Summary of Definitions and Theorems from Exercise Sheets

Jonas Neuschäfer

May 21, 2025

Disclaimer: I do *not* attend the lecture and therefore cannot guarantee that the content presented here aligns with the official material of the course. These notes are based solely on my interpretation and understanding of the assignments. This can **not replace the professor's lecture notes**, and it is **not binding** for the exam in any form. Use at your own discretion.

Contents

Sheet 1 Set Operations in Probability

Topic: Set Operations

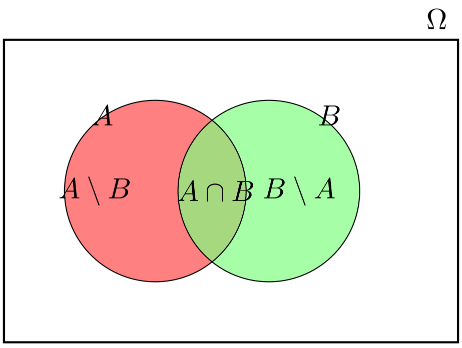
Definition 1.1 (Sample Space)**.** *The* sample space Ω *is the set of all possible outcomes of a random experiment.*

Definition 1.2 (Event)**.** *An* event *is a subset* $A \subset \Omega$*.*

Example 1.1 (Die Roll)**.** *Let* $\Omega = \{1, 2, 3, 4, 5, 6\}$*. The event "a die shows an odd face" is represented by the subset* $A = \{1, 3, 5\}$*.*

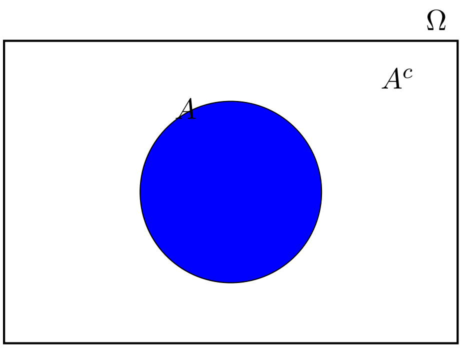
Basic Set Operations (for $A, B \subset \Omega$)

- **Union:** $A \cup B$ – all elements in A **or** in B
- **Intersection:** $A \cap B$ – all elements in both A **and** B
- **Complement:** $A^c = \Omega \setminus A$ – all elements **not** in A
- **Set Difference:** $A \setminus B = A \cap B^c$ – elements in A but not in B



Properties of Set Operations

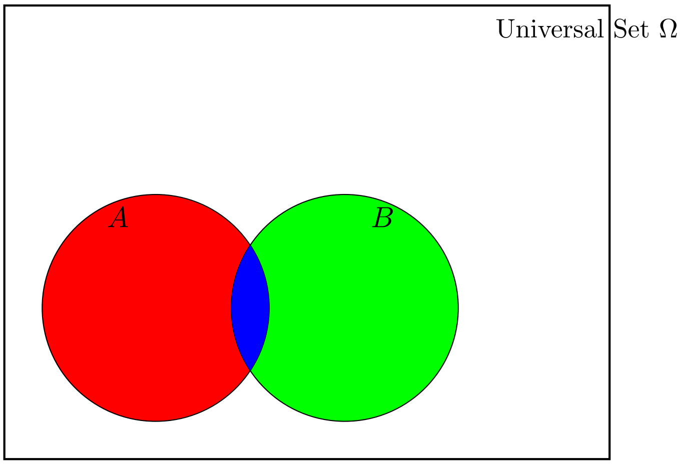
- $(A^c)^c = A$



Obviously is the complement of A^c, the white area, again blue, i.e. A.

- $(A \cup B)^c = A^c \cap B^c$

  In order to find $(A \cup B)^c$, the white area, just take all elements which are not in A, hence A^c. But this includes the green area, so just subtract B from A^c to get the white area. This implies $(A \cup B)^c = A^c \cap B^c$.



- $(A \cap B)^c = A^c \cup B^c$ (De Morgan's laws)

  We want to catch everything which is not in blue, i.e white + green + red. So take all elements in A^c (which is white + green) and add all elements in B^c (which is white + red) in order to achieve that $\implies (A \cap B)^c = A^c \cup B^c$.

- Distributivity:

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$
$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

Basic Observations

- $A \cap \Omega = A$
- $A \cap \emptyset = \emptyset$
- $A \cup A^c = \Omega$
- If $A, B \subset \Omega$, then:

$$A \cap B \subset A, \quad A \subset A \cup B$$

Remark 1.1. *By taking unions or intersections of events (subsets of Ω), we generate new events.*

Topic: Infinite Unions and Intersections

Definition 1.3 (Infinite Intersection)**.** *Let* $\{A_i\}_{i\in\mathbb{N}} \subseteq \Omega$ *be a sequence of events. Then:*

$$x \in \bigcap_{i=1}^{\infty} A_i \quad \Leftrightarrow \quad x \in A_i \text{ for all } i \in \mathbb{N}$$

This means x *lies in every single set* A_i*.*

Observation: The intersection is always a subset of each of the individual sets:

$$\bigcap_{i=1}^{\infty} A_i \subset A_j \quad \text{for all } j \in \mathbb{N}$$

Definition 1.4 (Infinite Union)**.** *Let* $\{A_i\}_{i\in\mathbb{N}} \subset \Omega$*. Then:*

$$x \in \bigcup_{i=1}^{\infty} A_i \quad \Leftrightarrow \quad \text{there exists } j \in \mathbb{N} \text{ such that } x \in A_j$$

This means x *lies in at least one of the sets* A_i*, but it is not specified which.*

Remark 1.2. *The union is always at least as large as any single set:*

$$A_j \subset \bigcup_{i=1}^{\infty} A_i \quad \text{for all } j \in \mathbb{N}$$

Remark 1.3 (Distributive Laws with Infinite Collections)**.** *The distributive properties also hold in the infinite case:*

$$B \cup \bigcap_{i=1}^{\infty} A_i = \bigcap_{i=1}^{\infty} (B \cup A_i)$$
$$B \cap \bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} (B \cap A_i)$$

How to Solve Exercises Like Those in Sheet 1

Remark 1.4.

- *Set expressions like* $(A \cap B) \cup (A \cap C)$ *can be interpreted logically: for an outcome to be in this union, it must be in* A *and also in either* B *or* C*. Alternatively (in simple cases), draw a Venn diagram to see what the expression looks like.*

- *To check whether two sets (possibly even complicated unions/intersections) are equal, take an element from the left-hand side and verify, by applying the properties of set operations (De Morgan's laws, distributivity, ...), whether it also belongs to the right-hand side. Then proceed vice versa.*

- *Remember: For an element to be contained in a union of several sets, it suffices to find one set that contains it. On the other hand, for an element to be in the intersection, every set must contain that element.*

- *If elements of a set must fulfill some condition (e.g.* i *is restricted), then*

  $$A = \{(i,j) \in \Omega : i \in B\} \implies |A| = |B| \quad \text{or} \quad A = \{(i,j) \in \Omega : 0 < i \leq k\} \implies |A| = k$$

  Attention!: $|A \cup B| = |A| + |B|$ *only if* A, B *disjoint* $A \cap B = \emptyset$*.*

Sheet 2 Probability Measure

Definition 2.1 (Probability Measure). *A function* $P: A \subset \Omega \mapsto P(A) \in \mathbb{R}$ *is called a* probability measure *if it satisfies the following axioms for all events* $A \in \mathcal{A} \subset \Omega$*:*

(i) ***Non-negativity:*** $P(A) \geq 0$

(ii) ***Normalization:*** $P(\Omega) = 1$

(iii) σ***-additivity:*** *For any finite or countable sequence of pairwise disjoint events* $A_1, A_2, \dots$ *(i.e.,* $A_i \cap A_j = \emptyset$ *for* $i \neq j$*):*

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

Properties of a Probability Measure

Let P be a probability measure. Then the following properties hold:

a) (Finite) Additivity. If $A_1, \dots, A_n$ are pairwise disjoint, then:

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$$

b) Strong Additivity.

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

Inclusion-Exclusion Principle

For two events A and B (easy, we already know that), we have:

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

This also works for three events A, B, and C, where we must subtract the pairwise intersections, but then the intersection of all three is subtracted as well (which we do not want because it is clearly included) and must be added back(draw a Venn-diagram to verify the 3 steps (all three - subtract intersections - but then observe that $A \cap B \cap C$ is removed, hence add it back):

$$\begin{aligned} P(A \cup B \cup C) = {} & P(A) + P(B) + P(C) && (1)\\ & - P(A \cap B) - P(A \cap C) - P(B \cap C) && (2)\\ & + P(A \cap B \cap C). && (3) \end{aligned}$$

Btw: The same procedure can be applied to unions of 4 (where then a fourth line is added above with intersec. of all four sets), 5, 6, and so on...

c) Monotonicity. If $A \subseteq B$, then:

$$P(A) \leq P(B)$$

d) If $A \subset B$ and $P(A) < \infty$, then:

$$P(B \setminus A) = P(B) - P(A)$$

In particular $P(A^c) = P(\Omega \setminus A) = \underbrace{P(\Omega)}_{=1} - P(A)$

If A is not a subset of B, we still have

$$P(B \setminus A) = P(B) - P(A \cap B)$$

as $A \cap B \subset B$.

e) Finite Sub-additivity. For any finite sequence of events $A_1, \ldots, A_n$, not necessarily disjoint:

$$P\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} P(A_i)$$

f) Countable Sub-additivity. For any countable sequence of events $A_1, A_2, \ldots$, we have:

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} P(A_i)$$

How to Solve Exercises Like Those in Sheet 2

In many exercises, we assume a uniform probability measure when the number of outcomes is finite and each outcome is equally likely. In other words, if we want to determine the probability of an event, we simply count the number of outcomes that fulfill the event and divide by the total number of outcomes.

Definition 2.2 (Uniform Probability). *If Ω is finite and all outcomes are equally likely, then for any event $A \subset \Omega$:*

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

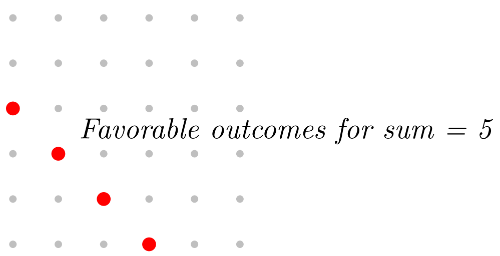
In particular, we sometimes encounter tuples: $A = \{(i,j) \in \Omega : i \in B, j \in C\} \Rightarrow |A| = |B||C|$.

Remark 2.1. *To simplify probability computations, we often split an event into smaller, disjoint sub-events:*

$$P(A) = P(A_1) + P(A_2), \quad \text{where } A = A_1 \cup A_2, \quad A_1 \cap A_2 = \emptyset$$

Example 2.1. *Let $\Omega = \{(i,j) \in \{1,\ldots,6\}^2\}$ be the sample space of two dice rolls. Then:*

$$P(\text{Sum is } 5) = P((1,4)) + P((2,3)) + P((3,2)) + P((4,1)) = \frac{4}{36}$$



Definition 2.3 (Multiplication Rule for Counting). *If a procedure consists of k sequential and independent steps, where the first step can be performed in a_1 ways, the second in a_2 ways, and so on up to the k-th step in a_k ways, then the total number of possible outcomes for the entire procedure is:*

$$a_1 \cdot a_2 \cdot \cdots \cdot a_k$$

This rule is frequently used in probability exercises where a task is broken into distinct sub-tasks, each with a known number of possible outcomes.

Sheet 3 Combinatorics and Conditional Probability

Motivation for the Binomial Coefficient

In many probability problems, we are interested not just in outcomes but in *how many* ways an event can occur. The binomial coefficient

$$\binom{n}{k}$$

counts the number of ways to choose k elements from a set of n distinct elements, without regard to order. Typical applications include:

- Determining how many subsets of a given size can be formed.
- Computing the number of favorable outcomes when events are defined by a number of chosen objects (e.g., choosing 3 red balls from 10).

Always keep in mind: "How many possibilities are there to choose k elements out of n?"

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

Some useful identities include:

$$\binom{n}{0} = \binom{n}{n} = 1, \qquad \binom{n}{1} = \binom{n}{n} = n + 1 \text{ (think about why)}$$

$$\binom{n}{k} + \binom{n}{k-1} = \binom{n+1}{k}$$

In exercise contexts, one common technique to compute probabilities is to count the number of favorable outcomes and divide by the total number of possibilities. The binomial coefficient often arises when the number of (favorable) cases or total possibilities involves choosing a subset of k elements from a total of n.

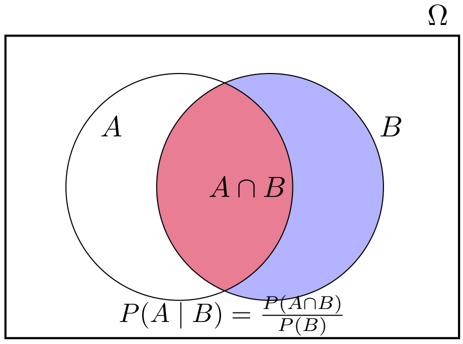
Example 3.1. *There are 15 professors in total: 3 in macroeconomics, 10 in microeconomics, and 2 in statistics. The university randomly selects 3 of them to give a lecture on Dies Academicus. What is the probability that exactly 2 micro and 1 statistics professor are chosen?*

$$\frac{\binom{3}{0} \cdot \binom{10}{2} \cdot \binom{2}{1}}{\binom{15}{3}}$$

Conditional Probability

Definition 3.1 (Conditional Probability)**.** *Let $A, B \subset \Omega$ with $P(B) > 0$. The conditional probability of A given B is defined as:*

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



Because we already know that event B has occurred, we restrict our sample space to the blue area. The conditional probability $P(A \mid B)$ is then the proportion of this area that also lies in A, i.e., the red region. Hence, we compute

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

From this definition, we obtain two equivalent forms:

$$\begin{aligned} P(A \cap B) &= P(A \mid B) \cdot P(B) \\ &= P(B \mid A) \cdot P(A) \end{aligned}$$

Remark 3.1. *The function $P(\cdot \mid B)$ itself satisfies all properties of a probability measure. Therefore, all the standard results apply, such as (and more, see properties of probability measure in chapter 2:*

$$P(A^c \mid B) = 1 - P(A \mid B)$$
$$P(A \cup C \mid B) = P(A \mid B) + P(C \mid B) - P(A \cap C \mid B)$$
$$\dots$$

Sheet 4 Independence and Bayes' Theorem

Definition: Independence of Events

Definition 4.1 (Independence)**.** *Two events A and B are called* independent *if and only if:*

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

Equivalently, using the definition of conditional probability:

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

This means that knowing B has occurred does not change the probability of A. The same reasoning applies in reverse for $P(B \mid A)$.

Remark 4.1. *If A and B are independent, then so are:*

$$A^c \text{ and } B, \quad A \text{ and } B^c, \quad A^c \text{ and } B^c$$

Independence of More Than Two Events

Definition 4.2 (Mutual Independence)**.** *Let $A_1, A_2, \dots, A_n \subset \Omega$ be events. They are called* mutually independent *if for every subset $\{i_1, i_2, \dots, i_k\} \subset \{1, 2, \dots, n\}$ with $k \geq 2$, it holds that:*

$$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \cdot P(A_{i_2}) \cdots P(A_{i_k})$$

*This means that **every nontrivial intersection of two or more events** must satisfy the product rule. In particular, pairwise independence does **not** imply mutual independence.*

Remark 4.2. *To check mutual independence, one must verify the product rule for:*

- *all $\binom{n}{2}$ pairwise intersections,*
- *all $\binom{n}{3}$ triple intersections,*
- *...*
- *and finally the full intersection $A_1 \cap A_2 \cap \dots \cap A_n$.*

Bayes' Theorem

Definition 4.3 (Bayes' Theorem)**.** *If $P(B) > 0$, then:*

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}$$

This is particularly useful when $P(B \mid A)$ and $P(A)$ are known, but $P(A \mid B)$ is needed.

Law of Total Probability

Definition 4.4 (Law of Total Probability)**.** *Let A and A^c partition the sample space, and let B be any event. Then:*

$$P(B) = P(B \mid A) \cdot P(A) + P(B \mid A^c) \cdot P(A^c)$$

*This is a common strategy when $P(B)$ must be **computed indirectly**.*

Tips for Solving Exercises

- Independence must hold for any finite intersection if stated. For example:

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

  if all three are mutually independent.

- Remember: if A and B are independent, so are their complements and combinations, as listed above.

- When asked for $P(A \mid B)$ but only $P(B \mid A)$ is known, use Bayes' theorem:

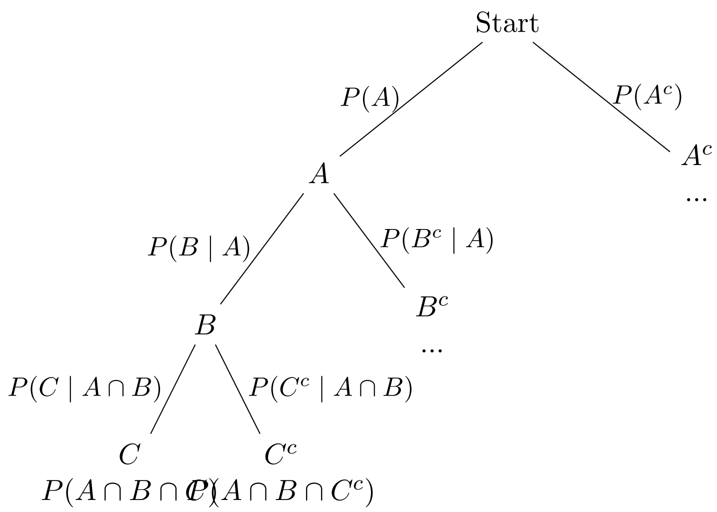
  $$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}$$

- When $P(B)$ is unknown but $P(B \mid A)$ and $P(B \mid A^c)$ are known, apply the law of total probability:

  $$P(B) = P(B \mid A) \cdot P(A) + P(B \mid A^c) \cdot P(A^c)$$

- You can chain conditional probabilities:

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



Sheet 5 (and 6) Distribution Functions of Random Variables

Topic: Distribution Function (Cumulative Distribution Function)

Definition 5.1 (Cumulative Distribution Function)**.** *Let $X \colon \Omega \to \mathbb{R}$ be a (real-valued) random variable. The function $F \colon \mathbb{R} \to [0,1]$ defined by*

$$F(x) = P(X \le x)$$

is called the (cumulative) distribution function *of X.*

Remark 5.1. *The distribution function $F(x)$ gives the probability that the random variable X takes a value less than or equal to x.*

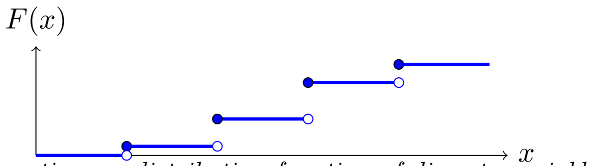
Properties of the distribution function

For any random variable X, the distribution function F satisfies the following:

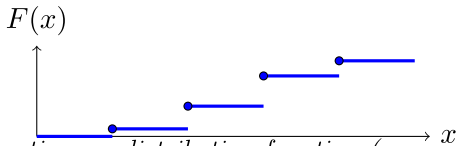
- **Non-decreasing:** $F(x_1) \le F(x_2)$ for all $x_1 < x_2$
- **Right-continuous:** $\lim_{x \downarrow x_0} F(x) = F(x_0)$
- **Limits:**

$$\lim_{x \to -\infty} F(x) = 0, \qquad \lim_{x \to \infty} F(x) = 1$$

Remark 5.2. *If X is a discrete random variable, then the distribution function is a* piecewise constant *function with jumps at the values that X can take.*



Right-continuous distribution function of discrete variable



Right-continuous distribution function (non-uniform)

Interpretation

This is the distribution function of a discrete random variable with support $\{1, 2, 3, 4\}$ and non-uniform probabilities. The jump heights correspond to the values:

$$\Delta F(x_i) = P(X = x_i)$$

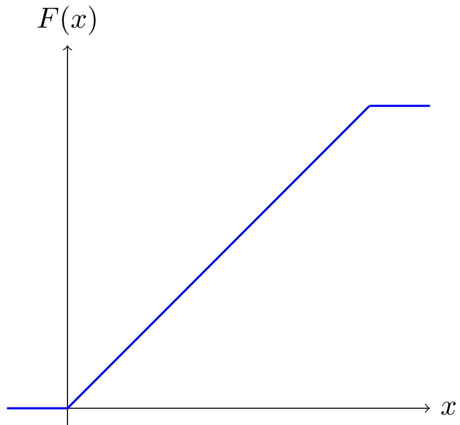
Horizontal segments represent intervals where the probability mass is zero.

Graphical Intuition

- *The distribution function is flat on intervals where X takes no values (in the discrete case).*
- *Jumps in the distribution function correspond to values with positive point mass:*

$$P(X = x_0) = \textit{height of jump at } x_0$$

If X is continuous, then F is a continuous function.



Continuous distribution function (Uniform[0,1])

Computing Probabilities Using the distribution function

Let X be a real-valued random variable with distribution function F_X. Then:

- $P(X \leq x) = F(x)$
- $P(X < x) = \lim_{t \uparrow x} F(t) = F(x^-)$
- $P(X = x) = F(x) - F(x^-)$
- $P(a < X \leq b) = F(b) - F(a)$

How to Solve Exercises Like Those in Sheet 5 and 6

Remark 5.3.

- ***Constructing the distribution function via Direct Summation:*** *If a probability mass function is known or the rule defining outcomes is discrete and explicit, compute:*

$$F(x) = P(X \leq x) = \sum_{k \leq x} P(X = k)$$

  This yields a right-continuous step function.

- ***Constructing the distribution function via Complement:*** *In many problems, computing $P(X > x)$ is simpler than summing $P(X = k)$ for all $k \leq x$. In such cases, use:*

$$F(x) = 1 - P(X > x)$$

  This is especially effective when the "bad" or "complement" events are structurally simpler or when cumulative counting is infeasible (e.g., Exercise 29(a)).

- ***Complement Tricks:*** *Probabilities of intersections or "at least one" often reduce via complements (**Attention: Only when the random variables are independent!**:*

$$P(\min\{X_1, X_2, X_3\} < a) = 1 - P(X_1 \geq a)^3$$

  Also effective for tail bounds: $P(X > x)$, then invert to get CDF.

- ***Independence Assumptions:*** *Factor joint probabilities only if independence is given. Mutual independence does not follow from triple intersections alone (cf. Exercise 23). Always verify pairwise independence separately if needed.*