IV. Statistical Learning

- Probability Basics
- Bayes Classification
- MAP versus ML Hypotheses
- Probability theory: probability measures, Kolmogorov axioms
- Mathematical statistics: application of probability theory, Naive Bayes
Probability Basics

Area Overview

- **Mathematics**
  - **Stochastics**
    - **Probability theory**
    - **Mathematical statistics**
    - **Inferential statistics**
    - **Descriptive statistics**
    - **Exploratory data analysis**
    - **Data mining**

- **Statistics**

- **ML**

- **Probability theory**: probability measures, Kolmogorov axioms
- **Mathematical statistics**: application of probability theory, Naive Bayes
- **Inferential statistics**: hypothesis tests, confidence intervals
- **Descriptive statistics**: variances, contingencies
- **Exploratory data analysis**: histograms, principal component analysis
- **Data mining**: anomaly detection, cluster analysis
Probability Basics

**Definition 1 (Random Experiment, Random Observation)**

A random experiment or random trial is a procedure that, at least theoretically, can be repeated infinite times. It is characterized as follows:

1. **Configuration.**
   
   A precisely specified system that can be reconstructed.

2. **Procedure.**
   
   An instruction of how to execute the experiment, based on the configuration.

3. **Unpredictability of the outcome.**
Probability Basics

Definition 1 (Random Experiment, Random Observation)

A random experiment or random trial is a procedure that, at least theoretically, can be repeated infinite times. It is characterized as follows:

1. Configuration.
   A precisely specified system that can be reconstructed.

2. Procedure.
   An instruction of how to execute the experiment, based on the configuration.

3. Unpredictability of the outcome.

Random experiments whose configuration and procedure are not designed artificially are called *natural random experiments* or *natural random observations*. 
Remarks:

- A procedure can be repeated several times using the same system, but also with different “copies” of the original system. In particular, a random experiment is called *ergodic* if its time average (= sequential analysis) is the same as its ensemble average (= parallel analysis). [Wikipedia]

- Note that random experiments are causal in the sense of cause and effect. The randomness of an experiment, i.e., the unpredictability of its outcome, is a consequence of the missing information about the causal chain. Hence a random experiment can turn into a deterministic process when new insights become known.
Probability Basics

**Definition 2 (Sample Space, Event Space)**

A set $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$ is called sample space of a random experiment, if each experiment outcome is associated with at most one element $\omega \in \Omega$. The elements in $\Omega$ are called outcomes.

Let $\Omega$ be a finite sample space. Each subset $A \subseteq \Omega$ is called an event; an event $A$ occurs iff the experiment outcome $\omega$ is a member of $A$. The set of all events, $\mathcal{P}(\Omega)$, is called the event space.
Probability Basics

**Definition 2 (Sample Space, Event Space)**

A set $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$ is called sample space of a random experiment, if each experiment outcome is associated with at most one element $\omega \in \Omega$. The elements in $\Omega$ are called outcomes.

Let $\Omega$ be a finite sample space. Each subset $A \subseteq \Omega$ is called an event; an event $A$ occurs iff the experiment outcome $\omega$ is a member of $A$. The set of all events, $\mathcal{P}(\Omega)$, is called the event space.

Examples:

<table>
<thead>
<tr>
<th>Experiment:</th>
<th>Rolling a dice.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample space:</td>
<td>$\Omega = {1, 2, 3, 4, 5, 6}$</td>
</tr>
<tr>
<td>Some event:</td>
<td>$A = {2, 4, 6}$</td>
</tr>
</tbody>
</table>
Probability Basics

**Definition 2 (Sample Space, Event Space)**

A set \( \Omega = \{\omega_1, \omega_2, \ldots, \omega_n\} \) is called sample space of a random experiment, if each experiment outcome is associated with at most one element \( \omega \in \Omega \). The elements in \( \Omega \) are called outcomes.

Let \( \Omega \) be a finite sample space. Each subset \( A \subseteq \Omega \) is called an event; an event \( A \) occurs iff the experiment outcome \( \omega \) is a member of \( A \). The set of all events, \( \mathcal{P}(\Omega) \), is called the event space.

**Examples:**

- **Experiment:** Rolling a dice.
  - **Sample space:** \( \Omega = \{1, 2, 3, 4, 5, 6\} \)
  - **Some event:** \( A = \{2, 4, 6\} \)

- **Experiment:** Rolling two dice at the same time.
  - **Sample space:** \( \Omega = \{\{1, 1\}, \{1, 2\}, \ldots, \{2, 2\}, \ldots, \{6, 6\}\} \)
  - **Some event:** \( B = \{\{1, 2\}\} \)
Probability Basics

**Definition 2 (Sample Space, Event Space)**

A set $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$ is called sample space of a random experiment, if each experiment outcome is associated with at most one element $\omega \in \Omega$. The elements in $\Omega$ are called outcomes.

Let $\Omega$ be a finite sample space. Each subset $A \subseteq \Omega$ is called an event; an event $A$ occurs iff the experiment outcome $\omega$ is a member of $A$. The set of all events, $\mathcal{P}(\Omega)$, is called the event space.

**Examples:**

- **Experiment:** Rolling a dice.
  - **Sample space:** $\Omega = \{1, 2, 3, 4, 5, 6\}$
  - **Some event:** $A = \{2, 4, 6\}$

- **Rolling two dice at the same time.**
  - $\Omega = \{\{1, 1\}, \{1, 2\}, \ldots, \{2, 2\}, \ldots, \{6, 6\}\}$
  - $B = \{\{1, 2\}\}$

- **Rolling two dice in succession.**
  - $\Omega = \{(1, 1), (1, 2), \ldots, (2, 1), \ldots, (6, 6)\}$
  - $B = \{(1, 2), (2, 1)\}$
**Definition 3 (Important Event Types)**

Let $\Omega$ be a finite sample space, and let $A \subseteq \Omega$ and $B \subseteq \Omega$ be two events. Then we agree on the following notation:

1. $\emptyset$ The impossible event.
2. $\Omega$ The certain event.
3. $\overline{A} := \Omega \setminus A$ The complementary event of $A$.  
4. $|A| = 1$ An elementary event.
5. $A \subseteq B \quad \iff \quad A$ is a sub-event of $B$, “$A$ entails $B$”, $A \Rightarrow B$
6. $A = B \quad \iff \quad A \subseteq B \quad \text{and} \quad B \subseteq A$
7. $A \cap B = \emptyset \quad \iff \quad A$ and $B$ are incompatible (otherwise, they are compatible).
Probability Basics

Definition 3 (Important Event Types)
Let $\Omega$ be a finite sample space, and let $A \subseteq \Omega$ and $B \subseteq \Omega$ be two events. Then we agree on the following notation:

1. $\emptyset$ The impossible event.
2. $\Omega$ The certain event.
3. $\overline{A} := \Omega \setminus A$ The complementary event of $A$.
4. $|A| = 1$ An elementary event.
5. $A \subseteq B$ $\iff$ $A$ is a sub-event of $B$, “$A$ entails $B$”, $A \Rightarrow B$
6. $A = B$ $\iff$ $A \subseteq B$ and $B \subseteq A$
7. $A \cap B = \emptyset$ $\iff$ $A$ and $B$ are incompatible (otherwise, they are compatible).

Example (Point 5):

“Roll a two.” $\subset$ “Even number roll.”
“2” entails “Even number roll.”
“2” $\Rightarrow$ “2 or 4 or 6”
Remarks:

- Alternative and semantically equivalent notations of the probability for the combined event “A and B”:
  1. $P(A, B)$
  2. $P(A \land B)$
  3. $P(A \cap B)$
Probability Basics
Approaches to Capture the Nature of Probability

1. Classical definition, symmetry principle
2. Frequentism
3. Subjectivism, Bayesian probability
4. Axiomatic probability
Probability Basics
Approaches to Capture the Nature of Probability

1. Classical definition, symmetry principle
2. Frequentism
3. Subjectivism, Bayesian probability
4. Axiomatic probability

Definition 4 (Classical / Laplace Probability [1749-1827])
If each elementary event in $\Omega$ gets assigned the same probability (equiprobable events), then the probability $P(A)$ of an event $A$ is defined as follows:

$$P(A) = \frac{|A|}{|\Omega|} = \frac{\text{number of cases favorable for } A}{\text{number of total outcomes possible}}$$
A random experiment whose configuration and procedure imply an equiprobable sample space, be it by definition or by construction, is called Laplace experiment. The probabilities of the outcomes are called Laplace probabilities. Since Laplace probabilities are defined by the experiment configuration along with the experiment procedure, they need not to be estimated.

The assumption that a given experiment is a Laplace experiment is called Laplace assumption. If the Laplace assumption cannot be presumed, the probabilities can only be obtained from a (possibly large) number of trials.

Strictly speaking, the Laplace probability as introduced above is not a definition but a circular definition: the probability concept is defined by means of the concept of equiprobability, i.e., another kind of probability.
Probability Basics
Approaches to Capture the Nature of Probability (continued)

1. Classical definition, symmetry principle
2. Frequentism
3. Subjectivism, Bayesian probability
4. Axiomatic probability

Basis is the empirical law of large numbers:

Given a random experiment, the average of the outcomes obtained from a large number of trials is close to the expected value, and it will become closer as more trials are performed.
Remarks:

- Inspired by the empirical law of large numbers, scientists have tried to develop a frequentist probability concept, which is completely based on the (fictitious) limit of the relative frequencies [von Mises, 1951].

  These attempts failed since such a limit formation is possible only within mathematical settings (infinitesimal calculus), where accurate repetitions unto infinity can be made.
Probability Basics
Approaches to Capture the Nature of Probability (continued)

1. Classical definition, symmetry principle
2. Frequentism
3. Subjectivism, Bayesian probability
4. Axiomatic probability

Integration of previous knowledge into a decision process:

\[ p(\text{hypothesis} \mid \text{data}) = \frac{p(\text{data} \mid \text{hypothesis}) \cdot p(\text{hypothesis})}{p(\text{data})} \]
Probability Basics
Approaches to Capture the Nature of Probability (continued)

1. Classical definition, symmetry principle
2. Frequentism
3. Subjectivism, Bayesian probability
4. Axiomatic probability

Integration of previous knowledge into a decision process:

\[
p(\text{hypothesis} \mid \text{data}) = \frac{p(\text{data} \mid \text{hypothesis}) \cdot p(\text{hypothesis})}{p(\text{data})}
\]

- Likelihood: How accurate does a hypothesis “explain” (fit) the data?
- Prior: How probable is a hypothesis a-priori?
Probability Basics
Approaches to Capture the Nature of Probability (continued)

1. Classical definition, symmetry principle
2. Frequentism
3. Subjectivism, Bayesian probability
4. Axiomatic probability

Integration of previous knowledge into a decision process:

\[ p(\text{hypothesis} \mid \text{data}) = \frac{p(\text{data} \mid \text{hypothesis}) \cdot p(\text{hypothesis})}{p(\text{data})} \]

- Likelihood: How accurate does a hypothesis “explain” (fit) the data?
- Prior: How probable is a hypothesis a-priori?

\[ p(\text{hypothesis} \mid \text{data}) \propto p(\text{data} \mid \text{hypothesis}) \cdot p(\text{hypothesis}) \]
Remarks:

- Likelihood is the hypothetical probability that an event that has already occurred would yield a specific outcome. The concept differs from that of a probability in that a probability refers to the occurrence of future events, while a likelihood refers to past events with known outcomes. [Mathworld]

- If applicable and if properly applied the frequentist and the Bayesian approach lead to the same result in most cases.

- The frequentism approach cannot handle singleton or rare events. Example: “What are the chances that the first human mission to Mars will become a success?”

- “It is unanimously agreed that statistics depends somehow on probability. But, as to what probability is and how it is connected with statistics, there has seldom been such complete disagreement and breakdown of communication since the Tower of Babel. Doubtless, much of the disagreement is merely terminological and would disappear under sufficiently sharp analysis.” [Savage, 1954]
Probability Basics
Approaches to Capture the Nature of Probability (continued)

1. Classical definition, symmetry principle
2. Frequentism
3. Subjectivism, Bayesian probability
4. Axiomatic probability

Axiomatic approach to phenomena modeling:

(a) Postulate a function that assigns a “probability” to each element in $\mathcal{P}(\Omega)$.

(b) Specify the required properties of this function in the form of axioms.
Probability Basics
Approaches to Capture the Nature of Probability (continued)

1. Classical definition, symmetry principle
2. Frequentism
3. Subjectivism, Bayesian probability
4. Axiomatic probability

Axiomatic approach to phenomena modeling:

(a) Postulate a function that assigns a “probability” to each element in $\mathcal{P}(\Omega)$.

(b) Specify the required properties of this function in the form of axioms.
Probability Basics
Axiomatic Approach to Probability

**Definition 5 (Probability Measure)** [Kolmogorov 1933]

Let $\Omega$ be a set, called sample space, and let $\mathcal{P}(\Omega)$ be the set of all events, called event space. A function $P$, $P : \mathcal{P}(\Omega) \rightarrow \mathbb{R}$, which maps each event $A \in \mathcal{P}(\Omega)$ onto a real number $P(A)$, is called probability measure if it has the following properties:

1. $P(A) \geq 0$ (Axiom I)
2. $P(\Omega) = 1$ (Axiom II)
3. $A \cap B = \emptyset$ implies $P(A \cup B) = P(A) + P(B)$ (Axiom III)
Probability Basics
Axiomatic Approach to Probability

**Definition 5 (Probability Measure)** [Kolmogorov 1933]

Let $\Omega$ be a set, called sample space, and let $\mathcal{P}(\Omega)$ be the set of all events, called event space. A function $P, P : \mathcal{P}(\Omega) \rightarrow \mathbb{R}$, which maps each event $A \in \mathcal{P}(\Omega)$ onto a real number $P(A)$, is called probability measure if it has the following properties:

1. $P(A) \geq 0$ (Axiom I)
2. $P(\Omega) = 1$ (Axiom II)
3. $A \cap B = \emptyset \rightarrow P(A \cup B) = P(A) + P(B)$ (Axiom III)
Probability Basics
Axiomatic Approach to Probability (continued)

Definition 5 (Probability Measure) [Kolmogorov 1933])
Let \( \Omega \) be a set, called sample space, and let \( \mathcal{P}(\Omega) \) be the set of all events, called event space. A function \( P, P : \mathcal{P}(\Omega) \to \mathbb{R} \), which maps each event \( A \in \mathcal{P}(\Omega) \) onto a real number \( P(A) \), is called probability measure if it has the following properties:

1. \( P(A) \geq 0 \) (Axiom I)
2. \( P(\Omega) = 1 \) (Axiom II)
3. \( A \cap B = \emptyset \quad \rightarrow \quad P(A \cup B) = P(A) + P(B) \) (Axiom III)

Definition 6 (Probability Space)
Let \( \Omega \) be a sample space, let \( \mathcal{P}(\Omega) \) be an event space, and let \( P : \mathcal{P}(\Omega) \to \mathbb{R} \) be a probability measure. Then the tuple \( (\Omega, P) \), as well as the triple \( (\Omega, \mathcal{P}(\Omega), P) \), is called probability space.
Probability Basics
Axiomatic Approach to Probability (continued)

**Definition 5 (Probability Measure)** [Kolmogorov 1933]

Let $\Omega$ be a set, called sample space, and let $\mathcal{P}(\Omega)$ be the set of all events, called event space. A function $P, P : \mathcal{P}(\Omega) \rightarrow \mathbb{R}$, which maps each event $A \in \mathcal{P}(\Omega)$ onto a real number $P(A)$, is called probability measure if it has the following properties:

1. $P(A) \geq 0$ (Axiom I)
2. $P(\Omega) = 1$ (Axiom II)
3. $A \cap B = \emptyset \rightarrow P(A \cup B) = P(A) + P(B)$ (Axiom III)

**Definition 6 (Probability Space)**

Let $\Omega$ be a sample space, let $\mathcal{P}(\Omega)$ be an event space, and let $P : \mathcal{P}(\Omega) \rightarrow \mathbb{R}$ be a probability measure. Then the tuple $(\Omega, P)$, as well as the triple $(\Omega, \mathcal{P}(\Omega), P)$, is called probability space.

We can work with probabilities without interpreting them.
Theorem 7 (Implications of Kolmogorov Axioms)

1. \( P(A) + P(\overline{A}) = 1 \) (from Axioms II, III)

2. \( P(\emptyset) = 0 \) (from 1. with \( A = \Omega \))

3. Monotonicity law of the probability measure:
   \[ A \subseteq B \iff P(A) \leq P(B) \] (from Axioms I, II)

4. “Sum rule” or “addition rule”:
   \[ P(A \cup B) = P(A) + P(B) - P(A \cap B) \] (from Axiom III)

5. Let \( A_1, A_2, \ldots, A_k \) be mutually exclusive (incompatible), then holds:
   \[ P(A_1 \cup A_2 \cup \ldots \cup A_k) = P(A_1) + P(A_2) + \ldots + P(A_k) \]
Remarks:

- The three axioms are also called the Axiom System of Kolmogorov.
- $P(A)$ is called “probability of the occurrence of $A$.”
- Observe that nothing is said about how to interpret the probabilities $P$. An axiomatic approach does not explain but “only” specifies properties.
- Also observe that nothing is said about the distribution of the probabilities $P$.
- A function that provides the three properties of a probability measure is called a non-negative, normalized, and additive measure.
**Probability Basics**

**Conditional Probability**

**Definition 8 (Conditional Probability)**

Let $(\Omega, \mathcal{P}(\Omega), P)$ be a probability space and let $A, B \in \mathcal{P}(\Omega)$ be two events. Then the probability of the occurrence of event $A$ given that event $B$ is known to have occurred is defined as follows:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}, \quad \text{if } P(B) > 0$$

$P(A \mid B)$ is called “probability of $A$ under condition $B$.”
Probability Basics
Conditional Probability (continued)

Definition 8 (Conditional Probability)
Let \((\Omega, \mathcal{P}(\Omega), P)\) be a probability space and let \(A, B \in \mathcal{P}(\Omega)\) be two events. Then the probability of the occurrence of event \(A\) given that event \(B\) is known to have occurred is defined as follows:

\[
P(A \mid B) = \frac{P(A \cap B)}{P(B)}, \quad \text{if } P(B) > 0
\]

\(P(A \mid B)\) is called “probability of \(A\) under condition \(B\).”
Definition 8 (Conditional Probability)

Let \((\Omega, \mathcal{P}(\Omega), P)\) be a probability space and let \(A, B \in \mathcal{P}(\Omega)\) be two events. Then the probability of the occurrence of event \(A\) given that event \(B\) is known to have occurred is defined as follows:

\[
P(A \mid B) = \frac{P(A \cap B)}{P(B)}, \quad \text{if } P(B) > 0
\]

\(P(A \mid B)\) is called “probability of \(A\) under condition \(B\).”

\(A :\) The road is wet.

\[A \equiv A \mid \Omega\]
**Definition 8 (Conditional Probability)**

Let $(\Omega, \mathcal{P}(\Omega), P)$ be a probability space and let $A, B \in \mathcal{P}(\Omega)$ be two events. Then the **probability of the occurrence of event $A$ given that event $B$ is known to have occurred** is defined as follows:

\[
P(A \mid B) = \frac{P(A \cap B)}{P(B)}, \quad \text{if } P(B) > 0
\]

$P(A \mid B)$ is called “**probability of $A$ under condition $B$**.”

**Example:**

$A$: The road is wet.
$B$: It’s raining.

$A \cap B$: The road is wet and it’s raining.

$A \mid B$: The road is wet when it’s raining.

$B \equiv B \mid \Omega$
**Definition 8 (Conditional Probability)**

Let \((\Omega, \mathcal{P}(\Omega), P)\) be a probability space and let \(A, B \in \mathcal{P}(\Omega)\) be two events. Then the **probability of the occurrence of event \(A\) given that event \(B\) is known to have occurred** is defined as follows:

\[
P(A \mid B) = \frac{P(A \cap B)}{P(B)}, \quad \text{if } P(B) > 0
\]

\(P(A \mid B)\) is called “probability of \(A\) under condition \(B\).”

\[A \cap B \equiv A \cap B \mid \Omega \]

\(A\) : The road is wet.

\(B\) : It’s raining.

\(A \cap B\) : The road is wet and it’s raining.
**Probability Basics**

**Conditional Probability** (continued)

**Definition 8 (Conditional Probability)**

Let \((Ω, P(Ω), P)\) be a probability space and let \(A, B \in P(Ω)\) be two events. Then the probability of the occurrence of event \(A\) given that event \(B\) is known to have occurred is defined as follows:

\[
P(A \mid B) = \frac{P(A \cap B)}{P(B)}, \quad \text{if } P(B) > 0
\]

\(P(A \mid B)\) is called “probability of \(A\) under condition \(B\).”

- \(A\): The road is wet.
- \(B\): It’s raining.
- \(A \cap B\): The road is wet and it’s raining.
- \(A \mid B\): The road is wet when it’s raining.
Remarks:

- Important consequences (deductions) from the conditional probability definition:

1. \( P(A \cap B) = P(B) \cdot P(A \mid B) \) (see multiplication rule for statistical independence)

2. \( P(A \cap B) = P(B \cap A) = P(A) \cdot P(B \mid A) \)

3. \( P(B) \cdot P(A \mid B) = P(A) \cdot P(B \mid A) \iff P(A \mid B) = \frac{P(A \cap B)}{P(B)} \overset{(*)}{=} \frac{P(A) \cdot P(B \mid A)}{P(B)} \)

4. \( P(\overline{A} \mid B) = 1 - P(A \mid B) \) (see Point 1 in Kolmogorov implications)
   or \( P_B(\overline{A}) = 1 - P_B(A) \)

\((*)\) The identity shows the (simple) Bayes rule.

- Considered as a function in the parameter \( A \) and the constant \( B \), the conditional probability \( P(A \mid B) \) fulfills the Kolmogorov axioms and in turn defines a probability measure, denoted as \( P_B \) here.
While Deduction 4 is obvious since $P_B$ is a probability measure, the interpretation of complementary events when used as conditions may be confusing. In particular, the following inequality must be assumed: $P(A \mid \overline{B}) \neq 1 - P(A \mid B)$

For illustrating purposes, consider the probability $P(A \mid B) = 0.9$ for the event “The road is wet” ($A$) under the event “It’s raining” ($B$). Observe that this information doesn’t give us any knowledge regarding the wetness of the road under the complementary event $\overline{B}$ “It’s not raining”.
Theorem 9 (Total Probability)

Let \((\Omega, \mathcal{P}(\Omega), P)\) be a probability space, and let \(A_1, \ldots, A_k\) be mutually exclusive events with \(\Omega = A_1 \cup \ldots \cup A_k\), \(P(A_i) > 0, i = 1, \ldots, k\). Then for each \(B \in \mathcal{P}(\Omega)\) holds:

\[
P(B) = \sum_{i=1}^{k} P(A_i) \cdot P(B \mid A_i)
\]
Theorem 9 (Total Probability)

Let \((\Omega, \mathcal{P}(\Omega), P)\) be a probability space, and let \(A_1, \ldots, A_k\) be mutually exclusive events with \(\Omega = A_1 \cup \ldots \cup A_k\), \(P(A_i) > 0\), \(i = 1, \ldots, k\). Then for each \(B \in \mathcal{P}(\Omega)\) holds:

\[
P(B) = \sum_{i=1}^{k} P(A_i) \cdot P(B \mid A_i)
\]
Theorem 9 (Total Probability)

Let \((\Omega, \mathcal{P}(\Omega), P)\) be a probability space, and let \(A_1, \ldots, A_k\) be mutually exclusive events with \(\Omega = A_1 \cup \ldots \cup A_k\), \(P(A_i) > 0, i = 1, \ldots, k\). Then for each \(B \in \mathcal{P}(\Omega)\) holds:

\[
P(B) = \sum_{i=1}^{k} P(A_i) \cdot P(B \mid A_i)
\]
Theorem 9 (Total Probability)

Let $(\Omega, \mathcal{P}(\Omega), P)$ be a probability space, and let $A_1, \ldots, A_k$ be mutually exclusive events with $\Omega = A_1 \cup \ldots \cup A_k$, $P(A_i) > 0$, $i = 1, \ldots, k$. Then for each $B \in \mathcal{P}(\Omega)$ holds:

$$P(B) = \sum_{i=1}^{k} P(A_i) \cdot P(B \mid A_i)$$
Theorem 9 (Total Probability)

Let \((\Omega, \mathcal{P}(\Omega), P)\) be a probability space, and let \(A_1, \ldots, A_k\) be mutually exclusive events with \(\Omega = A_1 \cup \ldots \cup A_k, P(A_i) > 0, i = 1, \ldots, k\). Then for each \(B \in \mathcal{P}(\Omega)\) holds:

\[
P(B) = \sum_{i=1}^{k} P(A_i) \cdot P(B \mid A_i)
\]
Theorem 9 (Total Probability)

Let $(\Omega, \mathcal{P}(\Omega), P)$ be a probability space, and let $A_1, \ldots, A_k$ be mutually exclusive events with $\Omega = A_1 \cup \ldots \cup A_k$, $P(A_i) > 0$, $i = 1, \ldots, k$. Then for each $B \in \mathcal{P}(\Omega)$ holds:

$$P(B) = \sum_{i=1}^{k} P(A_i) \cdot P(B \mid A_i)$$
**Theorem 9 (Total Probability)**

Let \((\Omega, \mathcal{P}(\Omega), P)\) be a probability space, and let \(A_1, \ldots, A_k\) be mutually exclusive events with \(\Omega = A_1 \cup \ldots \cup A_k\), \(P(A_i) > 0, i = 1, \ldots, k\). Then for each \(B \in \mathcal{P}(\Omega)\) holds:

\[
P(B) = \sum_{i=1}^{k} P(A_i) \cdot P(B \mid A_i)
\]

**Proof**

\[
P(B) = P(\Omega \cap B) = P((A_1 \cup \ldots \cup A_k) \cap B)
\]

(exploitation of completeness of the \(A_i\))

\[
= P((A_1 \cap B) \cup \ldots \cup (A_k \cap B))
\]

(exploitation of exclusiveness of the \(A_i\))

\[
= \sum_{i=1}^{k} P(A_i \cap B) = \sum_{i=1}^{k} P(B \cap A_i) = \sum_{i=1}^{k} P(A_i) \cdot P(B \mid A_i)
\]
Remarks:

- The theorem of total probability states that the probability of an arbitrary event equals the sum of the probabilities of the sub-events into which the event has been partitioned.
Probability Basics

Independence of Events

**Definition 10 (Statistical Independence of two Events)**

Let \((\Omega, \mathcal{P}(\Omega), P)\) be a probability space, and let \(A, B \in \mathcal{P}(\Omega)\) be two events. Then \(A\) and \(B\) are called statistically independent iff the following equation holds:

\[
P(A \cap B) = P(A) \cdot P(B)\]

“multiplication rule”
**Definition 10 (Statistical Independence of two Events)**

Let \((\Omega, \mathcal{P}(\Omega), P)\) be a probability space, and let \(A, B \in \mathcal{P}(\Omega)\) be two events. Then \(A\) and \(B\) are called statistically independent iff the following equation holds:

\[
P(A \cap B) = P(A) \cdot P(B)
\]

"multiplication rule"

If statistical independence is given for \(A, B,\) and \(0 < P(B) < 1,\) the following equivalences hold:

\[
\begin{align*}
P(A \cap B) &= P(A) \cdot P(B) \\
\Leftrightarrow P(A \mid B) &= P(A \mid \overline{B}) \\
\Leftrightarrow P(A \mid B) &= P(A)
\end{align*}
\]
**Definition 10 (Statistical Independence of two Events)**

Let \((\Omega, \mathcal{P}(\Omega), P)\) be a probability space, and let \(A, B \in \mathcal{P}(\Omega)\) be two events. Then \(A\) and \(B\) are called statistically independent iff the following equation holds:

\[
P(A \cap B) = P(A) \cdot P(B) \quad \text{“multiplication rule”}
\]

If statistical independence is given for \(A, B,\) and \(0 < P(B) < 1,\) the following equivalences hold:

\[
P(A \cap B) = P(A) \cdot P(B) \iff P(A \mid B) = P(A \mid \bar{B})
\]

\[
\iff P(A \mid B) = P(A)
\]
Definition 10 (Statistical Independence of two Events)

Let \((\Omega, \mathcal{P}(\Omega), P)\) be a probability space, and let \(A, B \in \mathcal{P}(\Omega)\) be two events. Then \(A\) and \(B\) are called statistically independent iff the following equation holds:

\[
P(A \cap B) = P(A) \cdot P(B)
\]

“multiplication rule”

If statistical independence is given for \(A, B\), and \(0 < P(B) < 1\), the following equivalences hold:

\[
\begin{align*}
P(A \cap B) &= P(A) \cdot P(B) \\
\Leftrightarrow P(A \mid B) &= P(A \mid \overline{B}) \\
\Leftrightarrow P(A \mid B) &= P(A)
\end{align*}
\]
**Definition 10 (Statistical Independence of two Events)**

Let \((\Omega, \mathcal{P}(\Omega), P)\) be a probability space, and let \(A, B \in \mathcal{P}(\Omega)\) be two events. Then \(A\) and \(B\) are called statistically independent iff the following equation holds:

\[ P(A \cap B) = P(A) \cdot P(B) \]

“multiplication rule”

If statistical independence is given for \(A, B\), and \(0 < P(B) < 1\), the following equivalences hold:

\[
\begin{align*}
P(A \cap B) &= P(A) \cdot P(B) \\
\iff P(A \mid B) &= P(A \mid \overline{B}) \\
\iff P(A \mid B) &= P(A)
\end{align*}
\]

[dependent events]
Definition 10 (Statistical Independence of two Events)

Let $(\Omega, \mathcal{P}(\Omega), P)$ be a probability space, and let $A, B \in \mathcal{P}(\Omega)$ be two events. Then $A$ and $B$ are called statistically independent iff the following equation holds:

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

“multiplication rule”

If statistical independence is given for $A$, $B$, and $0 < P(B) < 1$, the following equivalences hold:

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

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

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

[dependent events]
Definition 10 (Statistical Independence of two Events)

Let \((\Omega, \mathcal{P}(\Omega), P)\) be a probability space, and let \(A, B \in \mathcal{P}(\Omega)\) be two events. Then \(A\) and \(B\) are called statistically independent iff the following equation holds:

\[
P(A \cap B) = P(A) \cdot P(B)
\]

"multiplication rule"

If statistical independence is given for \(A, B,\) and \(0 < P(B) < 1,\) the following equivalences hold:

\[
P(A \cap B) = P(A) \cdot P(B)
\]

\[
\iff P(A \mid B) = P(A \mid \overline{B})
\]

\[
\iff P(A \mid B) = P(A)
\]
**Definition 10 (Statistical Independence of two Events)**

Let $(\Omega, \mathcal{P}(\Omega), P)$ be a probability space, and let $A, B \in \mathcal{P}(\Omega)$ be two events. Then $A$ and $B$ are called statistically independent iff the following equation holds:

\[ P(A \cap B) = P(A) \cdot P(B) \]

“multiplication rule”

If statistical independence is given for $A$, $B$, and $0 < P(B) < 1$, the following equivalences hold:

\[ P(A \cap B) = P(A) \cdot P(B) \]
\[ \Leftrightarrow P(A \mid B) = P(A \mid \overline{B}) \]
\[ \Leftrightarrow P(A \mid B) = P(A) \]

[dependent events]
**Definition 10 (Statistical Independence of two Events)**

Let \((\Omega, \mathcal{P}(\Omega), P)\) be a probability space, and let \(A, B \in \mathcal{P}(\Omega)\) be two events. Then \(A\) and \(B\) are called statistically independent iff the following equation holds:

\[
P(A \cap B) = P(A) \cdot P(B)
\]

“multiplication rule”

If statistical independence is given for \(A, B,\) and \(0 < P(B) < 1,\) the following equivalences hold:

\[
P(A \cap B) = P(A) \cdot P(B)
\]

\[
\Leftrightarrow P(A \mid B) = P(A \mid \overline{B})
\]

\[
\Leftrightarrow P(A \mid B) = P(A)
\]
Definition 11 (Statistical Independence of $k$ Events)

Let $(\Omega, \mathcal{P}(\Omega), P)$ be a probability space, and let $A_1, \ldots, A_k \in \mathcal{P}(\Omega)$ be $k$ events. Then the $A_1, \ldots, A_k$ are called jointly statistically independent at $P$ iff for all subsets $\{A_{i_1}, \ldots, A_{i_l}\} \subseteq \{A_1, \ldots, A_k\}$ the multiplication rule holds:

$$P(A_{i_1} \cap \ldots \cap A_{i_l}) = P(A_{i_1}) \cdot \ldots \cdot P(A_{i_l}),$$

where $i_1 < i_2 < \ldots < i_l$ and $2 \leq l \leq k.$