{THTH}, & \text{TTTH}, & \text{TTTT} \end{array}$$

$|A_2| = 6.$
$|A_2| = \binom{4}{2} = \frac{4\times 3}{2}.$
4 positions, choose 2 of them to be heads, and the rest are tails.

Multiplication: table

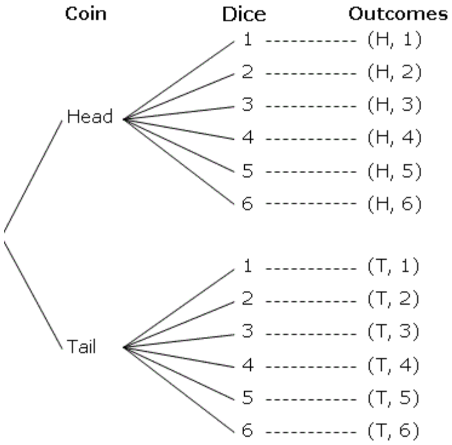
Ordered pair: roll a die twice

	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

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 $n_1 \times n_2$.

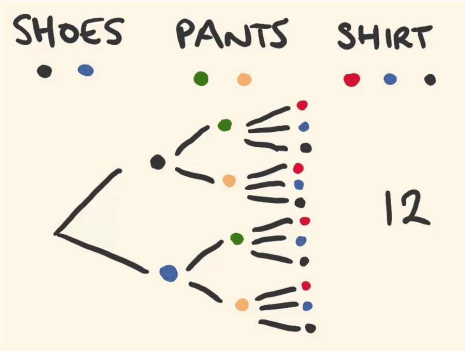
Multiplication

Ordered pair: flip a coin and roll a die



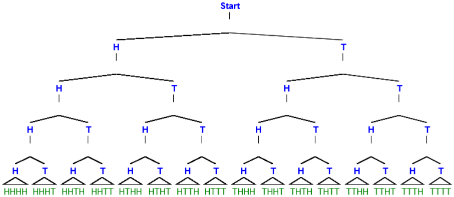
Multiplication

Ordered triplet



Sample space of sequences: coin

Flip a fair coin n times independently.
Sample space Ω_n: all possible sequences of heads and tails.



$|\Omega_n| = 2^n$.
$\Omega_n = \Omega_1 \times \Omega_1 \times ... \times \Omega_1 = \Omega_1^n$.
Ω_1: base sample space of flipping the fair coin once.
Ω_n: hyper sample space of flipping n times independently.
Population of sequences.

Sample space of sequences: die

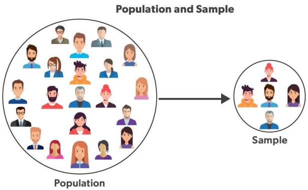
Roll a fair die n times independently.
Sample space Ω_n: all possible sequences of numbers.

	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

$|\Omega_n| = 6^n$.
$\Omega_n = \Omega_1 \times \Omega_1 \times ... \times \Omega_1 = \Omega_1^n$.
Ω_1: base sample space of rolling the fair die once.
Ω_n: hyper sample space of rolling n times independently.
Population of sequences.

Sample space of sequences: population

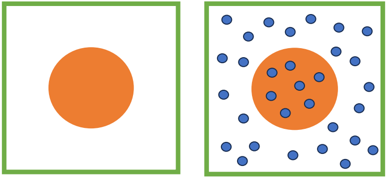
Randomly sample a person from a population of N (e.g., 300 million) people.
Repeat random sampling n (e.g., 1000) times independently.
Sample space Ω_n: all possible sequences of people.



$|\Omega_n| = N^n$ (e.g., $300m^{1000}$).
$\Omega_n = \Omega_1 \times \Omega_1 \times ... \times \Omega_1 = \Omega_1^n$.
Ω_1: base sample space, the population of people.
Ω_n: hyper sample space, the hyper-population of sequences.
Population of sequences.

Sample space of sequences: region

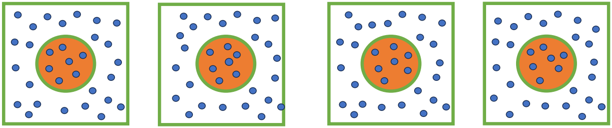
Randomly sample a point from a region.
Repeat the above n times independently.
Sample space Ω_n: all possible sequences of points.



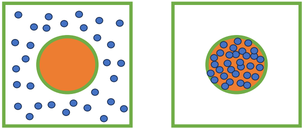
$\Omega_n = \Omega_1 \times \Omega_1 \times ... \times \Omega_1 = \Omega_1^n$.
Ω_1: base sample space, unit square $[0,1]^2$.
Ω_n: hyper sample space, unit hyper-cube $[0,1]^{2n}$.
$(x_1, y_1, x_2, y_2, ..., x_n, y_n)$: a point in Ω_n.
Population of sequences.

Population of sequences

Equally likely outcomes in Ω_1 + independent repetitions
= equally likely sequences in Ω_n.
m: number of times A happens.
m fluctuates over all sequences.
Among all equally likely possibilities, 99.999% are like below, where m/n is close to $P(A)$.

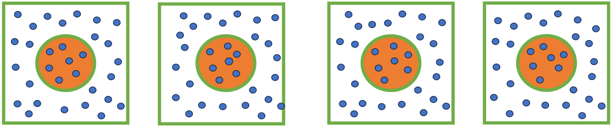


.00000001% are like below, where m/n are far from $P(A)$.



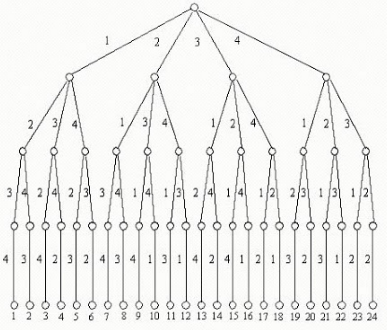
Convergence in probability, concentration of measure, law of large number

All sequences in Ω_n are equally likely.
Among all equally likely possibilities, 99.999% are like below, where m/n is close to $P(A)$.



Can prove $P(|\frac{m}{n} - P(A)| \leq .01) \to 1$ as $n \to \infty$.
A representative sequence: $|m(\text{sequence})/n - P(A)| \leq .01$.
A non-representative sequence: $|m/n - P(A)| > .01$.
Among all possible sequences, the proportion of representative sequences $\to 1$ as $n \to \infty$.
(1) Population setting: count the number of sequences.
(2) Region setting: measure the volume of set of sequences.

Permutation



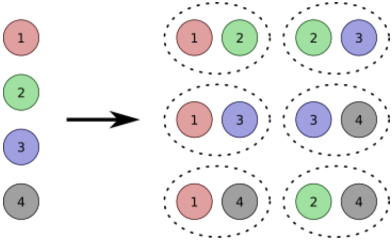
n different cards. Choose k of them. Order matters. Number of different sequences:

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

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

How many different ways to permute things.

Combination



n different balls. Choose k of them. Order does NOT matters. Number of different combinations:

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

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

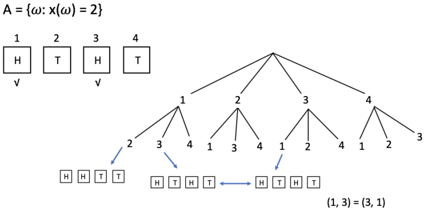
Combination

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



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

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

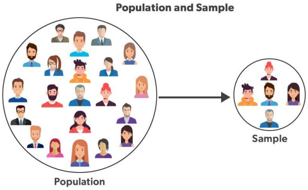
Do not confuse order of picking blanks with order of coin flippings.

In general, flip a fair coin n times independently,

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

Survey sampling

Population of N people, M males.
Repeat random sampling n times independently



$\to N^n$ equally likely sequences.
For a sequence ω, $X(\omega) =$ number of males in ω.
$A_m = \{\omega : X(\omega) = m\}$: sequences with m males.
$|A_m| = \binom{n}{m} M^m (N-M)^{n-m}$. n blanks. Choose m blanks for males, the rest $n-m$ blanks for females. Each male blank has M choices. Each female blank has $N-M$ choices.

Survey sampling

Population of N people. M males.
Sample a person, $p = M/N =$ Prob(male).

$$P(A_m) = P(X = m) = \frac{|A_m|}{|\Omega_n|}$$
$$= \frac{\binom{n}{m} M^m (N - M)^{n-m}}{N^n}$$
$$= \binom{n}{m} p^m (1 - p)^{n-m}.$$

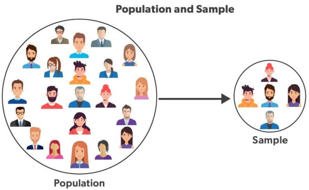
Most sequences are representative, $X/n \approx M/N = p$.

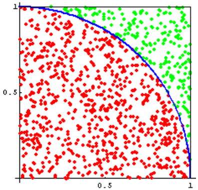
Binomial distribution, probability mass function

Flip a coin n times independently, p = probability of head.

$$p(x) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$
$$x = 0, 1, ..., n.$$

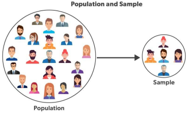
$p(x)$: probability mass function, probability distribution.





Survey sampling, poll before election, $p = M/N$.
Monte Carlo, $p = \pi/4$.

Survey sampling, poll before election, $p = M/N$.



Among all N^n sequences in the hyper-population of sequences Ω_n, let

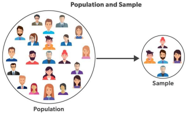
$$A = \left\{ \omega : \left| \frac{X(\omega)}{n} - p \right| \leq .01 \right\}.$$

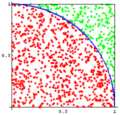
consist of representative sequences.

$$P(A) = \frac{|A|}{|\Omega_n|} = \sum_{x \in [n(p-.01), n(p+.01)]} p(x) \to 1,$$

($x \in [49, 51]$ ($n = 100$), [490, 510] ($n = 1000$), ...)
$X/n \to p$ in probability.

Definition of probability





Right: Population proportion, $P(A) = \frac{|A|}{|\Omega|}$, normalized measure, subjective belief or common sense of uncertainty.
Wrong: Long run frequency, $P(A) = \lim_{n\to} \frac{X}{n}$, under independent repetitions of the same experiments.
Limit does not always exist nor is the same for any sequence of repetitions. Independence not defined.
Right: Hyper-population of sequences of repetitions.
Uniform + Independence: all sequences are equally likely.
Proportion of representative sequences within hyper-population $\to 1$ as $n \to \infty$.

Special case: flip fair coin

$p = 1/2$, or $N = 2$.

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

$$p(x) = \frac{\binom{n}{x}}{2^n}.\ x = 0, 1, ..., n.$$

Among all 2^n sequences, let

$$A = \left\{\omega : \left|\frac{X(\omega)}{n} - \frac{1}{2}\right| \leq .01\right\}.$$

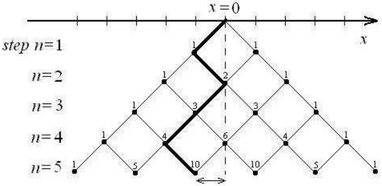
consist of representative sequences.

$$P(A) = \sum_{x \in [n(p-.01), n(p+.01)]} \frac{\binom{n}{x}}{2^n} \to 1,$$

($x \in [49, 51]$ ($n = 100$), [490, 510] ($n = 1000$), ...)
$X/n \to 1/2$ in probability.

Random walk based on coin flipping

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



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
T	H	H	T	2 heads
T	H	T	H	2 heads
T	T	H	H	2 heads
H	T	T	T	1 heads
T	H	T	T	1 heads
T	T	H	T	1 heads
T	T	T	H	1 heads
T	T	T	T	0 heads

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

$$p_Y(y) = P(Y = y) = P(X = x) = p_X(x) = \frac{\binom{n}{x}}{2^n} = \frac{\binom{n}{(y+n)/2}}{2^n}.$$

Random walk

Either go forward or backward by flipping a fair coin.

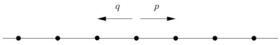


Figure 1: Simple random walk

The Symmetric Random Walk

$n \setminus x$	-5	-4	-3	-2	-1	0	1	2	3	4	5
0						1					
1					$\frac{1}{2}$	0	$\frac{1}{2}$				
2				$\frac{1}{4}$	0	$\frac{2}{4}$	0	$\frac{1}{4}$			
3			$\frac{1}{8}$	0	$\frac{3}{8}$	0	$\frac{3}{8}$	0	$\frac{1}{8}$		
4		$\frac{1}{16}$	0	$\frac{4}{16}$	0	$\frac{6}{16}$	0	$\frac{4}{16}$	0	$\frac{1}{16}$	
5	$\frac{1}{32}$	0	$\frac{5}{32}$	0	$\frac{10}{32}$	0	$\frac{10}{32}$	0	$\frac{5}{32}$	0	$\frac{1}{32}$

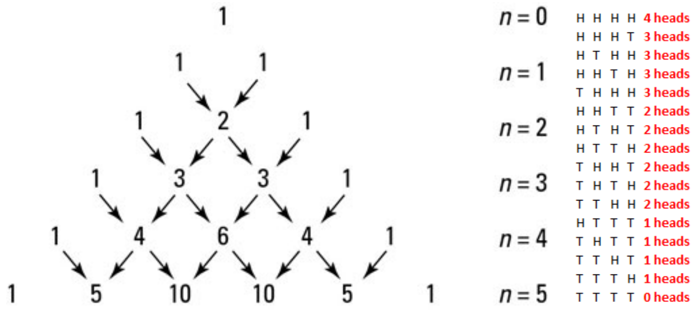
H H H H 4 heads
H H H T 3 heads
H T H H 3 heads
H H T H 3 heads
T H H H 3 heads
H H T T 2 heads
H T H T 2 heads
H T T H 2 heads
T H H T 2 heads
T H T H 2 heads
T T H H 2 heads
H T T T 1 heads
T H T T 1 heads
T T H T 1 heads
T T T H 1 heads
T T T T 0 heads

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

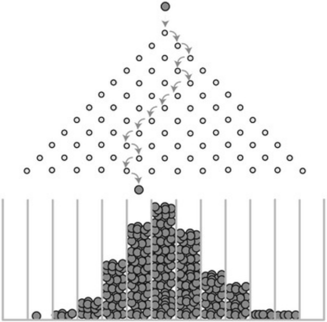
$$p_Y(y) = P(Y = y) = P(X = x) = p_X(x) = \frac{\binom{n}{x}}{2^n} = \frac{\binom{n}{(y+n)/2}}{2^n}.$$

Pascal triangle

Example 4: Coin flipping
Pascal triangle



Example 4: Coin flipping



All 2^n paths are equally likely (population of trajectories)
Number of paths that end up in x-th bin $= \binom{n}{x}$.
X: destination. $p(x) = P(X = x) = \binom{n}{x}/2^n$.
Drop 1 million balls, how often the balls fall into x-th bin.

Transition probability

100A

Ying Nian Wu

Basics

Population

Region

Coin

Markov

Reasoning

Either go forward or backward

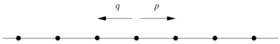


Figure 1: Simple random walk

The Symmetric Random Walk

$n \setminus x$	-5	-4	-3	-2	-1	0	1	2	3	4	5
0						1					
1					$\frac{1}{2}$	0	$\frac{1}{2}$				
2				$\frac{1}{4}$	0	$\frac{2}{4}$	0	$\frac{1}{4}$			
3			$\frac{1}{8}$	0	$\frac{3}{8}$	0	$\frac{3}{8}$	0	$\frac{1}{8}$		
4		$\frac{1}{16}$	0	$\frac{4}{16}$	0	$\frac{6}{16}$	0	$\frac{4}{16}$	0	$\frac{1}{16}$	
5	$\frac{1}{32}$	0	$\frac{5}{32}$	0	$\frac{10}{32}$	0	$\frac{10}{32}$	0	$\frac{5}{32}$	0	$\frac{1}{32}$

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
T H H T 2 heads
T H T H 2 heads
T T H H 2 heads
H T T T 1 heads
T H T T 1 heads
T T H T 1 heads
T T T H 1 heads
T T T T 0 heads

$$X_t = Z_1 + Z_2 + ... + Z_t.$$

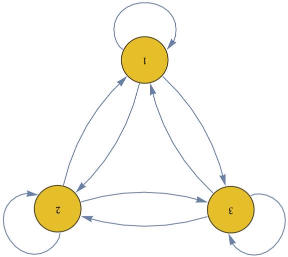
$Z_k = 1$ or -1 with probability $1/2$ each.

$$X_{t+1} = X_t + Z_{t+1}.$$

$$P(X_{t+1} = x + 1 | X_t = x) = P(X_{t+1} = x - 1 | X_t = x) = 1/2.$$

Markov chain

Example 5: Random walk over three states



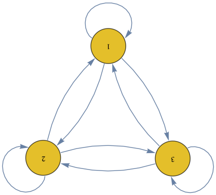
With probability 1/2, stay. With probability 1/4, go to either states.

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

Markov property: past history before X_t does not matter.

Population migration

Example 5: Random walk over three states

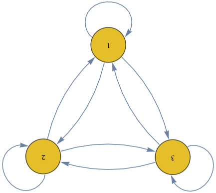


With probability $1/2$, stay. With probability $1/4$, go to either of the other two states.

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

Forward conditional probability, from cause to effect.
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. 1 million trajectories.

Example 5: Random walk over three states



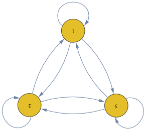
With probability 1/2, stay. With probability 1/4, go to either of the other two states.

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

$$\mathbf{K} = \begin{bmatrix} 1/2 & 1/4 & 1/4 \\ 1/4 & 1/2 & 1/4 \\ 1/4 & 1/4 & 1/2 \end{bmatrix}$$

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.

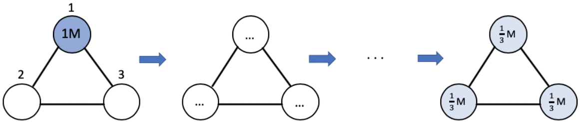
$$K_{ij} = P(X_{t+1} = j | X_t = i).$$

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

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

$$\mathbf{p}^{(t)} = (p_1^{(t)}, p_2^{(t)}, p_3^{(t)}).$$

Example 5: Random walk over three states



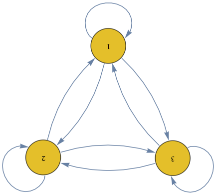
$$p_i^{(t)} = P(X_t = i).$$

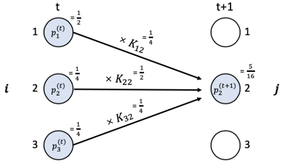
Imagine 1 million people migrating. $p_i^{(t)}$ is the number of people (in million) in state i at time t.

$$\mathbf{p}^{(t)} = (p_1^{(t)}, p_2^{(t)}, p_3^{(t)}).$$

Population migration

Example 5: Random walk over three states





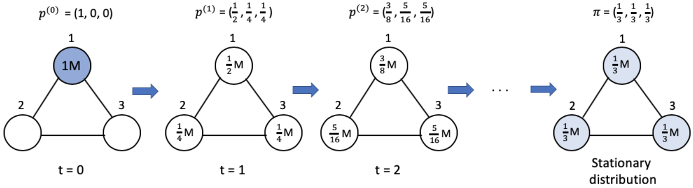
$$K_{ij} = P(X_{t+1} = j | X_t = i).$$

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

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

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



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

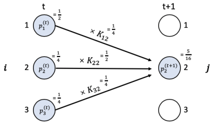
$$p_i^{(t)} \to \pi_i.$$

$$\pi_j = \sum_i \pi_i K_{ij}.$$

Stationary distribution, arrow of time.

Matrix multiplication

Example 5: Random walk over three states



$p^{(t+1)}$

1	2	3
	$\frac{5}{16}$	

=

$p^{(t)}$

1	2	3
$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$

K

$i \backslash j$	1	2	3
1		$\frac{1}{4}$	
2		$\frac{1}{2}$	
3		$\frac{1}{4}$	

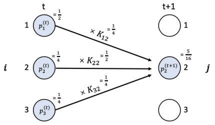
$$p_j^{(t+1)} = \sum_i p_i^{(t)} K_{ij}.$$

$$p^{(t+1)} = p^{(t)} K.$$

$$p^{(t)} = p^{(0)} K^t \to \pi.$$

Diagonalization and eigen-analysis

Example 5: Random walk over three states



$p^{(t+1)}$

1	2	3
	$\frac{5}{16}$	

=

$p^{(t)}$

1	2	3
$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$

K

$i \backslash j$	1	2	3
1		$\frac{1}{4}$	
2		$\frac{1}{2}$	
3		$\frac{1}{4}$	

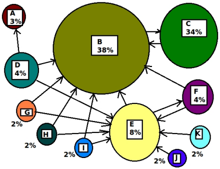
Diagonalization and eigen-analysis: $K = PDP^{-1}$, D diagonal, eigenvalues.

$$K^t = PDP^{-1}PDP^{-1}...PDP^{-1} = PD^tP^{-1}.$$

$$p^{(t)} = p^{(0)}K^t \to \pi.$$

Largest eigenvalue $= 1$, $1^t = 1$.
Second largest eigenvalue < 1, e.g., $.99^t \to 0$.

Example 5: Random walk



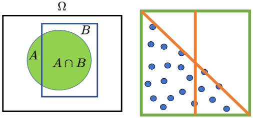
$$p_j^{(t+1)} = \sum_i p_i^{(t)} K_{ij}.$$

$$p_i^{(t)} \to \pi_i.$$

$$\pi_j = \sum_i \pi_i K_{ij}.$$

π_i: proportion of people who are in page i.
Popularity of i depends on the popularities of pages linked to i.

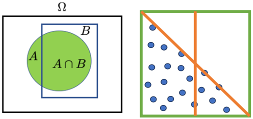
Conditional



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

(1) Counting population: Randomly sample from subpopulation B (e.g., males).
(2) Counting repetitions: When B happens, how often A happens.
Regular prob is conditional prob: $P(A) = P(A|\Omega)$.
Fixed condition (within the same subpopulation B), conditional prob behaves like regular prob.
e.g., $P(A^c) = 1 - P(A)$; $P(A^c|B) = 1 - P(A|B)$.

Chain rule

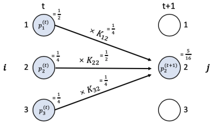


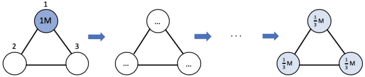
$$P(A \cap B) = P(B)P(A|B).$$

(1) Counting population: Population proportion of tall males = proportion of males × proportion of tall among males.
(2) Counting repetitions: B happens $1/2$ times. When B happens, A happens $3/4$ times. How often A and B happen together?
Generalize to chain of multiple events:

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

Chain rule and rule of total probability





Chain rule:

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

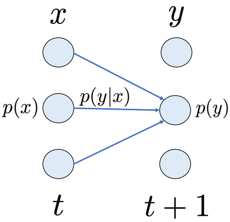
Rule of total probability:

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

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

Add up probabilities of alternative chains of events.

Marginal, conditional and joint distributions



Marginal: $p_t(x) = P(X_t = x)$, $p_{t+1}(y) = P(X_{t+1} = y)$.
Conditional: Forward $p(y|x) = P(X_{t+1} = y | X_t = x)$.
x: cause, y: effect. $p(y|x)$: cause $\rightarrow$ effect, given or learned.
Joint: $p(x, y) = P(X_t = x, X_{t+1} = y)$.
Chain rule: $p(x, y) = p_t(x)p(y|x)$.
Rule of total probability:
$p_{t+1}(y) = \sum_x p(x, y) = \sum_x p_t(x)p(y|x)$.
Add up probabilities of alternative chains of events.

Disease and symptom

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?
$P(D) = 1\%$.
Forward: $P(+|D) = 90\%$, $P(-|N) = 90\%$.
Backward: $P(D|+) = ?$

Cause and effect

Example 6: Rare disease example

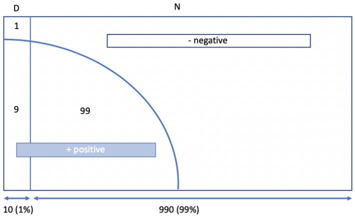
$P(D) = 1\%$.

Forward: from cause to effect.

$P(+|D) = 90\%$, $P(-|N) = 90\%$.

Backward: from effect to cause.

$P(D|+) = ?$



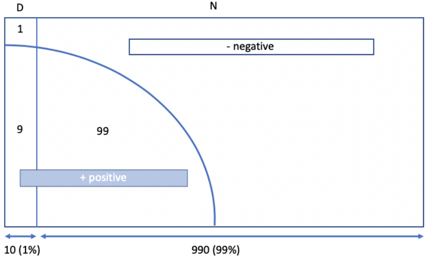
$P(D|+) = \frac{9}{9+99} = \frac{1}{12}$.

$P(\text{alarm} \mid \text{fire})$ vs $P(\text{fire} \mid \text{alarm})$.

Example 6: Rare disease example

$P(D) = 1\%$.

$P(+|D) = 90\%$, $P(-|N) = 90\%$.



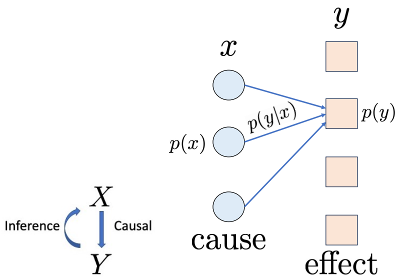
$P(D \cap +) = P(D)P(+|D) = 1\% \times 90\%$.

$P(N \cap +) = P(N)P(+|N) = 99\% \times 10\%$.

$P(+) = P(D \cap +) + P(N \cap +) = 1\% \times 90\% + 99\% \times 10\%$.

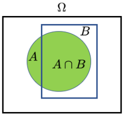
$P(D|+) = \frac{P(D \cap +)}{P(+)} = \frac{9}{9+99} = \frac{1}{12}$.

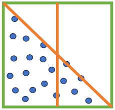
Random variables, probability mass functions

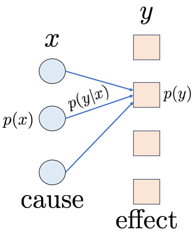


Marginal: prior $p(x) = P(X = x)$, marginal $p(y) = P(Y = y)$.
Conditional: forward generation $p(y|x) = P(Y = y|X = x)$
backward inference $p(x|y) = P(X = x|Y = y)$.
Chain rule: joint $p(x, y) = p(x)p(y|x)$.
Rule of total probability: marginal
$p(y) = \sum_x p(x, y) = \sum_x p(x)p(y|x)$.

Bayes rule







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

Bayes rule: backward inference, back tracing, posterior

$$p(x|y) = P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}$$

$$= \frac{p(x, y)}{p(y)} = \frac{p(x)p(y|x)}{\sum_{x'} p(x')p(y|x')}.$$

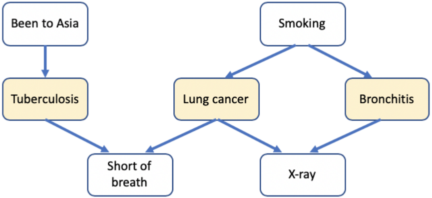
Cause, effect and conditioning

Conditional:

(1) **Forward:** cause $\rightarrow$ effect, physical, given. fire $\rightarrow$ alarm.

(2) **Backward:** effect $\rightarrow$ cause, mental, inferred. alarm $\rightarrow$ fire.

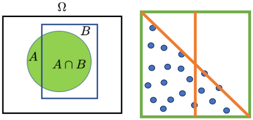
Bayes network, directed acyclic graph, graphic model



Conditional independence:

(1) Sibling nodes are independent given parent node.

(2) Child node is independent of grandparents given parent.



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

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

Independence

$$P(A|B) = P(A).$$

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

A and B have nothing to do with each other.

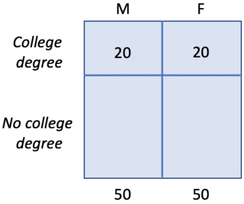
Independence

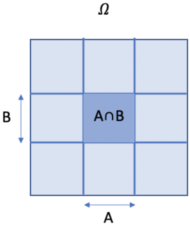
Definition 1:

$$P(A|B) = P(A).$$
$$p(y|x) = p(y).$$

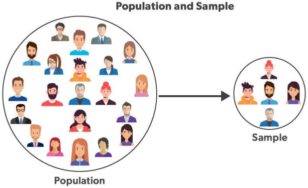
Definition 2:

$$P(A \cap B) = P(A)P(B).$$
$$p(x, y) = p(x)p(y).$$





Population of sequences



Sample a person from population Ω_1 of N people uniformly.
Repeat n times independently.
$\Omega_n = \{$ all N^n possible sequences $\}$.
equally likely outcomes in Ω_1 + independent repetitions
= equally likely sequences in Ω_n.
Let $\omega = (a_1, a_2, ..., a_n) \in \Omega_n$, each $a_i \in \Omega_1$.
$P(\omega) = P(a_1)P(a_2)...P(a_n) = \frac{1}{N} \times \frac{1}{N} \times ... \times \frac{1}{N} = \frac{1}{N^n}$.
Coin flipping: $\Omega_1 = \{$ head, tail $\}$.
Die rolling: $\Omega_1 = \{1, 2, ..., 6\}$.
Uniform random number $\Omega_1 = [0, 1]$.

Conditional independence

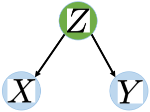
Markov chain: $C \to B \to A$, $Z \to X \to Y$.

$$P(A|B, C) = P(A|B).$$
$$p(y|x, z) = p(y|x).$$

Future is independent of the past given present.
Immediate cause (parent), remote cause (grandparent).
Meta rule: Insert same condition in a definition or equation.

Conditional independence

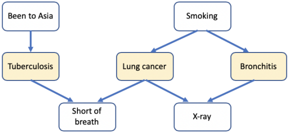
Shared cause: $C \leftarrow B \rightarrow A$.



$$P(A \cap C|B) = P(A|B)P(C|B).$$
$$p(x, y|z) = p(x|z)p(y|z).$$

Children given parent.
Meta rule: Insert same condition in a definition or equation.

Bayes net



a: Been to Asia; s: Smoking; t: Tuberculosis; l: Lung cancer; b: Bronchitis; d: Short of breath (Dyspnea); x: X-ray.

$$p(a, s, t, l, b, d, x) = p(a)p(s)p(t|a)p(l|s)p(b|s)p(d|t, l)p(x|b, l),$$

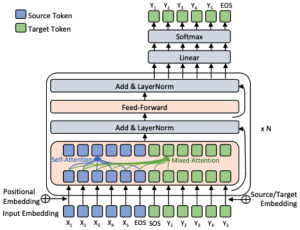
$$p(l|a, s, d, x) = \frac{p(l, a, s, d, x)}{p(a, s, d, x)},$$

$$p(l, a, s, d, x) = \sum_{t,b} p(a, s, t, l, b, d, x),$$

$$p(a, s, d, x) = \sum_{l} p(l, a, s, d, x).$$

Efficient calculation: message passing / belief propagation.

Generative Pre-trained Transformer (GPT)

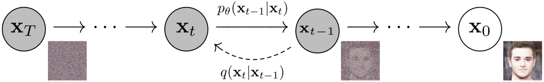


$x = (x_1, ..., x_{T_x})$ (e.g., "Can you write a poem?")
$y = (y_1, ..., y_{T_y})$ (e.g., "Certainly. Below is the poem...")
$p(y|x) = \prod_{t=1}^{T_y} p(y_t | y_{<t}, x)$.
Learn from training data $(x^{(i)}, y^{(i)}, i = 1, ..., n)$ by maximizing

$$\frac{1}{n}\sum_{i=1}^{n} \log p_\theta(y^{(i)}|x^{(i)}) = \frac{1}{n}\sum_{i=1}^{n}\sum_{t} \log p_\theta(y_t^{(i)}|y_{<t}^{(i)}, x^{(i)}).$$

memorize and generalize (interpolation).

Denoising Diffusion Probability Model



x_0: clean image.
$x_t = x_{t-1} + e_t$, e_t: small noise
Forward noising $q(x_t|x_{t-1})$, $t = 1, ..., T$. x_T: big noise.
Backward denoising $p(x_{t-1}|x_t)$.
Learn from training data $(x_0^{(i)}, i = 1, ..., n)$ by maximizing

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{t=T}^{1}\log p_\theta(x_{t-1}^{(i)}|x_t^{(i)}).$$

memorize and generalize (interpolation).

Take home message

As long as you can count
Count the population (of equally likely outcomes)
Count the repetitions (sequence of outcomes, fluctuation)
Population of sequences of repetitions (equally likely sequences)
Population of trajectories (random walk)

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
Random variables and probability functions, expectation
Continuous random variables, continuous time processes