4. Basic probability theory

Sample space, sample points, events

- **Sample space** $\Omega$ is the set of all possible sample points $\omega \in \Omega$
  - Example 0. Tossing a coin: $\Omega = \{H,T\}$
  - Example 1. Casting a die: $\Omega = \{1,2,3,4,5,6\}$
  - Example 2. Number of customers in a queue: $\Omega = \{0,1,2,...\}$
  - Example 3. Call holding time (e.g. in minutes): $\Omega = \{x \in \mathbb{R} \mid x > 0\}$
- **Events** $A,B,C,... \subset \Omega$ are measurable subsets of the sample space $\Omega$
  - Example 1. “Even numbers of a die”: $A = \{2,4,6\}$
  - Example 2. “No customers in a queue”: $A = \{0\}$
  - Example 3. “Call holding time greater than 3.0 (min)”: $A = \{x \in \mathbb{R} \mid x > 3.0\}$
- Denote by $\mathcal{F}$ the set of all events $A \in \mathcal{F}$
  - Sure event: The sample space $\Omega \in \mathcal{F}$ itself
  - Impossible event: The empty set $\emptyset \in \mathcal{F}$

Combination of events

- **Union** “$A$ or $B$”:
  $$A \cup B = \{\omega \in \Omega \mid \omega \in A \text{ or } \omega \in B\}$$
- **Intersection** “$A$ and $B$”:
  $$A \cap B = \{\omega \in \Omega \mid \omega \in A \text{ and } \omega \in B\}$$
- **Complement** “not $A$”:
  $$A^c = \{\omega \in \Omega \mid \omega \notin A\}$$
- Events $A$ and $B$ are disjoint if
  - $A \cap B = \emptyset$
- A set of events $\{B_1, B_2, ...\}$ is a partition of event $A$ if
  - (i) $B_i \cap B_j = \emptyset$ for all $i \neq j$
  - (ii) $\bigcup_{i} B_i = A$
- Example 1. Odd and even numbers of a die constitute a partition of the sample space:
  $B_1 = \{1,3,5\}$ and $B_2 = \{2,4,6\}$
4. Basic probability theory

**Probability**

- **Probability** of event $A$ is denoted by $P(A)$, $P(A) \in [0,1]$
  - Probability measure $P$ is thus a real-valued set function defined on the set of events $\mathcal{F}$, $P : \mathcal{F} \to [0,1]$
- **Properties**:
  - (i) $0 \leq P(A) \leq 1$
  - (ii) $P(\emptyset) = 0$
  - (iii) $P(\Omega) = 1$
  - (iv) $P(A^c) = 1 - P(A)$
  - (v) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
  - (vi) $A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$
  - (vii) $\{B_i\}$ is a partition of $A \Rightarrow P(A) = \sum_i P(B_i)$
  - (viii) $A \subset B \Rightarrow P(A) \leq P(B)$

4. Basic probability theory

**Conditional probability**

- Assume that $P(B) > 0$
- **Definition**: The conditional probability of event $A$ given that event $B$ occurred is defined as
  
  $$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

- It follows that
  
  $$P(A \cap B) = P(B)P(A \mid B) = P(A)P(B \mid A)$$

4. Basic probability theory

**Theorem of total probability**

- Let $\{B_i\}$ be a partition of the sample space $\Omega$
- It follows that $\{A \cap B_i\}$ is a partition of event $A$. Thus (by slide 5)
  
  $$P(A) = \sum_i P(A \cap B_i)$$

- Assume further that $P(B_i) > 0$ for all $i$. Then (by slide 6)
  
  $$P(A) = \sum_i P(B_i)P(A \mid B_i)$$

- This is the **theorem of total probability**

4. Basic probability theory

**Bayes’ theorem**

- Let $\{B_i\}$ be a partition of the sample space $\Omega$
- Assume that $P(A) > 0$ and $P(B_i) > 0$ for all $i$. Then (by slide 6)
  
  $$P(B_i \mid A) = \frac{P(A \cap B_i)}{P(A)} = \frac{P(B_i)P(A \mid B_i)}{P(A)}$$

- Furthermore, by the theorem of total probability (slide 7), we get
  
  $$P(B_i \mid A) = \frac{P(B_i)P(A \mid B_i)}{\sum_j P(B_j)P(A \mid B_j)}$$

- This is **Bayes’ theorem**
  - Probabilities $P(B_i)$ are called **a priori** probabilities of events $B_i$
  - Probabilities $P(B_i \mid A)$ are called **a posteriori** probabilities of events $B_i$ (given that the event $A$ occurred)
Statistical independence of events

- **Definition**: Events $A$ and $B$ are independent if
  \[ P(A \cap B) = P(A)P(B) \]
  It follows that
  \[ P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A) \]
  Correspondingly:
  \[ P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A)P(B)}{P(A)} = P(B) \]

Example

- A coin is tossed three times
  - Sample space:
    \[ \Omega = \{ (\omega_1, \omega_2, \omega_3) \mid \omega_i \in \{H, T\}, i = 1, 2, 3 \} \]
  - Let $X$ be the random variable that tells the total number of tails in these three experiments:

<table>
<thead>
<tr>
<th>$\omega$</th>
<th>HHH</th>
<th>HHT</th>
<th>HTH</th>
<th>THH</th>
<th>HTT</th>
<th>THT</th>
<th>TTH</th>
<th>TTT</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X(\omega)$</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

Random variables

- **Definition**: Real-valued random variable $X$ is a real-valued and measurable function defined on the sample space $\Omega$, $X : \Omega \to \mathbb{R}$
  - Each sample point $\omega \in \Omega$ is associated with a real number $X(\omega)$
  - **Measurability** means that all sets of type
    \[ \{ X \leq x \} := \{ \omega \in \Omega \mid X(\omega) \leq x \} \subset \Omega \]
    belong to the set of events $\mathcal{F}$, that is
    \[ \{ X \leq x \} \in \mathcal{F} \]
    The probability of such an event is denoted by $P\{ X \leq x \}$

Indicators of events

- Let $A \in \mathcal{F}$ be an arbitrary event
  - **Definition**: The indicator of event $A$ is a random variable defined as follows:
    \[ 1_A(\omega) = \begin{cases} 1, & \omega \in A \\ 0, & \omega \notin A \end{cases} \]
  - Clearly:
    \[ P\{ 1_A = 1 \} = P(A) \]
    \[ P\{ 1_A = 0 \} = P(A^c) = 1 - P(A) \]
4. Basic probability theory

Cumulative distribution function

- **Definition**: The cumulative distribution function (cdf) of a random variable $X$ is a function $F_X: \mathbb{R} \rightarrow [0,1]$ defined as follows:
  
  $$F_X(x) = P\{X \leq x\}$$

- Cdf determines the distribution of the random variable, that is: the probabilities $P\{X \in B\}$, where $B \subset \mathbb{R}$ and $\{X \in B\} \in \mathcal{F}$

- **Properties**:
  - (i) $F_X$ is non-decreasing
  - (ii) $F_X$ is continuous from the right
  - (iii) $F_X(-\infty) = 0$
  - (iv) $F_X(\infty) = 1$

Statistical independence of random variables

- **Definition**: Random variables $X$ and $Y$ are independent if for all $x$ and $y$
  $$P\{X \leq x, Y \leq y\} = P\{X \leq x\}P\{Y \leq y\}$$

- **Definition**: Random variables $X_1, \ldots, X_n$ are totally independent if for all $i$ and $x_i$
  $$P\{X_1 \leq x_1, \ldots, X_n \leq x_n\} = P\{X_1 \leq x_1\} \cdots P\{X_n \leq x_n\}$$

Maximum and minimum of independent random variables

- Let the random variables $X_1, \ldots, X_n$ be totally independent
- Denote: $X_{\text{max}} := \max\{X_1, \ldots, X_n\}$. Then
  $$P\{X_{\text{max}} \leq x\} = P\{X_1 \leq x, \ldots, X_n \leq x\} = P\{X_1 \leq x\} \cdots P\{X_n \leq x\}$$

- Denote: $X_{\text{min}} := \min\{X_1, \ldots, X_n\}$. Then
  $$P\{X_{\text{min}} > x\} = P\{X_1 > x, \ldots, X_n > x\} = P\{X_1 > x\} \cdots P\{X_n > x\}$$

Contents

- Basic concepts
- Discrete random variables
  - Discrete distributions (nbr distributions)
- Continuous random variables
  - Continuous distributions (time distributions)
- Other random variables
Discrete random variables

- **Definition:** Set \( A \subset \mathbb{R} \) is called *discrete* if it is
  - finite, \( A = \{x_1, \ldots, x_n\} \), or
  - countably infinite, \( A = \{x_1, x_2, \ldots\} \)

- **Definition:** Random variable \( X \) is discrete if there is a discrete set \( S_X \subset \mathbb{R} \) such that

\[
P(X \in S_X) = 1
\]

- It follows that
  - \( P(X = x) \geq 0 \) for all \( x \in S_X \)
  - \( P(X = x) = 0 \) for all \( x \not\in S_X \)
- The set \( S_X \) is called the **value set**

Point probabilities

- Let \( X \) be a discrete random variable
- The distribution of \( X \) is determined by the **point probabilities** \( p_i \)

\[
p_i := P\{X = x_i\}, \quad x_i \in S_X
\]

- **Definition:** The probability mass function (pmf) of \( X \) is a function \( p_X : \mathbb{R} \to [0,1] \) defined as follows:

\[
p_X(x) := P\{X = x\} = \begin{cases} p_i, & x = x_i \in S_X \vspace{1em} \\ 0, & x \not\in S_X \end{cases}
\]

- Cdf is in this case a step function:

\[
F_X(x) = P\{X \leq x\} = \sum_{i: x_i \leq x} p_i
\]

Example

\[
p_X(x) \quad F_X(x)
\]

- **Probability mass function (pmf)**
  - \( x_1, x_2, x_3, x_4 \)

- **Cumulative distribution function (cdf)**
  - \( x_1, x_2, x_3, x_4 \)

\[
S_X = \{x_1, x_2, x_3, x_4\}
\]

Independence of discrete random variables

- Discrete random variables \( X \) and \( Y \) are independent if and only if for all \( x_i \in S_X \) and \( y_j \in S_Y \)

\[
P\{X = x_i, Y = y_j\} = P\{X = x_i\}P\{Y = y_j\}
\]
4. Basic probability theory

**Expectation**

- **Definition**: The expectation (mean value) of $X$ is defined by
  \[
  \mu_X := E[X] := \sum_{x \in S_X} P\{X = x\} \cdot x = \sum_{i} p_i x_i
  \]
  
  - Note 1: The expectation exists only if $\sum p_i |x_i| < \infty$
  - Note 2: If $\sum p_i x_i = \infty$, then we may denote $E[X] = \infty$

- **Properties**:
  - (i) $c \in \mathbb{R} \Rightarrow E[cX] = cE[X]$
  - (ii) $E[X + Y] = E[X] + E[Y]$
  - (iii) $X$ and $Y$ independent $\Rightarrow E[XY] = E[X]E[Y]$

**Variance**

- **Definition**: The variance of $X$ is defined by
  \[
  \sigma^2_X := D^2[X] := \text{Var}[X] := E[(X - E[X])^2]
  \]
  
  - Useful formula (prove!):
    \[
    D^2[X] = E[X^2] - E[X]^2
    \]

- **Properties**:
  - (i) $c \in \mathbb{R} \Rightarrow D^2[cX] = c^2D^2[X]$
  - (ii) $X$ and $Y$ independent $\Rightarrow D^2[X + Y] = D^2[X] + D^2[Y]$

**Covariance**

- **Definition**: The covariance between $X$ and $Y$ is defined by
  \[
  \sigma^2_{XY} := \text{Cov}[X,Y] := E[(X - E[X])(Y - E[Y])]
  \]
  
  - Useful formula (prove!):
    \[
    \]

- **Properties**:
  - (i) $\text{Cov}[X,X] = \text{Var}[X]$
  - (ii) $\text{Cov}[X,Y] = \text{Cov}[Y,X]$
  - (iii) $\text{Cov}[X+Y,Z] = \text{Cov}[X,Z] + \text{Cov}[Y,Z]$
  - (iv) $X$ and $Y$ independent $\Rightarrow \text{Cov}[X,Y] = 0$

**Other distribution related parameters**

- **Definition**: The standard deviation of $X$ is defined by
  \[
  \sigma_X := D[X] := \sqrt{D^2[X]}
  \]

- **Definition**: The coefficient of variation of $X$ is defined by
  \[
  c_X := C[X] := \frac{D[X]}{E[X]}
  \]

- **Definition**: The $k$th moment, $k=1,2,...$, of $X$ is defined by
  \[
  \mu_X^{(k)} := E[X^k]
  \]
Average of IID random variables

• Let $X_1, \ldots, X_n$ be independent and identically distributed (IID) with mean $\mu$ and variance $\sigma^2$
• Denote the average (sample mean) as follows:
  $$\bar{X}_n := \frac{1}{n} \sum_{i=1}^{n} X_i$$
• Then (prove!)
  $$E[\bar{X}_n] = \mu$$
  $$D^2[\bar{X}_n] = \frac{\sigma^2}{n}$$
  $$D[\bar{X}_n] = \frac{\sigma}{\sqrt{n}}$$

Law of large numbers (LLN)

• Let $X_1, \ldots, X_n$ be independent and identically distributed (IID) with mean $\mu$ and variance $\sigma^2$
• Weak law of large numbers: for all $\varepsilon > 0$
  $$P(\left| \bar{X}_n - \mu \right| > \varepsilon) \to 0$$
• Strong law of large numbers: with probability 1
  $$\bar{X}_n \to \mu$$
• It follows that for large values of $n$
  $$\bar{X}_n \approx \mu$$

Bernoulli distribution

$$X \sim \text{Bernoulli}(p), \quad p \in (0,1)$$

• describes a simple random experiment with two possible outcomes: success (1) and failure (0); cf. coin tossing
• success with probability $p$ (and failure with probability $1-p$)
• Value set: $S_X = \{0,1\}$
• Point probabilities:
  $$P\{X = 0\} = 1 - p, \quad P\{X = 1\} = p$$
• Mean value: $E[X] = (1-p)\cdot 0 + p \cdot 1 = p$
• Second moment: $E[X^2] = (1-p)\cdot 0^2 + p \cdot 1^2 = p$
• Variance: $D^2[X] = E[X^2] - E[X]^2 = p - p^2 = p(1 - p)$
### Binomial distribution

\( X \sim \text{Bin}(n, p), \ n \in \{1, 2, \ldots\}, \ p \in (0, 1) \)

- number of successes in an independent series of simple random experiments (of Bernoulli type); \( X = X_1 + \ldots + X_n \) (with \( X_i \sim \text{Bernoulli}(p) \))
- \( n = \) total number of experiments
- \( p = \) probability of success in any single experiment

**Value set:** \( S_X = \{0, 1, \ldots, n\} \)

**Point probabilities:**

\[
P\{X = i\} = \binom{n}{i} p^i (1 - p)^{n-i}
\]

- Mean value: \( E[X] = E[X_1] + \ldots + E[X_n] = np \)
- Variance: \( D^2[X] = D^2[X_1] + \ldots + D^2[X_n] = np(1 - p) \) (independence!)

### Geometric distribution

\( X \sim \text{Geom}(p), \ p \in (0, 1) \)

- number of successes until the first failure in an independent series of simple random experiments (of Bernoulli type)
- \( p = \) probability of success in any single experiment

**Value set:** \( S_X = \{0, 1, \ldots\} \)

**Point probabilities:**

\[
P\{X = i\} = p^i (1 - p)
\]

- Mean value: \( E[X] = \sum_i i p^i (1 - p) = p/(1 - p) \)
- Second moment: \( E[X^2] = \sum_i i^2 p^i (1 - p) = 2(p/(1 - p))^2 + p(1 - p) \)
- Variance: \( D^2[X] = E[X^2] - E[X]^2 = p/(1 - p)^2 \)

### Memoryless property of geometric distribution

- Geometric distribution has so called **memoryless property:** for all \( i, j \in \{0, 1, \ldots\} \)

\[
P\{X \geq i + j \mid X \geq i\} = P\{X \geq j\}
\]

- Prove!
  - **Tip:** Prove first that \( P\{X \geq i\} = p^i \)

### Minimum of geometric random variables

- Let \( X_1 \sim \text{Geom}(p_1) \) and \( X_2 \sim \text{Geom}(p_2) \) be independent. Then

\[
X_{\min} := \min\{X_1, X_2\} \sim \text{Geom}(p_1 p_2)
\]

and

\[
P\{X_{\min} = i\} = \frac{1 - p_i}{1 - p_1 p_2}, \ i \in \{1, 2\}
\]

- Prove!
  - **Tip:** See slide 15
### Poisson distribution

\[ X \sim \text{Poisson}(a), \quad a > 0 \]

- Limit of binomial distribution as \( n \to \infty \) and \( p \to 0 \) in such a way that \( np \to a \)
- Value set: \( S_X = \{0,1,\ldots\} \)
- Point probabilities:
  \[ P\{X = i\} = \frac{a^i}{i!} e^{-a} \]
- Mean value: \( E[X] = a \)
- Second moment: \( E[X(X-1)] = a^2 \Rightarrow E[X^2] = a^2 + a \)
- Variance: \( D^2[X] = E[X^2] - E[X]^2 = a \)

### Example

- Assume that
  - 200 subscribers are connected to a local exchange
  - each subscriber’s characteristic traffic is 0.01 erlang
  - subscribers behave independently
- Then the number of active calls \( X \sim \text{Bin}(200,0.01) \)
- Corresponding Poisson-approximation \( X \approx \text{Poisson}(2.0) \)
- Point probabilities:

  \[
  \begin{array}{c|ccccccc}
  \text{Bin(200,0.01)} & 0 & 1 & 2 & 3 & 4 & 5 \\
  \text{Poisson(2.0)} & 0.1326 & 0.2679 & 0.2693 & 0.1795 & 0.0893 & 0.0354
  \end{array}
  \]

### Properties

- (i) **Sum**: Let \( X_1 \sim \text{Poisson}(a_1) \) and \( X_2 \sim \text{Poisson}(a_2) \) be independent. Then
  \[ X_1 + X_2 \sim \text{Poisson}(a_1 + a_2) \]
- (ii) **Random sample**: Let \( X \sim \text{Poisson}(a) \) denote the number of elements in a set, and \( Y \) denote the size of a random sample of this set (each element taken independently with probability \( p \)). Then
  \[ Y \sim \text{Poisson}(pa) \]
- (iii) **Random sorting**: Let \( X \) and \( Y \) be as in (ii), and \( Z = X - Y \). Then \( Y \) and \( Z \) are independent (given that \( X \) is unknown) and
  \[ Z \sim \text{Poisson}((1-p)a) \]

### Contents

- Basic concepts
- Discrete random variables
- Discrete distributions (nbr distributions)
- Continuous random variables
- Continuous distributions (time distributions)
- Other random variables
Continuous random variables

- **Definition:** Random variable $X$ is continuous if there is an integrable function $f_X: \mathbb{R} \rightarrow \mathbb{R}_+$ such that for all $x \in \mathbb{R}$

$$F_X(x) := P\{X \leq x\} = \int_{-\infty}^{x} f_X(y) \, dy$$

- The function $f_X$ is called the probability density function (pdf)
  - The set $S_X$, where $f_X > 0$, is called the value set
- Properties:
  - (i) $P\{X = x\} = 0$ for all $x \in \mathbb{R}$
  - (ii) $P\{a < X \leq b\} = P\{a \leq X \leq b\} = \int_{a}^{b} f_X(x) \, dx$
  - (iii) $P\{X \in A\} = \int_{A} f_X(x) \, dx$
  - (iv) $P\{X \in \mathbb{R}\} = \int_{-\infty}^{\infty} f_X(x) \, dx = 1$

Example

- **Probability density function (pdf)**
- **Cumulative distribution function (cdf)**

$$S_X = [x_1, x_3]$$

Expectation and other distribution related parameters

- **Definition:** The expectation (mean value) of $X$ is defined by

$$\mu_X := E[X] := \int_{-\infty}^{\infty} f_X(x) \, dx$$

- Note 1: The expectation exists only if $\int_{-\infty}^{\infty} f_X(x) |x| \, dx < \infty$
- Note 2: If $\int_{-\infty}^{\infty} f_X(x) |x| \, dx = \infty$, then we may denote $E[X] = \infty$
  - The expectation has the same properties as in the discrete case (see slide 21)
- The other distribution parameters (variance, covariance,...) are defined just as in the discrete case
  - These parameters have the same properties as in the discrete case (see slides 22-24)

Contents

- Basic concepts
- Discrete random variables
- Discrete distributions (nbr distributions)
- Continuous random variables
- Continuous distributions (time distributions)
- Other random variables
4. Basic probability theory

**Uniform distribution**

\[ X \sim U(a, b), \quad a < b \]

- continuous counterpart of “casting a die”
- Value set: \( S_X = (a, b) \)
- Probability density function (pdf):
  \[ f_X(x) = \frac{1}{b-a}, \quad x \in (a, b) \]
- Cumulative distribution function (cdf):
  \[ F_X(x) := P\{X \leq x\} = \frac{x-a}{b-a}, \quad x \in (a, b) \]
- Mean value: \( E[X] = \int_a^b x/(b-a) \, dx = (a+b)/2 \)
- Second moment: \( E[X^2] = \int_a^b x^2/(b-a) \, dx = (a^2 + ab + b^2)/3 \)
- Variance: \( D^2[X] = E[X^2] - E[X]^2 = (b-a)^2/12 \)

**Exponential distribution**

\[ X \sim \text{Exp}(\lambda), \quad \lambda > 0 \]

- continuous counterpart of geometric distribution (“failure” prob. \( \approx \lambda dt \))
- Value set: \( S_X = (0, \infty) \)
- Probability density function (pdf):
  \[ f_X(x) = \lambda e^{-\lambda x}, \quad x > 0 \]
- Cumulative distribution function (cdf):
  \[ F_X(x) = P\{X \leq x\} = 1 - e^{-\lambda x}, \quad x > 0 \]
- Mean value: \( E[X] = \int_0^\infty \lambda x \exp(-\lambda x) \, dx = 1/\lambda \)
- Second moment: \( E[X^2] = \int_0^\infty \lambda x^2 \exp(-\lambda x) \, dx = 2/\lambda^2 \)
- Variance: \( D^2[X] = E[X^2] - E[X]^2 = 1/\lambda^2 \)

**Memoryless property of exponential distribution**

- Exponential distribution has so called **memoryless property**: for all \( x, y \in (0, \infty) \)
  \[ P\{X > x + y \mid X > x\} = P\{X > y\} \]
- Prove!
  - Tip: Prove first that \( P\{X > x\} = e^{-\lambda x} \)
- Application:
  - Assume that the call holding time is exponentially distributed with mean \( h \) minutes.
  - Consider a call that has already lasted for \( x \) minutes.
  Due to memoryless property, this gives no information about the length of the remaining holding time: it is distributed as the original holding time and, on average, lasts still \( h \) minutes!

**Minimum of exponential random variables**

- Let \( X_1 \sim \text{Exp}(\lambda_1) \) and \( X_2 \sim \text{Exp}(\lambda_2) \) be independent. Then
  \[ X^\min := \min\{X_1, X_2\} \sim \text{Exp}(\lambda_1 + \lambda_2) \]

and

\[ P\{X^\min = X_i\} = \frac{\lambda_i}{\lambda_1 + \lambda_2}, \quad i \in \{1, 2\} \]

- Prove!
  - Tip: See slide 15
Standard normal (Gaussian) distribution

\[ X \sim N(0,1) \]

- limit of the “normalized” sum of IID r.v.s with mean 0 and variance 1 (cf. slide 48)
- Value set: \( S_X = (-\infty, \infty) \)
- Probability density function (pdf):
  \[ f_X(x) = \varphi(x) := \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} x^2} \]
- Cumulative distribution function (cdf):
  \[ F_X(x) := P\{X \leq x\} = \Phi(x) := \int_{-\infty}^{x} \varphi(y) \, dy \]
- Mean value: \( E[X] = 0 \) (symmetric pdf)
- Variance: \( D^2[X] = 1 \)

Normal (Gaussian) distribution

\[ X \sim N(\mu, \sigma^2) \], \( \mu \in \mathbb{R}, \ \sigma > 0 \)

- if \((X - \mu)/\sigma \sim N(0,1)\)
- Value set: \( S_X = (-\infty, \infty) \)
- Probability density function (pdf):
  \[ f_X(x) = f_{X'}(x) = \frac{1}{\sigma} \varphi\left(\frac{x-\mu}{\sigma}\right) \]
- Cumulative distribution function (cdf):
  \[ F_X(x) := P\{X \leq x\} = P\left\{ \frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma} \right\} = \Phi\left(\frac{x-\mu}{\sigma}\right) \]
- Mean value: \( E[X] = \mu + \sigma E[(X - \mu)/\sigma] = \mu \) (symmetric pdf around \( \mu \))
- Variance: \( D^2[X] = \sigma^2 D^2[(X - \mu)/\sigma] = \sigma^2 \)

Properties of the normal distribution

- (i) **Linear transformation**: Let \( X \sim N(\mu, \sigma^2) \) and \( \alpha, \beta \in \mathbb{R} \). Then
  \[ Y := \alpha X + \beta \sim N(\alpha \mu + \beta, \alpha^2 \sigma^2) \]
- (ii) **Sum**: Let \( X_1 \sim N(\mu_1, \sigma_1^2) \) and \( X_2 \sim N(\mu_2, \sigma_2^2) \) be independent. Then
  \[ X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2) \]
- (iii) **Sample mean**: Let \( X_i \sim N(\mu, \sigma^2) \), \( i = 1, \ldots, n \), be independent and identically distributed (IID). Then (cf. slide 25)
  \[ \bar{X}_n := \frac{1}{n} \sum_{i=1}^{n} X_i \sim N(\mu, \frac{1}{n} \sigma^2) \]

Central limit theorem (CLT)

- Let \( X_1, \ldots, X_n \) be independent and identically distributed (IID) with mean \( \mu \) and variance \( \sigma^2 \) (and the third moment exists)
- **Central limit theorem**:
  \[ \frac{1}{\sigma / \sqrt{n}} (\bar{X}_n - \mu) \xrightarrow{i.d.} N(0,1) \]
- It follows that for large values of \( n \)
  \[ \bar{X}_n \approx N(\mu, \frac{1}{n} \sigma^2) \]
4. Basic probability theory

Contents

• Basic concepts
• Discrete random variables
• Discrete distributions (nbr distributions)
• Continuous random variables
• Continuous distributions (time distributions)
• Other random variables

Other random variables

• In addition to discrete and continuous random variables, there are so called mixed random variables
  – containing some discrete as well as continuous portions
• Example:
  – The customer waiting time $W$ in an M/M/1 queue has an atom at zero
    $\{P\{W = 0\} = 1 - \rho > 0\}$ but otherwise the distribution is continuous