Probability

Chia-Ping Chen

Professor
Department of Computer Science and Engineering
National Sun Yat-sen University

Probability

Outline

- Set
- Random experiment
- Probability model
- Conditional probability

- Total probability theorem
- Bayes rule
- Independence
- Counting principle

Sets

Definition (set and member)

- A **set** is a collection (may be empty) of objects
- An object in a set is called a **member** or an **element** of the set

$$\begin{array}{ccc} & & \text{member} \\ & & \downarrow \\ \text{set} & \longrightarrow & \{\text{dog, bear, pig}\} \end{array}$$

Definition (universal set and empty set)

- A **universal set** contains all objects of interest
- An **empty set** does not contain any member

- A universal set is often denoted by Ω. For example

$$\Omega = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$$

  if we are discussing random outcome of a digit.
- We denote an empty set by $\emptyset$. That is

$$\emptyset = \{\,\}$$

Definition (subset)

Let T be a set. A **subset** of T is formed by member(s) of T. Note that, by definition, $\emptyset$ is a subset of every set.

Let S be a subset of T.

- This is denoted by

$$S \subset T$$

- Every member of S is a member of T
- For example, $\Omega = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$ and $S = \{2, 3, 5, 7\}$

$$S \subset \Omega$$

Sets in probability theory

Many concepts in probability theory are based on sets. In particular

- A **sample space** is a universal set
- A **field** is a set of subsets of the sample space with certain properties
- An **event** is a member of the field and a subset of the sample space

Definition (set operation)

Let Ω be a universal set and $A, B \subset \Omega$.

- **complement**

$$A^c = \{x \in \Omega \mid x \notin A\}$$

- **union**

$$A \cup B = \{x \mid x \in A \text{ or } x \in B\}$$

- **intersection**

$$A \cap B = \{x \mid x \in A \text{ and } x \in B\}$$

For the above example

$$S^c = \{0, 1, 4, 6, 8, 9\}$$

$$S \cup S^c = \Omega$$

$$S \cap S^c = \emptyset$$

Set algebra

Let Ω be a universal set and $S, T, U \subset \Omega$.

- **commutative law**

$$S \cap T = T \cap S, \ S \cup T = T \cup S$$

- **associative law**

$$S \cap (T \cap U) = (S \cap T) \cap U, \ S \cup (T \cup U) = (S \cup T) \cup U$$

- **distributive law**

$$S \cap (T \cup U) = (S \cap T) \cup (S \cap U), \ S \cup (T \cap U) = (S \cup T) \cap (S \cup U)$$

- **identity**

$$S \cup S^c = \Omega, \ S \cap S^c = \emptyset$$

$$S \cup \emptyset = S, \ S \cap \emptyset = \emptyset, \ S \cup \Omega = \Omega, \ S \cap \Omega = S$$

Boolean algebra (bit algebra)

Note the similarity between set algebra and bit algebra.

Let p, q, r be bits (0 and 1).

$$p \wedge q = q \wedge p, \;\; p \vee q = q \vee p$$

$$p \wedge (q \wedge r) = (p \wedge q) \wedge r, \;\; p \vee (q \vee r) = (p \vee q) \vee r$$

$$p \wedge (q \vee r) = (p \wedge q) \vee (p \wedge r), \;\; p \vee (q \wedge r) = (p \vee q) \wedge (p \vee r)$$

$$p \vee \neg p = 1, \;\; p \wedge \neg p = 0$$

$$p \vee 0 = p, \;\; p \wedge 0 = 0, \;\; p \vee 1 = 1, \;\; p \wedge 1 = p$$

De Morgan's law

Let Ω be a universal set and $S_1, \ldots, S_n$ be subsets of Ω. Then

$$\left(\bigcup_{i=1}^{n} S_i\right)^c = \bigcap_{i=1}^{n} S_i^c$$

$$\left(\bigcap_{i=1}^{n} S_i\right)^c = \bigcup_{i=1}^{n} S_i^c$$

- Complement of union is intersection of complements
- Complement of intersection is union of complements
- Base case

$$(A \cup B)^c = A^c \cap B^c$$

$$(A \cap B)^c = A^c \cup B^c$$

Proof.

$$
\begin{aligned}
x \in \left(\bigcup_{i=1}^{n} S_i\right)^c &\Leftrightarrow x \notin \left(\bigcup_{i=1}^{n} S_i\right) \\
&\Leftrightarrow x \notin S_i \text{ for every } i \\
&\Leftrightarrow x \in S_i^c \text{ for every } i \\
&\Leftrightarrow x \in \bigcap_{i=1}^{n} S_i^c
\end{aligned}
$$

$$
\begin{aligned}
x \in \left(\bigcap_{i=1}^{n} S_i\right)^c &\Leftrightarrow x \notin \left(\bigcap_{i=1}^{n} S_i\right) \\
&\Leftrightarrow x \notin S_i \text{ for some } i \\
&\Leftrightarrow x \in S_i^c \text{ for some } i \\
&\Leftrightarrow x \in \bigcup_{i=1}^{n} S_i^c
\end{aligned}
$$

Note

$$A = B \;\Leftrightarrow\; (x \in A \Leftrightarrow x \in B)$$

□

Definition (disjoint sets)

- Sets A and B are said to be **disjoint** if

$$A \cap B = \emptyset$$

- A group of sets is said to be disjoint if any pair of sets in the group are disjoint

Definition (partition of a set)

Let T be a set. A **partition** of T is a group of sets such that

- the group is disjoint
- the group union is T

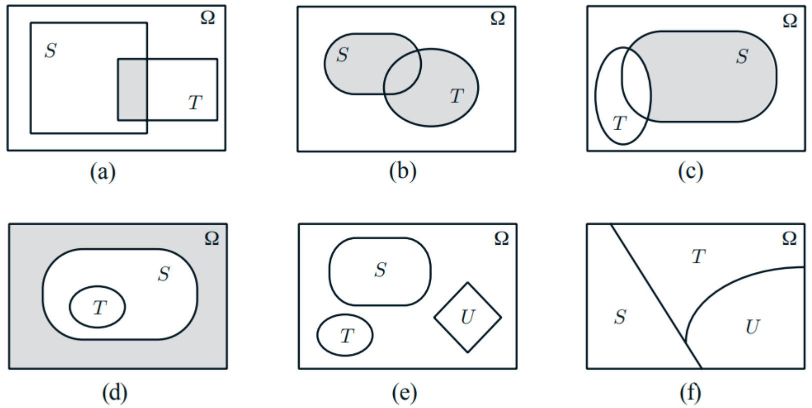
In the earlier example of digits, is (S, S^c) a partition of Ω?

Yes, since we have

$$S \cap S^c = \emptyset$$

$$S \cup S^c = \Omega$$

Venn diagrams



Useful for showing set relations and set operations

Random Experiment and Probability Model

Definition (random experiment)

- A procedure that produces a random outcome
- The stopping criteria and final outcome must be clear

Example

- Flip a coin 3 times and record the 3-long sequence
- Flip 3 coins and record the number of heads
- Flip a coin until back-to-back heads show up and record the number of flips
- Draw 3 cards from a deck with replacement and record the 3-long sequence
- Draw 3 cards from a deck without replacement and record the 3-long sequence

Definition (sample space)

- A set for the possible outcomes of a random experiment
- Members of a sample space must be **exclusive** and **exhaustive**

Example

Consider the random experiment of throwing a dice once.

- If we care about which face comes up, use $\{1, 2, 3, 4, 5, 6\}$ as sample space
- If we only care about whether the outcome is even or odd, use $\{\text{even}, \text{odd}\}$ as sample space
- Both sets are exclusive and exhaustive

Example (1.1 Choice of sample space)

Consider two games involving n successive coin tosses.

- Game A: We receive \$1 each time a head comes up.
- Game B: The reward for the next toss is doubled whenever a head comes up. Specifically, we receive \$1 for each coin toss, up to and including the first time a head comes up; after the first head, we receive \$2 for each coin toss, up to and including the second time a head comes up, etc.

Choice of sample space

- Game A: the number of heads ($n+1$ members)
- Game B: all length-n binary sequences (2^n members)

Definition (finite set, countable set)

- A set is **finite** if it has a finite number of members
- Otherwise, it is **infinite**
- A set is **countable** if there exists a one-to-one mapping from this set to (not necessarily onto) the set of natural numbers
- Otherwise, it is **uncountable**

Size of sample space

- A sample space may be finite or infinite
- An infinite sample space may be countable or uncountable

Example (Sample space sizes)

- Flip a coin once
  $\rightarrow$ A sample space is $\Omega_1 = \{H, T\}$. It is finite.
- Flip a coin until a head comes up
  $\rightarrow$ A sample space is $\Omega_2 = \{\omega_1, \omega_2, \dots\}$, where ω_n means that the first head shows up at the nth flip. Ω_2 is countable.
- Flip a coin forever
  $\rightarrow$ A sample space Ω_3 is the set of infinite sequences of heads and tails. Ω_3 is 1-to-1 to the set of real numbers in $[0, 1]$ if we let bit 1 denote head and bit 0 denote tail. It is uncountable.

Definition (probability model)

A probability model is a triplet

$$(\Omega, \mathcal{F}, P)$$

- Ω is a **sample space**
- $\mathcal{F}$, called **field**, is a set of subsets of Ω
- P, called **probability law**, is a mapping from $\mathcal{F}$ to $[0, 1]$

In other words

- Ω consists of exclusive and exhaustive outcomes
- Every member of $\mathcal{F}$ is a subset of Ω
- Every member of $\mathcal{F}$ is assigned a probability by P

random experiment: produce random outcome

$\Downarrow$

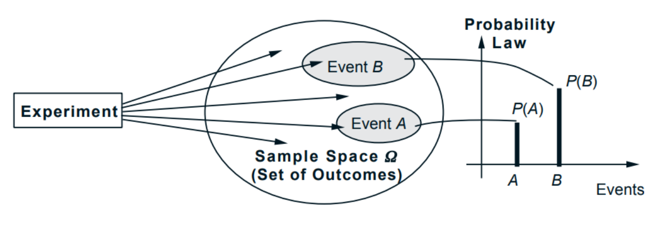
sample space Ω: exclusive and exhaustive set of outcomes

$\Downarrow$

field $\mathcal{F}$: set of subsets of Ω

$\Downarrow$

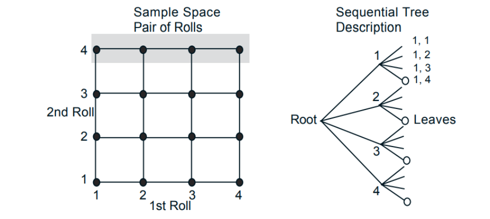
probability law: mapping $\mathcal{F}$ to $[0, 1]$



An **event** is a member of $\mathcal{F}$.

Sequential experiment

- A random experiment may be sequential
- Then the sample space may be structured by a tree
- Depth corresponds to stage



Field properties

Let $(\Omega, \mathcal{F}, P)$ be a probability model. By definition, $\mathcal{F}$ is a set of subsets of Ω. Furthermore

- $\mathcal{F}$ contains Ω

$$\Omega \in \mathcal{F}$$

- $\mathcal{F}$ is closed with respect to set intersection

$$(S_1 \in \mathcal{F}) \wedge (S_2 \in \mathcal{F}) \;\Rightarrow\; (S_1 \cap S_2 \in \mathcal{F})$$

- $\mathcal{F}$ is closed with respect to set union

$$(S_1 \in \mathcal{F}) \wedge (S_2 \in \mathcal{F}) \;\Rightarrow\; (S_1 \cup S_2 \in \mathcal{F})$$

- $\mathcal{F}$ is closed with respect to set complement

$$(S \in \mathcal{F}) \;\Rightarrow\; (S^c \in \mathcal{F})$$

Example (fields)

Let Ω be a sample space.

1. The power set of Ω defined by $\mathcal{P}(\Omega) = \{S \mid S \subset \Omega\}$ is a field
2. For any $S \subset \Omega$, $\{\emptyset, S, S^c, \Omega\}$ is a field
3. What is the smallest field with sample space Ω?

DEFINITION (EVENT)

Let $(\Omega, \mathcal{F}, P)$ be a probability model. Every member of $\mathcal{F}$ is called an **event** of the model.

For example, for the field $\mathcal{F} = \{\emptyset, S, S^c, \Omega\}$, the events are

$$\emptyset, \; S, \; S^c, \; \Omega$$

Probability law

Let $(\Omega, \mathcal{F}, P)$ be a probability model.

- P maps $\mathcal{F}$ to $[0,1]$

$$P : \mathcal{F} \ \mapsto \ [0,1]$$

- That is, for every event $A \in \mathcal{F}$, we have

$$0 \leq P(A) \leq 1$$

- $P(A)$ is called the **probability** of A

Probability axioms

Let $(\Omega, \mathcal{F}, P)$ be a probability model. The probability law P must have the following properties.

- non-negativity

$$P(A) \geq 0, \text{ for any event } A \in \mathcal{F}$$

- normalization

$$P(\Omega) = 1$$

- additivity

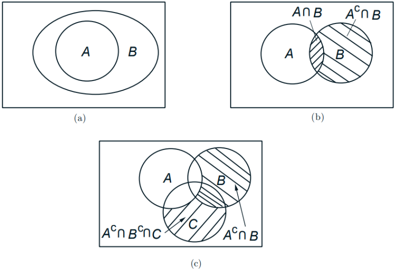
$$P(A \cup B) = P(A) + P(B), \text{ for any disjoint events } A \text{ and } B$$

Union probability (2 sets)

Let $(\Omega, \mathcal{F}, P)$ be a probability model. For any events A and B we have

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$\begin{aligned} P(A \cup B) &= P(A) + P(B \cap A^c) \\ &= P(A) + P(B) - P(A \cap B) \end{aligned}$$



Union probability (3 sets)

Let $(\Omega, \mathcal{F}, P)$ be a probability model. For any events A, B and C we have

$$\begin{aligned} P(A \cup B \cup C) = P(A) + P(B) + P(C) \\ - P(A \cap B) - P(A \cap C) - P(B \cap C) \\ + P(A \cap B \cap C) \end{aligned}$$

$$\begin{aligned} P(A \cup B \cup C) &= P(A \cup B) + P(C) - P(C \cap (A \cup B)) \\ &= P(A \cup B) + P(C) - P((C \cap A) \cup P(C \cap B)) \\ &= P(A \cup B) + P(C) - (P(C \cap A) + P(C \cap B) - P(C \cap A \cap B)) \\ &= P(A) + P(B) - P(A \cap B) + P(C) - (P(C \cap A) + P(C \cap B) - P(C \cap A \cap B)) \\ &= P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C) \end{aligned}$$

Inclusion-exclusion principle

Let $(\Omega, \mathcal{F}, P)$ be a probability model. For any events $A_1, \dots, A_n$, the probability of set union is related to the probabilities of set intersection by

$$
\begin{aligned}
P\left(\bigcup_{i=1}^{n} A_i\right) &= \sum_{i=1}^{n} P(A_i) \\
&\quad - \sum_{i=1}^{n} \sum_{j=i+1}^{n} P(A_i \cap A_j) \\
&\quad + \sum_{i=1}^{n} \sum_{j=i+1}^{n} \sum_{k=j+1}^{n} P(A_i \cap A_j \cap A_k) \\
&\quad + \cdots + (-1)^{n+1} P\left(\bigcap_{i=1}^{n} A_i\right)
\end{aligned}
$$

The equality holds for the base case $n = 2$

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$$

For the induction case, the induction assumption is that the equality holds for $n = k$

$$P\left(\bigcup_{i=1}^{k} A_i\right) = \sum_{i=1}^{k} P(A_i) - \sum_{i=1}^{k} \sum_{j=i+1}^{k} P(A_i \cap A_j) + \cdots + (-1)^{k+1} P\left(\bigcap_{i=1}^{k} A_i\right)$$

For $n = k + 1$, we have

$$\begin{aligned} P\left(\bigcup_{i=1}^{k+1} A_i\right) &= P\left(\left(\bigcup_{i=1}^{k} A_i\right) \cup A_{k+1}\right) \\ &= P\left(\bigcup_{i=1}^{k} A_i\right) + P(A_{k+1}) - P\left(\left(\bigcup_{i=1}^{k} A_i\right) \cap A_{k+1}\right) \\ &= P\left(\bigcup_{i=1}^{k} A_i\right) + P(A_{k+1}) - P\left(\bigcup_{i=1}^{k} (A_i \cap A_{k+1})\right) \end{aligned}$$

Using the induction assumption, we re-write the last term

$$
\begin{aligned}
P\left(\bigcup_{i=1}^{k}(A_i \cap A_{k+1})\right) &= \sum_{i=1}^{k} P(A_i \cap A_{k+1}) \\
&- \sum_{i=1}^{k}\sum_{j=i+1}^{k} P((A_i \cap A_{k+1}) \cap (A_j \cap A_{k+1})) + \cdots + (-1)^{k+1} P\left(\bigcap_{i=1}^{k}(A_i \cap A_{k+1})\right) \\
&= \sum_{i=1}^{k} P(A_i \cap A_{k+1}) - \sum_{i=1}^{k}\sum_{j=i+1}^{k} P(A_i \cap A_j \cap A_{k+1}) + \cdots + (-1)^{k+1} P\left(\bigcap_{i=1}^{k+1} A_i\right)
\end{aligned}
$$

and get

$$
P\left(\bigcup_{i=1}^{k+1} A_i\right) = \sum_{i=1}^{k+1} P(A_i) - \sum_{i=1}^{k+1}\sum_{j=i+1}^{k+1} P(A_i \cap A_j) + \cdots + (-1)^{k+2} P\left(\bigcap_{i=1}^{k+1} A_i\right)
$$

Thus, the equality holds for $n = k+1$ if it holds for $n = k$.

Definition (discrete probability model)

Let $(\Omega, \mathcal{F}, P)$ be a probability model. It is discrete if Ω is countable.

- P is often specified by the probabilities of the members of Ω
- For any event A

$$P(A) = \sum_{\omega \in A} P(\omega)$$

Definition (continuous probability model)

Let $(\Omega, \mathcal{F}, P)$ be a probability model. It is continuous if Ω is uncountable.

- P is often specified by the probabilities of infinitesimal subsets of Ω
- For an event $S \in \mathcal{F}$, partition S to infinitesimal subsets

$$S = \bigcup_i \Delta_i$$

Then

$$P(S) = \sum_i P(\Delta_i)$$

Uniform model

Let $(\Omega, \mathcal{F}, P)$ be a probability model. It is said to be uniform if P has no preference to any member of Ω.

- discrete uniform model

$$P(A) = \frac{|A|}{|\Omega|},\ A \in \mathcal{F}$$

- continuous uniform model

$$P(A) = \frac{\mathsf{vol}(A)}{\mathsf{vol}(\Omega)},\ A \in \mathcal{F}$$

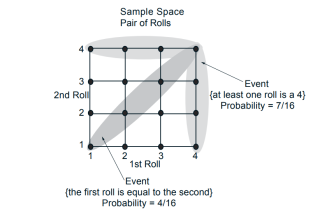
Example (1.2 Probability model)

- Construct a probability model for a single toss of a fair coin
- Construct a few probability models for 3 tosses of a fair coin

Example (1.3 Discrete uniform model)

Consider rolling a pair of dice of 4 sides. Assume the dice are fair. Calculate the probability of the following events.

- The sum of the rolls is even
- The sum of the rolls is odd
- The first equals the second
- The first $\geq$ the second
- At least one roll is 4



$$\Omega = \{1,2,3,4\} \times \{1,2,3,4\}$$

Example (1.4 Continuous uniform model)

A wheel of fortune is uniformly calibrated from 0 to 1. The outcome of a spin is a point in $\Omega = (0, 1)$.

- What is the probability of a single number?
- What is the probability of an interval $(a, b) \subset \Omega$?

Example (1.5 Continuous uniform model)

Romeo and Juliet have a date. Each arrives with a delay between 0 and 1 hour, all times of delay being equally likely, and waits for 15 minutes. What is the probability that they meet?

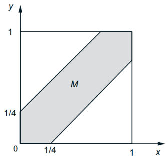
- X, Y: delays of R and J
- Sample space

$$\Omega = [0,1] \times [0,1]$$

- Event that they meet

$$M = \left\{ (X,Y) \;\middle|\; |X - Y| \le \frac{1}{4} \right\}$$

$$P(M) = 1 - \frac{9}{16} = \frac{7}{16}$$



Conditional Probability

Definition (conditional probability)

Let $(\Omega, \mathcal{F}, P)$ be a probability model and $A, B \in \mathcal{F}$ be events. The conditional probability of A given (the occurrence of) B is defined by

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

- Implicitly assume B has non-zero probability
- Probability is exclusively re-allocated to B
- $P(B)$ in denominator re-normalizes probability
- Only outcomes in B matter, so $P(A|B) \propto P(A \cap B)$

Definition (conditional probability model)

Let $(\Omega, \mathcal{F}, P)$ be a probability model and $B \in \mathcal{F}$ be an event. Define $P_B : \mathcal{F} \mapsto [0,1]$ with $P_B(S) = P(S|B)$ for all $S \in \mathcal{F}$. Then $(\Omega, \mathcal{F}, P_B)$ is a probability model.

- non-negativity

$$P_B(S) = \frac{P(S \cap B)}{P(B)} \geq 0$$

- normalization

$$P_B(\Omega) = \frac{P(\Omega \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$$

- additivity

$$\begin{aligned} S_1 \cap S_2 = \emptyset \ \Rightarrow \ P_B(S_1 \cup S_2) &= \frac{P((S_1 \cup S_2) \cap B)}{P(B)} \\ &= \frac{P((S_1 \cap B) \cup (S_2 \cap B))}{P(B)} \\ &= \frac{P(S_1 \cap B)}{P(B)} + \frac{P(S_2 \cap B)}{P(B)} - \frac{P(S_1 \cap S_2 \cap B)}{P(B)} \\ &= \frac{P(S_1 \cap B)}{P(B)} + \frac{P(S_2 \cap B)}{P(B)} \\ &= P_B(S_1) + P_B(S_2) \end{aligned}$$

Conditional probability of union and intersection

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

$$P(A \cup B|C) = P(A|C) + P(B|C) - P(A \cap B|C)$$

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

$$\begin{aligned} P(A \cup B|C) &= \frac{P((A \cup B) \cap C)}{P(C)} \\ &= \frac{P((A \cap C) \cup (B \cap C))}{P(C)} \\ &= \frac{P(A \cap C) + P(B \cap C) - P(A \cap B \cap C)}{P(C)} \\ &= P(A|C) + P(B|C) - P(A \cap B|C) \end{aligned}$$

Example (1.6 Conditional probability)

Toss a fair coin three times. Define events

$$\begin{aligned} A &= \{\text{more heads than tails come up}\} \\ B &= \{\text{1st toss is a head}\} \end{aligned}$$

Compute $P(A|B)$.

$$\begin{aligned} \Omega &= \{H,T\} \times \{H,T\} \times \{H,T\} \\ A &= \{HHH, HHT, HTH, THH\} \\ B &= \{HHH, HHT, HTH, HTT\} \\ A \cap B &= \{HHH, HHT, HTH\} \end{aligned}$$

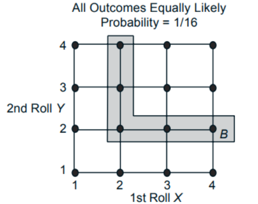
$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{3}{8}}{\frac{4}{8}} = \frac{3}{4}$$

Example (1.7 Conditional probability)

A fair four-face dice is rolled twice, with outcomes X and Y. Define

$$A_m = \{\max(X, Y) = m\}$$
$$B = \{\min(X, Y) = 2\}$$

Compute $P(A_m|B)$, $m = 1, 2, 3, 4$.



$$\Omega = \{(x_1, x_2) \,|\, x_{1,2} = 1, 2, 3, 4\}$$
$$A_1 = \{(1,1)\}, A_2 = \{(1,2), (2,2), (2,1)\}, \dots$$
$$B = \{(2,2), (2,3), (2,4), (3,2), (4,2)\}$$
$$A_1 \cap B = \emptyset, A_2 \cap B = \{(2,2)\}, A_2 \cap B = \{(3,2), (2,3)\}, \dots$$

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

Example (1.8 Conditional probability)

A conservative team and an innovative team are asked to design a new product within a month. Based on past experience

- the conservative team is successful with probability 2/3
- the innovative team is successful with probability 1/2
- at least one team is successful with probability 3/4

Given that exactly one team is successful, what is the probability that it is the innovative team?

Let A be the event that the conservative team succeeds and B be the event that the innovative team succeeds. From the problem statement, we have

$$P(A) = \frac{2}{3},\ P(B) = \frac{1}{2},\ P(A \cup B) = \frac{3}{4}$$

With the help of Venn diagram, we have

$$P(A \cap B) = P(A) + P(B) - P(A \cup B) = \frac{5}{12}$$
$$P(A \cap B^c) = P(A) - P(A \cap B) = \frac{2}{3} - \frac{5}{12} = \frac{1}{4}$$
$$P(A^c \cap B) = P(B) - P(A \cap B) = \frac{1}{2} - \frac{5}{12} = \frac{1}{12}$$

Thus

$$\begin{aligned}
P(B \mid (A \cap B^c) \cup (A^c \cap B)) &= \frac{P(B \cap ((A \cap B^c) \cup (A^c \cap B)))}{P((A \cap B^c) \cup (A^c \cap B))} \\
&= \frac{P((B \cap (A \cap B^c)) \cup (B \cap (A^c \cap B)))}{P(A \cap B^c) + P(A^c \cap B) - P((A \cap B^c) \cap (A^c \cap B))} \\
&= \frac{P(A^c \cap B)}{P(A \cap B^c) + P(A^c \cap B)} \\
&= \frac{1}{4}
\end{aligned}$$

Factorization

Let $(\Omega, \mathcal{F}, P)$ be a probability model and $A, B \in \mathcal{F}$ be events. Then

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

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

- One event at a time

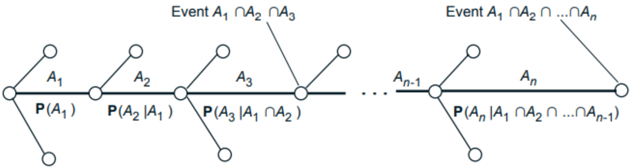
  $$\text{"}A \text{ and } B \text{ occur"} = \text{"}A \text{ occurs"} + \text{"}B \text{ occurs given } A \text{ occurs"}$$

- Factorization (a.k.a. multiplication rule or chain rule)

$$P\left(A_1 \cap \cdots \cap A_n\right) = P(A_1) \prod_{k=2}^{n} \overbrace{P\left(A_k | A_1 \cap \cdots \cap A_{k-1}\right)}^{A_k \text{ occurs given } A_1 \ldots A_{k-1} \text{ occur}}$$

Joint event as sequence of single events

- Arrange sequence of events in a tree
  - Start from root
  - An edge for every possible next event
  - Every node corresponds to a joint event
- Probability of an edge: conditional probability
- Probability of a node: joint probability



Example (1.9)

If an aircraft is present in a military zone, a radar detects it and generates a (true) alarm with probability 0.99. If an aircraft is not present, the radar generates a (false) alarm with probability 0.10. Suppose that an aircraft is present with probability 0.05. What is the probability of no aircraft presence and a (false) alarm? What is the probability of aircraft presence and no detection, i.e. a miss?

Define

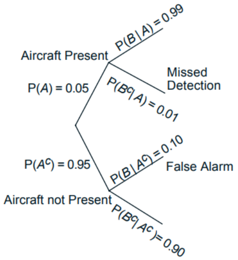
$$A = \{\text{an aircraft is present}\}$$
$$B = \{\text{radar generates alarm}\}$$

From the problem statement, we have

$$P(A) = 0.05, \; P(B|A) = 0.99, \; P(B|A^c) = 0.1$$

By factorization rule

$$P(B \cap A^c) = P(A^c)P(B|A^c) = 0.95 \times 0.1$$
$$P(B^c \cap A) = P(A)P(B^c|A) = 0.05 \times 0.01$$



Example (1.10)

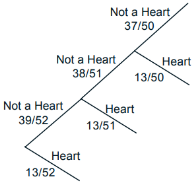
3 cards are drawn from a 52-card deck without replacement. What is the probability that not a Heart is drawn?

Define $A_i = \{\text{drawn card } i \text{ is not a Heart}\}$. From the definition, we have

$$P(A_1) = \frac{39}{52},\ P(A_2|A_1) = \frac{38}{51},\ P(A_3|A_1 \cap A_2) = \frac{37}{50}$$

Define $A = \{\text{not a Heart is drawn}\}$. Then $A = A_1 \cap A_2 \cap A_3$. By factorization

$$\begin{aligned} P(A) &= P(A_1 \cap A_2 \cap A_3) \\ &= P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \\ &= \frac{39}{52}\frac{38}{51}\frac{37}{50} \end{aligned}$$



Example (1.11)

Four graduate and 12 undergraduate students are divided into four groups of four students. What is the probability that every group includes a graduate student?

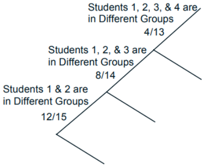
Define $A_i = \{\text{graduate students } 1, \ldots, i, i+1 \text{ in different groups}\}$. From the problem statement, we have

$$P(A_1) = \frac{12}{15},\ P(A_2|A_1) = \frac{8}{14},\ P(A_3|A_2) = \frac{4}{13}$$

Note $A_3 = \{\text{every group includes a graduate student}\}$.

Since $A_3 = A_1 \cap A_2 \cap A_3$, we have

$$\begin{aligned} P(A_3) &= P(A_1 \cap A_2 \cap A_3) \\ &= P(A_1)P(A_2|A_1)P(A_3|A_2 \cap A_1) \\ &= \frac{12}{15}\,\frac{8}{14}\,\frac{4}{13} \end{aligned}$$



Total Probability Theorem

Total probability through a partition

Let $(\Omega, \mathcal{F}, P)$ be a probability model, $A_1 \ldots A_n$ be events, and $(A_1, \ldots, A_n)$ be a partition of Ω. The for any event B

$$P(B) = \sum_{i=1}^{n} P(A_i) P(B|A_i)$$

$$B = \Omega \cap B = \left(\bigcup_{i=1}^{n} A_i \right) \cap B = \bigcup_{i=1}^{n} (A_i \cap B)$$

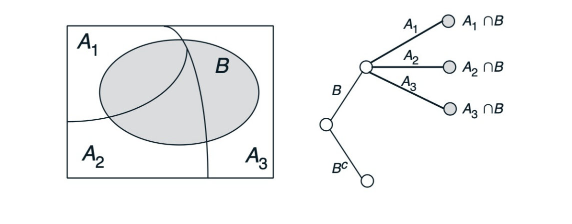
$$P(B) = P\left(\bigcup_{i=1}^{n} (A_i \cap B) \right) = \sum_{i=1}^{n} P(A_i \cap B)$$

Using factorization, we have

$$P(B) = \sum_{i=1}^{n} P(A_i) P(B|A_i)$$

How TPT works

- Use partition $(A_1, \ldots, A_n)$ of Ω such that $P(A_i)$ and $P(B|A_i)$ can be decided
- $(A_1 \cap B, \ldots, A_n \cap B)$ is a partition of B
- $(A_i \cap B)$ has probability $P(A_i)P(B|A_i)$



Example (Total probability)

You win if the sum of two rolls of a fair six-face dice is more than 9. What is the probability of winning?

Define $B = \{\text{you win}\}$ and $A_i = \{\text{first roll is } i\}$. Using partition $(A_1, \ldots, A_6)$, we have

$$\begin{aligned} P(B) &= \sum_{i=1}^{6} P(A_i \cap B) \\ &= \sum_{i=1}^{6} P(A_i) P(B|A_i) \\ &= \frac{1}{6} \cdot \frac{0}{6} + \frac{1}{6} \cdot \frac{0}{6} + \frac{1}{6} \cdot \frac{0}{6} + \frac{1}{6} \cdot \frac{1}{6} + \frac{1}{6} \cdot \frac{2}{6} + \frac{1}{6} \cdot \frac{3}{6} \\ &= \frac{6}{36} \end{aligned}$$

Example (1.13 Total probability)

In a chess tournament, your probability of winning a game is 0.3 against half the players, 0.4 against a quarter of the players, and 0.5 against the remaining players. You play a game against a randomly chosen opponent. What is the probability of winning?

Define $B = \{\text{you win}\}$ and $A_i = \{\text{an opponent of type } i \text{ is chosen}\}$. Using partition (A_1, A_2, A_3), we have

$$\begin{aligned} P(B) &= \sum_{i=1}^{3} P(A_i \cap B) \\ &= \sum_{i=1}^{3} P(A_i)P(B|A_i) \\ &= \frac{1}{2}(0.3) + \frac{1}{4}(0.4) + \frac{1}{4}(0.5) \end{aligned}$$

Example (1.14 Total probability)

You roll a fair four-face dice. If the result is 1 or 2, you roll once more, otherwise you stop. What is the probability that the total of the roll(s) is at least 4?

Define $B = \{\text{total is at least 4}\}$ and $A_i = \{\text{first roll is } i\}$. Using partition $(A_1, \ldots, A_4)$, we have

$$\begin{aligned} P(B) &= \sum_{i=1}^{4} P(A_i \cap B) \\ &= \sum_{i=1}^{4} P(A_i) P(B|A_i) \\ &= \frac{1}{4} \cdot \frac{2}{4} + \frac{1}{4} \cdot \frac{3}{4} + \frac{1}{4} \cdot 0 + \frac{1}{4} \cdot 1 \end{aligned}$$

Example (1.15 Total probability)

Alice is taking a course. At the end of a week, she is either up-to-date or behind. If she is up-to-date, she will be up-to-date the next week with probability 0.8. If she is behind, she will be behind the next week with probability 0.6. What is the probability that she is up-to-date after three weeks?

Define

$$U_i = \{\text{up-to-date at the end of week } i\}$$

$$B_i = \{\text{behind at the end of week } i\}$$

Note (U_i, B_i) is a partition of Ω for any i. Thus

$$\begin{aligned} P(U_{i+1}) &= P(U_{i+1}|U_i)\ P(U_i) + P(U_{i+1}|B_i)P(B_i) \\ &= 0.8\ P(U_i) + 0.4\ P(B_i) \\ P(B_{i+1}) &= P(B_{i+1}|U_i)\ P(U_i) + P(B_{i+1}|B_i)P(B_i) \\ &= 0.2\ P(U_i) + 0.6\ P(B_i) \end{aligned}$$

Total conditional probability

Let $(\Omega, \mathcal{F}, P)$ be a probability model, $(A_1, \ldots, A_n)$ be a partition of Ω. Then for any events B and C

$$P(B|C) = \sum_{i=1}^{n} P(A_i|C)P(B|A_i \cap C)$$

$$\begin{aligned}
P(B|C) &= P\left(\left(\cup_{i=1}^{n} A_i\right) \cap B|C\right) \\
&= P\left(\cup_{i=1}^{n} (A_i \cap B)\,|C\right) \\
&= \frac{P\left(\cup_{i=1}^{n} (A_i \cap B \cap C)\right)}{P(C)} \\
&= \sum_{i=1}^{n} \frac{P(A_i \cap B \cap C)}{P(C)} \\
&= \sum_{i=1}^{n} \frac{P(A_i \cap B \cap C)P(A_i \cap C)}{P(A_i \cap C)P(C)} \\
&= \sum_{i=1}^{n} P(A_i|C)P(B|A_i \cap C)
\end{aligned}$$

Bayes Rule

Bayes rule

Let $(\Omega, \mathcal{F}, P)$ be a probability model and A, B be events. Then $P(A)$ and $P(A|B)$ are related by

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

Applying the definition of conditional probability, we get

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

Let A be hypothesis and B be information (data). $P(A)$ is the prior probability of A, $P(A|B)$ is the posterior probability of A given B, $P(B|A)$ is the likelihood of B conditional on A. Bayes rule means

$$\text{posterior} \;\propto\; \text{prior} \times \text{likelihood}$$

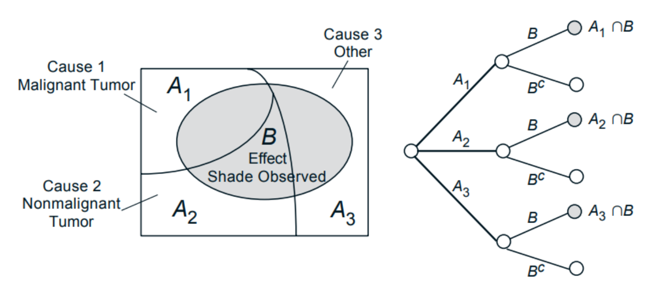
Bayes theorem

Let $(\Omega, \mathcal{F}, P)$ be a probability model and $(A_1, \dots, A_n)$ be a partition of Ω. Then for any event B

$$P(A_k|B) = \frac{P(A_k)P(B|A_k)}{\sum_{i=1}^{n} P(A_i)P(B|A_i)}, \quad k = 1, \dots, n$$

$$\begin{aligned} P(A_k|B) &= \frac{P(A_k \cap B)}{P(B)} \\ &= \frac{P(A_k \cap B)}{\sum_{i=1}^{n} P(A_i \cap B)} \\ &= \frac{P(A_k)P(B|A_k)}{\sum_{i=1}^{n} P(A_i)P(B|A_i)} \end{aligned}$$

Cause-effect scenario



- Event B denotes an *Effect* (Shade Observed)
- Event A_i denotes *Cause* i that could lead to the *Effect*
- Without observation, the prior probability of *Cause* i is $P(A_i)$
- After the *Effect* is observed (B occurs), the probability of *Cause* i is updated to $P(A_i|B)$ by Bayes rule

Example (1.12 Monty Hall problem)

You are told that a prize is equally likely to be placed behind any one of three closed doors in front of you. You first choose one of the doors. The host opens another door for you, after making sure that the prize is not behind it. At this point, you can stick with your choice, or switch to the other door. Will you switch?

Define $A = \{\text{Prize door is chosen}\}$ and $B = \{\text{Empty door is opened}\}$. By Bayes rule

$$P(A|B) = \frac{P(A)P(B|A)}{P(A)P(B|A) + P(A^c)P(B|A^c)} = \frac{\frac{1}{3}\cdot 1}{\frac{1}{3}\cdot 1 + \frac{2}{3}\cdot 1} = \frac{1}{3}$$

$$P(A^c|B) = \frac{P(A^c)P(B|A^c)}{P(A)P(B|A) + P(A^c)P(B|A^c)} = \frac{2}{3}$$

Believe it or not, the probability that the prize is behind the other door is **doubled** as soon as an empty door is open.

Example (1.16 Bayes)

Refer to Example 1.9. Given that an alarm has been generated by the radar, what is the probability that an aircraft is present?

Define $A = \{\text{aircraft is present}\}$ and $B = \{\text{radar generates alarm}\}$. By Bayes rule

$$P(A|B) = \frac{P(A)P(B|A)}{P(A)P(B|A) + P(A^c)P(B|A^c)}$$
$$= \frac{(0.05)(0.99)}{(0.05)(0.99) + (0.95)(0.1)}$$

Example (1.17 Bayes)

Refer to Example 1.13. Given that you have won a game, what is the probability that the opponent is of type 1?

Define $A_i = \{\text{play an opponent of type } i\}$ and $B = \{\text{you win}\}$. By Bayes rule

$$P(A_1|B) = \frac{P(A_1)P(B|A_1)}{\sum_{i=1}^{3} P(A_i)P(B|A_i)}$$

$$= \frac{\frac{1}{2}(0.3)}{\frac{1}{2}(0.3) + \frac{1}{4}(0.4) + \frac{1}{4}(0.5)}$$

Example (1.18 Bayes)

The test result of a certain rare disease is correct 95% of the time. A person has a probability of 0.001 of having the disease. Given that a person has just tested positive, what is the probability that the person actually has the disease?

Define $A = \{\text{has disease}\}$ and $B = \{\text{test positive}\}$. By Bayes rule

$$
\begin{aligned}
P(A|B) &= \frac{P(A)P(B|A)}{P(A)P(B|A) + P(A^c)P(B|A^c)} \\
&= \frac{(0.001)(0.95)}{(0.001)(0.95) + (0.999)(0.05)}
\end{aligned}
$$

Here the probability that a person has the disease is still below 5% after test positive, even the test result is 95% correct.

Example (Probability in bridge)

A newbie always leads a Spade. He knows not to lead an honor if he has a choice. Suppose the declarer has 10-card Spade suit missing K, 3, 2 and the newbie leads Spade 2 from West. What is the probability that East has Spade K? Spade 3? (3-0 split is 22%)

Define $B = \{\text{2 played}\}$, $A_1 = \{\text{West has 2}\}$, $A_2 = \{\text{West has 2, 3}\}$, and $A_3 = \{\text{West has 2, K}\}$, $A_4 = \{\text{West has 2, 3, K}\}$. By Bayes rule

$$P(A_k|B) = \frac{P(A_k)P(B|A_k)}{\sum_{i=1}^{4} P(A_i)P(B|A_i)}$$

Probability that East has Spade K

$$P(A_1 \cup A_2|B) = \frac{P(A_1)P(B|A_1) + P(A_2)P(B|A_2)}{\sum_{i=1}^{4} P(A_i)P(B|A_i)} = \frac{39}{76}$$

Probability that East has Spade 3

$$P(A_1 \cup A_3|B) = \frac{P(A_1)P(B|A_1) + P(A_3)P(B|A_3)}{\sum_{i=1}^{4} P(A_i)P(B|A_i)} = \frac{52}{76}$$

Independence

Definition (independent events)

Let $(\Omega, \mathcal{F}, P)$ be a probability model. Events $A, B \in \mathcal{F}$ are independent if

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

Note independence requires A, B to have non-zero probabilities.

Independence is denoted by

$$A \perp\!\!\!\perp B$$

Disjointness vs. Independence

Let $(\Omega, \mathcal{F}, P)$ be a probability model and $A, B \in \mathcal{F}$ have non-zero probabilities.

- $A \not\perp\!\!\!\perp B$ if they are disjoint
- $A \perp\!\!\!\perp B$ then they are not disjoint

$$A \cap B = \emptyset \ \Rightarrow \ 0 = P(A \cap B) \neq P(A)P(B) > 0 \ \Rightarrow \ A \not\perp\!\!\!\perp B$$

$$A \perp\!\!\!\perp B \ \Rightarrow \ P(A \cap B) = P(A)P(B) > 0 \ \Rightarrow \ A \cap B \neq \emptyset$$

Invariance of probability

Let $(\Omega, \mathcal{F}, P)$ be a probability model and $A \perp\!\!\!\perp B$. Then

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

By factorization

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

By independence

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

Hence

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

Similarly

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

Definition (conditional independence)

Let $(\Omega, \mathcal{F}, P)$ be a probability model and $A, B, C \in \mathcal{F}$ have non-zero probabilities. By definition, A and B are conditionally independent given C if

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

Conditional independence is denoted by

$$A \perp\!\!\!\perp B \mid C$$

Invariance of conditional probability

Let $(\Omega, \mathcal{F}, P)$ be a probability model and $A \perp\!\!\!\perp B \mid C$. Then

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

By factorization

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

By conditional independence

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

Hence

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

Similarly

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

Example (1.20 Independence $\not\Rightarrow$ CI)

For 2 tosses of a fair coin, define events $H_i = \{\text{toss } i \text{ is a head}\}$ and $D = \{\text{2 tosses have different results}\}$.

- $H_1 \perp\!\!\!\perp H_2$

$$P(H_1 \cap H_2) = \frac{1}{4} = \frac{1}{2}\,\frac{1}{2} = P(H_1)P(H_2)$$

- $H_1 \not\perp\!\!\!\perp H_2 \mid D$

$$P(H_1 \cap H_2 | D) = 0 \neq \frac{1}{2}\,\frac{1}{2} = P(H_1|D)P(H_2|D)$$

Example (1.21 CI $\not\Rightarrow$ independence)

The probability of head is 0.99 for a blue coin and 0.01 for a red coin. We choose one coin with equal probability, and toss the coin twice. Define $H_i = \{\text{toss } i \text{ is a head}\}$ and $B = \{\text{blue coin is chosen}\}$.

- $H_1 \perp\!\!\!\perp H_2 \mid B$

$$P(H_1 \cap H_2|B) = P(H_1|B)P(H_2|B)$$

- $H_1 \not\perp\!\!\!\perp H_2$

$$P(H_1) = P(H_1|B)P(B) + P(H_1|B^c)P(B^c) = \frac{1}{2} = P(H_2)$$

$$\begin{aligned} P(H_1 \cap H_2) &= P(H_1 \cap H_2 \cap B) + P(H_1 \cap H_2 \cap B^c) \\ &= P(H_1 \cap H_2|B)P(B) + P(H_1 \cap H_2|B^c)P(B^c) \\ &\neq P(H_1)P(H_2) \end{aligned}$$

Definition (joint independence)

Let $A_1, \ldots, A_n$ be events. By definition, they are **jointly independent** if and only if

$$P\left(\bigcap_{i \in S} A_i\right) = \prod_{i \in S} P(A_i)$$

holds for *every* $S \subset \{1, \ldots, n\}$.

In particular, joint independence implies joint factorization

$$P\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} P(A_i)$$

and pair-wise factorization

$$P(A_i \cap A_j) = P(A_i)P(A_j), \; i \neq j$$

Example (1.22 Pair-wise $\not\Rightarrow$ Joint factorization)

For 2 tosses of a fair coin, define $H_i = \{\text{toss } i \text{ is head}\}$ and $D = \{\text{2 tosses have different results}\}$. We have

$$P(H_1) = \frac{1}{2} = P(H_2) = P(D)$$

- Pair-wise factorization holds

$$P(H_1 \cap H_2) = \frac{1}{4} = \frac{1}{2}\frac{1}{2} = P(H_1)P(H_2)$$
$$P(H_1 \cap D) = \frac{1}{4} = \frac{1}{2}\frac{1}{2} = P(H_1)P(D)$$
$$P(H_2 \cap D) = \frac{1}{4} = \frac{1}{2}\frac{1}{2} = P(H_2)P(D)$$

- Joint factorization does not hold

$$P(H_1 \cap H_2 \cap D) = 0 \neq P(H_1)P(H_2)P(D)$$

Example (1.23 Joint $\nRightarrow$ Pair-wise factorization)

For 2 rolls of a fair dice, define $A = \{\text{first roll is 1, 2, or 3}\}$, $B = \{\text{first roll is 3, 4, or 5}\}$ and $C = \{\text{sum of the rolls is 9}\}$. We have

$$P(A) = \frac{1}{2},\ P(B) = \frac{1}{2},\ P(C) = \frac{1}{9}$$

- Joint factorization holds

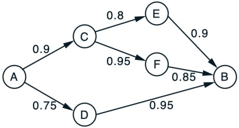
$$P(A \cap B \cap C) = \frac{1}{36} = \frac{1}{2}\frac{1}{2}\frac{1}{9} = P(A)P(B)P(C)$$

- Pair-wise factorization does not hold

$$P(A \cap B) = \frac{1}{6} \neq \frac{1}{2}\frac{1}{2} = P(A)P(B)$$
$$P(A \cap C) = \frac{1}{36} \neq \frac{1}{2}\frac{1}{9} = P(A)P(C)$$
$$P(B \cap C) = \frac{1}{12} \neq \frac{1}{2}\frac{1}{9} = P(B)P(C)$$

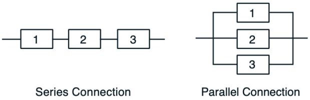
EXAMPLE (1.24)

A computer network connects two nodes A and B through nodes C, D, E, F. For every pair of directly connected nodes, say i and j, there is a given probability p_{ij} that the link from i to j is up. It is assumed that link failures are independent of each other. What is the probability that there is a path from A to B without link failures?



(a)

For the connectivity between A and B, the network is equivalent to a network with just one link.



(b)

- Two links in series can be replaced by an equivalent link. Since the equivalent link is up if and only if both links are up, the probabilities of being up are related by

$$r = pq$$

- Two links in parallel can be replaced by an equivalent link. Since the equivalent link is down if and only if both links are down, the probabilities of being up are related by

$$1 - r = (1 - p)(1 - q)$$

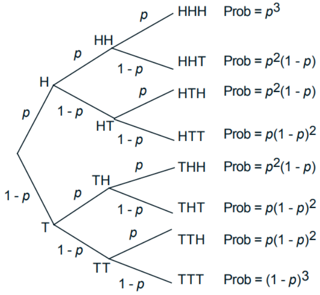
  That is

$$r = p + q - pq$$

Independent trials

A random experiment may consist of a sequence of independent stages called independent trials.

A trial of two possible outcomes is a Bernoulli trial. The following diagram is an experiment consisting of a sequence of 3 independent and identical Bernoulli trials with outcomes $\{H, T\}$.



Example (Binomial)

Consider an experiment that consists of n tosses of a coin with head probability p. What is the probability of event $B_k = \{k \text{ heads}\}$?

B_k consists of the length-n sequences with k heads. Every sequence in B_k has probability $p^k(1-p)^{n-k}$. Denote the number of sequences in B_k by $\binom{n}{k}$. We have

$$P(B_k) = \binom{n}{k} p^k (1-p)^{n-k}$$

We will decide $\binom{n}{k}$ by the counting principles shortly.

Example (1.25)

An internet service provider has installed c modems to serve n dialup customers. It is estimated that at a given time, each customer needs a connection with probability p, independent of the other customers. What is the probability that there are more customers needing a connection than there are modems?

Let $B_k = \{k \text{ customers need connection}\}$. Then

$$A = \bigcup_{k=c+1}^{n} B_k$$

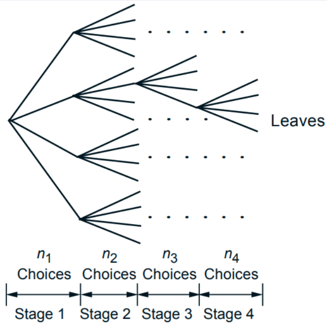
and

$$P(A) = \sum_{k=c+1}^{n} P(B_k) = 1 - \sum_{k=0}^{c} P(B_k)$$

Counting

Counting principles

- Suppose there are 2 stages to complete a task, m ways for stage 1 and n ways for stage 2. Then the number of ways to complete the task is $m \cdot n$.
- Suppose there are 2 options to complete a task, m ways for option 1 and n ways for option 2. Then the number of ways to complete the task is $m + n$.



Example (1.26)

A local telephone number is a 7-digit sequence, but the first digit cannot be 0 or 1. How many possible telephone numbers are there?

The task of forming a local telephone number can be completed in 7 stages. In stage i, a digit is selected for position i. By the counting principle, the number of possible telephone numbers is

$$N = (8)(10)\cdots(10) = 8 \times 10^6$$

Example (1.27)

Let S be a set with n members

$$S = \{s_1, \ldots, s_n\}$$

In how many possible ways can we form a subset of S?

The task of forming a subset of S can be completed in n stages. In stage i, the decision to include s_i or not is made. By the counting principle, the number of possible subsets is

$$N = (2) \cdots (2) = 2^n$$

Permutation

The number of ways we can arrange a sequence of k objects selected from n distinct objects is called (n, k)-permutation.

A sequence of k objects has k positions. An (n, k)-permutation can be completed in k stages. In stage i, a not-yet-selected object for position i is selected. By the counting principle, the number of arrangements in (n, k)-permutation is

$$p_k^n = (n)(n-1)\cdots(n-k+1) = \frac{n!}{(n-k)!}$$

Example (1.28)

How many possible 4-letter words are there without any repeated letter?

A 4-letter word has 4 positions. The task of forming a 4-letter word without any repeated letter can be completed in 4 stages, each selecting a new letter for the next position. This is an (n, k)-permutation with $n = 26$ and $k = 4$. Hence

$$N = p_4^{26} = \frac{26!}{22!}$$

Example (1.29)

You have n_1 classical music CDs, n_2 rock music CDs, and n_3 country music CDs. In how many possible ways can you arrange them so that the CDs of the same type are contiguous?

The overall task can be completed in 4 stages. The first 3 stages arrange classical music CDs, rock music CDs, and country music CDs. For music type i, there are (n_i, n_i)-permutation with $p_{n_i}^{n_i} = n_i!$ ways. Stage 4 is to arrange the order of the music types, which is an $(3, 3)$-permutation with $p_3^3 = 3!$ ways. Hence

$$N = n_1!n_2!n_3!3!$$

Combination

The number of ways we can select k objects from n distinct objects is called (n, k)-combination.

Consider the (n, k)-permutation task again. It can be completed in 2 stages: an (n, k)-combination followed by a (k, k)-permutation. Denote the number of ways in (n, k)-combination by $\binom{n}{k}$. By counting principle

$$p_k^n = \frac{n!}{(n-k)!} = \binom{n}{k} k!$$

Hence

$$\binom{n}{k} = \frac{n!}{(n-k)!k!}$$

EXAMPLE (1.31)

The members of a club are selected from a group of n persons. It must have a leader and it may have additional members. In how many possible ways can such a club be formed?

- There are n options for the number of club members, 1 to n. For each option, the task of forming a club can be completed in 2 stages: select the members and then select a leader among them. By counting principle

$$N = \binom{n}{1} 1 + \cdots + \binom{n}{n} n = \sum_{k=1}^{n} \binom{n}{k} k$$

- Alternatively, the task of forming a club can also be completed in 2 stages: select a leader and then select additional members. We have

$$N = n\, 2^{n-1}$$

Partition

The number of ways we can partition n distinct objects into r distinct groups of sizes $n_1, \ldots, n_r$ is called $(n, r, n_{1:r})$-partition.

$(n, r, n_{1:r})$-partition can be completed in r stages. Stage i is $(n - (n_1 + \cdots + n_{i-1}), n_i)$-combination. Denote the number of ways in $(n, r, n_{1:r})$-partition by $\binom{n}{n_1, \ldots, n_r}$. By counting principle

$$\begin{aligned}\binom{n}{n_1, \ldots, n_r} &= \binom{n}{n_1}\binom{n - n_1}{n_2}\cdots\binom{n - n_1 - \cdots - n_{r-1}}{n_r} \\ &= \frac{n!}{n_1! n_2! \ldots n_r!}\end{aligned}$$

Note (n, k)-combination is a special case of $(n, r, n_{1:r})$-partition with

$$r = 2,\ n_1 = k,\ n_2 = n - k$$

Example (1.32)

In how many ways can we form a word by rearranging the letters in the word TATTOO?

A re-arrangement of the letters corresponds to allocating 6 positions into 3 groups of sizes $1, 2, 3$ for letters A, O, T respectively. Therefore, it is $(6, 3, (1, 2, 3))$-partition. The number of ways is

$$N = \binom{6}{1, 2, 3} = \frac{6!}{1!\,2!\,3!}$$

Example (1.33)

Four graduate students and 12 undergraduate students are divided into four groups of four students. What is the probability that every group includes a graduate student?

The task of forming 4 groups of 4 students is a partition with

$$N_0 = \binom{16}{4, 4, 4, 4} = \frac{16!}{4!\,4!\,4!\,4!}$$

The task of forming 4 groups of 4 students and each group includes a graduate student can be completed in 2 stages: partition of 4 graduate students followed by partition of 12 undergraduate students. So

$$N = \binom{4}{1, 1, 1, 1} \binom{12}{3, 3, 3, 3} = \frac{4!\,12!}{3!\,3!\,3!\,3!}$$

The probability is $\frac{N}{N_0}$.

Summary 1

Basic set operations: intersection, union, complementation

Probability model

$$(\Omega, \mathcal{F}, P)$$

Conditional probability

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

Independence

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

Probability of set union (inclusion-exclusion principle)

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{i<j}^{n} P(A_i \cap A_j) + \dots + (-1)^{n+1} P(\bigcap_{i=1}^{n} A_i)$$

Factorization (probability of joint event)

$$P\left(\bigcap_{i=1}^{n} A_i\right) = P(A_1) \prod_{k=2}^{n} P\left(A_k | A_1 \cap \cdots \cap A_{k-1}\right)$$

Total probability theorem

$$P(B) = \sum_i P(B \cap A_i) = \sum_i P(A_i)P(B|A_i)$$

Bayes rule

$$P(A_k|B) = \frac{P(A_k)P(B|A_k)}{\sum_{i=1}^{n} P(A_i)P(B|A_i)}$$

Counting

$$N = n_1 \times \cdots \times n_S, \; N = n_1 + \cdots + n_C$$