STATS 100A: BASICS & EXAMPLES

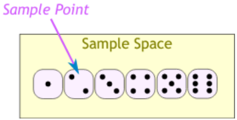
Ying Nian Wu

Department of Statistics
University of California, Los Angeles

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

Basic language and notation

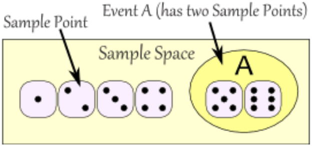
Experiment $\rightarrow$ **outcome** $\rightarrow$ **number**
Example 1: Roll a die



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

Basic language and notation

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

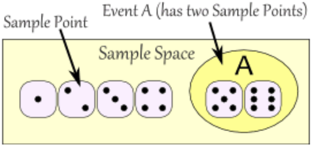


Sample space Ω: The set of all the outcomes.
Event A:
(1) A **statement** about the outcome, e.g., bigger than 4.
(2) A **subset** of sample space, e.g., $\{5, 6\}$.

Basic language and notation

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

Example 1: Roll a die



Probability: defined on event:

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

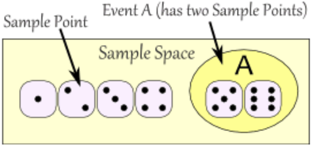
Assume the die is fair so that all the outcomes are **equally likely**.

$|A|$ 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



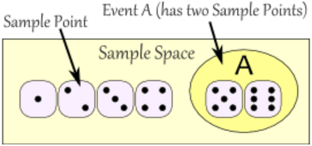
Random variable: Let X be the number:

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

An event is a **math statement** about the random variable.

Experiment $\to$ **outcome** $\to$ **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?

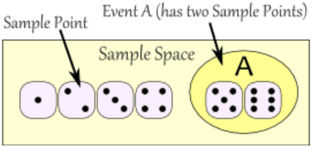
$$P(B|A) = \frac{1}{2}.$$

As if we randomly sample a number from A.
As if A is the sample space.

Basic language and notation

Experiment $\rightarrow$ outcome $\rightarrow$ number

Example 1: Roll a die



Random variable

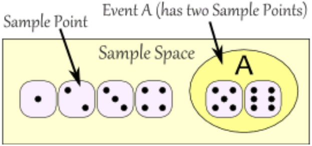
$$P(X = 6 | X > 4) = \frac{1}{2}.$$

Basic language and notation

100A

Ying Nian Wu

Example 1: Roll a die

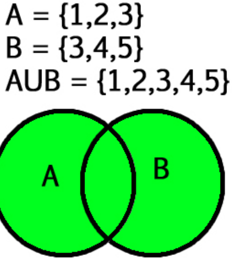


Complement

Statement: Not A

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

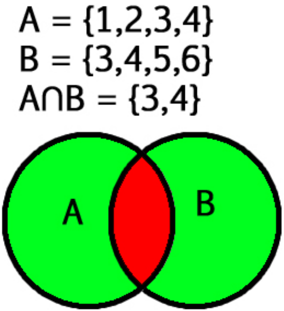
Example 1: Roll a die



Union
Statement: A or B.
Subset: $A \cup B$.

Basic language and notation

Example 1: Roll a die

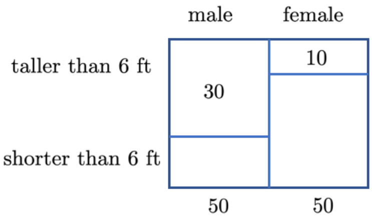


Intersection
Statement: A and B.
Subset: $A \cap B$.

Population proportion

Experiment $\rightarrow$ outcome $\rightarrow$ number

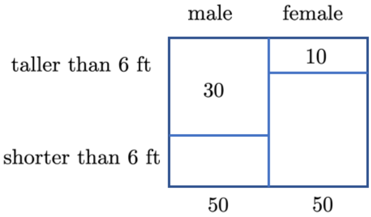
Example 2: Sample a random person from a population of 100 people, 50 males and 50 females. 30 males are taller than 6 ft, 10 females are taller than 6 ft. The sample space Ω is the whole population.



Population proportion

Experiment $\to$ outcome $\to$ number

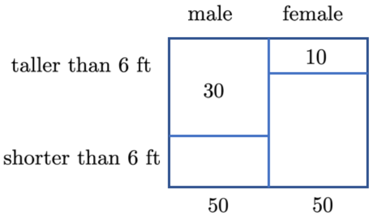
Example 2: Let A be the event that the person is male. Let B be the event that the person is taller than 6 feet (or simply the person is tall). A is the sub-population of males, and B is the sup-population of tall people.



Population proportion

Experiment $\rightarrow$ outcome $\rightarrow$ number

Example 2: A male, B tall.



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

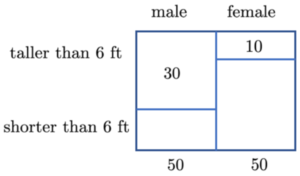
$$P(B) = \frac{|B|}{|\Omega|} = \frac{30+10}{100} = 40\%.$$

Probability = population proportion.

Population proportion

Experiment $\to$ outcome $\to$ 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.

Population proportion

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

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

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

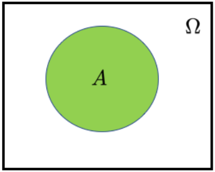
$$P(B) = P(\{\omega : Y(\omega) > 6\}) = P(Y > 6).$$

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

Link between event and random variable.

Population proportion

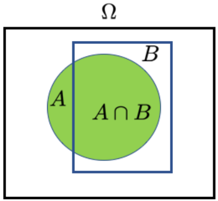
Equally likely scenario



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

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

Equally likely scenario



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

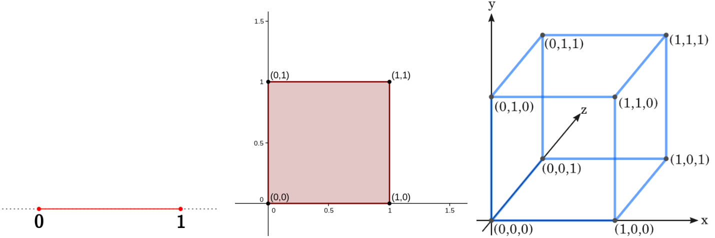
As if B is the sample space.

Axiom 4

Or definition of conditional probability.

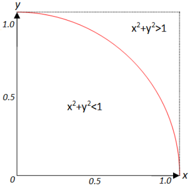
Area

Random point



(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 $[0, 1]$.

Random point
Example 3: throwing point into region

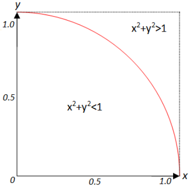


X and Y are independent uniform random numbers in [0, 1].
(X, Y) is a random point in $\Omega = [0, 1]^2$.
$A = \{(x, y) : x^2 + y^2 \leq 1\}$.

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

$|A|$ 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 $\Omega = [0, 1]^2$.
$A = \{(x, y) : x^2 + y^2 \le 1\}$.

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

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

Capital letters for random variables.

Measure

100A

Ying Nian Wu

Basics

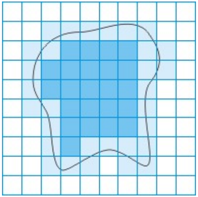
Population

Area

Coin

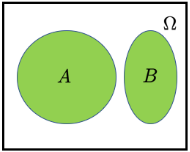
Random walk

Reasoning



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, σ-algebra.
Integral: area under curve.

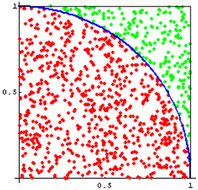
Probability as measure
Axiom 1: $P(\Omega) = 1$.
Axiom 2: $P(A) \geq 0$.



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

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

Example 3: π



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

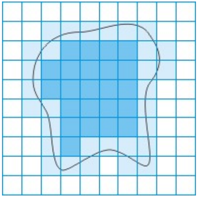
$$P(A) \approx \frac{m}{n}.$$

Monte Carlo method:

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

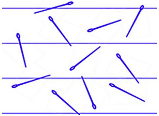
For fixed n, m is random.
As $n \to \infty$, $\frac{m}{n} \to P(A)$ in probability.
$P(A)$ can be interpreted as long run frequency.

Deterministic method



Go over all the $n = 100 = 10^2$ 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: π, buffon needle



Lazzarini threw $n = 3408$ times.

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

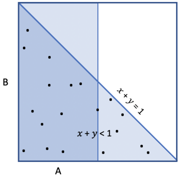
Monte Carlo method:

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

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.

Area

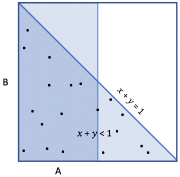
Example 3: throwing point into region



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

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

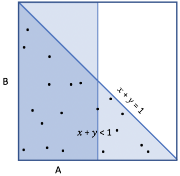
Example 3: throwing point into region



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



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

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

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 → outcome → number

EXAMPLE 4: Coin flipping

(4.1) Flip a coin → head or tail → 1 or 0

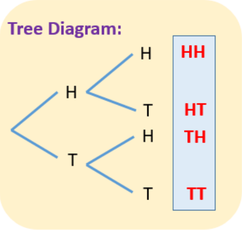
(4.2) Flip a coin twice → (head, head), or (head, tail), or (tail, head) or (tail, tail) → 11 or 10 or 01 or 00

List:

HH HT TH TT

Table:

	H	T
H	HH	HT
T	TH	TT

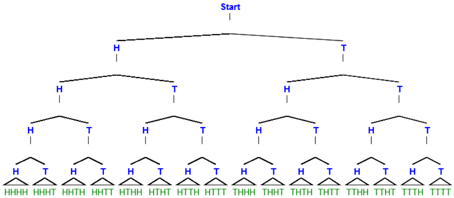


The sample space is {HH, HT, TH, TT}

Experiment $\rightarrow$ outcome $\rightarrow$ number

Experiment 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.

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

Coin flipping

Experiment 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.

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

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

Coin flipping

Experiment 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.

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

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

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

Experiment 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

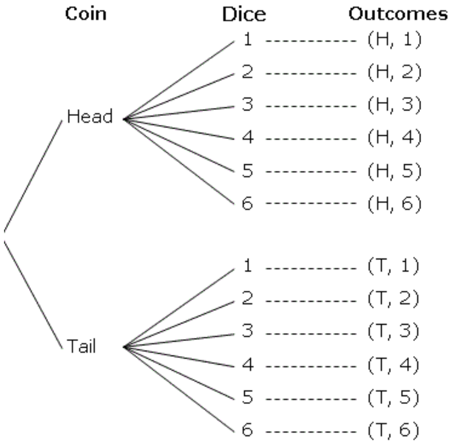
Ordered pair: roll a die twice

	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 $n_1 \times n_2$.

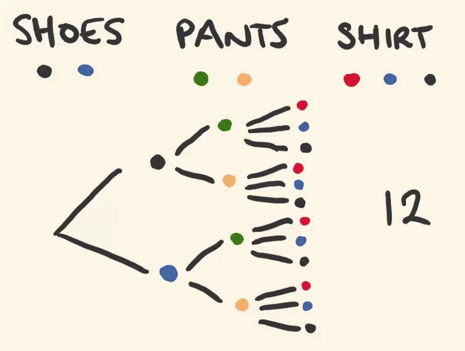
Multiplication

Ordered pair: roll a die twice

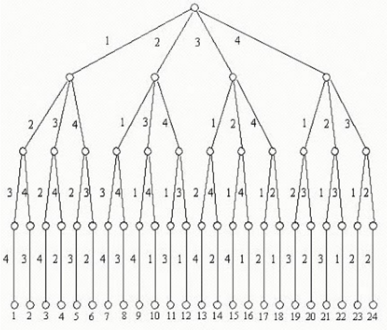


Multiplication

Ordered triplet



Permutation



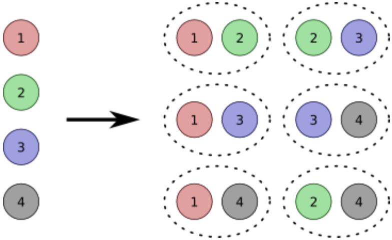
n different cards. Choose k of them. Order matters. Number of different sequences:

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

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

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.

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

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

Experiment 4: Coin flipping

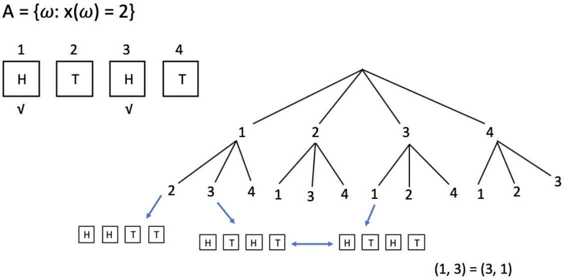
$$\begin{array}{llll} \text{HHHH}, & \text{THHH}, & \text{HTHT}, & \text{TTHT}, \\ \text{HHHT}, & \text{HHTT}, & \text{THHT}, & \text{THTT}, \\ \text{HHTH}, & \text{TTHH}, & \text{HTTH}, & \text{HTTT}, \\ \text{HTHH}, & \text{THTH}, & \text{TTTH}, & \text{TTTT} \end{array}$$

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

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

$$X(\omega) = \sum_{i=1}^{n} Z_i(\omega).$$

Experiment 4: Coin flipping



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

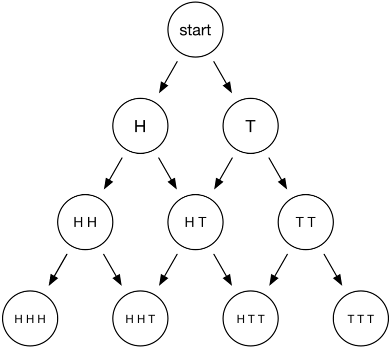
Experiment 4: Coin flipping

Combinations

$$\binom{0}{0}$$
$$\binom{1}{0} \quad \binom{1}{1}$$
$$\binom{2}{0} \quad \binom{2}{1} \quad \binom{2}{2}$$
$$\binom{3}{0} \quad \binom{3}{1} \quad \binom{3}{2} \quad \binom{3}{3}$$
$$\binom{4}{0} \quad \binom{4}{1} \quad \binom{4}{2} \quad \binom{4}{3} \quad \binom{4}{4}$$
$$\binom{5}{0} \quad \binom{5}{1} \quad \binom{5}{2} \quad \binom{5}{3} \quad \binom{5}{4} \quad \binom{5}{5}$$

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

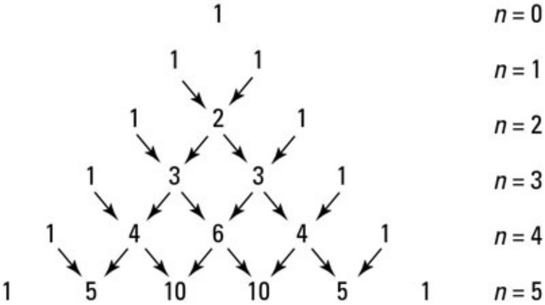
Experiment 4: Coin flipping
Combinations



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

Un-ordered sequence in each circle.

Experiment 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
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

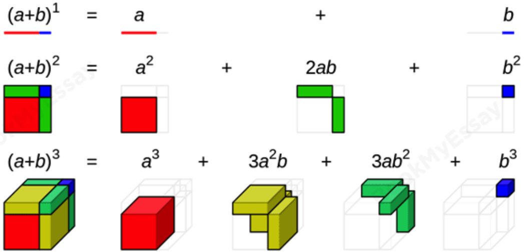
Experiment 4: Coin flipping
Binomial

$$(x+y)^0 = 1$$
$$(x+y)^1 = 1x + 1y$$
$$(x+y)^2 = 1x^2 + 2x^1y^1 + 1y^2$$
$$(x+y)^3 = 1x^3 + 3x^2y^1 + 3x^1y^2 + 1y^3$$
$$(x+y)^4 = 1x^4 + 4x^3y^1 + 6x^2y^2 + 4x^1y^3 + 1y^4$$
$$(x+y)^5 = 1x^5 + 5x^4y^1 + 10x^3y^2 + 10x^2y^3 + 5x^1y^4 + 1y^5$$

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

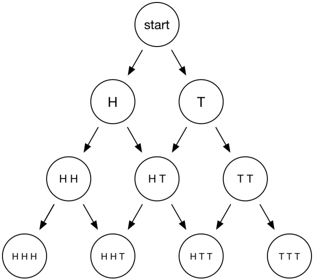
$(x+y)(x+y)(x+y)(x+y)$: each $(x+y)$ contributes either x or y, 2^n products, $\binom{n}{k}$ of them have k x's and $n-k$ y's.

Experiment 4: Coin flipping
Binomial



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

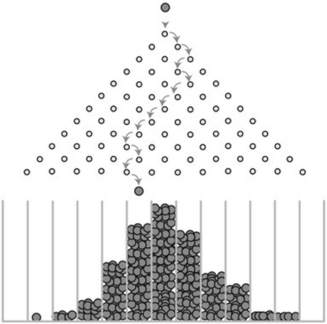
Experiment 4: Coin flipping



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

Order of sequence in each circle does not matter.
All 2^n paths are equally likely.
Counting the number of paths that end up in k-th bin.

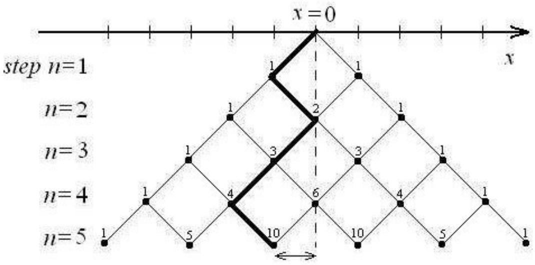
Experiment 4: Coin flipping



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.

Random walk

Either go forward or backward



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**

Random walk

Either go forward or backward

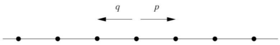


Figure 1: Simple random walk

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

$$P(X_t = m) = \binom{t}{(m+t)/2} / 2^t.$$

Transition probability

Either go forward or backward

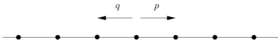


Figure 1: Simple random walk

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

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

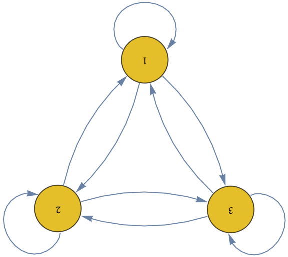
$Z_k = 1$ or -1 with probability $1/2$ each.

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

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

Markov chain

Example 5: Random walk over three states



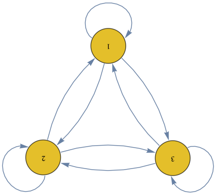
With probability 1/2, stay. With probability 1/4, go to either states.

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

Markov property: past history before X_t does not matter.

Population migration

Example 5: Random walk over three states

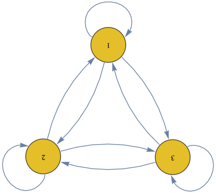


With probability $1/2$, stay. With probability $1/4$, go to either of the other two states.

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

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



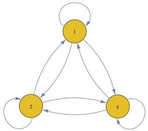
With probability 1/2, stay. With probability 1/4, go to either of the other two states.

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

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

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.

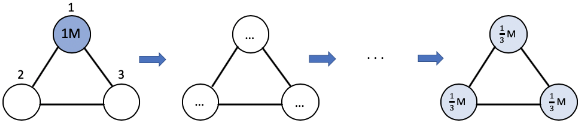
$$K_{ij} = P(X_{t+1} = j | X_t = i).$$

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

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



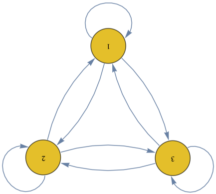
$$p_i^{(t)} = P(X_t = i).$$

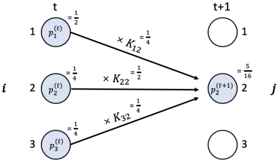
Imagine 1 million people migrating. $p_i^{(t)}$ is the number of people (in million) in state i at time t.

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

Population migration

Example 5: Random walk over three states





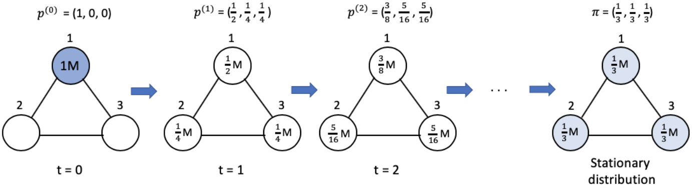
$$K_{ij} = P(X_{t+1} = j | X_t = i).$$

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

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

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



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

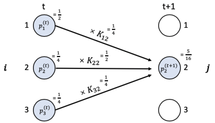
$$p_i^{(t)} \to \pi_i.$$

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

Stationary distribution.

Matrix multiplication

Example 5: Random walk over three states



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

	1	2	3
$p^{(t+1)}$		$\frac{5}{16}$	
$p^{(t)}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$

K

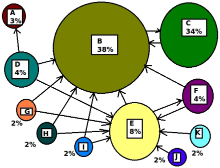
$i \backslash j$	1	2	3
1		$\frac{1}{4}$	
2		$\frac{1}{2}$	
3		$\frac{1}{4}$	

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

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

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

Example 5: Random walk over three states



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

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

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

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

Chain rule and rule of total probability

Example 5: Random walk over three states

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

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

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

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

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

Add up probabilities of alternative chains of events.

Independence

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

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

Independence

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

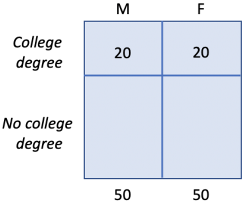
$$P(A \cap B) = P(A)P(B).$$

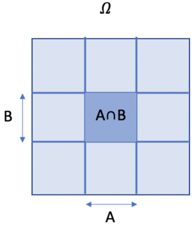
A and B have nothing to do with each other.

Independence

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

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





Conditional independence

100A

Ying Nian Wu

Basics

Population

Area

Coin

Random walk

Reasoning

Markov chain: $C \to B \to A$,

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

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

Future is independent of the past given present.
Immediate cause (parent), remote cause (grandparent).
Shared cause: $C \leftarrow B \to A$,

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

Children given parent.

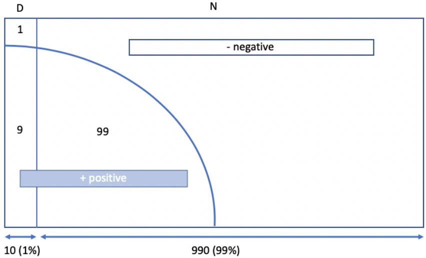
Reasoning

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

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

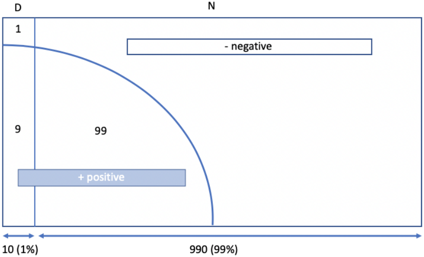


$P(D|+) = \frac{9}{9+99} = \frac{1}{12}$.
$P(alarm|fire)$ vs $P(fire|alarm)$.

Example 6: Rare disease example

$P(D) = 1\%$.

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



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

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

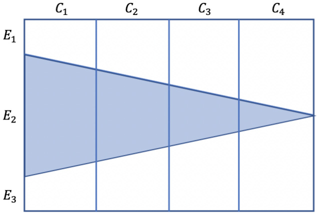
$P(+) = P(D \cap +) + P(N \cap +) = 1\% \times 90\% + 99\% \times 10\%$.

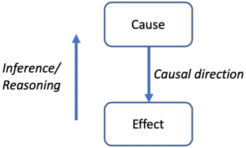
$P(D|+) = \frac{P(D \cap +)}{P(+)} = \frac{9}{9+99} = \frac{1}{12}$.

Chain rule, rule of total probability, Bayes rule

General formula

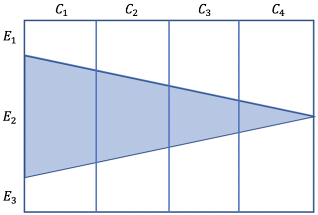
m causes: $C_1, ..., C_i, ..., C_m$.
n effects: $E_1, ..., E_j, ..., E_n$.
Given:
Prior: $P(C_i), i = 1, ..., m$.
Conditional: $P(E_j|C_i), j = 1, ..., n$.

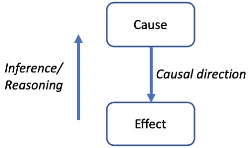




Chain rule, rule of total probability, Bayes rule

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

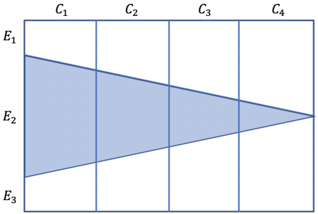


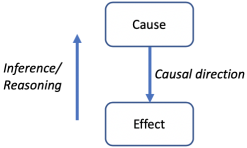


$$P(C_i|E).$$
$$P(C_i \cap E) = P(C_i)P(E|C_i).$$
$$P(E) = \sum_i P(C_i \cap E) = \sum_i P(C_i)P(E|C_i).$$
$$P(C_i|E) = \frac{P(C_i \cap E)}{P(E)} = \frac{P(C_i)P(E|C_i)}{\sum_{i'} P(C_{i'})P(E|C_{i'})}.$$

Chain rule, rule of total probability, Bayes rule

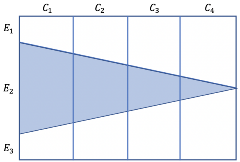
X = cause $\in \{1, ..., i, ..., m\}$.
Y = effect $\in \{1, ..., j, ..., n\}$.

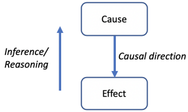


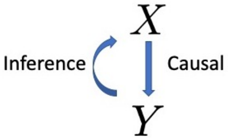


$$P(X = i|Y = j) = \frac{P(X = i \cap Y = j)}{P(Y = j)}$$
$$= \frac{P(X = i)P(Y = j|X = i)}{\sum_{i'} P(X = i')P(Y = j|X = i')}.$$

Chain rule, rule of total probability, Bayes rule



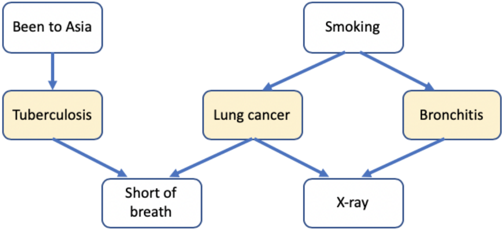




$$P(X=i|Y=j) = \frac{P(X=i\cap Y=j)}{P(Y=j)}$$
$$= \frac{P(X=i)P(Y=j|X=i)}{\sum_{i'} P(X=i')P(Y=j|X=i')}.$$

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

Bayes network, directed acyclic graph, graphic model



Conditional independence:

Sibling nodes are independent given parent node.

Child node is independent of grandparents given parent.

Take home message

As long as you can count
Count the number of people
Count the number of points or repetitions
Count the number of sequences
Two things
Intuition and visualization (and motivation)
Precise notation and formula
Accomplished
Most of the important concepts via intuitive examples
Next
Systematic and more in-depth treatments