CENG 222
Probability and Statistics

Probability

Nesli Erdoğmuş & Burak Galip Aslan

Contents

Contents

Events and their probabilities

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. However, if you toss a coin 1 million times, the proportion of heads is anticipated to be very close to 1/2.

Events and their probabilities

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. However, if you toss a coin 1 million times, the proportion of heads is anticipated to be very close to 1/2.

This example suggests that in a long run, probability of an event can be viewed as a proportion of times this event happens, or its **relative frequency**.

Events and their probabilities

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. However, if you toss a coin 1 million times, the proportion of heads is anticipated to be very close to 1/2.

This example suggests that in a long run, probability of an event can be viewed as a proportion of times this event happens, or its **relative frequency**.

In **forecasting**, it is common to speak about the probability as a **likelihood** (say, the company's profit is likely to rise during the next quarter).

Events and their probabilities

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. However, if you toss a coin 1 million times, the proportion of heads is anticipated to be very close to 1/2.

This example suggests that in a long run, probability of an event can be viewed as a proportion of times this event happens, or its **relative frequency**.

In **forecasting**, it is common to speak about the probability as a **likelihood** (say, the company's profit is likely to rise during the next quarter).

In **gambling and lottery**, probability is equivalent to **odds**. Having the winning odds of 1 to 99 (1:99) means that the probability to win is 0.01, and the probability to lose is 0.99. It also means, on a relative-frequency language, that if you play long enough, you will win about 1% of the time.

Odds vs Probability vs Chance (Likelihood)

- **Probability.** The fraction you would expect to see a **certain event occurring in many trials**. It is expressed as a real number within the interval [0,1]. For example, the probability of event A happening might be 0.75. This is more formally known as the normalized probability.
- **Chance or Likelihood.** Chance, a synonym of probability, is usually expressed as a percentage (e.g. 75%). Chance is used primarily in weather prediction, and typically indicates the **likelihood of a given weather event occurring** in a specific area over a certain period of time.
- **Odds.** Ratios of probabilities, they can be defined in **two different ways**: **odds in *favor* of an event** ("odds on") or ***against* an event** A ("Odds Against). Odds can be expressed as a ratio of the probability an event will happen divided by the probability an event won't happen: Odds in favor of A = A / (1 – A), usually simplified to lowest terms., For instance, if the probability of an event occurring is 0.75, then the *odds* for it happening are 0.75/0.25 = 3/1 = 3 to 1 for, while the probability that it doesn't occur is 1 to 3 against.

taken from: Odds vs Probability vs Chance - DataScienceCentral.com

***P.S.** it makes sense to have an idea about the different implications of those terms because almost all of them correspond to "olasılık" or "ihtimal" in Turkish by definition.*

Events and their probabilities

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.

Events and their probabilities

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.

Outcomes, events, and the sample space

Probabilities arise when one considers and weighs possible **results of some experiment**.

Some results are more likely than others.

An **experiment** may be as simple as a coin toss, or as complex as starting a new business.

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**.

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, events, and the sample space

Example 2.4. 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**.

Proof: To count all possible events, we shall see how many ways an event can be constructed. The first outcome can be included into our event or excluded, so there are two possibilities. Then, every next outcome is either included or excluded, so every time the number of possibilities doubles. Overall, we have

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

possibilities, leading to a total of 2^N possible events.

Outcomes, events, and the sample space

Example 2.5. Consider a football game between Washington and Dallas. The sample space consists of 3 outcomes: Ω = { Washington wins, Dallas wins, they tie }
Combining these outcomes in all possible ways, we obtain the following $\mathbf{2^3 = 8}$ events:
Washington

1. wins,
2. loses,
3. ties,
4. gets at least a tie,
5. gets at most a tie,
6. no tie,
7. gets some result
8. gets no result.

The event "some result" is the entire sample space Ω, 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.

Outcomes, events, and the sample space

Example 2.5. Consider a football game between Washington and Dallas. The sample space consists of 3 outcomes: Ω = { Washington wins, Dallas wins, they tie }
Combining these outcomes in all possible ways, we obtain the following $\mathbf{2^3 = 8}$ events:
Washington

1. wins,
2. loses,
3. ties,
4. gets at least a tie,
5. gets at most a tie,
6. no tie,
7. gets some result
8. gets no result.

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

The event "some result" is the entire sample space Ω, 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.

Outcomes, events, and the sample space

- **Outcome**: An element of the sample space Ω
- **Event**: A subset of the sample space Ω

Some experiments:

- Rolling a dice

  Ω = {1,2,3,4,5,6} ➔ finite

- Rolling a dice and counting until first 6

  Ω = {1,2,3,...} ➔ countable and infinite

- Turn on a light bulb and measure its lifetime

  Ω = [0,∞) ➔ uncountable and infinite

Set operations

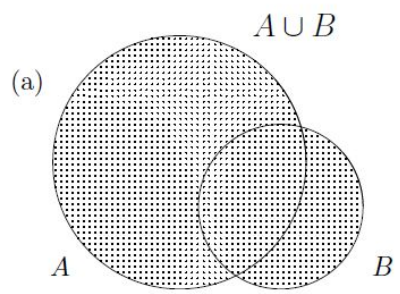
Events are sets of outcomes.

Following terms have to be defined;

- unions
- intersections
- differences
- complements

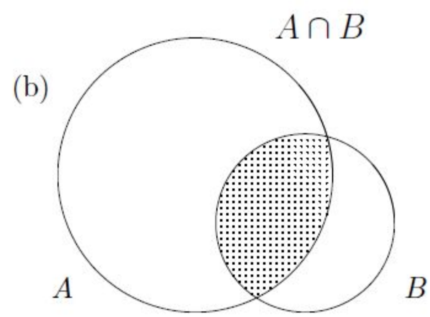
Set operations

DEFINITION 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 ...



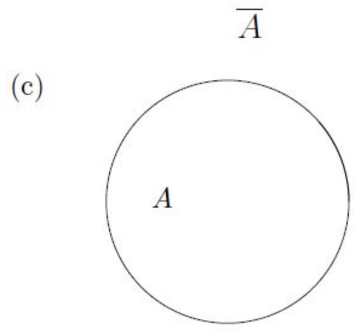
Set operations

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 ...



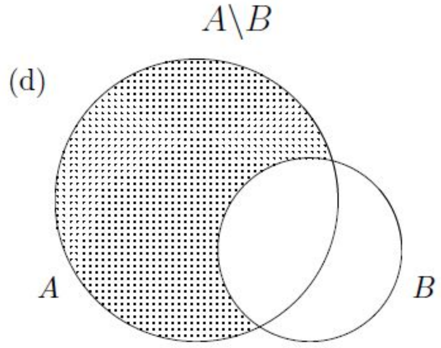
Set operations

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



Set operations

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



Set operations

NOTATION			
	$A \cup B$	$=$	union
	$A \cap B$	$=$	intersection
	$\overline{A}$ or A^c	$=$	complement
	$A \backslash B$	$=$	difference

Set operations

DEFINITION 2.7 Events A and B are **disjoint** if their intersection is empty,

$$\mathbf{A} \cap \mathbf{B} = \varnothing .$$

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

$$\mathbf{A_i} \cap \mathbf{A_j} = \varnothing \text{ for any } i = j .$$

Set operations

DEFINITION 2.7 Events A and B are **disjoint** if their intersection is empty,

$$\mathbf{A} \cap \mathbf{B} = \varnothing .$$

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

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

DEFINITION 2.8 Events A, B, C, . . . are **exhaustive** if their union equals the whole sample space, i.e.,

$$\mathbf{A} \cup \mathbf{B} \cup \mathbf{C} \cup \ldots = \Omega .$$

Set operations

DEFINITION 2.7 Events A and B are **disjoint** if their intersection is empty,

$$\mathbf{A} \cap \mathbf{B} = \varnothing .$$

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

$$\mathbf{A_i} \cap \mathbf{A_j} = \varnothing \text{ for any } i = j .$$

DEFINITION 2.8 Events A, B, C, . . . are **exhaustive** if their union equals the whole sample space, i.e.,

$$\mathbf{A} \cup \mathbf{B} \cup \mathbf{C} \cup \ldots = \Omega .$$

- **Mutually exclusive** events will never occur at the same time. Occurrence of any one of them eliminates the possibility for all the others to occur.
- **Exhaustive events** cover the entire Ω, so that “there is nothing left.” In other words, among any collection of exhaustive events, at least one occurs for sure.

Set operations

Example 2.6. When a card is pooled from a deck at random, the four suits are at the same time disjoint and exhaustive.

Hearts, spades, clubs, and diamonds: they are disjoint; the card can be only one of them. They also make up all possible outcomes that can occur, i.e. there are no other cards that are neither hearts, spades, clubs or diamonds. Hence they are also exhaustive.

Set operations

Example 2.6. When a card is pooled from a deck at random, the four suits are at the same time disjoint and exhaustive.

Hearts, spades, clubs, and diamonds: they are disjoint; the card can be only one of them. They also make up all possible outcomes that can occur, i.e. there are no other cards that are neither hearts, spades, clubs or diamonds. Hence they are also exhaustive.

Example 2.7. Any event A and its complement A^C represent a classical example of disjoint and exhaustive events.

Set operations

Example 2.6. When a card is pooled from a deck at random, the four suits are at the same time disjoint and exhaustive.

Hearts, spades, clubs, and diamonds: they are disjoint; the card can be only one of them. They also make up all possible outcomes that can occur, i.e. there are no other cards that are neither hearts, spades, clubs or diamonds. Hence they are also exhaustive.

Example 2.7. Any event A and its complement A^C represent a classical example of disjoint and exhaustive events.

Example 2.8. Receiving a grade of “AA”, “BA”, “BB”, “CB”, “CC”, “DC”, “DD”, “FD” or “FF” for CENG 222 course in IYTE?

Set operations

Example 2.6. When a card is pooled from a deck at random, the four suits are at the same time disjoint and exhaustive.

Hearts, spades, clubs, and diamonds: they are disjoint; the card can be only one of them. They also make up all possible outcomes that can occur, i.e. there are no other cards that are neither hearts, spades, clubs or diamonds. Hence they are also exhaustive.

Example 2.7. Any event A and its complement A^C represent a classical example of disjoint and exhaustive events.

Example 2.8. Receiving a grade of “AA”, “BA”, “BB”, “CB”, “CC”, “DC”, “DD”, “FD” or “FF” for CENG 222 course in IYTE are mutually exclusive events, but they are not exhaustive because of possible “I” and “NA” grades.

Set operations

It is often easier to compute probability of an intersection than probability of a union. Taking complements converts unions into intersections.

$$\boxed{\overline{E_1 \cup \ldots \cup E_n} = \overline{E}_1 \cap \ldots \cap \overline{E}_n, \qquad \overline{E_1 \cap \ldots \cap E_n} = \overline{E}_1 \cup \ldots \cup \overline{E}_n} \tag{2.2}$$

Proof: Since the union $E_1 \cup \ldots \cup E_n$ represents the event "at least one event occurs," its complement has the form

$$\begin{aligned}
\overline{E_1 \cup \ldots \cup E_n} &= \{ \text{ none of them occurs } \} \\
&= \{ E_1 \text{ does not occur} \cap \ldots \cap E_n \text{ does not occur } \} \\
&= \overline{E}_1 \cap \ldots \cap \overline{E}_n.
\end{aligned}$$

Set operations

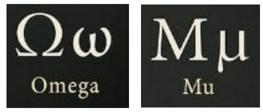
Example 2.9. Graduating with a GPA of 4.0 is an intersection of getting an A in each course. Its complement, graduating with a GPA below 4.0, is a union of receiving a grade below A at least in one course.

Set operations

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.

- A complement to "nothing" is "something"
- "Not everything" means "at least one is missing"

Contents



Sigma-algebra

Definition 2.9:

A collection M ($\mathfrak{M}$) of events is a **sigma-algebra** on sample space Ω if

a. it includes the sample space,

$$\Omega \in M$$

b. every event in M is contained along with its complement,

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

c. every finite or countable collection of events in M is contained along with its union.

$$E_1, E_2, \dots \in M \Rightarrow E_1 \cup E_2 \cup \dots \in M$$

Sigma-algebra examples

- **Example 2.10**: (Degenerate sigma-algebra). By conditions (a) and (b) in Definition 2.9, every sigma-algebra has to contain the sample space Ω and the empty event ∅.

  This minimal collection M = {Ω, ∅} 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**.

Axioms of probability

Definition 2.10:

Assume a sample space Ω and a sigma-algebra of events M on it. **Probability**

$$\mathbf{P} : M \to [0, 1]$$

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,

$$\mathbf{P}(\Omega) = 1.$$

- (Sigma-additivity) For any finite or countable collection of mutually exclusive events $E_1, E_2, \ldots \in M$,

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

All the rules of probability are consequences from this definition.

Axioms of probability

All propositions below can be proved using the 3 axioms of probability given in the previous definition:

- $P(\emptyset) = 0$
- $P(\overline{E}) = 1 - P(E)$
- $P(E_1 \setminus E_2) = P(E_1) - P(E_1 \cap E_2)$
- $P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$
- If $E_1 \subseteq E_2$, then $P(E_1) \leq P(E_2)$

A premature example

- The possibility of rain on Saturday is 0.5 and the probability of rain on Sunday is also 0.5. What is the probability of having raing on both days?

Think:

- Are they disjoint?
- Are they independent?

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 $\varnothing$ never occurs.

$$P\{\Omega\} = 1 \text{ and } P\{\varnothing\} = 0. \qquad (2.3)$$

Proof: Probability of Ω is given by the definition of probability. By the same definition, $P\{\Omega\} = P\{\Omega \cup \varnothing\} = P\{\Omega\} + P\{\varnothing\}$, because Ω and $\varnothing$ are <u>mutually exclusive</u>. Therefore, $P\{\varnothing\} = 0$.

Computing probabilities of events

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,

$$P(E) = P(\omega_1) + P(\omega_2) + P(\omega_3) + \ldots$$

Computing probabilities of events

Union

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.

It is crucial to notice that **only mutually exclusive events** (those with empty intersections) **satisfy the sigma-additivity**.

If events intersect, their probabilities cannot be simply added.

Computing probabilities of events

Union

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.

Computing probabilities of events

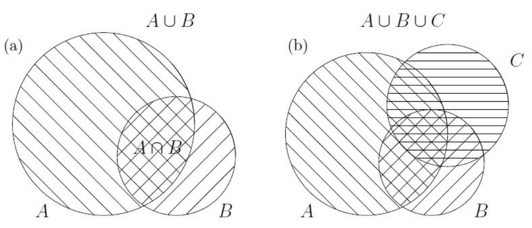


FIGURE 2.2: *(a) Union of two events. (b) Union of three events.*

Probability of a union

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

For mutually exclusive events,

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

Generalization of this formula is not straightforward. For 3 events,

$$\begin{aligned} \boldsymbol{P}\{A \cup B \cup C\} &= \boldsymbol{P}\{A\} + \boldsymbol{P}\{B\} + \boldsymbol{P}\{C\} - \boldsymbol{P}\{A \cap B\} - \boldsymbol{P}\{A \cap C\} \\ &\quad - \boldsymbol{P}\{B \cap C\} + \boldsymbol{P}\{A \cap B \cap C\}. \end{aligned}$$

Computing probabilities of events

The **inclusion–exclusion principle** is a counting technique which generalizes the familiar method of obtaining the number of elements in the union of two finite sets; symbolically expressed as $|A \cup B| = |A| + |B| - |A \cap B|$.

Generalizing the results of these examples gives the principle of inclusion–exclusion. To find the cardinality of the union of n sets:

- Include the cardinalities of the sets.
- Exclude the cardinalities of the pairwise intersections.
- Include the cardinalities of the triple-wise intersections.
- Exclude the cardinalities of the quadruple-wise intersections.
- Continue, until the cardinality of the n-tuple-wise intersection is included (if n is odd) or excluded (n even).

Computing probabilities of events

Complement

Recall that events A and $\bar{A}$ are exhaustive, hence $A \cup \bar{A} = \Omega$. Also, they are disjoint, hence:

$$P(A) + P(\bar{A}) = P(A \cup \bar{A}) = P(\Omega) = 1.$$

Solving this for $P(\bar{A})$, we obtain a rule that perfectly agrees with the common sense,

$$P(\bar{A}) = 1 - P(A)$$

This is called the **complement rule**.

Computing probabilities of events

Complement

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.

Computing probabilities of events

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.

The following basic formula can serve as the criterion of independence:

$$P(E_1 \cap \ldots \cap E_n) = P(E_1) \cdot \ldots \cdot P(E_n)$$

Computing probabilities of events

Example 2.14. During some construction, a network blackout occurs on Monday with probability 0.7 and on Tuesday with probability 0.5. Then, the probability of a blackout on Monday or Tuesday is;

$$P(M) + P(T) - P(M \cap T)$$

$$0.7 + 0.5 - 0.7 \cdot 0.5 = 0.85$$

Computing probabilities of events

Example 2.14. During some construction, a network blackout occurs on Monday with probability 0.7 and on Tuesday with probability 0.5. Then, the probability of a blackout on Monday or Tuesday is;

$$P(M) + P(T) - P(M \cap T)$$

$$0.7 + 0.5 - 0.7 \cdot 0.5 = 0.85$$

or

$$1 - P(\overline{M \cup T}) = 1 - P(\bar{M} \cap \bar{T})$$

$$1 - 0.3 \cdot 0.5 = 0.85.$$

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?

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: First organize the data!

- H = { hard drive crashes }
- B_1 = { first backup crashes }
- B_2 = { second backup crashes } .
- H, B_1, and B_2 are independent (which makes sense indeed).

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: Next, write down what you have and what you need!

- $P(H) = 0.01$, $P(B_1) = P(B_2) = 0.02$
- $P(\text{saved}) = 1 - P(\text{lost}) = 1 - P(H \cap B1 \cap B2)$
- $P(\text{saved}) = 1 - P(H) \cdot P(B1) \cdot P(B2) = 1 - (0.01)(0.02)(0.02)$
- $P(\text{saved}) = 0.999996$

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?

Comments: The risk of failure drops from 1/100 to 1/250000, a significant improvement with two backups. When the system's components are connected in parallel, it is sufficient for at least one component to work in order for the whole system to function.

Applications in reliability

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.

Comments: This is a system whose components are connected in sequel. Failure of one component inevitably causes the whole system to fail. Such a system is more "vulnerable." In order to function with a high probability, it needs each component to be reliable.

Applications in reliability

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: First organize the data!

- C_1 = { 1st device crashes}
- C_2 = { 2nd device crashes}
- C_3 = { 3rd device crashes}
- C_1, C_2 and C_3 are independent.

Applications in reliability

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: Next, write down what you have and what you need!

- $P(C_1) = 0.01$ $P(C_2) = 0.02$ $P(C_3) = 0.02$
- $P(\text{on time}) = P(\text{all devices function}) = P(\overline{C}_1 \cap \overline{C}_2 \cap \overline{C}_3)$
- $P(\text{on time}) = P(\overline{C}_1) \cdot P(\overline{C}_2) \cdot P(\overline{C}_3) = (1 - 0.01)(1 - 0.02)(1 - 0.02)$
- $P(\text{on time}) = 0.9508$

Contents

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, $\omega_1, \ldots, \omega_n$, each having the same probability. Since

$$\Sigma_{k=1\ldots n} P(\omega_k) = P(\Omega) = 1$$

we have in this case **$P\{\omega_k\} = 1/n$** for all k. Further, a probability of any event E consisting of t outcomes, equals

$$P(E) = \Sigma_{\omega \in E}(1/n) = t\,(1/n) = t/n = \frac{\text{number of outcomes in } E}{\text{number of outcomes in } \Omega}$$

The outcomes forming event E are often called "**favorable**."

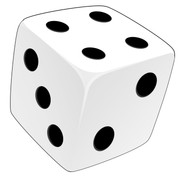
Equally likely outcomes

The formula of an event for an experiment with equally likely outcomes is

$$P(E) = \frac{\text{number of favorable outcomes}}{\text{number of outcomes in } \Omega} = \frac{N_F}{N_T}$$

where index "F" means "favorable" and "T " means "total."

Equally likely outcomes



The formula of an event for an experiment with equally likely outcomes is

$$P(E) = \frac{\text{number of favorable outcomes}}{\text{number of outcomes in } \Omega} = \frac{N_F}{N_T}$$

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. P(1), P(odd), P(less than 5) can be calculated using this formula.

Equally likely outcomes

Example 2.22. A card is drawn from a 52-card deck at random. Compute the probability that the selected card is a spade.

Equally likely outcomes

Example 2.22. A card is drawn from a 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(spade) = 13/52 = 1/4.

Equally likely outcomes

Example 2.22. A card is drawn from a 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(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(spade) = 1/4.

Different sample spaces - both equally likely!

Equally likely outcomes

Example 2.23. A young family plans to have two children. What is the probability of two girls?

Equally likely outcomes

Example 2.23. A young family plans to have two children. What is the probability of two girls?

Solution 1. 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.

Equally likely outcomes

Example 2.23. A young family plans to have two children. What is the probability of two girls?

~~**Solution 1**. 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. Each child is (supposedly) equally likely to be a boy or a girl. Genders of the two children are (supposedly) independent.

$$P(2 \text{ girls}) = (½)*(½) = ¼$$

Equally likely outcomes

Example 2.23. A young family plans to have two children. What is the probability of two girls?

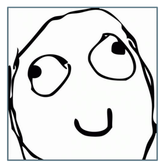
Solution 2. Each child is (supposedly) equally likely to be a boy or a girl. Genders of the two children are (supposedly) independent.

$$P(2 \text{ girls}) = (½)*(½) = ¼$$

The second solution implies that the sample space consists of four, not three, equally likely outcomes: two boys, two girls, a boy and a girl, a girl and a boy. Each outcome in this sample has probability 1/4. This was **NOT** the case for Solution 1.

Equally likely outcomes

Example 2.24 (Paradox). A family has two children. You met one of them, Lev, and he is a boy. What is the probability that the other child is also a boy?

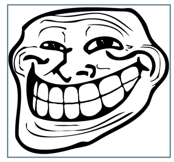


Equally likely outcomes

Example 2.24 (Paradox). A family has two children. You met one of them, Lev, 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 Lev? Lev should have a brother or a sister with probabilities 1/2 and 1/2.

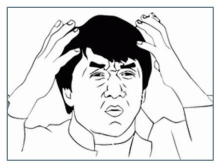
Equally likely outcomes



Example 2.24 (Paradox). A family has two children. You met one of them, Lev, and he is a boy. What is the probability that the other child is also a boy?

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, Lev has a brother in one case and a sister in two cases. Thus, isn't the probability of a boy equal 1/3?

Equally likely outcomes



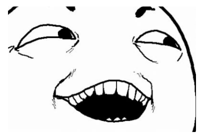
Example 2.24 (Paradox). A family has two children. You met one of them, Lev, and he is a boy. What is the probability that the other child is also a boy?

Where is the catch? Apparently, the sample space Ω 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 Lev, do the outcomes {BB,BG,GB} remain equally likely?

Equally likely outcomes

Example 2.24 (Paradox). A family has two children. You met one of them, Lev, and he is a boy. What is the probability that the other child is also a boy?

"**Boy or Girl Paradox**"



To be further discussed in 2.4 - Conditional probability and independence

Permutations and combinations

$$P(E) = \frac{\text{number of favorable outcomes}}{\text{number of outcomes in } \Omega} = \frac{N_F}{N_T}$$

This formula is simple, as long as its numerator and denominator can be easily evaluated.

This is rarely the case; often the sample space consists of a multitude of outcomes. **Combinatorics** provides special techniques for the computation of N_T and N_F , the total number and the number of favorable outcomes.

Permutations and combinations

$$P(E) = \frac{\text{number of favorable outcomes}}{\text{number of outcomes in } \Omega} = \frac{N_F}{N_T}$$

For computations of N_T and N_F, we will consider a generic situation when objects are selected **at random** from **a set of n**. This general model has a number of useful applications.

- The objects may be selected *with replacement* or *without replacement*.
- They may also be *distinguishable* or *indistinguishable*.

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.

Permutations and combinations

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.

Permutations and combinations

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 (P_r)

Possible selections of k **distinguishable** objects from a set of n are called **permutations**.

When we sample with **replacement**, each time there are n possible selections, hence the total number of permutations is

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

Permutations with replacement (P_r)

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.

Permutations with replacement (P_r)

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. It can test 604,800,000,000 passwords within 1 week. The probability that it guesses your password in 1 week is **0.00277**.

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

Permutations with replacement (P_r)

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.

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 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**.

Why it is recommended to include all different types of characters in our passwords and to change them once a year is perfectly should be clear now.

Permutations without replacement (P)

During sampling without replacement, the number of possible selections reduces by 1 each time an object is sampled:

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

where n! = 1 · 2 · . . . n (n-factorial) denotes the product of all integers from 1 to n.

The number of permutations without replacement also equals the number of possible allocations of k distinguishable objects among n available slots.

Permutations without replacement (P)

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

Permutations without replacement (P)

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 needs 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.

Combinations without replacement (C)

Possible selections of k indistinguishable objects from a set of n are called **combinations**. The number of combinations without replacement is also called "**n choose k**" and is denoted by

$$C(n,k) \text{ or } \binom{n}{k}$$

Combinations without replacement (C)

Possible selections of k indistinguishable objects from a set of n are called **combinations**. The number of combinations without replacement is also called "**n choose k**" and is denoted by

$$C(n,k) \text{ or } \binom{n}{k}$$

The only difference from P(n, k) is disregarding the order. Now the same objects sampled in a different order produce the same outcome. Thus, P(k,k) = k! different permutations (rearrangements) of the same objects yield only 1 combination.

Combinations without replacement

$$\boxed{C(n,k) = \binom{n}{k} = \frac{P(n,k)}{P(k,k)} = \frac{n!}{k!(n-k)!}} \qquad (2.6)$$

Combinations without replacement (C)

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

Combinations without replacement (C)

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

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

Combinations without replacement (C)

Computational shortcuts

Instead of computing C(n, k) directly by the formula, we can simplify the fraction. At least, the numerator and denominator can both be divided by either k! or (n − k)! (choose the larger of these for greater reduction). As a result,

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

the top and the bottom of this fraction being products of k terms. It is also handy to notice that C(n,k)=C(n,n-k) for any k and n, C(n,0)=1 and C(n,1)=n.

Combinations without replacement (C)

Example 2.30. There are 20 computers in a store. Among them, 15 are brand new and 5 are refurbished (yenilenmiş). 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.

Combinations without replacement (C)

Example 2.30. There are 20 computers in a store. Among them, 15 are brand new and 5 are refurbished (yenilenmiş). 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:

$$\mathcal{N}_T = \binom{20}{6} = \frac{20 \cdot 19 \cdot 18 \cdot 17 \cdot 16 \cdot 15}{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}.$$

Combinations without replacement (C)

Example 2.30. There are 20 computers in a store. Among them, 15 are brand new and 5 are refurbished (yenilenmiş). 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. Next, for the number of favorable outcomes, 2 refurbished computers are selected from a total of 5, and the remaining 4 new ones are selected from a total of 15:

$$\mathcal{N}_F = \binom{5}{2}\binom{15}{4} = \left(\frac{5 \cdot 4}{2 \cdot 1}\right)\left(\frac{15 \cdot 14 \cdot 13 \cdot 12}{4 \cdot 3 \cdot 2 \cdot 1}\right)$$

Combinations without replacement (C)

Example 2.30. There are 20 computers in a store. Among them, 15 are brand new and 5 are refurbished (yenilenmiş). 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. Finally, with further reduction of fractions, the probability equals:

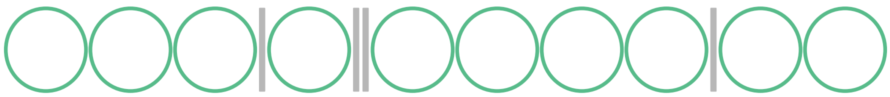
$$\boldsymbol{P}\{\text{ two refurbished computers }\} = \frac{\mathcal{N}_F}{\mathcal{N}_T} = \frac{7 \cdot 13 \cdot 5}{19 \cdot 17 \cdot 4} = 0.3522.$$

Combinations with replacement (C_r)

For combinations with replacement, the order is not important, and **each object may be sampled more than once**. Then each outcome consists of counts, how many times each of **n objects** appears in the sample. In total, we select **k objects**.

A graphical approach:

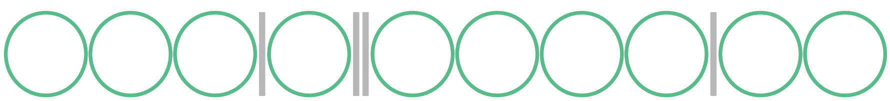
We can draw a circle for each time object 1 is sampled, then draw a separating bar, then a circle for each time object 2 is sampled, etc.



Combinations with replacement (C_r)

A graphical approach:

We can draw a circle for each time object 1 is sampled, then draw a separating bar, then a circle for each time object 2 is sampled, etc.

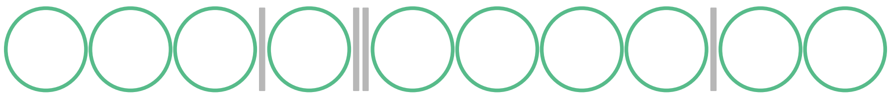


This figure shows an example where out of 5 objects 10 are selected with replacement. Number of selected objects are 3, 1, 0, 4 and 2 for objects 1, 2, 3, 4 and 5, respectively.

Combinations with replacement (C_r)

A graphical approach:

We can draw a circle for each time object 1 is sampled, then draw a separating bar, then a circle for each time object 2 is sampled, etc.



Total number of such configurations can be calculated using:

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

Combinations with replacement (C_r)

Example*: There are five kinds of frozen yogurt: banana, chocolate, lemon, strawberry and vanilla. You can have three scoops. What number of varieties will there be?

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

*from: [Statistics - Combination with replacement (tutorialspoint.com)]

Combinations with replacement (C_r)

Example*: There are five kinds of frozen yogurt: banana, chocolate, lemon, strawberry and vanilla. You can have three scoops. What number of varieties will there be?

Solution: Here n = 5 and k = 3. Substitute the values in formula:

$$\frac{(5+3-1)!}{3!\,(5-1)!} = \frac{7!}{3!\,4!} = \frac{5\cdot 6\cdot 7}{2\cdot 3} = 35$$

Combinations with replacement

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

*from: [Statistics - Combination with replacement (tutorialspoint.com)]

Permutations and combinations

NOTATION

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

Exercise

Matching Problem

Consider a well shuffled deck of n cards labelled 1 through n. You flip over the cards one by one, saying the numbers 1 through n as you do so.

You win the game if, at some point, the number you say aloud is the same as the number on the number on the card being flipped over (for example, if the 8th card in the deck has the label 8).

What is the probability of losing?

Exercise

Matching Problem

Consider a well shuffled deck of n cards labelled 1 through n. You flip over the cards one by one, saying the numbers 1 through n as you do so.

You win the game if, at some point, the number you say aloud is the same as the number on the number on the card being flipped over (for example, if the 8th card in the deck has the label 8).

What is the probability of losing?

Exercise

Matching Problem

P(losing) = 1 - P(winning) = 1 - P(at least 1 match)

= 1 - (P(1 match) + P(2 matches) + … +P(n matches))

For 1 match:

- We can have a match at nC1 (n choose 1) cards.
- The rest of the cards can have (n-1)! different configurations.

For 2 matches:

- We can have a match at nC2 (n choose 2) cards.
- The rest of the cards can have (n-2)! different configurations.

Exercise

Matching Problem

$$P(\text{winning}) = \frac{\text{number of favorable outcomes}}{\text{number of outcomes in } \Omega}$$

$$= \frac{nC1(n-1)!+nC2(n-2)!+ \dots +nCn0!}{n!}$$

Exercise

Matching Problem

$$P(\text{winning}) = \frac{nC1(n-1)!+nC2(n-2)!+ \ldots +nCn0!}{n!}$$

Let’s test for cards 1,2,3. - We have 6 permutations:

(1,2,3), (1,3,2), (2,1,3), **(2,3,1)**, **(3,1,2)**, (3,2,1) with 2 zero matches marked as bold. So, probability of winning should be $4/6 \cong 0.667$.

Exercise

Matching Problem

$$P(\text{winning}) = \frac{nC1(n-1)!+nC2(n-2)!+ \ldots +nCn0!}{n!}$$

Let’s test for cards 1,2,3. - We have 6 permutations:

(1,2,3), (1,3,2), (2,1,3), **(2,3,1)**, **(3,1,2)**, (3,2,1) with 2 zero matches marked as bold. So, probability of winning should be $4/6 \approx 0.667$.

Using the formula:

$P(\text{winning}) = (3 \cdot 2! + 3 \cdot 1! + 1 \cdot 0!) / 3! = 10 / 6 =$ **1.67 ?!!?**

Exercise

Matching Problem

Tackling one problem at a time:

P(1 match) = nC1 * (n-1)!/n! = 3*2/3! = 1 (**Not good!**)

3C1 =3 is correct. 3 sets with 1 match {(1,3,2), (3,2,1), (2,1,3)}

How about the (n-1)! part?

Exercise

Matching Problem

How about the (n-1)! part?

(n-1)!=2!=2 means there are 2 possible permutations for remaining 2 cards to perform after fixing one card.

- Fix 1st card, permute other 2 cards: (1,2,3), (1,3,2)
- Fix 2nd card, permute other 2 cards: (1,2,3), (3,2,1)
- Fix 3rd card, permute other 2 cards: (1,2,3), (2,1,3)

We are adding (1,2,3) in every permute.

We aren't exclusively fetching sets with 1 match but fetching sets with 1 or more matches.

Exercise

Matching Problem

Using the Inclusion-Exclusion Principle:

P(Winning) = (3*2 -3*1 +1*1)/6 = (6-3+1)/6 = 4/6 = **0.667**

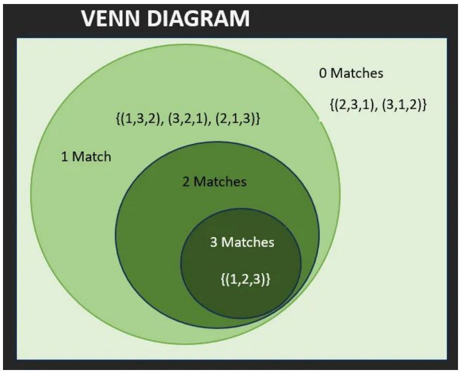


*diagram from: https://medium.com/@gjvidya08/de-montmorts-matching-problem-92261906282f

Exercise

Matching Problem

Let's generalize:

$$P(\text{winning}) = \frac{nC1(n-1)!-nC2(n-2)!+ \ldots nCn0!}{n!}$$

$$= 1 - (1/2!) + (1/3!) - (1/4!)+ \ldots (1/n!)$$

P(losing) = (1/2!) + (1/3!) - (1/4!)+ … (1/n!)

(Taylor's series)

As n➡∞, P(losing)➡1/e≈0.36788

Contents

Conditional probability

Suppose you are meeting someone at an airport. The flight is likely to arrive on time; the probability of that is 0.8. Suddenly it is announced that the flight departed one hour behind the schedule. Now it has the probability of only 0.05 to arrive on time.

New information affected the probability of meeting this flight on time. The new probability is called **conditional probability**, where the new information, that the flight departed late, is a condition.

Conditional probability

Suppose you are meeting someone at an airport. The flight is likely to arrive on time; the probability of that is 0.8. Suddenly it is announced that the flight departed one hour behind the schedule. Now it has the probability of only 0.05 to arrive on time.

New information affected the probability of meeting this flight on time. The new probability is called **conditional probability**, where the new information, that the flight departed late, is a condition.

DEFINITION 2.15 Conditional probability of event A given event B is the probability that A occurs when B is known to occur.

Notation: **P {A | B}** = conditional probability of A given B

Conditional probability

unconditional probability of A

$$P(A) = \frac{\text{number of outcomes in } A}{\text{number of outcomes in } \Omega}$$

conditional probability of A **given** B

$$P(A \mid B) = \frac{\text{number of outcomes in } A \cap B}{\text{number of outcomes in } B} = \frac{P(A \cap B)}{P(B)}$$

Conditional probability

The general formula.

Conditional probability

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

The general formula for the probability of intersection.

Intersection, general case

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

Conditional Probability

With B known, we operate in a new sample space:

- $0 \le P(A|B) \le 1$
- $P(\Omega|B) = 1$
- For a sequence of disjoint events A_k, $P(\cup A_k|B) = \sum P(A_k|B)$

- $P(A^c|B) = 1 - P(A|B)$
- $P(B \setminus A|C) = P(B|C) - P(A \cap B|C)$
- $P(A \cup B|C) = P(A|C) + P(B|C) - P(A \cap B|C)$
- If $A \subseteq B$, then $P(A|C) \le P(B|C)$

Conditional Probability

With B known, we operate in a new sample space:

- $0 \le P(A|B) \le 1$
- $P(\Omega|B) = 1$
- For a sequence of disjoint events A_k, $P(\cup A_k|B) = \sum P(A_k|B)$

- $P(A^c|B) = 1 - P(A|B)$
- $P(B \setminus A|C) = P(B|C) - P(A \cap B|C)$
- $P(A \cup B|C) = P(A|C) + P(B|C) - P(A \cap B|C)$
- If $A \subseteq B$, then $P(A|C) \le P(B|C)$

just an update in degree of belief!

Independence

DEFINITION 2.16 Events A and B are independent if occurrence of B does not affect the probability of A, i.e.,

$$\mathbf{P(A \mid B) = P(A)\,.}$$

According to this definition, conditional probability equals unconditional probability in case of independent events.

Substituting this into the intersection equation:

$$\mathbf{P(A \cap B) = P(A)P(B)\,.}$$

Independence

Example: Experiment is drawing a card from a deck:

- A={Drawing an ace}
- H={Drawing a heart}

Are A and H independent?

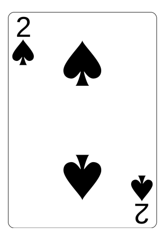
Independence

Example: Experiment is drawing a card from a deck:

- A={Drawing an ace}
- H={Drawing a heart}

Are A and H independent?

What if we remove 2 of spades from the deck and repeat the same experiment?



Independence

- If A and B are independent, A and B^c are also independent.

  We can prove that $P(A \cap B^c) = P(A)P(B^c)$

Independence

- If A and B are independent, A and B^c are also independent.

  We can prove that $P(A \cap B^c) = P(A)P(B^c)$:

$$P(A \cap B^c) = P(A \setminus B) = P(A) - P(A \cap B) = P(A) - P(A)P(B) = P(A)(1 - P(B)) = P(A)P(B^c)$$

- Are disjoint events independent?

Independence

- If A and B are independent, A and B^c are also independent.

  We can prove that $P(A \cap B^c) = P(A)P(B^c)$:

$$P(A \cap B^c) = P(A \setminus B) = P(A) - P(A \cap B) = P(A) - P(A)P(B) = P(A)(1 - P(B)) = P(A)P(B^c)$$

- Are disjoint events independent?

$$P(A \cap B) = 0 \rightarrow \left.\begin{array}{l} P(A|B) = 0 \ ?= P(A) \\ P(A)P(B) \ ?= 0 \end{array}\right\}$$

If neither of $P(A)$ or $P(B)$ is zero, they are absolutely **NOT** independent.

- Dependence goes in either direction, it is symmetric.

Independence

- Three events A, B and C are independent if:
  - They are pairwise independent.
  - $P(A \cap B \cap C) = P(A)P(B)P(C)$

Example: Experiment is 2 coin flips:

- A={heads in 1st flip}
- B={heads in 2nd flip}
- C={different results in 1st and 2nd flips}

Are A, B and C independent?

Conditional probability and independence

Example 2.31. Ninety percent of flights depart on time. Eighty percent of flights arrive on time. Seventy-five percent of flights depart on time and arrive on time.

Denote the events,

$$A = \{\text{arriving on time}\},$$

$$D = \{\text{departing on time}\}.$$

We have:

$$P(A) = 0.8,\ P(D) = 0.9,\ P(A \cap D) = 0.75.$$

Conditional probability and independence

Example 2.31. $A = \{\text{arriving on time}\}$, $D = \{\text{departing on time}\}$

$$P(A) = 0.8,\ P(D) = 0.9,\ P(A \cap D) = 0.75.$$

(a) Eric is meeting Alyssa's flight, which departed on time. What is the probability that Alyssa will arrive on time?

$$P(A \mid D) = \frac{P(A \cap D)}{P(D)} = \frac{0.75}{0.9} = 0.8333$$

Conditional probability and independence

Example 2.31. A = {arriving on time} , D = {departing on time}

$$P(A) = 0.8,\ P(D) = 0.9,\ P(A \cap D) = 0.75.$$

(a) Eric is meeting Alyssa's flight, which departed on time. What is the probability that Alyssa will arrive on time?

$$P(A|D) = \frac{P(A \cap D)}{P(D)} = \frac{0.75}{0.9} = 0.8333$$

(b) Eric has met Alyssa, and she arrived on time. What is the probability that her flight departed on time?

$$P(D|A) = \frac{P(D \cap A)}{P(A)} = \frac{0.75}{0.8} = 0.9375$$

Conditional probability and independence

Example 2.31. A = {arriving on time} , D = {departing on time}

$$P(A) = 0.8,\ P(D) = 0.9,\ P(A \cap D) = 0.75.$$

(c) Are the events, departing on time and arriving on time, independent?

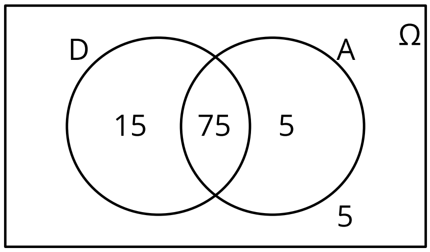
NO, because $P(A|D) \neq P(A)$, $P(D|A) \neq P(D)$, $P(A \cap D) \neq P(A)P(D)$

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

Bayes rule

The last example shows that two conditional probabilities, $P(A|D) \neq P(D|A)$, in general.

Bayes rule tries to find a relation between these two conditional probabilities.



Bayes rule

Example 2.32 There exists a test for a certain viral infection. It is 95% reliable for infected patients and 99% reliable for the healthy ones. That is, if a patient has the virus (V), the test shows that (S) with probability $P(S|V) = 0.95$, and if the patient does not have the virus, the test shows that with probability $P(\overline{S}|\overline{V}) = 0.99$.

Consider a patient whose test result is positive. 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, $P(V|S)$, is not stated among the given characteristics of this test.

Bayes rule

This example is applicable in numerous fields. The problem is to connect the given $P(S|V)$ and the quantity in question, $P(V|S)$. This was done in the 18th century by English minister **Thomas Bayes** (1702–1761) in the following way:

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

If we solve this for $P(B|A)$, we obtain the **Bayes Rule**:

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

Bayes rule

This example is applicable in numerous fields. The problem is to connect the given $P(S|V)$ and the quantity in question, $P(V|S)$. This was done in the 18th century by English minister **Thomas Bayes** (1702–1761) in the following way:

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

If we solve this for $P(B|A)$, we obtain the **Bayes Rule**:

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

likelihood of evidence

prior (before evidence)

posterior (after evidence)

normalizing constant:

$P(A)=P(A\cap B) + P(A\cap \overline{B})$

$P(A)=P(A|B)P(B)+P(A|\overline{B})P(\overline{B})$

Bayes rule

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?

Bayes rule

Example 2.33 - 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 $P(G|Z)=0.5$. If we apply the Bayes Rule, we obtain:

$$P(ZB|GGB)=\frac{P(GGB|ZB)P(ZB)}{P(GGB)}$$

Bayes rule

Example 2.33 - 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 $P(G|Z)=0.5$. If we apply the Bayes Rule, we obtain:

$$P(ZB|GGB)=\frac{P(GGB|ZB)P(ZB)}{P(GGB)}$$

Given ZB, event GGB occurs only when both X and Y write good exams. Thus,

$P(GGB|ZB)=(0.8)(0.7)=0.56$

Bayes rule

Example 2.33 - 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 $P(G|Z)=0.5$. If we apply the Bayes Rule, we obtain:

$$P(ZB|GGB)=\frac{\overset{\mathbf{0.56}}{P(GGB|ZB)}\,\overset{\mathbf{0.5}}{P(ZB)}}{P(GGB)}$$

Bayes rule

Example 2.33 - 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. If we apply the Bayes Rule, we obtain:

$$P(ZB|GGB)=\frac{\overset{\mathbf{0.56}}{P(GGB|ZB)}\overset{\mathbf{0.5}}{P(ZB)}}{P(GGB)}$$

Event GGB consists of three outcomes depending on the student who wrote the bad exam:

Law of Total Probability

$P(GGB)=P(XG\cap YG\cap ZB)+P(XG\cap YB\cap ZG)+P(XB\cap YG\cap ZG)$

$P(GGB)=(0.8)(0.7)(0.5)+(0.8)(0.3)(0.5)+(0.2)(0.7)(0.5)=0.47$

Bayes rule

Example 2.33 - 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. If we apply the Bayes Rule, we obtain:

$$P(ZB|GGB)=\frac{\overset{\mathbf{0.56}}{P(GGB|ZB)}\overset{\mathbf{0.5}}{P(ZB)}}{\underset{\mathbf{0.47}}{P(GGB)}}=0.5957$$

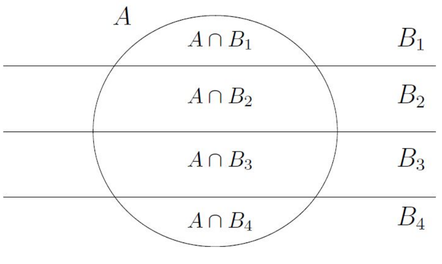
Law of total probability

This law relates the unconditional probability of an event A with its conditional probabilities. It is helpful when we have conditional probabilities of A given additional information.

Law of total probability

Consider some partition of Ω with **mutually exclusive and exhaustive** events $B_1,\ldots,B_k$. It means $B_i \cap B_j = \varnothing$ for any $i \neq j$ 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,

$$\mathbf{P(A) = \sum_{j=1:k} P(A \cap B_j)}$$



Law of total probability

$P(A) = \sum_{j=1:k} P(A \cap B_j) = \sum_{j=1:k} P(A|B_j)P(B_j)$

In case of 2 complementary events:

$P(A) = P(A|B)P(B) + P(A|\overline{B})P(\overline{B})$

Together with the Bayes Rule, it makes the following popular formula:

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

Back to Example 2.32

Example 2.32 There exists a test for a certain viral infection. It is 95% reliable for infected patients and 99% reliable for the healthy ones. That is, if a patient has the virus (V), the test shows that (S) with probability $P(S|V) = 0.95$, and if the patient does not have the virus, the test shows that with probability $P(\overline{S}|\overline{V}) = 0.99$.

Suppose that 4% of all the patients are infected with the virus, $P(V)=0.04$. $P(V|S)$ can be computed as **0.7983** as the following:

$$P(V|S) = \frac{P(S|V)P(V)}{P(S|V)P(V) + P(S|\overline{V})P(\overline{V})} = \frac{(0.95)(0.04)}{(0.95)(0.04)+(1-0.99)(1-0.04)}$$

Bayes rule and the law of total probability

Return to example 2.24 A family has two children. You met one of them, Lev, and he is a boy. What is the probability that the other child is also a boy?

Exercise 2.31. Let us define the sample space clearly. Suppose that one child is older, and the other is younger, their gender is independent of their age, and the child you met is one or the other with probabilities 1/2 and 1/2.

Bayes rule and the law of total probability

Return to example 2.24 A family has two children. You met one of them, Lev, and he is a boy. What is the probability that the other child is also a boy?

(a) List all the outcomes in this sample space. Each outcome should tell the children's gender, which child is older, and which child you have met.
(b) Show that **unconditional** probabilities of outcomes BB, BG, and GB are **equal**.
(c) Show that **conditional** probabilities of BB, BG, and GB, after you met Lev, are **not equal**.
(d) Show that Lev has a brother with conditional probability 1/2.

Bayes rule and the law of total probability

Exercise 2.31. (a) All outcomes are listed in the table below. According to the problem, they are equally likely.

Outcome	The older child	The younger child	Who is met
1	girl	girl	the older girl
2	girl	girl	the younger girl
3	girl	boy	the girl
4	girl	boy	the boy
5	boy	girl	the girl
6	boy	girl	the boy
7	boy	boy	the older boy
8	boy	boy	the younger boy

Bayes rule and the law of total probability

Exercise 2.31. (b) unconditional probabilities of outcomes BB, BG, and GB are equal.

Outcome	The older child	The younger child	Who is met
1	girl	girl	the older girl
2	girl	girl	the younger girl
3	girl	boy	the girl
4	girl	boy	the boy
5	boy	girl	the girl
6	boy	girl	the boy
7	boy	boy	the older boy
8	boy	boy	the younger boy

P(BB) = P{outcomes 7, 8} = 1/4
P(BG) = P{outcomes 5, 6} = 1/4
P(GB) = P{outcomes 3, 4} = 1/4

Bayes rule and the law of total probability

Exercise 2.31. (c) Conditional probabilities of BB, BG, and GB, after you met Lev, are not equal. Meeting Lev automatically eliminates outcomes 1, 2, 3 and 5. The remaining outcomes are:

Outcome	The older child	The younger child	Who is met
4	girl	boy	the boy
6	boy	girl	the boy
7	boy	boy	the older boy
8	boy	boy	the younger boy

P(BB|met Lev) = P{outcomes 7, 8 | met Lev } = 1/2
P(BG|met Lev) = P{outcome 6 | met Lev } = 1/4
P(GB|met Lev) = P{outcome 4 | met Lev } = 1/4

Bayes rule and the law of total probability

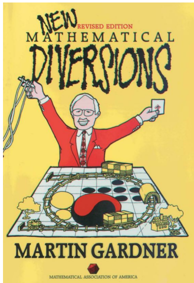
Exercise 2.31. (d) Lev has a brother with conditional probability 1/2.

Outcome	The older child	The younger child	Who is met
4	girl	boy	the boy
6	boy	girl	the boy
7	boy	boy	the older boy
8	boy	boy	the younger boy

P(Lev has a brother|met Lev) = P{outcome 7, 8 | met Lev } = 1/2

Bayes rule and the law of total probability

Exercise 2.31. (d) Lev has a brother with conditional probability 1/2.



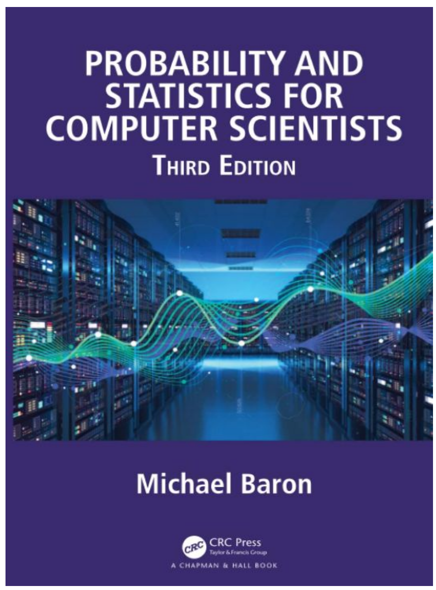
Outcome	The older child	The younger child	Who is met
4	girl	boy	the boy
6	boy	girl	the boy
7	boy	boy	the older boy
8	boy	boy	the younger boy

P(Lev has a brother|met Lev) = P{outcome 7, 8 | met Lev } = 1/2

References

WIKIPEDIA

Google images



Appendix


THE GREEK ALPHABET
Α α
Alpha
Β β
Beta
Γ γ
Gamma
Δ δ
Delta
Ε ε
Epsilon
Ζ ζ
Zeta
Η η
Eta
Θ θ
Theta
Ι ι
Iota
Κ κ
Kappa
Λ λ
Lambda
Μ μ
Mu
Ν ν
Nu
Ξ ξ
Xi
Ο ο
Omicron
Π π
Pi
Ρ ρ
Rho
Σ σ ς
Sigma
Τ τ
Tau
Υ υ
Upsilon
Φ φ
Phi
Χ χ
Chi
Ψ ψ
Psi
Ω ω
Omega