Probability 1

The ASTA team

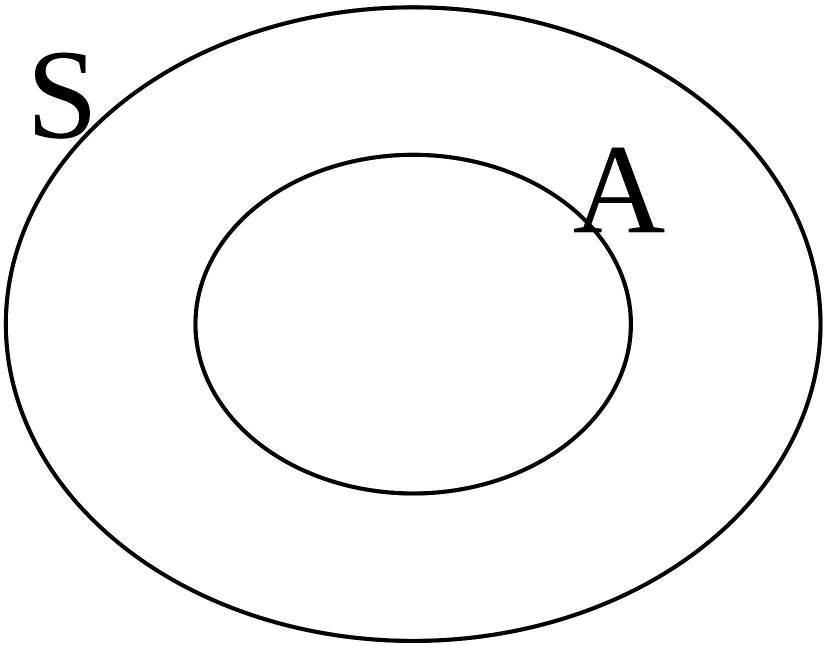
Contents

1 Introduction to probability

1.1 Events

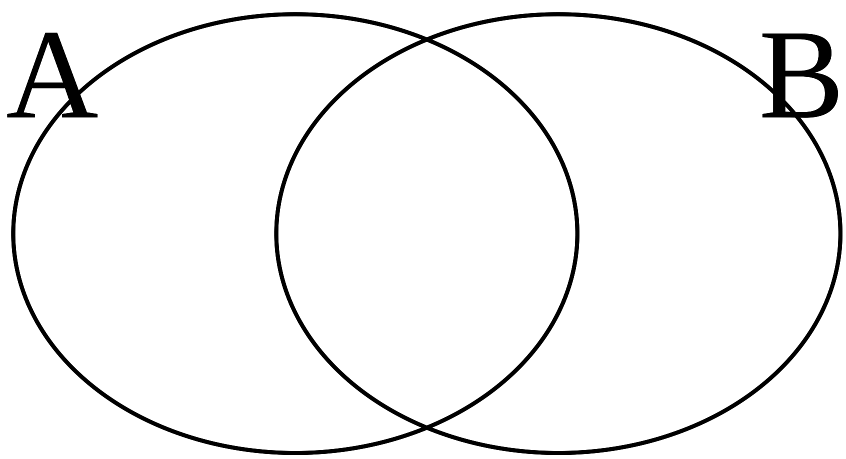
- Consider an experiment.
- The **state space** S is the set of all possible outcomes.
  - **Example:** We roll a die. The possible outcomes are $S = \{1, 2, 3, 4, 5, 6\}$.
  - **Example:** We measure wind speed (in m/s). The state space is $[0, \infty)$.
- An **event** is a subset $A \subseteq S$ of the sample space.

- **Example:** Rolling a die and getting an even number is the event $A = \{2, 4, 6\}$.
- **Example:** Measuring a wind speed of at least 5m/s is the event $[5, \infty)$.

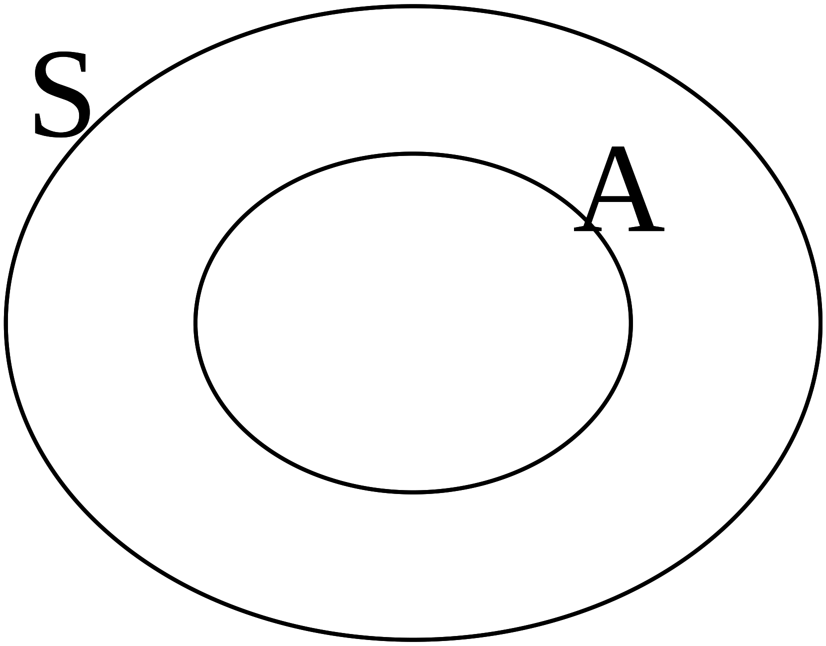


1.2 Combining events

- Consider two events A and B.
  - The **union** $A \cup B$ of is the event that either A or B occurs.
  - The **intersection** $A \cap B$ of is the event that both A and B occurs.



- The **complement** A^c of A of is the event that A does not occur.



- **Example:** We roll a die and consider the events $A = \{2, 4, 6\}$ that we get an even number and $B = \{4, 5, 6\}$ that we get at least 4. Then
  - $A \cup B = \{2, 4, 5, 6\}$
  - $A \cap B = \{4, 6\}$
  - $A^c = \{1, 3, 5\}$

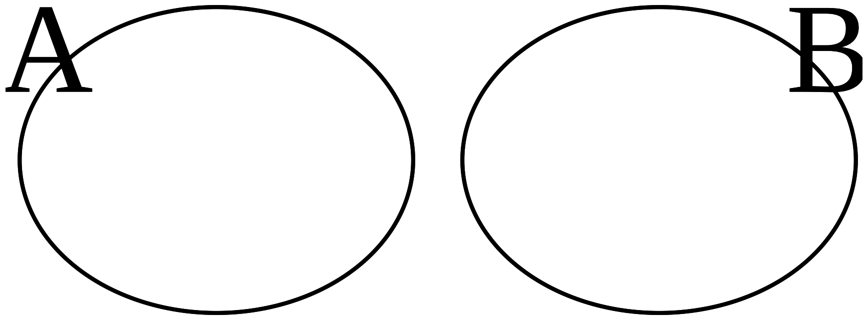
1.3 Probability of event

- The **probability** of an event is the proportion of times the event A would occur when the experiment is repeated many times.
- The probability of the event A is denoted $P(A)$.
  - **Example:** We throw a coin and consider the outcome $A = \{Head\}$. We expect to see the outcome Head half of the time, so $P(Head) = \frac{1}{2}$.
  - **Example:** We throw a die and consider the outcome $A = \{4\}$. Then $P(4) = \frac{1}{6}$.
- Properties:
  1. $P(S) = 1$
  2. $P(\emptyset) = 0$
  3. $0 \leq P(A) \leq 1$ for all events A

1.4 Probability of mutually exclusive events

- Consider two events A and B.
- If A and B are **mutually exclusive** (never occur at the same time, i.e. $A \cap B = \emptyset$), then

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



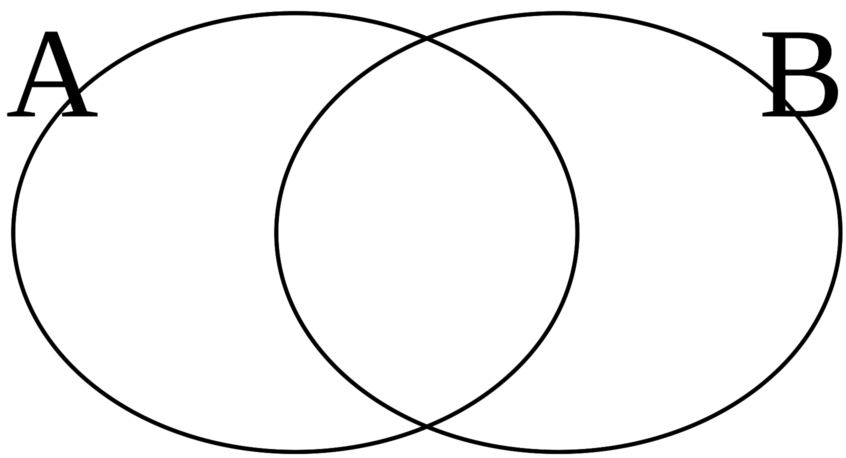
- **Example:** We roll a die and consider the events $A = \{1\}$ and $B = \{2\}$. Then

$$P(A \cup B) = P(A) + P(B) = \tfrac{1}{6} + \tfrac{1}{6} = \tfrac{1}{3}.$$

1.5 Probability of union

- For general events A an B,

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



- **Example:** We roll a die and consider the events $A = \{1, 2\}$ and $B = \{2, 3\}$. Then $A \cap B = \{2\}$, so

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \tfrac{1}{3} + \tfrac{1}{3} - \tfrac{1}{6} = \tfrac{1}{2}.$$

1.6 Probability of complement

- Since A and A^c are mutually exclusive with $A \cup A^c = S$, we get

$$1 = P(S) = P(A \cup A^c) = P(A) + P(A^c),$$

so

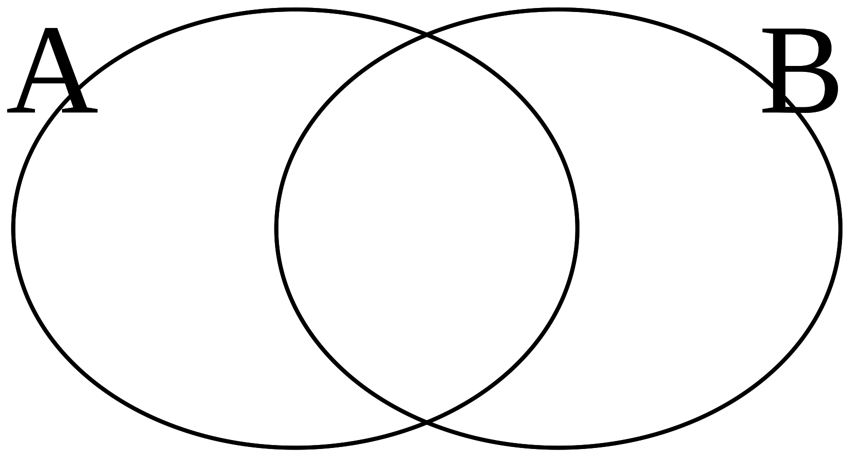
$$P(A^c) = 1 - P(A).$$

1.7 Conditional probability

- Consider events A and B.
- The **conditional probability** of A given B is defined by

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

if $P(B) > 0$.



- **Example:** We toss a coin two times. The possible outcomes are $S = \{HH, HT, TH, TT\}$. Each outcome has probability $\frac{1}{4}$. What is the probability of at least one head if we know there was at least one tail?
  - Let $A = \{\text{at least one H}\}$ and $B = \{\text{at least one T}\}$. Then

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

1.8 Independent events

- Two events A and B are said to be **independent** if

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

  - **Example:** Consider again a coin tossed two times with possible outcomes HH, HT, TH, TT.
    - Let $A = \{\text{at least one H}\}$ and $B = \{\text{at least one T}\}$.
    - We found that $P(A|B) = \frac{2}{3}$ while $P(A) = \frac{3}{4}$, so A and B are not independent.

1.9 Independent events - equivalent definition

- Two events A and B are said to be **independent** if and only if

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

- Proof: A and B are independent if and only if

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

  Multiplying by $P(B)$ we get $P(A)P(B) = P(A \cap B)$.

- **Example:** Roll a die and let $A = \{2, 4, 6\}$ be the event that we get an even number and $B = \{1, 2\}$ the event that we get at most 2. Then,
  - $P(A \cap B) = P(2) = \frac{1}{6}$
  - $P(A)P(B) = \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}$.
  - So A and B are independent.

2 Stochastic variables

2.1 Definition of stochastic variables

- A **stochastic variable** is a function that assigns a real number to every element of the state space.
  - **Example:** Throw a coin three times. The possible outcomes are

$$S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}.$$

    - The random variable X assigns to each outcome the number of heads, e.g.

$$X(HHH) = 3, \quad X(HTT) = 1.$$

  - **Example:** Consider the question whether a certain machine is defect. Define
    - $X = 0$ if the machine is not defect,
    - $X = 1$ if the machine is defect.
  - **Example:** X is the temperature in the lecture room.

2.2 Discrete or continuous stochastic variables

- A stochastic variable X may be
- **Discrete:** X can take a finite or infinite list of values.
  - **Examples:**
    - Number of heads in 3 coin tosses (can take values $0, 1, 2, 3$)
    - Number of machines that break down over a year (can take values $0, 1, 2, 3, \ldots$)
- **Continuous:** X takes values on a continuous scale.
  - **Examples:**
    - Temperature, speed, mass,...

3 Discrete random variables

3.1 Discrete random variables

- Let X be a discrete stochastic variable which can take the values $x_1, x_2, \ldots$
- The distribution of X is given by the **probability function**, which is given by

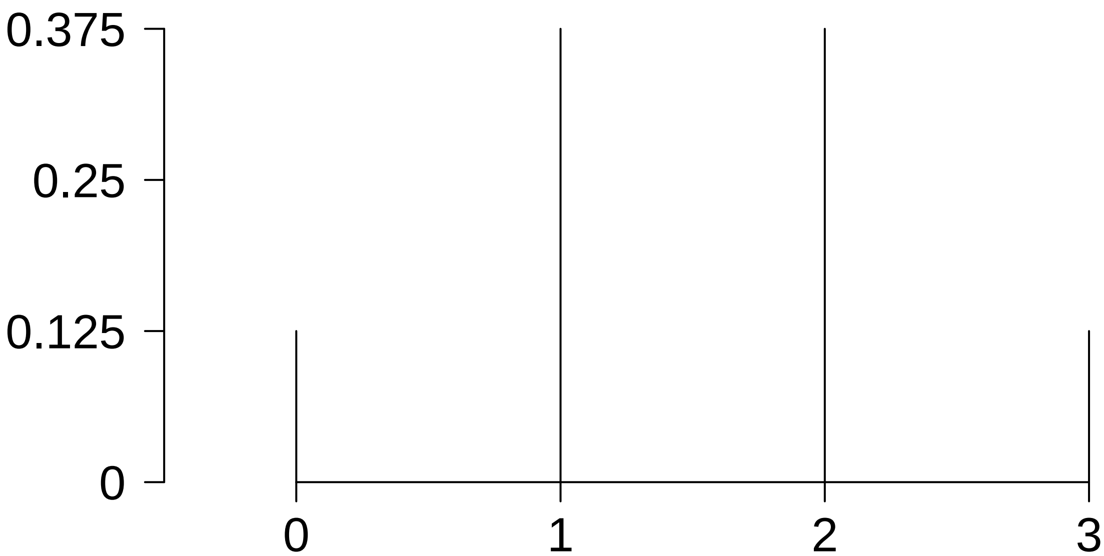
$$f(x_i) = P(X = x_i), \quad i = 1, 2, \ldots$$

- **Example:** We throw a coin three times and let X be the number of heads. The possible outcomes are

$$S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}.$$

The probability function is

  * $f(0) = P(X = 0) = \frac{1}{8}$
  * $f(1) = P(X = 1) = \frac{3}{8}$
  * $f(2) = P(X = 2) = \frac{3}{8}$
  * $f(3) = P(X = 3) = \frac{1}{8}$

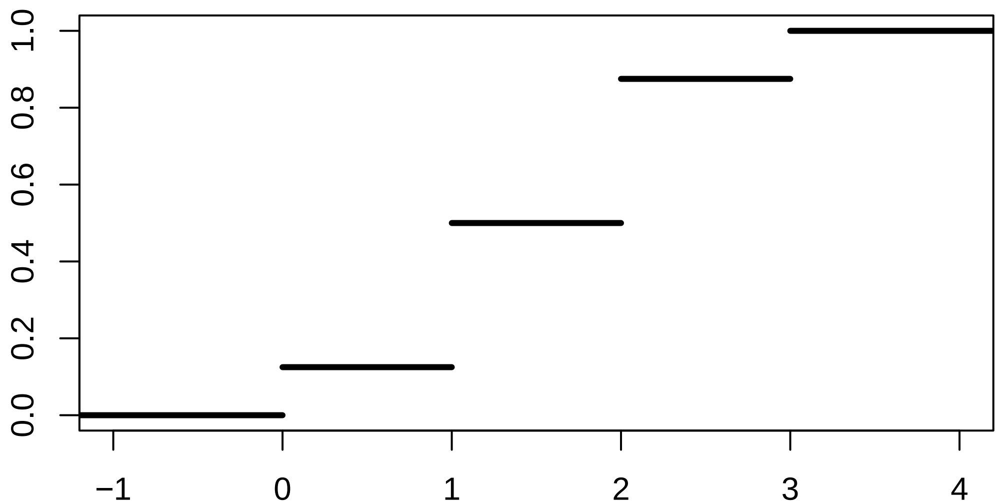


3.2 The distribution function

- Let X be a discrete random variable with probability function f. The **distribution function** of X is given by

$$F(x) = P(X \le x) = \sum_{y \le x} f(y), \quad x \in \mathbb{R}.$$

  - **Example:** For the three coin tosses, we have
    * $F(0) = P(X \le 0) = \frac{1}{8}$
    * $F(1) = P(X \le 1) = P(X = 0) + P(X = 1) = \frac{1}{2}$
    * $F(2) = P(X \le 2) = P(X = 0) + P(X = 1) + P(X = 2) = \frac{7}{8}$
    * $F(3) = P(X \le 3) = 1$



- For a discrete variable, the result is an increasing step function.

3.3 Mean of a discrete variable

- The **mean** or **expected value** of a discrete random variable X with values $x_1, x_2, \ldots$ and probability function $f(x_i)$ is
$$\mu = E(X) = \sum_i x_i P(X = x_i) = \sum_i x_i f(x_i).$$
- Interpretation: A weighted average of the possible values of X, where each value is weighted by its probability. A sort of "center" value for the distribution.
  - **Example:** Toss a coin 3 times. What are the expected number of heads?

    $E(X) = 0 \cdot P(X = 0) + 1 \cdot P(X = 1) + 2 \cdot P(X = 2) + 3 \cdot P(X = 3) = 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = 1.5.$

3.4 Variance of a discrete variable

- The **variance** is the mean squared distance between the values of the variable and the mean value. More precisely,
$$\sigma^2 = \sum_i (x_i - \mu)^2 P(X = x_i) = \sum_i (x_i - \mu)^2 f(x_i).$$
- A high variance indicates that the values of X have a high probability of being far from the mean values.
- The **standard deviation** is the square root of the variance
$$\sigma = \sqrt{\sigma^2}.$$
- The advantage of the standard deviation over the variance is that it is measured in the same units as X.
  - **Example** Let X be the number of heads in 3 coin tosses. What is the variance and standard deviation?
    - Solution: The mean was found to be 1.5. Thus,

      $\sigma^2 = (0-1.5)^2 \cdot f(0) + (1-1.5)^2 \cdot f(1) + (2-1.5)^2 \cdot f(2) + (3-1.5)^2 \cdot f(3) = (0-1.5)^2 \cdot \frac{1}{8} + (1-1.5)^2 \cdot \frac{3}{8} + (2-1.5)^2 \cdot \frac{3}{8} + (3-$

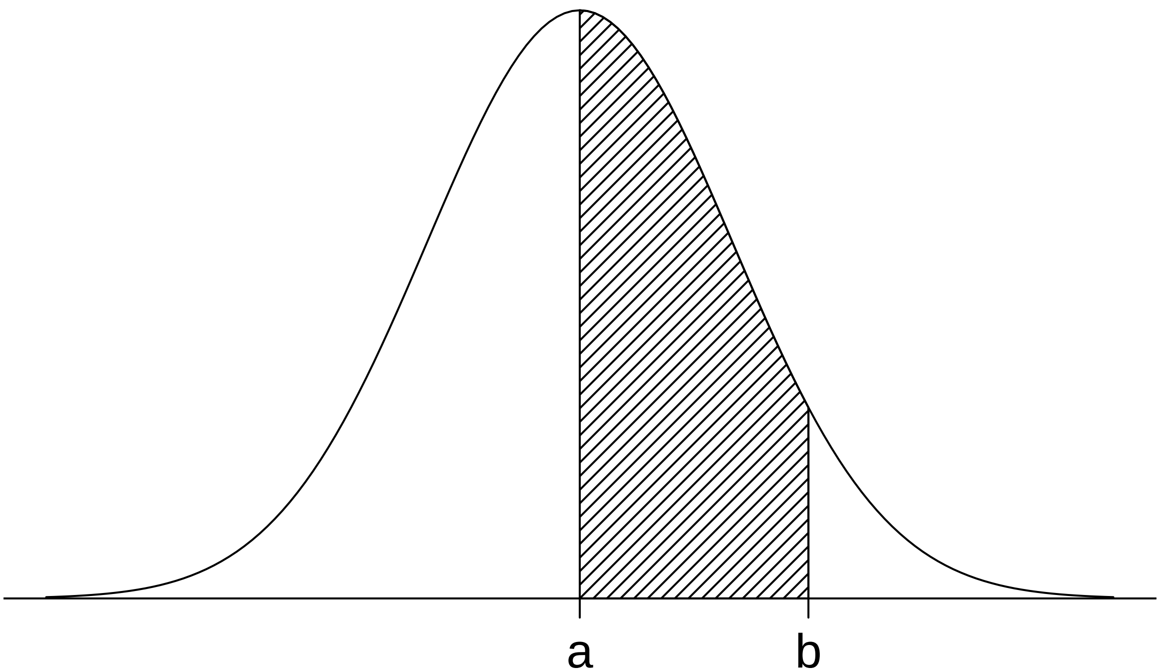
      The standard deviation is $\sigma = \sqrt{0.75} \approx 0.866$.

4 Continuous random variables

4.1 Distribution of continuous random variables

- The distribution of a continuous random variable X is given by a **probability density function** f, which is a function satisfying
  1. $f(x)$ is defined for all x in $\mathbb{R}$,
  2. $f(x) \geq 0$ for all x in $\mathbb{R}$,
  3. $\int_{-\infty}^{\infty} f(x)dx = 1$.
- The probability that X lies between the values a and b is given by

$$P(a < X < b) = \int_a^b f(x)dx.$$

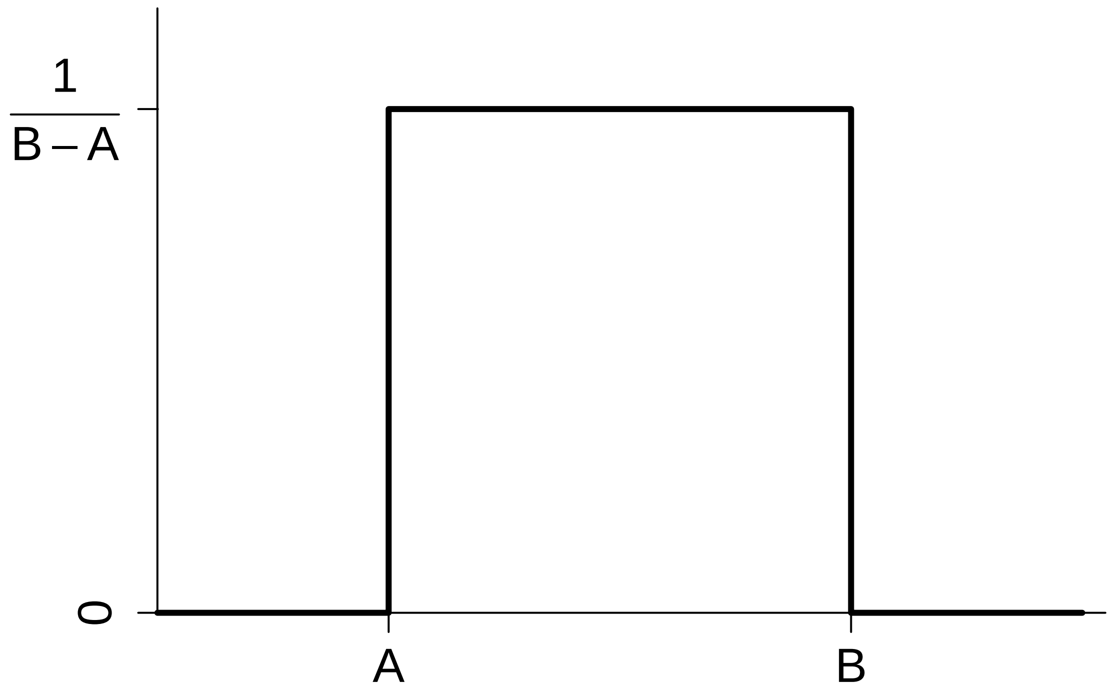


- Notes:
  - Condition 3. ensures that $P(-\infty < X < \infty) = 1$.
  - The probability of X assuming a specific value a is zero, i.e. $P(X = a) = 0$.

4.2 Example: The uniform distribution

- The **uniform distribution** on the interval (A, B) has density

$$f(x) = \begin{cases} \frac{1}{B-A} & A \leq x \leq B \\ 0 & \text{otherwise} \end{cases}$$

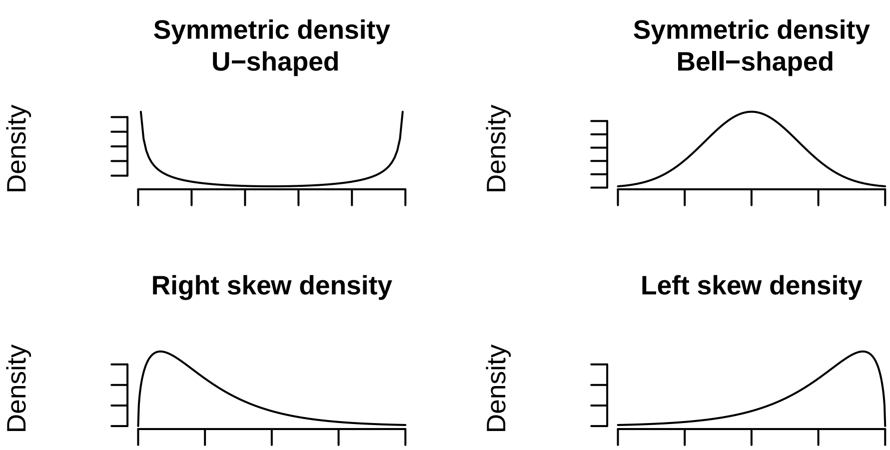


– **Example:** If X has a uniform distribution on $(0,1)$, find $P(\frac{1}{3} < X \leq \frac{2}{3})$.

  ∗ Solution:

$$P\left(\tfrac{1}{3} < X \leq \tfrac{2}{3}\right) = P\left(\tfrac{1}{3} < X < \tfrac{2}{3}\right) + P\left(X = \tfrac{2}{3}\right) = \int_{1/3}^{2/3} f(x)dx + 0 = \int_{1/3}^{2/3} 1dx = \frac{1}{3}.$$

4.3 Density shapes



4.4 Distribution function of continuous variable

- Let X be a continuous random variable with probability density f. The **distribution function** of X is given by

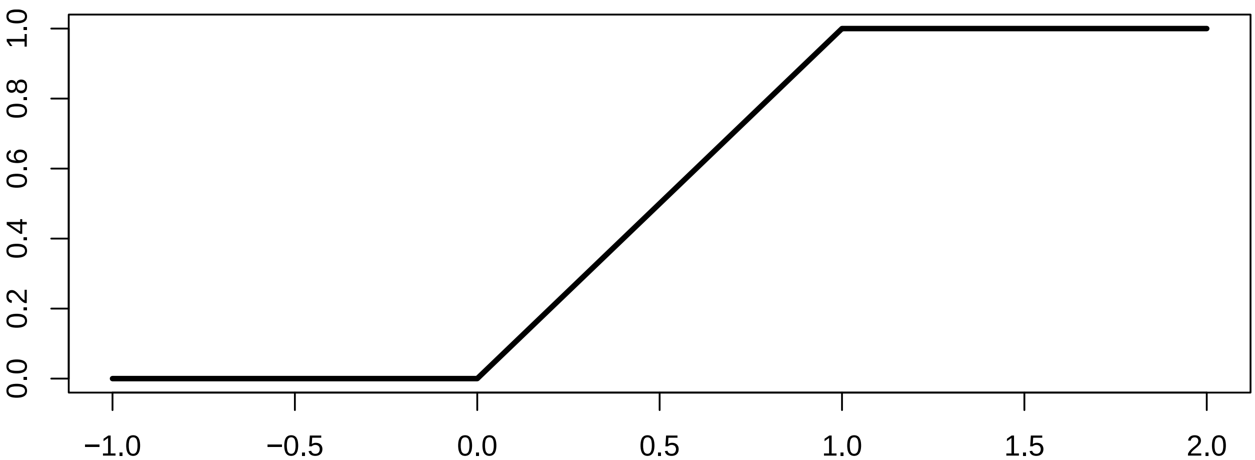
$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(y)dy, \quad x \in \mathbb{R}.$$

– **Example:** For the uniform distribution on $[0,1]$, the density was

$$f(x) = \begin{cases} 1, & 0 \le x \le 1, \\ 0, & \text{otherwise.} \end{cases}$$

Hence,

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(y)dy = \int_{0}^{x} 1dy = x, \quad x \in [0,1].$$



4.5 Mean and variance of a continuous variable

- The **mean** or **expected value** of a continuous random variable X is

$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x)dx.$$

- The **variance** is given by

$$\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx.$$

  – In calculations, it is often more convenient to use the formula

$$\sigma^2 = E(X^2) - E(X)^2 = \int_{-\infty}^{\infty} x^2 f(x)dx - \mu^2.$$

4.5.1 Example: Mean and variance in the uniform distribution

- Consider again the uniform distribution on the interval $(0,1)$ with density

$$f(x) = \begin{cases} 1 & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

  Find the mean and variance.

- **Solution:** The mean is

$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x)dx = \int_{0}^{1} x \cdot 1dx = \left[\tfrac{1}{2}x^2\right]_0^1 = \tfrac{1}{2},$$

  and the variance is computed using the formula

$$\sigma^2 = E(X^2) - E(X)^2 = \int_{-\infty}^{\infty} x^2 f(x)dx - \mu^2 = \int_{0}^{1} x^2 dx - \mu^2 = \left[\tfrac{1}{3}x^3\right]_0^1 - \left(\tfrac{1}{2}\right)^2 = \tfrac{1}{3} - \tfrac{1}{4} = \tfrac{1}{12}.$$

4.6 Rules for computing mean and variance

- Let X be a random variable and a, b be constants. Then,

  1. $E(aX + b) = aE(X) + b$.
  2. $\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$.

  - **Example:** If X has mean μ and variance σ^2, then

    * $E\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma}E(X-\mu) = \frac{1}{\sigma}(E(X)-\mu) = 0$,
    * $\mathrm{Var}\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma^2}\mathrm{Var}(X-\mu) = \frac{1}{\sigma^2}\mathrm{Var}(X) = \frac{1}{\sigma^2}\sigma^2 = 1$.
    * So $\frac{X-\mu}{\sigma}$ is a standardization of X that has mean 0 and variance 1.