확률과 통계

Class 1

Chapter 2: Probability

2.1 Events and their probabilities

Probability

- The **probability** of an event is understood as a **chance** that this event will happen.

Example 2.1

- If a fair coin is tossed,
  we say that it has a 50-50 (equal) chance of turning up heads or tails.
- Hence, the probability of each side equals 1/2.
- It does not mean
  that a coin tossed 10 times will always produce exactly 5 heads and 5 tails.
- If you don't believe, try it!
- However, if you toss a coin 1 million times,
  the proportion of heads is anticipated to be very close to 1/2.

Example 2.2

- If there are 5 communication channels in service,
  and a channel is selected at random
  when a telephone call is placed,
  then each channel has a probability $1/5 = 0.2$ of being selected.

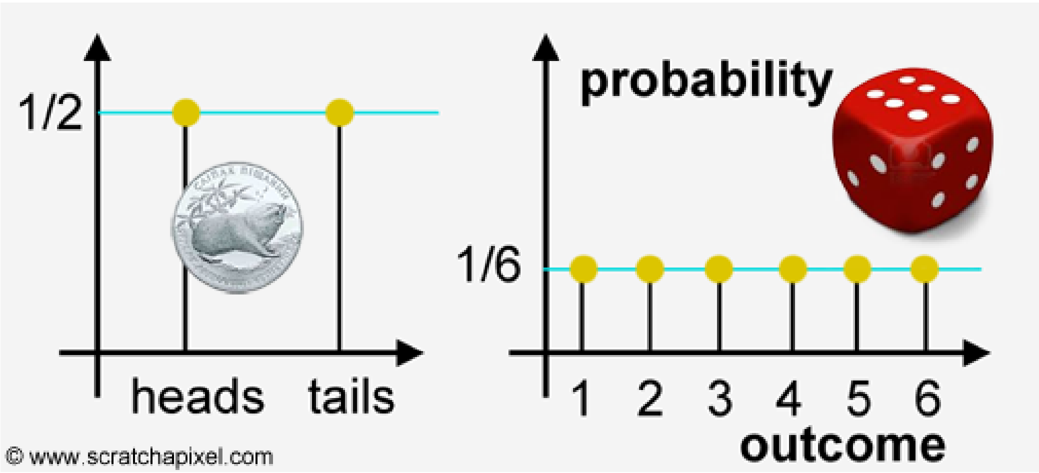
Example 2.3

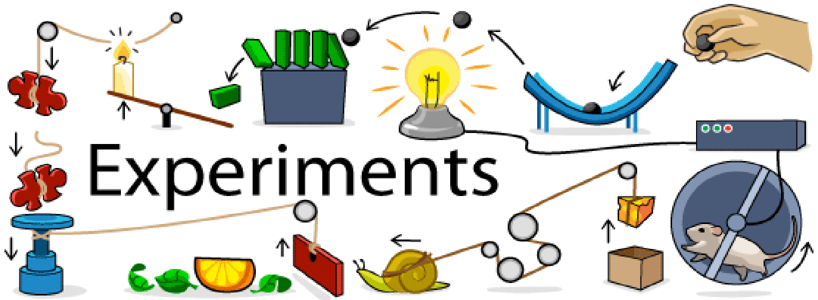
- Two competing software companies are after an important contract.
- Company A is twice as likely to win this competition as company B.
- Hence, the probability to win the contract equals 2/3 for A and 1/3 for B.

2.1.1 Outcomes, events, and the sample space

- Definition 2.1:
  - A collection of all elementary results, or **outcomes** of an experiment, is called a **sample space**.
- Definition 2.2:
  - Any set of outcomes is an **event**.
  - Thus, events are subsets of the sample space.

Outcomes and experiments





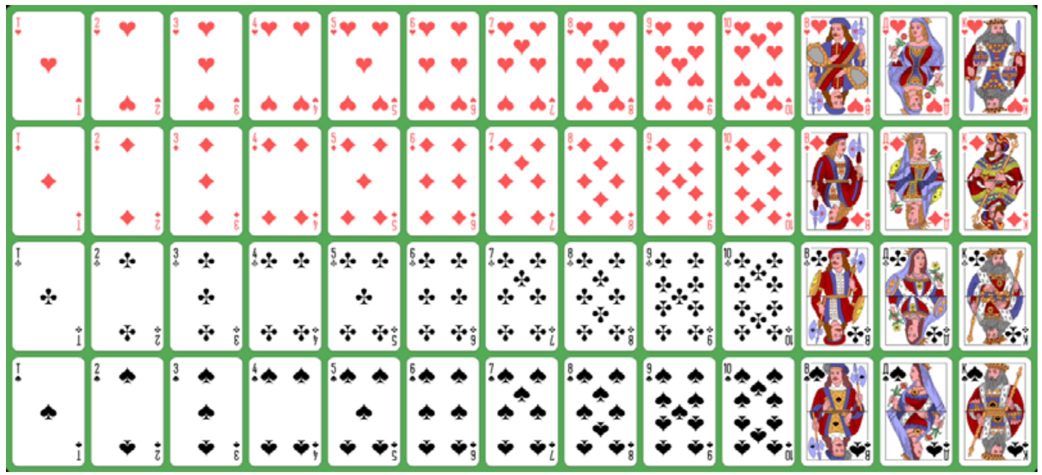
https://askabiologist.asu.edu/activities/experiments

- https://www.youtube.com/watch?v=M0I-xm7iCBU

Sample space

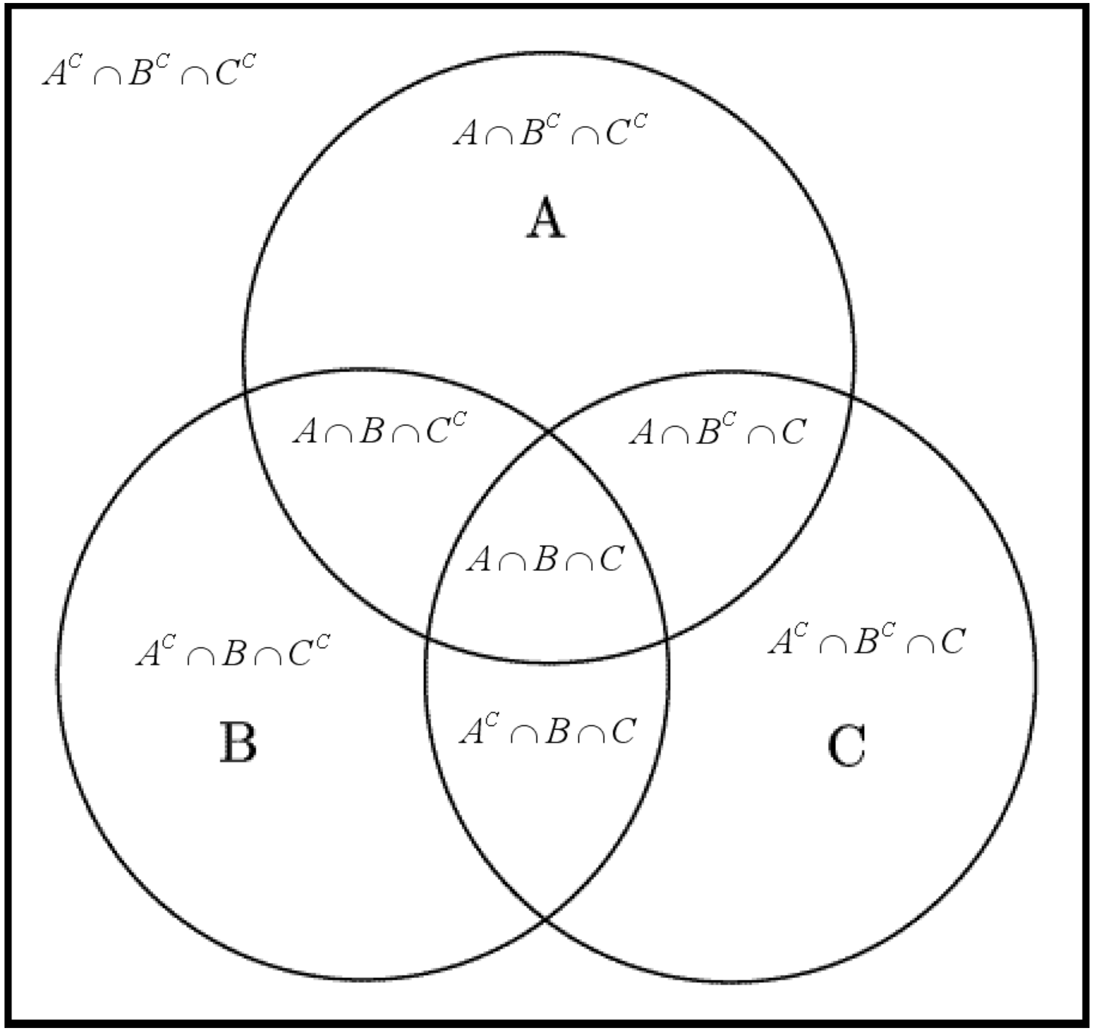
- Sample spaces are sets
  whose elements describe the outcomes of the experiments.

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



Events

- Events: subsets of the sample space



Example 2.3

- A tossed die can produce one of 6 possible outcomes: 1 dot through 6 dots.
- Each outcome is an event.
- There are other events:
  - observing an even number of dots,
  - an odd number of dots,
  - a number of dots less than 3,
  - etc.

- A sample space of N possible outcomes yields 2^N possible events.

- $$\overbrace{2 \cdot 2 \cdot \ldots \cdot 2}^{N \text{ times}} = 2^N \tag{2.1}$$

Example 2.5

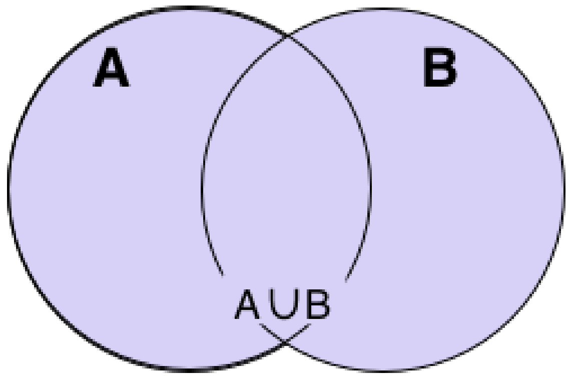
- Consider a football game between the Dallas Cowboys and the New York Giants.
- The sample space consists of 3 outcomes, $\Omega = \{\text{Cowboys win, Giants win, they tie}\}$
- Combining these outcomes in all possible ways, we obtain the following $2^3 = 8$ events:
    - Cowboys win, lose, tie, get at least a tie, get at most a tie, no tie, get some result, and get no result.
- The event “some result” is the entire sample space, and by common sense, it should have probability 1.
- The event “no result” is empty, it does not contain any outcomes, so its probability is 0.

Notation

- Ω = sample space
- $\varnothing$ = empty event
- $\boldsymbol{P}\{E\}$ = probability of event E

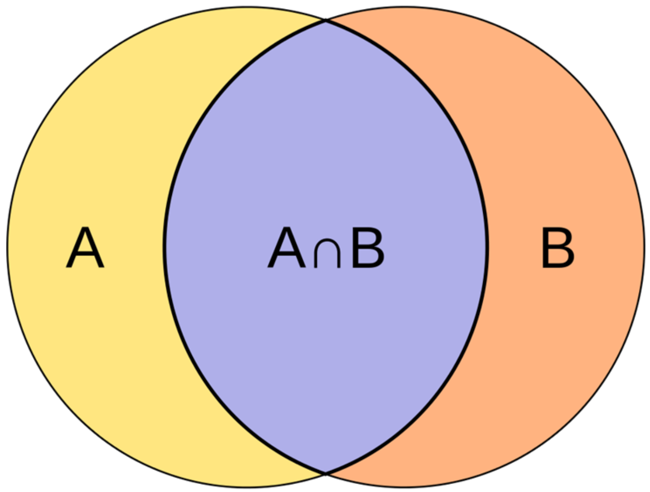
2.1.2 Set operations

- Def. 2.3:
  - A **union** of events A, B, C, . . . is an event consisting of all the outcomes in all these events.
  - It occurs if any of A, B, C, . . . occurs, and therefore, corresponds to the word "or": A or B or C or ...



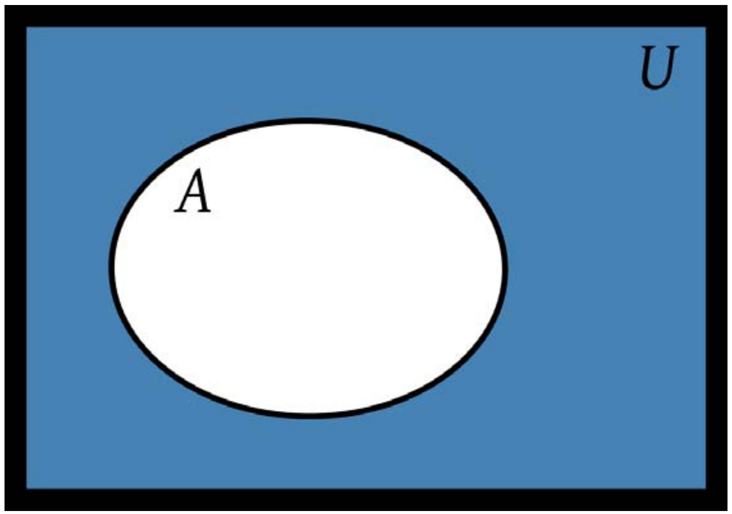
Definition 2.4

- An **intersection** of events A, B, C, . . .
  is an event consisting of outcomes that are common in all these events.
- It occurs if each A, B, C, . . . occurs,
  and therefore, corresponds to the word “and”:
  A and B and C and ...



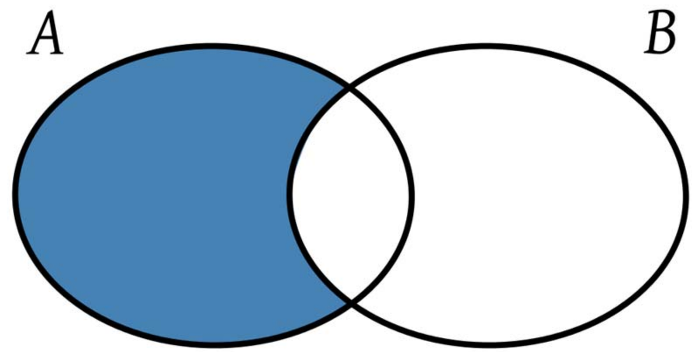
Definition 2.5

- A **complement** of an event A is an event that occurs every time when A does not occur.
- It consists of outcomes excluded from A, and therefore, corresponds to the word "not": not A



Definition 2.6

- A **difference** of events *A* and *B* consists of all outcomes included in *A* but excluded from *B*.
- It occurs when *A* occurs and *B* does not,
  and corresponds to "BUT NOT": *A* but not *B*



Notation

- $A \cup B$ = union
- $A \cap B$ = intersection
- $\bar{A}$ or Ac = complement
- $A \setminus B$ = difference

Definition 2.7

- Events A and B are **disjoint** if their intersection is empty,

$$A \cap B = \emptyset.$$

- Events $A_1, A_2, A_3, \ldots$ are **mutually exclusive** or **pairwise disjoint** if any two of these events are disjoint, i.e.,

$$A_i \cap A_j = \emptyset \text{ for any } i \neq j .$$

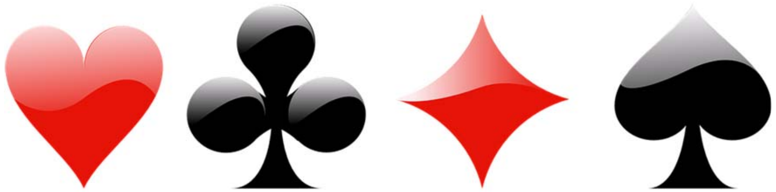
Definition 2.8

- Events A, B, C, . . . are **exhaustive** if their union equals the whole sample space, i.e.,

$$A \cup B \cup C \cup \ldots = \Omega.$$

Example 2.6

- When a card is pooled from a deck at random,
  the four suits are at the same time disjoint and exhaustive.



Example 2.7

- Any event A and its complement A represent a classical example of disjoint and exhaustive events.

Example 2.8

- Receiving a grade of A, B, or C for some course are mutually exclusive events,
  but unfortunately, they are not exhaustive.

Equation 2.2

- $\overline{E_1 \cup \cdots \cup E_n} = \overline{E_1} \cap \cdots \cap \overline{E_n}$
- $\overline{E_1 \cap \cdots \cap E_n} = \overline{E_1} \cup \cdots \cup \overline{E_n}$

Example 2.9

- Graduating with a GPA of 4.0 is an intersection of getting an A in *each* course.
- Its complement, graduating with a GPA below 4.0, is a union of receiving a grade below A *at least in one* course.

2.2 Rules of Probability

2.2.1 Axioms of Probability

- Axiom (公理) = "statement of self-evident truth"

- Def. 2.9:

  A collection M of events is a sigma-algebra on sample space if

  (a) it includes the sample space,

  $$\Omega \in M$$

  (b) every event in M is contained along with its complement; that is,

  $$E \in M \Rightarrow \bar{E} \in M$$

  (c) every finite or countable collection of events in M is contained along with its union; that is,

  $$E_1, E_2, \ldots \in M \Rightarrow E_1 \cup E_2 \cup \ldots \in M.$$

Example 2.10

- By conditions **(a)** and **(b)** in **Definition 2.9**, every sigma-algebra has to contain the sample space and the empty event $\emptyset$.
- This minimal collection

$$M = \{\Omega, \emptyset\}$$

forms a sigma-algebra that is called *degenerate*.

Example 2.11

- (Power set). On the other extreme,
  what is the richest sigma-algebra on a sample space Ω?
- It is the collection of all the events,

$$M = 2^{\Omega} = \{E, E \subset \Omega\} \ .$$

- As we know from **(2.1)**,
  there are 2^N events on a sample space of N outcomes.
- This explains the notation 2^{Ω}.
- This sigma-algebra is called a *power* set.

- $\overbrace{2 \cdot 2 \cdot \ldots \cdot 2}^{N \text{ times}} = 2^N$

Definition 2.10

- Assume a sample space Ω and a sigma-algebra of events M on it.
- **Probability**
  $$\boldsymbol{P} : M \longrightarrow [0, 1]$$
  is a function of events with the domain M and the range $[0, 1]$ that satisfies the following two conditions,
  - (Unit measure) The sample space has unit probability, $\boldsymbol{P}(\Omega) = 1$.
  - (Sigma-additivity) For any finite or countable collection of mutually exclusive events $E_1, E_2, \ldots \in M$,

$$\boldsymbol{P}\left\{E_1 \cup E_2 \cup \ldots\right\} = \boldsymbol{P}(E_1) + \boldsymbol{P}(E_2) + \ldots.$$

2.2.2 Computing probabilities of events

- **Extreme cases**
  - A sample space Ω consists of all possible outcomes,
    therefore, it occurs for sure.
  - On the contrary,
    an empty event $\emptyset$ never occurs.
  - So,

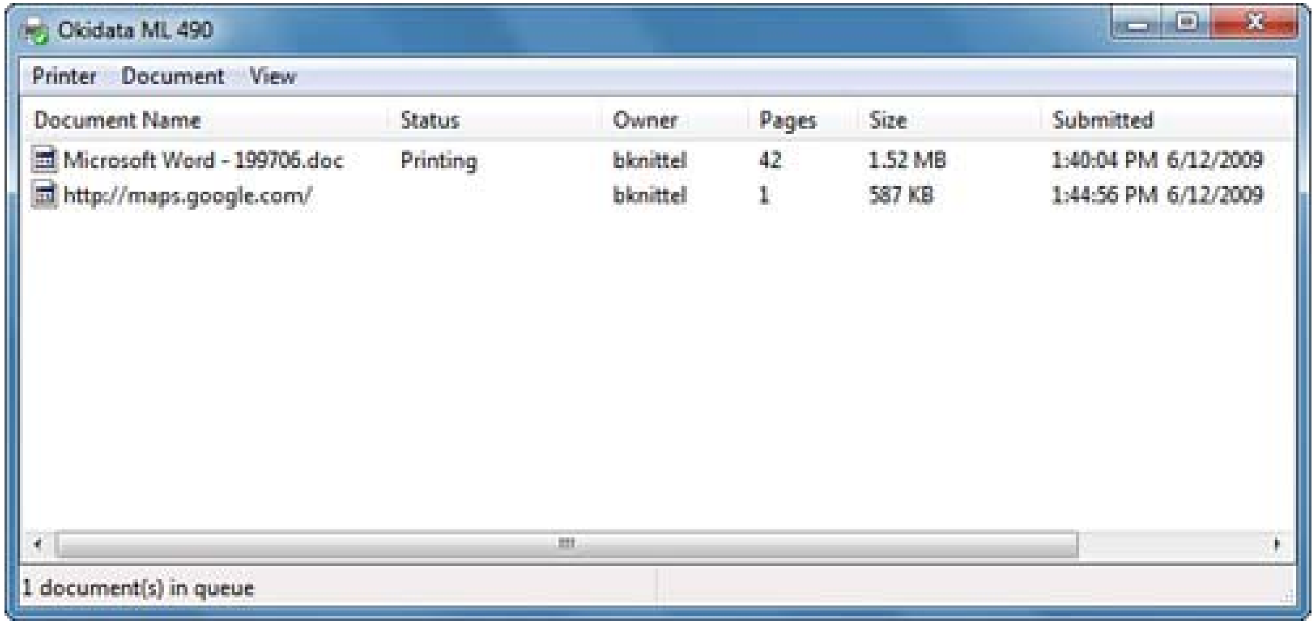
$$\boldsymbol{P}\{\Omega\} = 1 \text{ and } \boldsymbol{P}\{\emptyset\} = 0. \quad (2.3)$$

- **Union**
  - Consider an event that consists of some finite or countable collection of *mutually exclusive* outcomes,
    $E = \{\omega_1, \omega_2, \omega_3, \ldots\}$.
  - Summing probabilities of these outcomes, we obtain the probability of the entire event,

$$\boldsymbol{P}\{E\} = \sum_{\omega_k \in E} \boldsymbol{P}\{\omega_k\}$$
$$= \boldsymbol{P}\{\omega_1\} + \boldsymbol{P}\{\omega_2\} + \boldsymbol{P}\{\omega_3\} + \cdots$$

Example 2.13

- If a job sent to a printer appears first in line with probability 60%, and second in line with probability 30%, then with probability 90% it appears either first or second in line.



Example 2.14

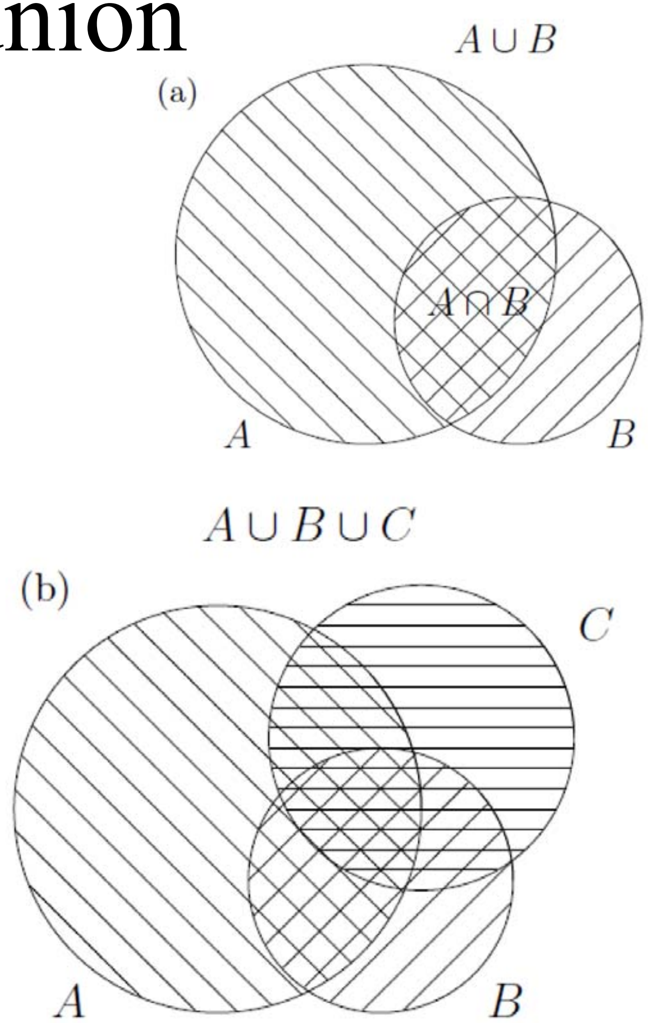
- During some construction,
  a network blackout occurs on Monday with probability 0.7
  and on Tuesday with probability 0.5.
- Then, does it appear on Monday or Tuesday
  with probability $0.7 + 0.5 = 1.2$?
- Obviously not,
  because probability should always be between 0 and 1!
- Probabilities are not additive here
  because blackouts on Monday and Tuesday are not mutually exclusive.
- In other words,
  it is not impossible to see blackouts on both days.

Equation 2.4: Probability of union

- $\boldsymbol{P}\{A \cup B\} = \boldsymbol{P}\{A\} + \boldsymbol{P}\{B\} - \boldsymbol{P}\{A \cap B\}$

- For mutually exclusive events,
  $$\boldsymbol{P}\{A \cup B\} = \boldsymbol{P}\{A\} + \boldsymbol{P}\{B\}$$

- $\boldsymbol{P}\{A \cup B \cup C\} = \boldsymbol{P}\{A\} + \boldsymbol{P}\{B\} + \boldsymbol{P}\{C\} - \boldsymbol{P}\{A \cap B\} - \boldsymbol{P}\{A \cap C\} - \boldsymbol{P}\{B \cap C\} + \boldsymbol{P}\{A \cap B \cap C\}$



Example 2.15

- In **Example 2.14**, suppose there is a probability 0.35 of experiencing network blackouts on both Monday and Tuesday.
- Then the probability of having a blackout on Monday or Tuesday equals

$$0.7 + 0.5 - 0.35 = 0.85.$$

Complement

- Recall that events A and $\bar{A}$ are exhaustive, hence $A \cup \bar{A} = \Omega$.
- Also, they are disjoint, hence
$$\boldsymbol{P}\{A\} + \boldsymbol{P}\{\bar{A}\} = \boldsymbol{P}\{A \cup \bar{A}\} = \boldsymbol{P}\{\Omega\} = 1$$
- Solving this for $\boldsymbol{P}\{\bar{A}\}$, we obtain a rule that perfectly agrees with the common sense,
$$\boldsymbol{P}\{\bar{A}\} = 1 - \boldsymbol{P}\{A\}$$

Example 2.16

- If a system appears protected against a new computer virus with probability 0.7, then it is exposed to it with probability $1 - 0.7 = 0.3$.

Example 2.17

- Suppose a computer code has no errors with probability 0.45.
- Then, it has at least one error with probability 0.55.

Intersection of independent events

- Definition 2.11
  - Events $E_1, \ldots, E_n$ are independent
    if they occur independently of each other,
    i.e., occurrence of one event does not affect the probabilities of others.

Independent events

$$\boxed{\boldsymbol{P}\{E_1 \cap \ldots \cap E_n\} = \boldsymbol{P}\{E_1\} \cdot \ldots \cdot \boldsymbol{P}\{E_n\}}$$

2.2.3 Applications in reliability

Example 2.18 (Reliability of backups).

- There is a 1% probability for a hard drive to crash.
- Therefore, it has two backups,
  each having a 2% probability to crash,
  and all three components are independent of each other.
- The stored information is lost only in an unfortunate situation when all three devices crash.
- What is the probability that the information is saved?
- Solution.
  - Organize the data.
  - Denote the events, say,
    $H = \{\text{hard drive crashes}\}$,
    $B_1 = \{\text{first backup crashes}\}$,
    $B_2 = \{\text{second backup crashes}\}$.
- It is given that H, B_1, and B_2 are independent,
  $$P\{H\} = 0.01, \text{ and } P\{B_1\} = P\{B_2\} = 0.02.$$
- Applying rules for the complement and for the intersection of independent events,
  $$\begin{aligned} P\{\text{saved}\} &= 1 - P\{\text{lost}\} = 1 - P\{H \cap B_1 \cap B_2\} \\ &= 1 - P\{H\}P\{B_1\}P\{B_2\} \\ &= 1 - (0.01)(0.02)(0.02) = 0.999996. \end{aligned}$$
- This is precisely the reason of having backups, isn't it?
- Without backups,
  the probability for information to be saved is only 0.99.)

Example 2.19

- Suppose that a shuttle's launch depends on three key devices that operate independently of each other and malfunction with probabilities 0.01, 0.02, and 0.02, respectively.
- If any of the key devices malfunctions, the launch will be postponed.
- Compute the probability for the shuttle to be launched on time, according to its schedule.

- Solution.
  - In this case,
    $P\{\text{on time}\} = P\ \{\text{all devices function}\}$
    $= P\{\overline{H} \cap \overline{B_1} \cap \overline{B_2}\}$
    $= P\{\overline{H}\}P\{\overline{B_1}\}P\{\overline{B_2}\}$ (independence)
    $= (1 - 0.01)(1 - 0.02)(1 - 0.02)$
    (complement rule)
    $= 0.9508.$
- Notice how with the same probabilities of individual components as in **Example 2.18**, the system's reliability decreased because the components were connected sequentially.

Example 2.20 (Techniques for solving reliability problems)

- Calculate reliability of the system in **Figure 2.3** if each component is operable with probability 0.92 independently of the other components.
- Solution.
  - This problem can be simplified and solved "step by step."

  1. The upper link A-B works if both A and B work, which has probability $P\{A \cap B\} = (0.92)2 = 0.8464$. We can represent this link as one component F that operates with probability 0.8464.
  2. By the same token, components D and E, connected in parallel, can be replaced by component G, operable with probability $P\{D \cup E\} = 1 - (1 - 0.92)2 = 0.9936$, as shown in **Figure 2.4a**.

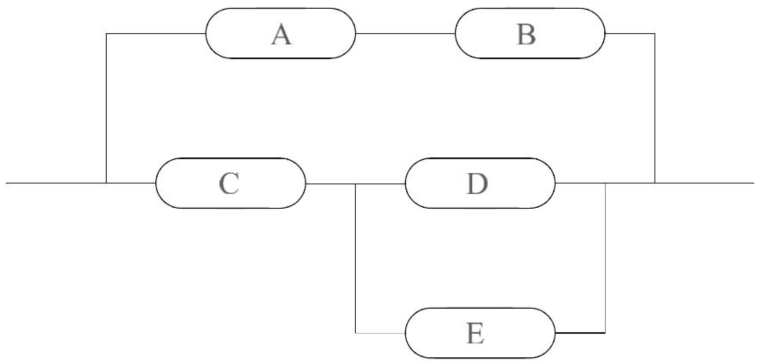
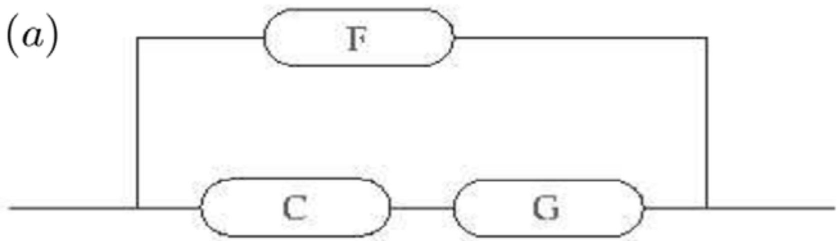


FIGURE 2.3: *Calculate reliability of this system (Example 2.20).*

(a)



3. Components C and G, connected sequentially,
   can be replaced by component H, operable with probability
   $P\{C \cap G\} = 0.92 \cdot 0.9936 = 0.9141,$
   as shown in **Figure 2.4b**.

4. Last step.
   The system operates with probability
   $P\{F \cup H\} = 1 - (1 - 0.8464)(1 - 0.9141) = 0.9868,$
   which is the final answer.

- In fact, the event "the system is operable" can be represented as $(A \cap B) \cup \{C \cap (D \cup E)\}$, whose probability we found step by step.

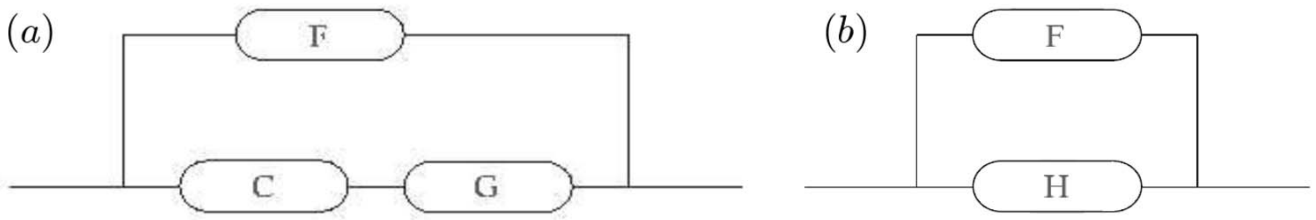
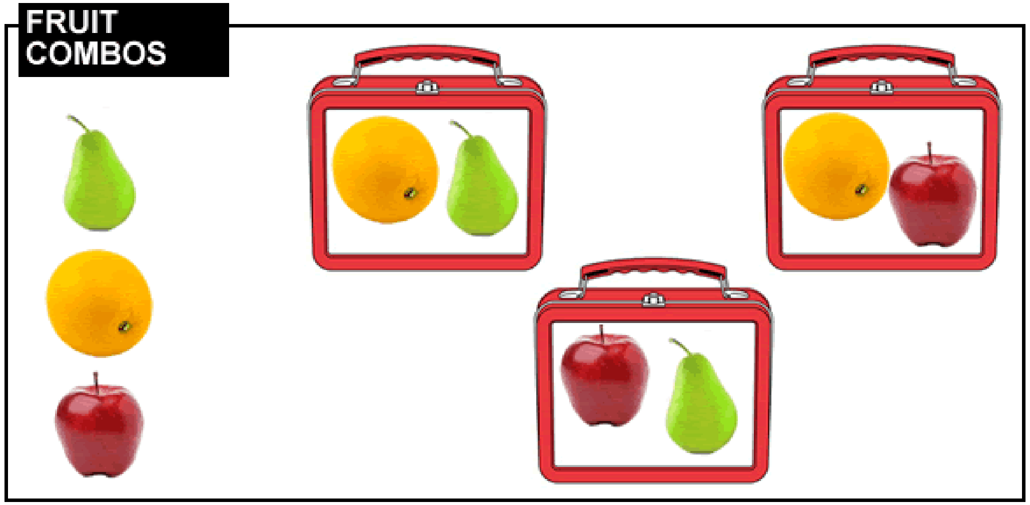


FIGURE 2.4: *Step by step solution of a system reliability problem.*



2.3 Combinatorics

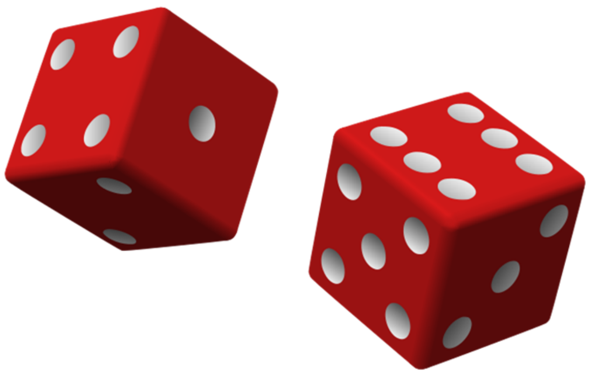
2.3.1 Equally likely outcomes

- A simple situation for computing probabilities is the case of *equally likely outcomes*.
- That is,
  when the sample space Ω consists of
  n possible outcomes, $\omega_1, \dots, \omega_n$,
  each having the same probability.
- Since
  $$\sum_1^n P\{w_k\} = P\{\Omega\} = 1$$
  we have in this case $P\{w_k\} = 1/n$ for all k.
- Further,
  a probability of any event E consisting of t outcomes, equals
  $$P\{E\} = \sum_{\omega_k \in E} \left(\frac{1}{n}\right) = t\left(\frac{1}{n}\right)$$
  $$= \frac{\text{number of outcomes in } E}{\text{number of outcomes in}}.$$
- The outcomes forming event E are often called "favorable."
- Thus we have a formula:
  - Equally likely outcomes
    $$P\{E\} = \frac{\text{number of favorable outcomes}}{\text{total number of outcomes}} = \frac{N_F}{N_T} \quad (2.5)$$
    where index "F" means "favorable"
    and "T" means "total."

Example 2.21

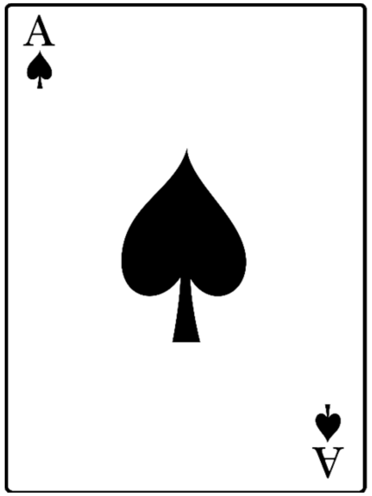
- Tossing a die results in 6 equally likely possible outcomes, identified by the number of dots from 1 to 6.
- Applying **(2.5)**, we obtain,

$$P\{1\} = 1/6,\ P\{\text{odd number of dots}\} = 3/6,\ P\{\text{less than } 5\} = 4/6.$$



Example 2.22

- A card is drawn from a bridge 52-card deck at random.
- Compute the probability that the selected card is a spade.
- First solution.
  - The sample space consists of 52 equally likely outcomes—cards.
  - Among them, there are 13 favorable outcomes—spades.
  - Hence, $P\{\text{spade}\} = 13/52 = 1/4$.
- Second solution.
  - The sample space consists of 4 equally likely outcomes —suits: clubs, diamonds, hearts, and spades.
  - Among them, one outcome is favorable—spades.
  - Hence, $P\{\text{spade}\} = 1/4$.



Example 2.23

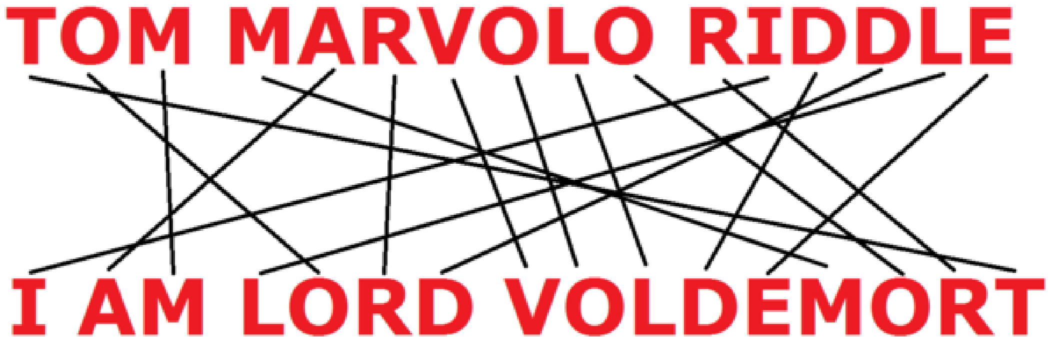
- A young family plans to have two children.
- What is the probability of two girls?
- Solution 1 (wrong).
  - There are 3 possible families with 2 children: two girls, two boys, and one of each gender.
  - Therefore, the probability of two girls is 1/3.
- Solution 2 (right).
  - Each child is (supposedly) equally likely to be a boy or a girl.
  - Genders of the two children are (supposedly) independent.
  - Therefore, $P\{\text{two girls}\} = (1/2)(1/2) = 1/4$.

0	0
0	1
1	0
1	1

Example 2.24 (Paradox)

- There is a simple but controversial "situation" about a family with two children.
- A family has two children.
- You met one of them, Leo, and he is a boy.
- What is the probability that the other child is also a boy?
- On one hand, why would the other child's gender be affected by Leo?
- Leo should have a brother or a sister with probabilities 1/2 and 1/2.
- On the other hand, see **Example 2.23**.
- The sample space consists of 4 equally likely outcomes, $\{GG,BB,BG,GB\}$.
- You have already met one boy, thus the first outcome is automatically eliminated: $\{BB,BG,GB\}$.
- Among the remaining three outcomes, Jimmy has a brother in one case and a sister in two cases.
- Thus, isn't the probability of a boy equal 1/3?
- Where is the catch?
- Apparently, the sample space has not been clearly defined in this example.
- The experiment is more complex than in **Example 2.23** because we are now concerned not only about the gender of children but also about meeting one of them.
- What is the mechanism, what are the probabilities for you to meet one or the other child?
- And once you met Leo, do the outcomes $\{BB,BG,GB\}$ remain equally likely?

2.3.2 Permutations and combinations



DEFINITION 2.12

Sampling **with replacement** means that every sampled item is replaced into the initial set, so that any of the objects can be selected with probability $1/n$ at any time. In particular, the same object may be sampled more than once.

DEFINITION 2.13

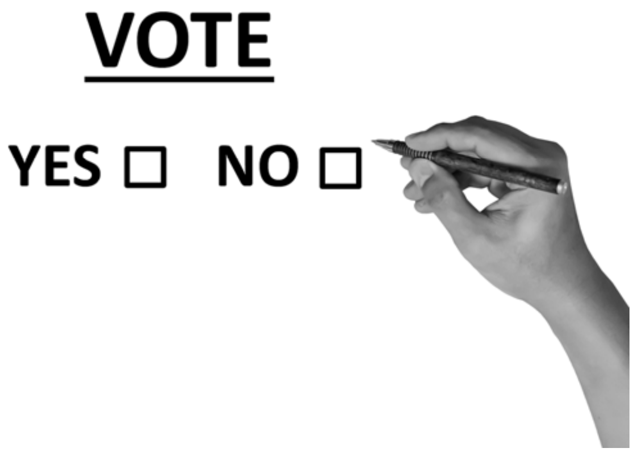
Sampling **without replacement** means that every sampled item is removed from further sampling, so the set of possibilities reduces by 1 after each selection.

DEFINITION 2.14

Objects are **distinguishable** if sampling of exactly the same objects *in a different order* yields a different outcome, that is, a different element of the sample space. For **indistinguishable** objects, the order is not important, it only matters which objects are sampled and which ones aren't. Indistinguishable objects arranged in a different order do not generate a new outcome.

Example 2.25 (Computer-generated passwords). When random passwords are generated, the order of characters is important because a different order yields a different password. Characters are distinguishable in this case. Further, if a password has to consist of different characters, they are sampled from the alphabet without replacement. ◊

Example 2.26 (Polls). When a sample of people is selected to conduct a poll, the same participants produce the same responses regardless of their order. They can be considered indistinguishable. ◊



Permutations with replacement

Permutations with replacement

$$\boxed{P_r(n,k) = \overbrace{n \cdot n \cdot \ldots \cdot n}^{k \text{ terms}} = n^k}$$

Example 2.27 (Breaking passwords)

- From an alphabet consisting of 10 digits, 26 lower-case and 26 capital letters, one can create

$$P_r(62, 8) = 218{,}340{,}105{,}584{,}896$$

(over 218 trillion) different 8-character passwords.

- At a speed of 1 million passwords per second, it will take a spy program almost 7 years to try all of them.
- Thus, on the average, it will guess your password in about 3.5 years.
- At this speed, the spy program can test 604,800,000,000 passwords within 1 week.
- The probability that it guesses your password in 1 week is

$$\frac{N_F}{N_T} = \frac{\text{number of favorable outcomes}}{\text{total number of outcomes}} = \frac{604{,}800{,}000{,}000}{218{,}340{,}105{,}584{,}896} = 0.00277.$$

- However, if capital letters are not used, the number of possible passwords is reduced to

$$P_r(36, 8) = 2{,}821{,}109{,}907{,}456.$$

- On the average, it takes the spy only 16 days to guess such a password!
- The probability that it will happen in 1 week is 0.214.
- A wise recommendation to include all three types of characters in our passwords and to change them once a year is perfectly clear to us now...

Permutations without replacement

Permutations without replacement

$$P(n,k) = \overbrace{n(n-1)(n-2)\cdot\ldots\cdot(n-k+1)}^{k \text{ terms}} = \frac{n!}{(n-k)!}$$

Example 2.28. In how many ways can 10 students be seated in a classroom with 15 chairs?

Solution. Students are distinguishable, and each student has a separate seat. Thus, the number of possible allocations is the number of permutations without replacement, $P(15,10) = 15 \cdot 14 \cdot \ldots \cdot 6 = 1.09 \cdot 10^{10}$. Notice that if students enter the classroom one by one, the first student has 15 choices of seats, then one seat is occupied, and the second student has only 14 choices, etc., and the last student takes one of 6 chairs available at that time. $\Diamond$

Combinations without replacement

Combinations without replacement

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

Example 2.29. An antivirus software reports that 3 folders out of 10 are infected. How many possibilities are there?

Solution. Folders A, B, C and folders C, B, A represent the same outcome, thus, the order is not important. A software clearly detected 3 different folders, thus it is sampling without replacement. The number of possibilities is

$$\binom{10}{3} = \frac{10!}{3!\,7!} = \frac{10 \cdot 9 \cdot \ldots \cdot 1}{(3 \cdot 2 \cdot 1)(7 \cdot \ldots \cdot 1)} = \frac{10 \cdot 9 \cdot 8}{3 \cdot 2 \cdot 1} = 120.$$

$\Diamond$

Shortcut

Combinations without replacement

$$\boxed{C(n,k) = \binom{n}{k} = \frac{P(n,k)}{P(k,k)} = \frac{n!}{k!(n-k)!}} = \frac{n \cdot (n-1) \cdot \ldots \cdot (n-k+1)}{k \cdot (k-1) \cdot \ldots \cdot 1},$$

Example 2.30. There are 20 computers in a store. Among them, 15 are brand new and 5 are refurbished. Six computers are purchased for a student lab. From the first look, they are indistinguishable, so the six computers are selected at random. Compute the probability that among the chosen computers, two are refurbished.

Solution. Compute the total number and the number of favorable outcomes. The *total* number of ways in which 6 computers are selected from 20 is

$$\mathcal{N}_T = \binom{20}{6} = \frac{20 \cdot 19 \cdot 18 \cdot 17 \cdot 16 \cdot 15}{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}. \quad \mathcal{N}_F = \binom{5}{2}\binom{15}{4} = \left(\frac{5 \cdot 4}{2 \cdot 1}\right)\left(\frac{15 \cdot 14 \cdot 13 \cdot 12}{4 \cdot 3 \cdot 2 \cdot 1}\right)$$

$$\boldsymbol{P}\{\text{ two refurbished computers }\} = \frac{\mathcal{N}_F}{\mathcal{N}_T} = \frac{7 \cdot 13 \cdot 5}{19 \cdot 17 \cdot 4} = 0.3522.$$

Combinations with replacement

Combinations with replacement

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

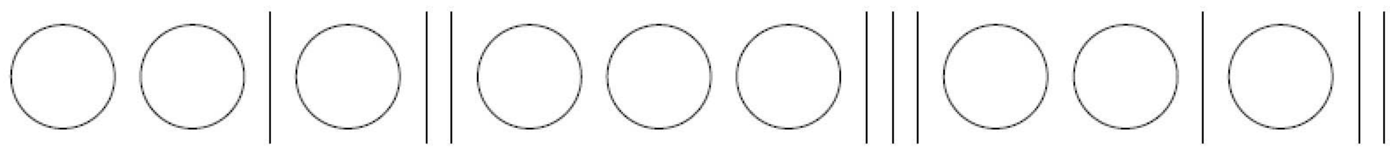


FIGURE 2.5: *Counting combinations with replacement. Vertical bars separate different classes of items.*

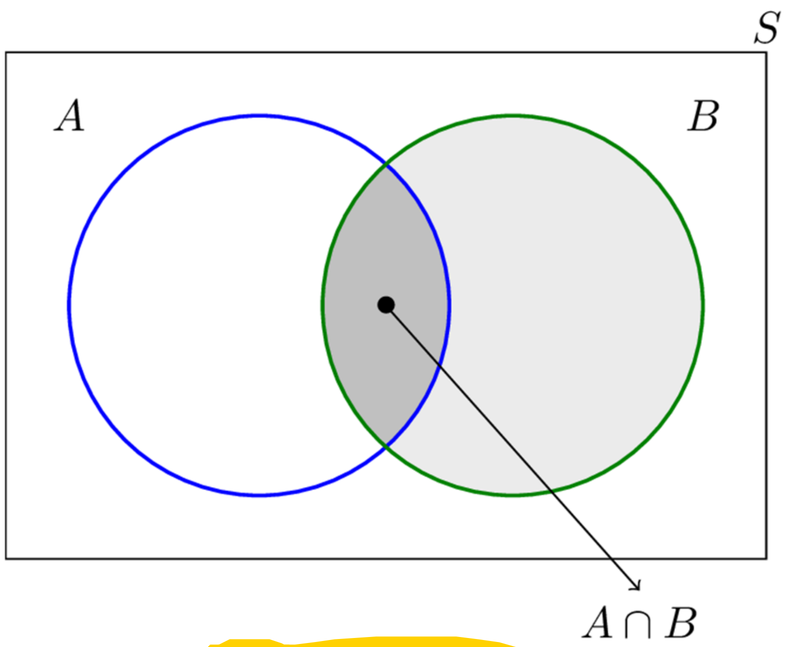
<u>NOTATION</u>

$$
\left\|\begin{array}{ccl}
P_r(n,k) & = & \text{number of permutations with replacement} \\
P(n,k) & = & \text{number of permutations without replacement} \\
C_r(n,k) & = & \text{number of combinations with replacement} \\
\left.\begin{array}{c} C(n,k) \\ \dbinom{n}{k} \end{array}\right\} & = & \text{number of combinations without replacement}
\end{array}\right\|
$$

2.4 Conditional probability and independence

Conditional probability

- Definition 2.15:
  - **Conditional probability** of event A given event B is the probability that A occurs when B is known to occur.



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

Conditional probability

$$\boxed{\boldsymbol{P}\{A \mid B\} = \frac{\boldsymbol{P}\{A \cap B\}}{\boldsymbol{P}\{B\}}}$$

Intersection, general case

$$\boxed{\boldsymbol{P}\{A \cap B\} = \boldsymbol{P}\{B\}\,\boldsymbol{P}\{A \mid B\}}$$

Independence

- Definition 2.16:
  - Events A and B are **independent**
    if occurrence of B does not affect the probability of A, i.e.,

$$\boldsymbol{P}\{A|B\} = \boldsymbol{P}\{A\}.$$

- $\boldsymbol{P}\{A \cap B\} = \boldsymbol{P}\{A\}\boldsymbol{P}\{B\}$.

Example 2.31

- 90 percent of flights depart on time.
- 80 percent of flights arrive on time.
- 75 percent of flights depart on time and arrive on time.

a. You are meeting a flight that departed on time. What is the probability that it will arrive on time?
   - Denote the events, $A = \{\text{arriving on time}\}$, $D = \{\text{departing on time}\}$.
   - We have: $P\{A\} = 0.8$, $P\{D\} = 0.9$, $P\{A \cap D\} = 0.75$.
   - $P\{A|D\} = \dfrac{P\{A \cap D\}}{P\{D\}} = \dfrac{0.75}{0.9} = 0.8333.$

b. You have met a flight, and it arrived on time. What is the probability that it departed on time?
   - $P\{D|A\} = \dfrac{P\{A \cap D\}}{P\{A\}} = \dfrac{0.75}{0.8} = 0.9375.$

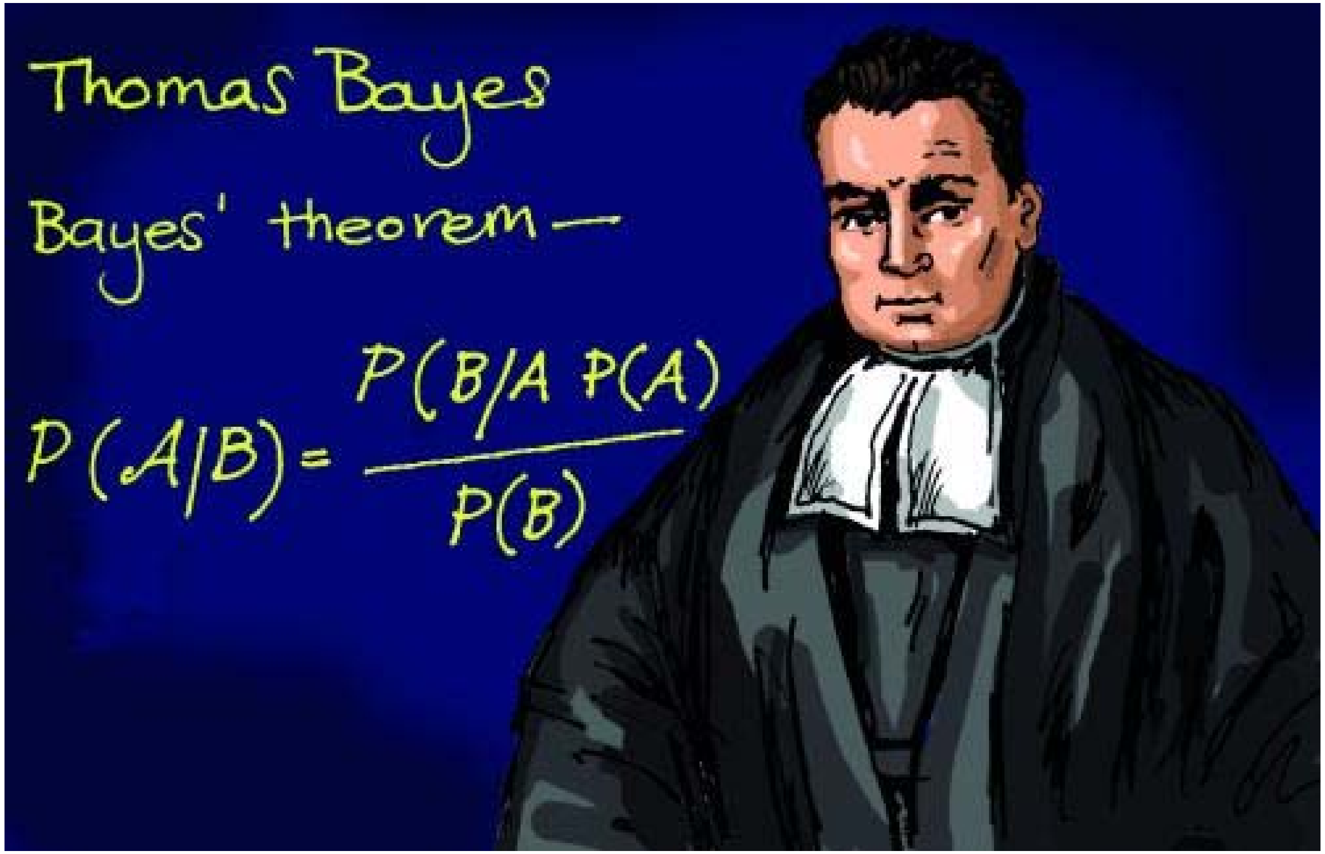
c. Are the events, departing on time and arriving on time, independent?
   - Events are not independent because
   - $P\{A|D\} \neq P\{A\}$, $P\{D|A\} \neq P\{D\}$, $P\{A \cap D\} \neq P\{A\}P\{D\}$.

- Actually, any one of these inequalities is sufficient to prove that A and D are dependent.
- Further, we see that $P\{A|D\} > P\{A\}$ and $P\{D|A\} > P\{D\}$.
- In other words, departing on time increases the probability of arriving on time, and vice versa.
- This perfectly agrees with our intuition.

Bayes' rule

- Example 2.32 (Reliability of a test).
  - There exists a test for a certain viral infection (including a virus attack on a computer network).
  - It is 95% reliable for infected patients and 99% reliable for the healthy ones.
  - That is, if a patient has the virus (event V), the test shows that (event S) with probability
    $$\boldsymbol{P}\{S|V\} = 0.95,$$
    and if the patient does not have the virus, the test shows that with probability
    $$\boldsymbol{P}\{\bar{S}|\bar{V}\} = 0.99$$
  - Consider a patient whose test result is positive (i.e., the test shows that the patient has the virus).
  - Knowing that sometimes the test is wrong, naturally, the patient is eager to know the probability that he or she indeed has the virus.
  - However, this conditional probability, $\boldsymbol{P}\{V|S\}$, is not stated among the given characteristics of this test.

English minister Thomas Bayes (1702–1761)



Example 2.33 (Situation on a midterm exam)

- On a midterm exam, students X, Y, and Z forgot to sign their papers.
- Professor knows that they can write a good exam with probabilities 0.8, 0.7, and 0.5, respectively.
- After the grading, he notices that two unsigned exams are good and one is bad.
- Given this information, and assuming that students worked independently of each other, what is the probability that the bad exam belongs to student Z?
- Solution.
  - Denote good and bad exams by G and B.
  - Also, let GGB denote two good and one bad exams, XG denote the event "student X wrote a good exam," etc.
  - We need to find $P\{ZB|GGB\}$ given that $P\{G|X\} = 0.8$, $P\{G|Y\} = 0.7$, and $P\{G|Z\} = 0.5$.
- By the Bayes Rule,
$$P\{ZB|GGB\} = \frac{P\{GGB|ZB\}P\{ZB\}}{P\{GGB\}}.$$
- Given ZB, event GGB occurs only when both X and Y write good exams.
- Thus, $P\{GGB|ZB\} = (0.8)(0.7)$.
- Event GGB consists of three outcomes depending on the student who wrote the bad exam.
- Adding their probabilities, we get
$$\begin{aligned} P\{GGB\} &= P\{XG \cap YG \cap ZB\} + P\{XG \cap YB \cap ZG\} \\ &\quad + P\{XB \cap YG \cap ZG\} \\ &= (0.8)(0.7)(0.5) + (0.8)(0.3)(0.5) \\ &\quad + (0.2)(0.7)(0.5) = 0.47. \end{aligned}$$
- Then
$$P\{ZB|GGB\} = \frac{(0.8)(0.7)(0.5)}{0.47} = 0.5957.$$

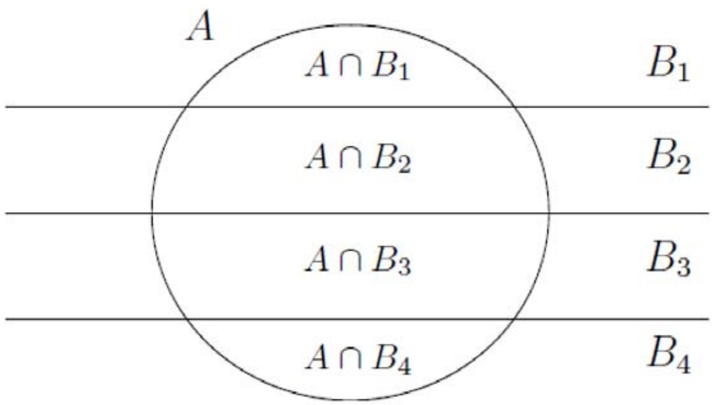
Law of total probability

- This law relates the unconditional probability of an event A with its conditional probabilities.
- It is used every time when it is easier to compute conditional probabilities of A given additional information.
- Consider some partition of the sample space with mutually exclusive and exhaustive events $B_1, \ldots, B_k$.
- It means that

$$B_i \cap B_j = \emptyset \text{ for any } i \neq j \text{ and } B_1 \cup \ldots \cup B_k = \Omega .$$

- These events also partition the event A,
  $$A = (A \cap B_1) \cup \ldots \cup (A \cap B_k),$$
  and this is also a union of mutually exclusive events.
- Hence,

$$\boldsymbol{P}\{A\} = \sum_{j=1}^{k} \boldsymbol{P}\{A \cap B_j\}$$



Eq. 2.10: Law of total probability

$$\boldsymbol{P}\{A\} = \sum_{j=1}^{k} \boldsymbol{P}\{A|B_j\}\boldsymbol{P}\{B_j\}$$

- In case of two events,

$$\boldsymbol{P}\{A\} = \boldsymbol{P}\{A|B\}\boldsymbol{P}\{B\} + \boldsymbol{P}\{A|\bar{B}\}\boldsymbol{P}\{\bar{B}\}$$

Bayes rule for two events

$$\boldsymbol{P}\{B|A\} = \frac{\boldsymbol{P}\{A|B\}\boldsymbol{P}\{B\}}{\boldsymbol{P}\{A|B\}\boldsymbol{P}\{B\} + \boldsymbol{P}\{A|\bar{B}\}\boldsymbol{P}\{\bar{B}\}}$$

Example 2.34 (Reliability of a test, continued)

- Continue **Example 2.32**.
- Suppose that 4% of all the patients are infected with the virus,
  $\boldsymbol{P}\{V\} = 0.04$.
- Recall that $\boldsymbol{P}\{S \mid V\} = 0.95$ and $\boldsymbol{P}\{\bar{S} \mid \bar{V}\} = 0.99$.
- If the test shows positive results,
  the (conditional) probability that a patient has the virus equals

$$\boldsymbol{P}\{V|S\} = \frac{\boldsymbol{P}\{S|V\}\boldsymbol{P}\{V\}}{\boldsymbol{P}\{S|V\}\boldsymbol{P}\{V\} + \boldsymbol{P}\{S|\bar{V}\}\boldsymbol{P}\{\bar{V}\}}$$

$$= \frac{(0.95)(0.04)}{(0.95)(0.04) + (1 - 0.99)(1 - 0.04)} = 0.7983$$

Example 2.35 (Diagnostics of computer codes)

- A new computer program consists of two modules.
- The first module contains an error with probability 0.2.
- The second module is more complex;
  it has a probability of 0.4 to contain an error,
  independently of the first module.
- An error in the first module alone causes the program to crash with probability 0.5.
- For the second module, this probability is 0.8.
- If there are errors in both modules,
  the program crashes with probability 0.9.
- Suppose the program crashed.
- What is the probability of errors in both modules?

- Solution.
  - Denote the events,
    A = {errors in module I},
    B = {errors in module II},
    C = {crash} .
  - Further,
    {errors in module I alone} = $A \backslash B = A \backslash (A \cap B)$
    {errors in module II alone} = $B \backslash A = B \backslash (A \cap B)$.
  - It is given that $\boldsymbol{P}\{A\} = 0.2$, $\boldsymbol{P}\{B\} = 0.4$,
    $\boldsymbol{P}\{A \cap B\} = (0.2)(0.4) = 0.08$,
    by independence,
    $\boldsymbol{P}\{C | A \backslash B\} = 0.5$, $\boldsymbol{P}\{C | B \backslash A\} = 0.8$,
    and $\boldsymbol{P}\{C | A \cap B\} = 0.9$.
  - We need to compute $\boldsymbol{P}\{A \cap B | C\}$.

- Since A is a union of disjoint events $A\backslash B$ and $A\cap B$, we compute $\boldsymbol{P}\{A\backslash B\} = \boldsymbol{P}\{A\} - \boldsymbol{P}\{A\cap B\} = 0.2 - 0.08 = 0.12$.
- Similarly, $\boldsymbol{P}\{B\backslash A\} = 0.4 - 0.08 = 0.32$.
- Events $(A\backslash B)$, $(B\backslash A)$, $A\cap B$, and $(\overline{A\cup B})$ form a partition of Ω, because they are mutually exclusive and exhaustive.
- The last of them is the event of no errors in the entire program.
- Given this event, the probability of a crash is 0.
- Notice that A, B, and $(A\cap B)$ are neither mutually exclusive nor exhaustive, so they cannot be used for the Bayes Rule.
- Now organize the data.

Location of errors			Probability of a crash		
$\boldsymbol{P}\{A\backslash B\}$	$=$	0.12	$\boldsymbol{P}\{C \mid A\backslash B\}$	$=$	0.5
$\boldsymbol{P}\{B\backslash A\}$	$=$	0.32	$\boldsymbol{P}\{C \mid B\backslash A\}$	$=$	0.8
$\boldsymbol{P}\{A\cap B\}$	$=$	0.08	$\boldsymbol{P}\{C \mid A\cap B\}$	$=$	0.9
$\boldsymbol{P}\{\overline{A\cup B}\}$	$=$	0.48	$\boldsymbol{P}\{C \mid \overline{A\cup B}\}$	$=$	0

- Combining the Bayes Rule and the Law of Total Probability,

$$\boldsymbol{P}\{A\cap B|C\} = \frac{\boldsymbol{P}\{C|A\cap B\}\boldsymbol{P}\{A\cap B\}}{\boldsymbol{P}\{C\}},$$

where

$$\begin{aligned}\boldsymbol{P}\{C\} = {} & \boldsymbol{P}\{C|A\backslash B\}P\{A\backslash B\} \\ & +\boldsymbol{P}\{C|B\backslash A\}P\{B\backslash A\} \\ & +\boldsymbol{P}\{C|A\cap B\}P\{A\cap B\} \\ & +\boldsymbol{P}\{C|\overline{A\cup B}\}\boldsymbol{P}\{\overline{A\cup B}\}.\end{aligned}$$

- Then $\boldsymbol{P}\{A\cap B|C\} =$

$$\frac{(0.9)(0.08)}{(0.5)(0.12) + (0.8)(0.32) + (0.9)(0.08) + 0} = 0.1856.$$

Q&A