Lecture Notes for Math 447 - Probability

Skeletal version: proofs omitted

Michael Fochler
Department of Mathematics
Binghamton University

This skeletal version of the document is meant to serve as a help to review the material for upcoming exams. It not only omits the proofs, but also many motivational paragraphs and examples, and even some propositions and theorems. The references are out of sync with the full student edition, so

Do NOT use this edition to find an item referenced, e.g., in your homework assignment!

Last update: September 4, 2025

Contents

1 Some Preliminaries **5**
1.1 About This Document 5
1.2 A First Look at Probability 5

2 Sets, Numbers, Sequences and Functions **6**
2.1 Sets – The Basics 6
2.2 The Proper Use of Language in Mathematics: Any vs All, etc 10
2.3 Numbers 10
2.4 Functions and Sequences 13
2.5 Preimages 17
2.6 Infimum and Supremum: Generalized Minimum and Maximum 19
2.7 Cartesian Products 21
2.8 Indicator Functions 21

3 Calculus Revisited **23**
3.1 Absolute Convergence of Series 23
3.2 Integration – The Riemann Integral 24
3.2.1 The Riemann Integral of a Step Function 24
3.2.2 The Riemann Integral as the Limit of Riemann Sums 25
3.3 Improper Integrals and Integrals Over Subsets 27
3.4 Series and Integrals as Tools to Compute Probabilities 28
3.4.1 Series and Sums 28
3.4.2 Integrals 28

4 Calculus Extensions **29**
4.1 Extension of Lebesgue Measure to the Borel sets of $\mathbb{R}^d$ 29
4.2 The Lebesgue Integral 30

5 The Probability Model **35**
5.1 Probability Spaces 35
5.2 Conditional Probability and Independent Events 39
5.3 Random Elements and their Probability Distributions 42
5.4 Independence of Random Elements 43

6 Advanced Topics – Measure and Probability $\boxed{\star}$ **47**
6.1 Random Variables as Measurable Functions 47
6.2 Measures 50
6.3 Abstract Integrals 52
6.4 The ILMD Method 56
6.5 Expectation and Variance as Probability Measure Integrals 56

7 Combinatorial Analysis **58**
7.1 The Multiplication Rule 58
7.2 Permutations 58
7.3 Combinations, Binomial and Multinomial Coefficients 59

8 More on Probability **62**
8.1 Total Probability and Bayes Formula . . . 62
8.2 Sampling and Urn Models With and Without Replacement . . . 62

9 Discrete Random Variables and Random Elements **64**
9.1 Probability Mass Function and Expectation . . . 64
9.2 Bernoulli Variables and the Binomial Distribution . . . 66
9.3 Geometric + Negative Binomial + Hypergeometric Distributions . . . 67
9.4 The Poisson Distribution . . . 70
9.5 Moments, Central Moments and Moment Generating Functions . . . 71

10 Continuous Random Variables **73**
10.1 Cumulative Distribution Function of a Random Variable . . . 73
10.2 Continuous Random Variables and Probability Density Functions . . . 73
10.3 Expected Value, Variance and MGF of a Continuous Random Variable . . . 75
10.4 The Uniform Probability Distribution . . . 78
10.5 The Normal Probability Distribution . . . 79
10.6 The Gamma Distribution . . . 80
10.7 The Beta Distribution . . . 82
10.8 Inequalities for Probabililities . . . 82
10.9 Mixed Random Variables . . . 83

11 Multivariate Probability Distributions **84**
11.1 Multivariate CDFs, PMFs and PDFs . . . 84
11.2 Marginal and Conditional Probability Distributions . . . 86
11.3 Independence of Random Variables and Discrete Random Elements . . . 87
11.4 The Mulitivariate Uniform Distribution . . . 89
11.5 The Expected Value of a Function of Several Random Variables . . . 89
11.6 Covariance . . . 90
11.7 Conditional Expectations and Conditional Variance . . . 92
11.7.1 The Conditional Expectation With Respect to an Event $\boxed{\star}$ . . . 92
11.7.2 The Conditional Expectation w.r.t a Random Variable or Random Element . . . 93
11.7.3 Conditional Expectations as Optimal Mean Squared Distance Approximations 95
11.8 The Multinomial Probability Distribution . . . 95
11.9 Order Statistics . . . 96
11.10 The Bivariate Normal Distribution . . . 98
11.11 Blank Page after Ch.11 . . . 99

12 Functions of Random Variables and their Distribution **100**
12.1 The Method of Distribution Functions . . . 100
12.2 The Method of Transformations in One Dimension . . . 100
12.3 The Method of Transformations in Multiple Dimension . . . 100
12.4 The Method of moment–generating Functions . . . 101

13 Limit Theorems **104**
13.1 Four Kinds of Limits for Sequences of Random Variables . . . 104
13.2 Two Laws of Large Numbers . . . 105

13.3 Sampling Distributions 105
13.4 The Central Limit Theorem 108

14 Sample Problems for Exams **111**
14.1 Practice Midterm 1 for Math 447 - Chris Haines 111

15 Other Appendices **112**
15.1 Greek Letters 112
15.2 Notation 112

References **114**

List of Symbols **115**

Index **117**

History of Updates:

Date	Topic
2020-12-23	Created.

1 Some Preliminaries

1.1 About This Document

1.2 A First Look at Probability

Definition 1.1 (Probability measure - Preliminary Definition, version I)**.** A **probability measure** $\mathbb{P}$ on a set Ω is a function which assigns to each subset A of Ω a real number $\mathbb{P}(A)$ between 0 and 1 as follows.

(a) $\mathbb{P}(\emptyset) = 0$ and $\mathbb{P}(\Omega) = 1$. Here $\emptyset$ denotes the empty set which contains no elements.

(b) If the subsets A, B of Ω have no elements in common, then probability is **additive**:

$$\mathbb{P}(A \cup B) \;=\; \mathbb{P}(A) \,+\, \mathbb{P}(B)\,.$$

This last formula makes disjoint unions so important that we have reserved the special symbol "$\uplus$" as a visual aid. Henceforth, we usually write $U \uplus V$ for $U \cup V$ if we know that $U \cap V = \emptyset$:

$$\mathbb{P}(A \uplus B) \;=\; \mathbb{P}(A) \,+\, \mathbb{P}(B)\,. \;\square$$

Definition 1.2 (Probability measure - Preliminary Definition, version II)**.** A **probability measure** $\mathbb{P}$ on a set Ω is a function which assigns to each subset A of Ω a real number $\mathbb{P}(A)$ between 0 and 1 as follows.

(a) $\mathbb{P}(\emptyset) = 0$ and $\mathbb{P}(\Omega) = 1$.

(b) If the subsets $A_1, A_2 \dots$ of Ω are mutually disjoint, then probability is σ**–additive**:

$$\mathbb{P}(A_1 \uplus A_2 \uplus \cdots) \;=\; \mathbb{P}(A_1) \,+\, \mathbb{P}(A_2) \,+ \cdots \;=\; \sum_{j=1}^{\infty} \mathbb{P}(A_j)\,. \tag{1.1}$$

- The combination $(\Omega, \mathbb{P})$ is called a **probability space** aka **sample space**.
- An element ω of Ω is called an **outcome** aka **sample point**
- A subset of Ω is called an **event**. $\square$

2 Sets, Numbers, Sequences and Functions

2.1 Sets – The Basics

Definition 2.1 (Sets)**.**

- A **set** is a collection of stuff called **members** or **elements** which satisfies the following rules: The order in which you write the elements does not matter and if you list an element two or more times then **it only counts once**.
- We write $x_1 \in X$ to denote that an item x_1 is an element of the set X and $x_2 \notin X$ to denote that an item x_2 is not an element of the set X.
- Occasionally we are less formal and write x_1 **in** X for $x_1 \in X$ and x_2 **not in** X for $x_2 \notin X$.

Definition 2.2 (empty set)**.** $\emptyset$ denotes the **empty set**. It is the set that does not contain any elements. $\square$

Definition 2.3 (subsets and supersets)**.**

- We say that a set A is a **subset** of the set B and we write $A \subseteq B$ if any element of A also belongs to B. Equivalently we say that B is a **superset** of the set A and we write $B \supseteq A$. We also say that B includes A or A is included by B. Note that $A \subseteq A$ and $\emptyset \subseteq A$ is true for any set A.
- If $A \subseteq B$ but $A \neq B$, i.e., there is at least one $x \in B$ such that $x \notin A$, then we say that A is a **strict subset** or a **proper subset** of B. We write "$A \subsetneq B$" Alternatively we say that B is a **strict superset** or a **proper superset** of A and we write "$B \supsetneq A$") $\square$

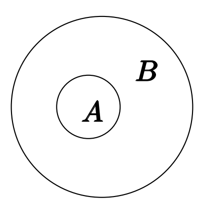

Figure 2.1: Set inclusion: $A \subseteq B,\ B \supseteq A$

Definition 2.4 (unions, intersections and disjoint unions of two sets)**.** Given are two sets A and B. No assumption is made that either one is contained in the other or that either one is not empty!

- The **union** $A \cup B$ (pronounced "A union B") is defined as the set of all elements which belong to at least one of A, B.
- The **intersection** $A \cap B$ (pronounced "A intersection B") is defined as the set of all elements which belong to both A and B.
- We call A and B **disjoint** , also **mutually disjoint** , if $A \cap B = \emptyset$. We then often write $A \uplus B$ (pronounced "A disjoint union B") rather than $A \cup B$. $\square$

Definition 2.5 (Arbitrary unions, intersections and disjoint unions of sets)**.** Let J be an arbitrary, nonempty set. J may be finite or infinite. J may or may not be a set of numbers.

Assume that each $j \in J$ is associated with a set A_j. [1] For $J = \{\diamond, 3, \boldsymbol{\mathcal{X}}\}$, the sets are $A_\diamond, A_3, A_{\boldsymbol{\mathcal{X}}}$; and $J = \{1, 2, \dots\}$, yields the infinite sequence (of sets!) $A_1, A_2, \dots$.

- The **union** $\bigcup\limits_{j \in J} A_j$ is defined as the set of all elements which belong to at least one A_j, where $j \in J$.
- The **intersection** $\bigcap\limits_{j \in J} A_j$ is defined as the set of all elements which belong to each A_j, where $j \in J$.
- We call this collection of sets **disjoint** , also **mutually disjoint** , if $A_i \cap A_j = \emptyset$ whenever $i, j \in J$ and $i \neq j$. We then often write $\biguplus\limits_{j \in J} A_j$ rather than $\bigcup\limits_{j \in J} A_j$. $\square$

Remark 2.1. Convince yourself that for any sets A, B and C.

$$A \cap B \subseteq A \subseteq A \cup B, \tag{2.1}$$

$$A \subseteq B \Rightarrow A \cap B = A \text{ and } A \cup B = B, \tag{2.2}$$

$$A \subseteq B \Rightarrow A \cap C \subseteq B \cap C \text{ and } A \cup C \subseteq B \cup C. \tag{2.3}$$

The symbol $\Rightarrow$ stands for "allows us to conclude that". So $A \subseteq B \Rightarrow A \cap B = A$ means "From the truth of $A \subseteq B$ we can conclude that $A \cap B = A$ is true". Shorter: "From $A \subseteq B$ we can conclude that $A \cap B = A$". Shorter: "If $A \subseteq B$, then it follows that $A \cap B = A$". Shorter: "If $A \subseteq B$, then $A \cap B = A$". More technical: $A \subseteq B$ implies $A \cap B = A$. $\square$

Definition 2.6 (set differences and symmetric differences)**.** Given are two arbitrary sets A and B. No assumption is made that either one is contained in the other or contains any elements!

- The **difference set** or **set difference** $A \setminus B$ (pronounced "A minus B") is defined as the set of all elements which belong to A but not to B:

  (2.4) $$A \setminus B := \{x \in A : x \notin B\}$$

- The **symmetric difference** $A \triangle B$ (pronounced "A delta B") is defined as the set of all elements which belong to either A or B but not to both A and B:

  (2.5) $$A \triangle B := (A \cup B) \ \setminus \ (A \cap B) \ \square$$

Definition 2.7 (Universal set)**.** Usually there always is a big set Ω that contains everything we are interested in and we then deal with all kinds of subsets $A \subseteq \Omega$. Such a set is called a **"universal" set**. $\square$

Definition 2.8 (Complement of a set)**.** Let Ω be a universal set. The **complement** of a set $A \subseteq \Omega$ consists of all elements of Ω which do not belong to A. We write A^{C}. In other words:

(2.6) $$A^{\mathsf{C}} \ = \ \Omega \setminus A \ = \ \{\omega \in \Omega : x \notin A\} \ \square$$

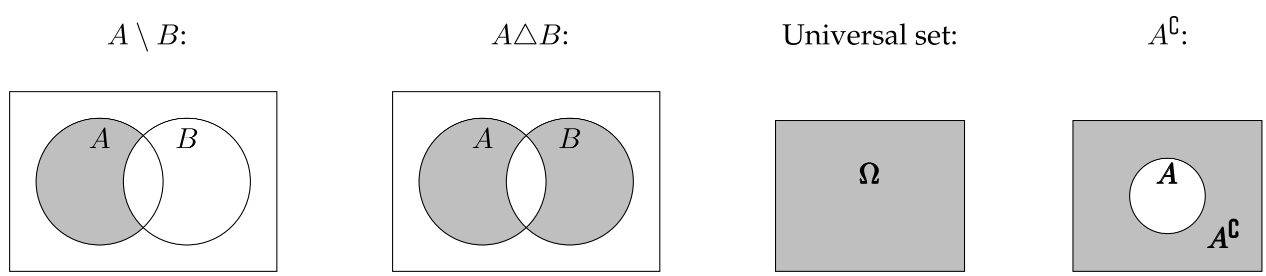

Figure 2.2: Difference, symmetric difference, universal set, complement

Proposition 2.1. *Let A, B, X be subsets of a universal set Ω and assume $A \subseteq X$. Then*

$$A \cup \emptyset = A; \qquad A \cap \emptyset = \emptyset \tag{2.7a}$$
$$A \cup \Omega = \Omega; \qquad A \cap \Omega = A \tag{2.7b}$$
$$A \cup A^{\complement} = \Omega; \qquad A \cap A^{\complement} = \emptyset \tag{2.7c}$$
$$A \triangle B = (A \setminus B) \uplus (B \setminus A) \tag{2.7d}$$
$$A \setminus A = \emptyset \tag{2.7e}$$
$$A \triangle \emptyset = A; \qquad A \triangle A = \emptyset \tag{2.7f}$$
$$X \triangle A = X \setminus A \tag{2.7g}$$
$$A \cup B = (A \triangle B) \uplus (A \cap B) \tag{2.7h}$$
$$A \cap B = (A \cup B) \setminus (A \triangle B) \tag{2.7i}$$
$$A \triangle B = \emptyset \text{ \textit{if and only if} } B = A \tag{2.7j}$$

Proposition 2.2 (Distributivity of unions and intersections for two sets)**.** *Let A, B, C be sets. Then*

$$(A \cup B) \cap C = (A \cap C) \cup (B \cap C), \tag{2.8}$$
$$(A \cap B) \cup C = (A \cup C) \cap (B \cup C). \tag{2.9}$$

Proposition 2.3 (De Morgan's Law for two sets)**.** *Let $A, B \subseteq \Omega$. Then the complement of the union is the intersection of the complements, and the complement of the intersection is the union of the complements:*

$$\textbf{a. } (A \cup B)^{\complement} = A^{\complement} \cap B^{\complement} \qquad \textbf{b. } (A \cap B)^{\complement} = A^{\complement} \cup B^{\complement} \tag{2.10}$$

Definition 2.9 (Power set)**.** The **power set**

$$2^{\Omega} := \{A : A \subseteq \Omega\}$$

of a set Ω is the set of all its subsets. Note that many older texts also use the notation $\mathfrak{P}(\Omega)$ for the power set. $\square$

Remark 2.2. Note that $\emptyset \in 2^{\Omega}$ for any set Ω, even if $\Omega = \emptyset$: $2^{\emptyset} = \{\emptyset\}$. It follows that the power set of the empty set is not empty. $\square$

Definition 2.10 (Partition)**.** Let Ω be a set and $\mathfrak{A} \subseteq 2^{\Omega}$, i.e., the elements of $\mathfrak{A}$ are subsets of Ω. We call $\mathfrak{A}$ a **partition** or a **partitioning** of Ω if

(a) If $A, B \in \mathfrak{A}$ such that $A \neq B$ then $A \cap B = \emptyset$. In other words, $\mathfrak{A}$ consists of mutually disjoint subsets of Ω.

(b) Each $x \in \Omega$ is an element of some $A \in \mathfrak{A}$. $\square$

Definition 2.11 (Size of a set)**.**

a. Let X be a finite set, i.e., a set which only contains finitely many elements. We write $|X|$ for the number of its elements, and we call $|X|$ the **size** of the set X.

b. For infinite, i.e., not finite sets Y, we define $|Y| := \infty$. $\square$

2.2 The Proper Use of Language in Mathematics: Any vs All, etc

2.2.0.1 OR vs. EITHER ... OR

Note that "OR" in mathematics always is an **inclusive or**, i.e., "A OR B" means "A OR B OR BOTH". More generally, "A OR B OR ..." means "at least one of A, B, ...".

To rule out that more than one of the choices is true you must use a phrase like "EXACTLY ONE OF A, B, C, ..." or "EITHER A OR B OR C OR ...". We refer to this as an **exclusive or**.

2.2.0.2 Some Convenient Shorthand Notation

- $\forall x \dots$ For all x...
- $\exists x$ s.t. $\dots$ There exists an x such that $\dots$
- $\exists! x$ s.t. $\dots$ There exists a UNIQUE x such that $\dots$
- $P \Rightarrow Q$ If P then Q
- $P \Leftrightarrow Q$ P iff Q, i.e., P if and only if Q

2.3 Numbers

Definition 2.12 (Types of numbers)**.** Here is a definition of the various kinds of numbers in a nutshell.

$\mathbb{N} := \{1, 2, 3, \dots\}$ denotes the set of **natural numbers**.

$\mathbb{Z} := \{0, \pm 1, \pm 2, \pm 3, \dots\}$ denotes the set of all **integers**.

$\mathbb{Q} := \{n/d : n \in \mathbb{Z}, d \in \mathbb{N}\}$ (fractions of integers) denotes the set of all **rational numbers**.

$\mathbb{R} :=$ {all integers or decimal numbers with finitely or infinitely many decimal digits} denotes the set of all **real numbers**.

$\mathbb{R} \setminus \mathbb{Q} =$ {all real numbers which cannot be written as fractions of integers} denotes the set of all **irrational numbers**. There is no special symbol for irrational numbers. Example: $\sqrt{2}$ and π are irrational. $\square$

$\mathbb{N}_0 := \mathbb{Z}_+ := \mathbb{Z}_{\geq 0} := \{0, 1, 2, 3, \dots\}$ denotes the set of nonnegative integers,
$\mathbb{R}_+ := \mathbb{R}_{\geq 0} := \{x \in \mathbb{R} : x \geq 0\}$ denotes the set of all nonnegative real numbers,
$\mathbb{R}^+ := \mathbb{R}_{>0} := \{x \in \mathbb{R} : x > 0\}$ denotes the set of all positive real numbers,
$\mathbb{R}_{\neq 0} := \{x \in \mathbb{R} : x \neq 0\}$. $\square$

Definition 2.13 (Intervals of Numbers). For $a, b \in \mathbb{R}$ we have the following intervals.

- $[a, b] := \{x \in \mathbb{R} : a \leq x \leq b\}$ is the **closed interval** with endpoints a and b.
- $]a, b[:= \{x \in \mathbb{R} : a < x < b\}$ is the **open interval** with endpoints a and b.
- $[a, b[:= \{x \in \mathbb{R} : a \leq x < b\}$ and $]a, b] := \{x \in \mathbb{R} : a < x \leq b\}$ are **half-open intervals** with endpoints a and b.

$$\begin{aligned}]-\infty, a] &:= \{x \in \mathbb{R} : x \leq a\}, \quad]-\infty, a[:= \{x \in \mathbb{R} : x < a\}, \\ ]a, \infty[&:= \{x \in \mathbb{R} : x > a\}, \quad [a, \infty[:= \{x \in \mathbb{R} : x \geq a\}, \quad]-\infty, \infty[:= \mathbb{R} \end{aligned} \tag{2.11}$$

- $[a, a] = \{a\}$; $[a, a[=]a, a[=]a, a] = \emptyset$
- $[a, b] = [a, b[=]a, b[=]a, b] = \emptyset$ for $a \geq b$ $\square$

Definition 2.14 (Extended real numbers). $\boxed{\star}$ It is sometimes convenient to refer to the set

$$\overline{\mathbb{R}} := [-\infty, \infty] := \mathbb{R} \cup \{-\infty\} \cup \{\infty\} \tag{2.12}$$

as the **extended real numbers**. and to work with intervals such as

$$[-\infty, a] := \{-\infty\} \cup]-\infty, a], \quad]b, \infty] :=]b, \infty[\cup \{\infty\}, \dots \quad \square \tag{2.13}$$

Definition 2.15 (Extended real numbers arithmetic). **Rules for Addition:**

$$c \pm \infty = \infty \pm c = \infty, \tag{2.14}$$
$$c \pm (-\infty) = -\infty \pm c = -\infty, \tag{2.15}$$
$$\infty + \infty = \infty, \tag{2.16}$$
$$-\infty - \infty = -\infty, \tag{2.17}$$
$$(\pm\infty) \mp \infty = \textbf{UNDEFINED}. \tag{2.18}$$

Rules for Multiplication:

$$p \cdot (\pm\infty) = (\pm\infty) \cdot p = \pm\infty, \tag{2.19}$$
$$(-p) \cdot (\pm\infty) = (\pm\infty) \cdot (-p) = \mp\infty, \tag{2.20}$$
$$0 \cdot (\pm\infty) = (\pm\infty) \cdot 0 = \frac{0}{0} = 0, \quad \text{and} \quad \frac{1}{\infty} = 0, \tag{2.21}$$
$$(\pm\infty) \cdot (\pm\infty) = \infty, \tag{2.22}$$
$$(\pm\infty) \cdot (\mp\infty) = -\infty, \tag{2.23}$$

Notation 2.1 (Notation Alert for intervals of integers or rational numbers)**.** It is at times convenient to also use the notation $[\dots]$, $]\dots[$, $[\dots[$, $]\dots]$, for intervals of integers or rational numbers. We will subscript them with $\mathbb{Z}$ or $\mathbb{Q}$. For example,

$$[\,3, n\,]_{\mathbb{Z}} = [\,3,\ n] \cap \mathbb{Z} = \{k \in \mathbb{Z} : 3 \leq k \leq n\},$$
$$]-\infty, 7\,]_{\mathbb{Z}} =]-\infty, 7\,] \cap \mathbb{Z} = \{k \in \mathbb{Z} : k \leq 7\} = \mathbb{Z}_{\leq 7},$$
$$]a, b[_{\mathbb{Q}} =]a, b[\cap\mathbb{Q} = \{q \in \mathbb{Q} : a < q < b\}.$$

An interval which is not subscripted always means an interval of real numbers, but we will occasionally write, e.g., $[a, b]_{\mathbb{R}}$ rather than $[a, b]$, if the focus is on integers or rational numbers and an explicit subscript helps to avoid confusion. $\square$

Definition 2.16 (Absolute value, positive and negative part)**.** For a real number x we define its

absolute value: $|x| = \begin{cases} x & \text{if } x \geq 0, \\ -x & \text{if } x < 0. \end{cases}$

positive part: $x^+ = \max(x, 0) = \begin{cases} x & \text{if } x \geq 0, \\ 0 & \text{if } x < 0. \end{cases}$

negative part: $x^- = \max(-x, 0) = \begin{cases} -x & \text{if } x \leq 0, \\ 0 & \text{if } x > 0. \end{cases}$

If f is a real–valued function then we define the functions $|f|, f^+, f^-$ argument by argument:

$$|f|(x) := |f(x)|, \qquad f^+(x) := \big(f(x)\big)^+, \qquad f^-(x) := \big(f(x)\big)^-. \ \square$$

Definition 2.17 (Minimum and maximum)**.** For two real number x, y we define

$$\textbf{maximum:}\quad x \vee y \;=\; \max(x,y) \;=\; \begin{cases} x & \text{if } x \geq y, \\ y & \text{if } x \leq y. \end{cases}$$

$$\textbf{minimum:}\quad x \wedge y \;=\; \min(x,y) \;=\; \begin{cases} y & \text{if } x \geq y, \\ x & \text{if } x \leq y. \end{cases}$$

If f and g is are real–valued function then we define the functions $f \vee g = \max(f,g)$ and $f \wedge g = \min(f,g)$ argument by argument:

$$f \vee g(x) \;:=\; f(x) \vee g(x) \;=\; \max\big(f(x), g(x)\big), \quad f \wedge g(x) \;:=\; f(x) \wedge g(x) \;=\; \min\big(f(x), g(x)\big).\ \square$$

Assumption 2.1 (Square roots are always assumed nonnegative)**.** We will always assume that "$\sqrt{b}$" is the **positive** value unless the opposite is explicitly stated. $\square$

Proposition 2.4 (Triangle Inequality for real numbers)**.**

$$\text{(2.24)}\qquad \textbf{\textit{Triangle Inequality}}: \qquad |a_1 + a_2 + \cdots + a_n| \;\leq\; |a_1| + |a_2| + \cdots + |a_n|$$

2.4 Functions and Sequences

Definition 2.18 (Function)**.** A **function** f consists of two nonempty sets X and Y and an assignment rule $x \mapsto f(x)$ which assigns any $x \in X$ uniquely to some $y \in Y$. We write $f(x)$ for this assigned value and call it the **function value** of the **argument** x. X is called the **domain** and Y is called the **codomain** of f. We write

$$\text{(2.25)}\qquad f: X \to Y, \qquad x \mapsto f(x).$$

We read "$a \mapsto b$" as "a is assigned to b" or "a maps to b" and refer to $\mapsto$ as the **maps to operator** or **assignment operator**. The **graph** of such a function is the collection of pairs

$$\text{(2.26)}\qquad \Gamma_f \;:=\; \{\big(x, f(x)\big) : x \in X\},$$

and the subset $f(X) := \{f(x) : x \in X\}$ of Y is called the **range** of the function f. $\square$

Definition 2.19 (Inverse function)**.** Given are two nonempty sets X and Y and a function $f: X \to Y$ with domain X and codomain Y. We say that f has an **inverse function** if it satisfies all of the following conditions which uniquely determine this inverse function, so that we are justified to give it the symbol f^{-1}:

(a) $f^{-1} : Y \to X$, i.e., f^{-1} has domain Y and codomain X.
(b) $f^{-1}\big(f(x)\big) = x$ for all $x \in X$, and $f\big(f^{-1}(y)\big) = y$ for all $y \in Y$. $\square$

Definition 2.20 (Surjective, injective and bijective functions)**.** Given are two nonempty sets X and Y and a function $f : X \to Y$ with domain X and codomain Y. We say that

(a) f is "one–one" or **injective**, if for each $y \in Y$ there is at most one $x \in X$ such that $f(x) = y$.
(b) f is "onto" or **surjective**, if for each $y \in Y$ there is at least one $x \in X$ such that $f(x) = y$.
(c) f is **bijective**, f is both injective and surjective. $\square$

Remark 2.3. One can show: A function f has an inverse f^{-1} if and only if f is bijective. $\square$

Remark 2.4. If the inverse function f^{-1} exists and if $x \in X$ and $y \in Y$, then we have the relation

$$y = f(x) \quad \Leftrightarrow \quad x = f^{-1}(y)\,.$$

Definition 2.21 (Restriction/Extension of a function)**.** $\boxed{\star}$ Given are three nonempty sets A, X and Y such that $A \subseteq X$, and a function $f : X \to Y$ with domain X. We define the **restriction of f to A** as the function

(2.27) $$f\big|_A : A \to Y \quad \text{defined as} \quad f\big|_A(x) := f(x) \text{ for all } x \in A.$$

Conversely let $f : A \to Y$ and $\varphi : X \to Y$ be functions such that $f = \varphi\,|_A$. We then call φ an **extension** of f to X. $\square$

Definition 2.22. Let $n_\star$ be an integer and assume that an item x_j associated

- **either** with each integer $j \geq n_\star$, In other words, we have an item x_j assigned to each $j = n_\star, n_\star + 1, n_\star + 2, \dots$.
- **or** with each integer j such that $n_\star \leq j \leq n^\star$. In this case an item x_j is assigned to each $j = n_\star, n_\star + 1, \dots, n^\star$.

Such items can be anything, but we usually deal with numbers or outcomes or sets of outcomes of an experiment.

- In the first case we usually write $x_{n_\star}, x_{n_\star+1}, x_{n_\star+2}, \dots$ or $(x_n)_{n \geq n_\star}$ for such a collection of items and we call it a **sequence** with **start index** $n_\star$.
- In the second case we speak of a **finite sequence**, which starts at $n_\star$ and ends at $n^\star$. We write $(x_n)_{n_\star \leq n \leq n^\star}$ or $x_{n_\star}, x_{n_\star+1}, \dots, x_{n^\star}$ for such a finite collection of items.
- If we refer to a sequence $(x_n)_n$ without qualifying it as finite then we imply that we deal with an **infinite sequence**, $x_{n_\star}, x_{n_\star+1}, x_{n_\star+2}, \dots$. $\square$

Definition 2.23.
- If $(x_n)_n$ is a finite or infinite sequence and one pares down the full set of indices to a subset $\{n_1,\ n_2,\ n_3, \dots\}$ such that $n_1 < n_2 < n_3 < \dots$, then we call the corresponding thinned out sequence $(x_{n_j})_{j \in \mathbb{N}}$ a **subsequence** of that sequence.
- If this subset of indices is finite, i.e., we have $n_1 < n_2 < \cdots < n_K$ for some suitable $K \in \mathbb{N}$, then we call $(x_{n_j})_{j \leq K}$ a **finite subsequence** of the original sequence. $\square$

Definition 2.24. We give some convenient definitions and notations for monotone sequences of numbers, functions and sets.

(a) Let x_n be a sequence of extended real–valued numbers.
- We call x_n a **nondecreasing** or **increasing** sequence, if $j < n \Rightarrow x_j \leq x_n$.
- We call x_n a **strictly increasing** sequence, if $j < n \Rightarrow x_j < x_n$.
- We call x_n a **nonincreasing** or **decreasing** sequence, if $j < n \Rightarrow x_j \geq x_n$.
- We call x_n a **strictly decreasing** sequence, if $j < n \Rightarrow x_j > x_n$.
- We write $x_n \uparrow$ for nondecreasing x_n, and $x_n \uparrow x$ to indicate that $\lim\limits_{n\to\infty} x_n = x$,
- We write $x_n \downarrow$ for nonincreasing x_n, $x_n \downarrow x$ to indicate that $\lim\limits_{n\to\infty} x_n = x$. $\square$

(b) Let A_n be a sequence of sets.
- We call A_n a **nondecreasing** or **increasing** sequence, if $j < n \Rightarrow A_j \subseteq A_n$.
- We call A_n a **strictly increasing** sequence, if $j < n \Rightarrow A_j \subsetneq A_n$.
- We call A_n a **nonincreasing** or **decreasing** sequence, if $j < n \Rightarrow A_j \supseteq A_n$.
- We call A_n a **strictly decreasing** sequence, if $j < n \Rightarrow A_j \subsetneq A_n$.
- We write $A_n \uparrow$ for nondecreasing A_n, and $A_n \uparrow A$ to indicate that $\bigcup_n A_n = A$,
- We write $A_n \downarrow$ for nonincreasing A_n, $A_n \downarrow A$ to indicate that $\bigcap_n A_n = A$. $\square$

Definition 2.25 (Countable and uncountable sets)**.** Let X be a set.

(a) We call X **countable** if its elements can be written as a finite sequence (those are the finite sets) $X = \{x_1, x_2, \dots, x_n\}$ or as an infinite sequences. $X = \{x_1, x_2, \dots\}$.

(b) We call X **countably infinite** X is both countable and infinite, i.e., there is an infinite sequence. $X = \{x_1, x_2, \dots\}$. of distinct items x_j.

(c) We call a nonempty set **uncountable** if it is not countable, i.e., its elements cannot be sequenced.

(d) By convention the empty set, $\emptyset$, is countable. $\square$

Fact 2.1. *One can prove the following important facts:*

(a) *The integers are countable. (Easy: $\mathbb{Z} = \{0, -1, 1, -2, 2, -3, 3, \dots\}$) lists all elements of $\mathbb{Z}$ in a sequence.*
(b) *Subsets of countable sets are countable. (Easy: If $X = \{x_1, x_2, \dots\}$ and $A \subseteq X$, then remove all x_j that are not in A. That subsequence lists the elements of A.*
(c) *Countable unions of countable sets are countable: If $A_1, A_2, \dots$ is a finite or infinite sequence of sets, then $A_1 \cup A_2 \cup \cdots$ is countable.*
(d) *The rational numbers $\mathbb{Q}$ are countable. A proof is given below.*
(e) *The real numbers $\mathbb{R}$ are uncountable!* $\square$

Definition 2.26 (Families)**.** $\boxed{\star}$ Let I and X be nonempty sets such that each $i \in I$ is associated with some $x_i \in X$. Then

a. $(x_i)_{i \in I}$ is called an **indexed family** or simply a **family** in X.
b. I is called the **index set** of the family.
c. For each $i \in IJ$, x_i is called a **member of the family** $(x_i)_{i \in I}$. $\square$

Definition 2.27 (Arbitrary unions and intersections of families of sets)**.** Let J be an arbitrary, nonempty set and $(A_j)_{j \in J}$ a family of sets with index set J. We define

- The **union** $\bigcup\limits_{j \in J} A_j := \{x : \exists\, i_0 \in J \text{ s.t. } x \in A_{i_0}\}$.
- The **intersection** $\bigcap\limits_{j \in J} A_j = \{x : \forall\, i \in J : x \in A_i\}$.
- If the sets A_i are disjoint, we often write $\biguplus\limits_{j \in J} A_j$ rather than $\bigcup\limits_{j \in J} A_j$.
- Let $(B_j)_{j \in J}$ be a family of subsets of a set X. We call this family a **partition** or a **partitioning** of X if the corresponding set of sets $\{B_i : i \in J\}$ is a partition of X:
(a) $i \neq j \Rightarrow B_i \cap B_j = \emptyset$ **(b)** $X = \biguplus\limits_{j \in J} B_j$. See Definition 2.10 on p.9. $\square$

Notation 2.2. Empty unions and intersections:

$$\bigcup_{i \in \emptyset} A_i := \emptyset, \text{ always}; \qquad \bigcap_{i \in \emptyset} A_i := \Omega, \text{ if there is a universal set, } \Omega\,. \tag{2.28}$$

Theorem 2.1 (De Morgan's Law)**.** *Let J be an arbitrary, nonempty set. Let $(A_j)_{j \in J}$ be a collection of subsets of a set Ω. Then the complement of the union is the intersection of the complements, and the complement of the intersection is the union of the complements:*

$$(2.29)\qquad \textbf{(a)}\ \left(\bigcup_{j\in J} A_j\right)^{\complement} = \bigcap_{j\in J} A_j^{\complement}; \qquad \textbf{(b)}\ \left(\bigcap_{j\in J} A_j\right)^{\complement} = \bigcup_{k} A_k^{\complement};$$

Remark 2.5. Note that (2.29) holds true for ANY index set J. In particular, for finite and infinite sequences of sets. □

Proposition 2.5 (Distributivity of unions and intersections)**.** *Let $(A_n)_n$ be a finite or infinite sequence of sets and let B be a set. Then*

$$\bigcup_j (B\cap A_j) = B\cap\bigcup_j A_j, \tag{2.30}$$

$$\bigcap_{j\in I} (B\cup A_j) = B\cup\bigcap_j A_j. \tag{2.31}$$

Proposition 2.6 (Rewrite unions as disjoint unions)**.** *Let $(A_j)_{j\in\mathbb{N}}$ be a sequence of sets which all are contained within the universal set Ω. Let*

$$B_n := \bigcup_{j=1}^n A_j = A_1\cup A_2\cup\dots\cup A_n\ (n\in\mathbb{N}),$$
$$C_1 := A_1 = B_1, \quad C_{n+1} := A_{n+1}\setminus B_n\ (n\in\mathbb{N}).$$

Then

- ***(a)*** *The sequence $(B_j)_j$ is nondecreasing: $m<n \Rightarrow B_m\subseteq B_n$.*
- ***(b)*** *For each $n\in\mathbb{N}$, $\bigcup\limits_{j=1}^n A_j = \bigcup\limits_{j=1}^n B_j$. Further, $\bigcup\limits_{j=1}^\infty A_j = \bigcup\limits_{j=1}^\infty B_j$.*
- ***(c)*** *The sets C_j are mutually disjoint, $\bigcup\limits_{j=1}^n A_j = \biguplus\limits_{j=1}^n C_j$ for all n, and $\bigcup\limits_{j=1}^\infty A_j = \bigcup\limits_{j=1}^\infty C_j$.*
- ***(d)*** *The sets $C_j\ (j\in\mathbb{N})$ form a partition of the set $\bigcup\limits_{j=1}^\infty A_j$.*

2.5 Preimages

Definition 2.28. Let X,Y be two nonempty sets. Let $f: X\to Y$ and $B\subseteq Y$. Then

$$f^{-1}(B) := \{x\in X : f(x)\in B\} \tag{2.32}$$

is a subset of X which we call the **preimage** of B under f. □

Notation 2.3 (Notational conveniences for preimages)**.** If we have a set that is written as $\{\dots\}$ then we may write $f^{-1}\{\dots\}$ instead of $f^{-1}(\{\dots\})$. Specifically for singletons $\{y\}$ such that $y \in Y$, it is OK to write $f^{-1}\{y\}$.

- You are **NOT** allowed to write $f^{-1}(y)$ instead of $f^{-1}\{y\}$, since it is a very bad idea to confound elements y and subsets $\{y\}$ of Y. $\square$

Proposition 2.7. *Some simple properties:*

$$f^{-1}(\emptyset) \;=\; \emptyset \tag{2.33}$$

$$B_1 \subseteq B_2 \subseteq Y \Rightarrow f^{-1}(B_1) \subseteq f^{-1}(B_2) \quad (\textbf{\textit{monotonicity of}}\ f^{-1}\{\dots\}) \tag{2.34}$$

$$f^{-1}(Y) \;=\; X \quad \textit{always!} \tag{2.35}$$

Notation 2.4 (Notational conveniences for preimages II)**. REMOVED: duplicate to Notation 2.3 on p.18.**

Theorem 2.2 (f^{-1} is compatible with all basic set ops)**.** *Assume that X, Y be nonempty, $f : X \to Y$, J is an arbitrary index set.* [2] *Further assume that $B \subseteq Y$ and that $B_j \subseteq Y$ for all j. Then*

$$f^{-1}\Big(\bigcap_{j\in J} B_j\Big) \;=\; \bigcap_{j\in J} f^{-1}(B_j) \tag{2.36}$$

$$f^{-1}\Big(\bigcup_{j\in J} B_j\Big) \;=\; \bigcup_{j\in J} f^{-1}(B_j) \tag{2.37}$$

$$f^{-1}(B^{\complement}) \;=\; \big(f^{-1}(B)\big)^{\complement} \tag{2.38}$$

$$B_1 \cap B_2 = \emptyset \;\Rightarrow\; f^{-1}(B_1) \cap f^{-1}(B_2) = \emptyset. \tag{2.39}$$

$$f^{-1}(B_1 \setminus B_2) \;=\; f^{-1}(B_1) \setminus f^{-1}(B_2) \tag{2.40}$$

$$f^{-1}(B_1 \Delta B_2) \;=\; f^{-1}(B_1) \Delta f^{-1}(B_2) \tag{2.41}$$

Note that (2.39) implies that the preimages of a disjoint family form a disjoint family.

Proposition 2.8 (Preimages of function composition)**.** *Let X, Y, Z be arbitrary, nonempty sets. Let $f : X \to Y$ and $g : Y \to Z$ and $h : X \to Z$ the composition*

$$h(x) \;=\; g \circ f(x) \;=\; g\big(f(x)\big).$$

Let $U \subseteq X$ and $W \subseteq Z$. Then

$$(g \circ f)^{-1} \;=\; f^{-1} \circ g^{-1},\ \textit{i.e.,}\quad (g \circ f)^{-1}(W) \;=\; f^{-1}\big(g^{-1}(W)\big) \ \textit{for all } W \subseteq Z. \tag{2.42}$$

Definition 2.29 (Direct image)**.** $\boxed{\star}$ Let X, Y be two nonempty sets and $f : X \to Y$. Let $A \subseteq X$. We call the set

$$f(A) := \{f(a) : a \in A\}. \tag{2.43}$$

which consists of all function values of arguments in A, the **direct image** of A under f. ☐

Notation 2.5 (Notational conveniences for direct images)**.** As we do for preimages, if we deal with a set that is written as $\{\dots\}$, then we may write $f\{\dots\}$ instead of $f(\{\dots\})$. In particular, we can write $f\{x\}$ for singletons $\{x\} \subseteq X$. ☐

The same symbol f is used for the original function $f : X \to Y$ and the direct image which we can think of as a function

$$2^X \to 2^Y; \qquad A \mapsto f(A) = \{f(a) : a \in A\}, \quad (A \subseteq X).$$

Be careful not to let this confuse you! ☐

2.6 Infimum and Supremum: Generalized Minimum and Maximum

Definition 2.30 (Minimum, maximum, infimum, supremum)**.** $\boxed{\star}$ Let $A \subseteq \mathbb{R}$, $A \neq \emptyset$, and let l and u be real numbers.

(a) We call l a **lower bound** of A if $l \leq a$ for all $a \in A$.
(b) We call u an **upper bound** of A if $u \geq a$ for all $a \in A$.
(c) We call A **bounded above** if this set has an upper bound.
(d) We call A **bounded below** if A has a lower bound.
(e) We call A **bounded** if A is both bounded above and bounded below.

(f) The **minimum** of A, if it exists, is the unique lower bound l of A such that $l \in A$.
(g) A **maximum** of A, if it exists, is the unique upper bound u of A such that $u \in A$.

Since they are uniquely determined by A, we may write $\min(A)$ for the minimum of A and $\max(A)$ for the maximum of A.

(h) If A is bounded below (i.e., A has lower bounds), we call the maximum of those bounds the **infimum** of A. Thus, it is the **greatest lower bound** of A. We write $\inf(A)$ or g.l.b.(A). Otherwise (A is not bounded below), we define $\inf(A) := -\infty$.
(i) If A is bounded above (i.e., A has upper bounds), we call the minimum of those bounds the **supremum** of A. Thus, it is the **least upper bound** of A. We write $\sup(A)$ or l.u.b.(A). Otherwise (A is not bounded above), we define $\sup(A) := \infty$. ☐

Remark 2.6. Here is the cookbook approach to infima and suprema. (NOT OPTIONAL!)

- Infima are generalized minima and suprema are generalized maxima.
- Think of $\inf(A)$ as a minimum that does not need to belong to A.
- Traverse the lower bounds of A from the left (from $-\infty$) to the rigt until you "hit" A. That's the greatest lower bound. That's $\inf(A)$.
- Think of $\sup(A)$ as a maximum that does not need to belong to A.
- Traverse the upper bounds of A from the right ($+\infty$) to the left until you "hit" A. That's the least (smallest) upper bound. That's $\sup(A)$. ☐

Definition 2.31. $\boxed{\star}$ Let X be an arbitrary set (need not be numbers or elements of $\mathbb{R}^d$!) and $A \subseteq X$.
Let $f : X \to \mathbb{R}$ be real–valued. The **supremum** and **infimum** of f on A are defined as

$$\sup_A f \ := \ \sup_{x\in A} f(x) \ := \ \sup\{f(x) : x \in A\} \tag{2.44}$$

$$\inf_A f \ := \ \inf_{x\in A} f(x) \ := \ \inf\{f(x) : x \in A\}. \tag{2.45}$$

The **supremum** and **infimum** of a family of real numbers $(x_i)_{i\in I}$ $(x_i)_{i\in I}$ are defined as

$$\sup\,(x_i) \ := \ \sup_i\,(x_i) \ := \ \sup\,(x_i)_i \ := \ \sup\,(x_i)_{i\in I} \ := \ \sup_{i\in I} x_i \ := \ \sup\,\{x_i : i \in I\}. \tag{2.46}$$

$$\inf\,(x_i) \ := \ \inf_i\,(x_i) \ := \ \inf\,(x_i)_i \ := \ \inf\,(x_i)_{i\in I} \ := \ \inf_{i\in I} x_i \ := \ \inf\,\{x_i : i \in I\}. \tag{2.47}$$ ☐

The definition above for families extends to sequences x_n, defined for $n = n_*, n_*+1, n_*+2, \dots$.
The **supremum** and **infimum** of a sequence of real numbers $(x_n)_{n\geq n_*}$ are defined as

$$\sup\,(x_n) \ := \ \sup\,(x_n)_{n\geq n_*} \ := \ \sup_{n\geq n_*} x_n \ = \ \sup\,\{x_n : n = n_*, n_*+1, n_*+2, \dots\} \tag{2.48}$$

$$\inf\,(x_n) \ := \ \inf\,(x_n)_{n\geq n_*} \ := \ \inf_{n\geq n_*} x_n \ = \ \inf\,\{x_n : n = n_*, n_*+1, n_*+2, \dots\} \tag{2.49}$$ ☐

Theorem 2.3. $\boxed{\star}$ *Let $\alpha_1 \geq \alpha_2 \geq \cdots$ be a nonincreasing sequence and $\beta_1 \leq \beta_2 \leq \cdots$ a nondecreasing sequence of real numbers. Then*

(a) $\lim\limits_{n\to\infty} \alpha_n$ *exists (might be $-\infty$) and equals* $\inf\limits_{n\in\mathbb{N}} \alpha_n$.
(b) $\lim\limits_{n\to\infty} \beta_n$ *exists (might be ∞) and equals* $\sup\limits_{n\in\mathbb{N}} \beta_n$.

Let $\emptyset \neq A \subseteq \mathbb{R}$ and $f_n, g_n : A \to \mathbb{R}$ two sequences of real–valued functions on A, such that

$(f_n)_n$ *is nonincreasing, i.e.,* $f_1 \geq f_2 \geq \cdots$, *i.e.,* $f_1(x) \geq f_2(x) \geq \cdots$, *for all $x \in A$,*
$(g_n)_n$ *is nonincreasing, i.e.,* $g_1 \leq g_2 \leq \cdots$, *i.e.,* $g_1(x) \leq g_2(x) \leq \cdots$, *for all $x \in A$,*

Then

(c) $x \to \lim\limits_{n\to\infty} f_n(x)$ *exists (might be* $-\infty$ *for some or all* $x \in A$*) and equals* $x \to \inf\limits_{n\in\mathbb{N}} f_n(x)$.

(d) $x \to \lim\limits_{n\to\infty} g_n(x)$ *exists (might be* ∞ *for some or all* $x \in A$*) and equals* $x \to \sup\limits_{n\in\mathbb{N}} g_n(x)$.

2.7 Cartesian Products

Definition 2.32 (Cartesian Product)**.** Let X and Y be two sets The set

$$X \times Y := \{(x,y) : x \in X, y \in Y\} \tag{2.50}$$

is called the **cartesian product** of X and Y. We write X^2 as an abbreviation for$X \times X$.

Note that the order is important: (x,y) and (y,x) are different unless $x = y$.

This definition generalizes to more than two sets as follows:

Let $X_1, X_2, \dots, X_n$ be sets. The set

$$X_1 \times X_2 \cdots \times X_n := \{(x_1, x_2, \dots, x_n) : x_j \in X_j \text{ for each } j = 1, 2, \dots n\} \tag{2.51}$$

is called the cartesian product of $X_1, X_2, \dots, X_n$.
We write X^n as an abbreviation for$X \times X \times \cdots \times X$. $\square$

Proposition 2.9. *Let* X_1, X_2, X_n *be finite, nonempty sets. Then,*
The size of the cartesian product is the product of the sizes of its factors, i.e.,

$$\left| X_1 \times X_2 \times \cdots \times X_n \right| = \left|X_1\right| \cdot \left|X_2\right| \cdot \left|X_3\right| \cdots \left|X_n\right|. \tag{2.52}$$

2.8 Indicator Functions

Definition 2.33 (Indicator function of a set)**.** Let Ω be a nonempty set and $A \subseteq \Omega$. Let $\mathbf{1}_A : \Omega \to \{0,1\}$ be the function defined as

$$\mathbf{1}_A(\omega) := \begin{cases} 1 & \text{if } \omega \in A, \\ 0 & \text{if } \omega \notin A. \end{cases} \tag{2.53}$$

$\mathbf{1}_A$ is called the **indicator function** [3] of the set A. $\square$

Proposition 2.10. *Let* $A_1, A_2, \dots$ *be subsets of* Ω. *Then*

$$A_1 \subseteq A_2 \;\Rightarrow\; \mathbf{1}_{A_1} \leq \mathbf{1}_{A_2}, \tag{2.54}$$

$$\mathbf{1}_{A_1 \cap A_2} = \min(\mathbf{1}_{A_1}, \mathbf{1}_{A_2}), \qquad \mathbf{1}_{\bigcap [A_n \,:\, n \in \mathbb{N}]} = \inf_{n \in \mathbb{N}} \mathbf{1}_{A_n}, \tag{2.55}$$

$$\mathbf{1}_{A_1 \cup A_2} = \max(\mathbf{1}_{A_1}, \mathbf{1}_{A_2}), \qquad \mathbf{1}_{\bigcup [A_n \,:\, n \in \mathbb{N}]} = \sup_{n \in \mathbb{N}} \mathbf{1}_{A_n}, \tag{2.56}$$

$$\mathbf{1}_{A_1^\complement} = 1 - \mathbf{1}_{A_1}, \tag{2.57}$$

$$\mathbf{1}_{A_1 \uplus A_2} = \mathbf{1}_{A_1} + \mathbf{1}_{A_2}\,, \qquad \mathbf{1}_{\biguplus [A_n \,:\, n \in \mathbb{N}]} = \sum_{n \in \mathbb{N}} \mathbf{1}_{A_n}\,, \quad (A_1, A_2, \dots \textit{ disjoint}). \tag{2.58}$$

[3] In abstract algebra $\mathbf{1}_A$ is often called the **characteristic function** of A. Some authors write χ_A or $\mathbb{1}_A$ instead of $\mathbf{1}_A$.

3 Calculus Revisited

3.1 Absolute Convergence of Series

Definition 3.1 (Absolute Convergence)**.** We say that an infinite series $\sum a_j (a_j \in \mathbb{R})$ is **absolutely convergent** and also, that it **converges absolutely**, if

$$\sum_{j=1}^{\infty} |a_j| \;=\; |a_1| + |a_2| + |a_3| + \cdots \;<\; \infty\,, \quad \square$$

Theorem 3.1. *If the series $\sum a_j (a_j \in \mathbb{R})$ is absolutely convergent, then the following holds true:*

(a) *The series $\sum a_j$ itself converges, i.e., there is $-\infty < a < \infty$ such that $\sum_{j=1}^{\infty} a_j \;=\; a$,*

(b) *ANY rearrangement $\sum_{j=1}^{\infty} a_{n_j} \;=\; a_{n_1} + a_{n_2} + \cdots$ converges to the same limit as $\sum a_j$.*

*We speak of a **rearrangement** of a sequence $(a_j)_{n\in\mathbb{N}}$ (a series $\sum a_j$) if its members are reshuffled into a sequence $(b_j)_{n\in\mathbb{N}}$ (a series $\sum b_j$) as follows: There are indices $n_j \in \mathbb{N}$ such that*

$$b_1 = a_{n_1},\; b_2 = a_{n_2},\; b_3 = a_{n_3},\; \dots,$$

and those indices satisfy the following:

(1) *They are distinct: $i \neq j \;\Rightarrow\; n_i \neq n_j$.*

(2) *They leave no gaps in the set $\mathbb{N}$ of all indices: For each $k \in \mathbb{N}$ there is $j \in \mathbb{N}$ such that $k = n_j$.* [4]

Theorem 3.2. *If the series $\sum a_j (a_j \in \mathbb{R})$ satisfies $a_j \geq 0$ for all j, then*

- *ANY rearrangement $\sum_{j=1}^{\infty} a_{n_j}$ possesses the same limit, finite or infinite, as $\sum_{j=1}^{\infty} a_j$.*
- *In particular, if $\sum a_j$ is not convergent, then $\sum_{j=1}^{\infty} a_{n_j} = \infty$ for each rearrangement.*

Proposition 3.1.
(1) *A series which only has finitely many nonzero terms converges absolutely.*
(2) *If $|a_n| \leq |b_n|$ for all n and $\sum b_n$ converges absolutely, then $\sum a_n$ converges absolutely.*

Theorem 3.3. *Let S be some (abstract) nonempty set and $f : S \to \mathbb{R}$ some real–valued function on S. Assume that $S^* := \{x \in S : f(x) \neq 0\}$ is countable, i.e. $S^* = \{x_1, x_2, \cdots\}$ for some finite or infinite sequence $x_1, x_2, \cdots$ of elements of S and that at least one of the following two is true:*

(a) *$f(x_j) \geq 0$, for all j,* ***(b)*** *the series $\sum f(x_j)$ is absolutely convergent.*

- *Then, ANY rearrangement $\sum_{j=1}^{\infty} f(x_{n_j})$ of the $f(x_j)$ possesses the same value as $\sum_{j=1}^{\infty} f(x_j)$.*

Theorem 3.4. *Assume that $J_1, J_2, \dots$ is a countable collection of disjoint subsets of $\mathbb{N}$. and $J := J_1 \uplus J_2 \uplus \cdots$. Let $\sum_{j \in J_1} a_j$, $\sum_{j \in J_2} a_j, \dots$ be a corresponding collection of series such that*

- *$a_j \geq 0$, for all $j \in J$ or • $\sum_{j \in J} a_j$ is absolutely convergent.*

Then
$$\sum_{j \in J_1} a_j + \sum_{j \in J_2} a_j + \cdots = \sum_{j \in J} a_j \, .$$

3.2 Integration – The Riemann Integral

3.2.1 The Riemann Integral of a Step Function

Definition 3.2 (d dimensional rectangles)**.**
For $a, b \in \mathbb{R}$, $a \prec b$ here denotes either $a < b$ or $a \leq b$.
Let $a_1 \leq b_1,\, a_2 \leq b_2,\, \dots,\, a_d \leq b_d$, be d pairs of numbers ($d \in \mathbb{N}$). We call the set

$$\{\vec{x} = (x_1, \dots, x_d) \in \mathbb{R}^d : a_1 \prec x_1 \prec b_1,\, a_2 \prec x_2 \prec b_2, \dots,\, a_d \prec x_d \prec b_d\}$$

a d**–dimensional rectangle** (simply **rectangles**, if there is no confusion about d).

The set $\{\vec{x} \in \mathbb{R}^d : a_1 \prec x_1 \prec b_1,\, a_2 \prec x_2 \prec b_2, \dots,\, a_d \prec x_d \prec b_d\}$ has alternate (and more familiar) notation in the following special cases. We also write

- $]a_1, b_1[\times \cdots \times]a_d, b_d[$, if $a_j < x_j < b_j$ for all j: **open rectangles**,
- $]a_1, b_1] \times \cdots \times]a_d, b_d]$, if $a_j < x_j \leq b_j$ for all j, or
- $[a_1, b_1[\times \cdots \times [a_d, b_d[$, if $a_j \leq x_j < b_j$ for all j:
  half open rectangles, also called **half closed rectangles**),
- $[a_1, b_1] \times \cdots \times [a_d, b_d]$, if $a_j \leq x_j \leq b_j$ for all j: **closed rectangles**).

Usually, onedimensional rectangles are called **intervals** and 3 dimensional rectangles are called **quads** or **boxes**. $\square$

Definition 3.3 (Lebesgue measure of d dimensional rectangles)**.** Let $a \prec b$ again stand for either $a < b$ or $a \leq b$. Given are $d \in \mathbb{N}$ and $a_j, b_j \in \mathbb{R}$ such that $a_j \leq b_j$, for $j = 1, 2, \dots, d$. Let

$$R := \{\vec{x} = (x_1, \dots, x_d) \in \mathbb{R}^d : a_1 \prec x_1 \prec b_1, a_2 \prec x_2 \prec b_2, \dots, a_d \prec x_d \prec b_d\}$$

be a d–dimensional rectangle. We call

$$\lambda^d(R) := (b_1 - a_1)(b_2 - a_2) \dots (b_d - a_d) \tag{3.1}$$

the d**–dimensional Lebesgue measure** of R. We also simply speak of the **Lebesgue measure** of R, if there is no confusion about d).

We extend λ^d as follows.

- If $a_j < b_j$ for all j and $a_j = -\infty$ and/or $b_j = \infty$ for at least one j, then $\lambda^d(R) := \infty$.
- If $a_j = b_j$ for at least one j, then $\lambda^d(R) := 0$, even if not all a_j and b_j are finite.
- $\lambda^d(\emptyset) := 0$.
- If $R_1, R_2, \dots$ is a finite or infinite sequence of disjoint rectangles, i.e., $R_i \cap R_j = \emptyset$ for $i \neq j$, then we define the **Lebesgue measure** of the union by "σ**–addititivity**" as follows:

$$\lambda^d(R_1 \uplus R_2 \uplus \cdots) := \lambda^d(R_1) + \lambda^d(R_2) + \cdots \quad \square \tag{3.2}$$

Definition 3.4. A function $\varphi : \mathbb{R}^d \to \mathbb{R}$ is called a **step function** if there is $n \in \mathbb{N}$, a list of d–dimensional rectangles $A_1, \dots, A_n$, and a list of real numbers $c_1, \dots, c_n$, such that

$$\varphi(\vec{x}) = \sum_{j=1}^{n} c_j \mathbf{1}_{A_j}(\vec{x}). \tag{3.3}$$

We call

$$\int \varphi(\vec{x})\, d\vec{x} := \int_{\mathbb{R}^d} \varphi(\vec{x})\, d\vec{x} := \iint \cdots \int_{\mathbb{R}^d} f(\vec{x})\, d\vec{x} := \sum_{j=1}^{n} c_j\, \lambda^d(A_j) \tag{3.4}$$

the (d dimensional) **Riemann integral** of the step function φ.

Here,

$$\vec{x} \mapsto \mathbf{1}_{A_j}(\vec{x}) = \begin{cases} 1 & \text{if } \vec{x} \in A_j, \\ 0 & \text{else}, \end{cases}$$

is the indicator function [5] of the subset A_j of $\mathbb{R}^d$. $\square$

3.2.2 The Riemann Integral as the Limit of Riemann Sums

3.2.2.1 The Riemann Integral in Dimension 1 $\boxed{\star}$

[5] see Definition 2.33 (indicator function for a set) on p.21.

Definition 3.5. Let Π be defined as in (**??**), and let $f : [a,b] \to \mathbb{R}$ be a function on $[a,b]$. We call

$$\mathcal{RS}(f;\Pi) := \sum_{j=1}^{n} f(u_j)(y_j - y_{j-1})$$

the **Riemann sum** of f with respect to Π, and we call

$$\int_a^b f(x)dx := \lim_{\|\Pi\|\to 0} \mathcal{RS}(f;\Pi)$$

the **Riemann integral** of f on $[a,b]$, provided that this limit exists.

3.2.2.2 The Riemann Integral in Dimension 2 $\boxed{\star}$

Definition 3.6. Let Π be defined as in (**??**). Consider the rectangle

$$R := [\,a^{(1)}, b^{(1)}\,] \times [\,a^{(2)}, b^{(2)}\,]\,.$$

Let $f : R \to \mathbb{R}\,;\ \vec{y} \mapsto f(\vec{y}),$ be a real–valued function on R. We call

$$\mathcal{RS}(f;\Pi) := \sum_{j_1=1}^{n}\sum_{j_2=1}^{n} f(\vec{u}(j_1,j_2))\,(y^{(1)}_{j_1} - y^{(1)}_{j_1-1}) \cdot (y^{(2)}_{j_2} - y^{(2)}_{j_2-1}) \tag{3.5}$$

the **Riemann sum** of f with respect to Π, and we call

$$\iint_R f(\vec{y})\,d\vec{y} := \lim_{\|\Pi\|\to 0} \mathcal{RS}(f;\Pi) \tag{3.6}$$

the **Riemann integral** of f on R, provided that this limit exists. □

3.2.2.3 The Riemann Integral in d Dimensions $\boxed{\star}$

Definition 3.7. Let Π be as in (**??**) and $R := [\,a^{(1)}, b^{(1)}\,] \times [\,a^{(2)}, b^{(2)} \times \cdots \times [\,a^{(d)}, b^{(d)}\,]$.
Let $f : R \to \mathbb{R}\,;\ \vec{y} \mapsto f(\vec{y}),$ be a real–valued function on R. We call

$$\mathcal{RS}(f;\Pi) := \sum_{j_1,\dots,j_d=1}^{n} f(\vec{u}(j_1,j_2,\dots,j_d))\,(y^{(1)}_{j_1} - y^{(1)}_{j_1-1}) \cdot (y^{(2)}_{j_2} - y^{(2)}_{j_2-1}) \cdots (y^{(d)}_{j_d} - y^{(d)}_{j_d-1}) \tag{3.7}$$

the **Riemann sum** of f with respect to Π, and we call

$$\iint\cdots\int_R f(\vec{y})\,d\vec{y} := \lim_{\|\Pi\|\to 0} \mathcal{RS}(f;\Pi) \tag{3.8}$$

the **Riemann integral** aka **proper Riemann integral** of f on R, provided that this limit exists.
□

3.3 Improper Integrals and Integrals Over Subsets

Definition 3.8 (Improper Riemann integral)**.** Let $f : [a, \infty[\to \mathbb{R}, \quad g :]-\infty, b] \to \mathbb{R}, \quad h :]-\infty, \infty[\to \mathbb{R}$.
Their **improper Riemann integrals** are defined as follows:

$$\begin{aligned} \int_a^\infty f(x)\,dx &= \lim_{b\to\infty} \int_a^b f(x)\,dx\,, \\ \int_{-\infty}^b f(x)\,dx &= \lim_{a\to-\infty} \int_a^b f(x)\,dx\,. \\ \int_{-\infty}^\infty f(x)\,dx &= \lim_{a\to-\infty} \lim_{b\to\infty} \int_a^b f(x)\,dx\,. \quad \square \end{aligned} \tag{3.9}$$

Definition 3.9 (Riemann integrability)**.**
- **(a)** Let $A \subseteq \mathbb{R}^d$ be a d dimensional rectangle and $\varphi : A \to \mathbb{R}$, a real–valued function on A. We say that φ is **Riemann integrable**, if its proper Riemann integral, as specified (for general d) in Definition 3.7 on p.26, exists <u>and is finite</u>.
- **(b)** Let ψ be one of the functions f, g, h specified in Definition 3.8 (Improper Riemann integral) above. We say that ψ is **Riemann integrable**, if its improper integral, as specified in Definition 3.8 above, exists <u>and is finite</u>.
- **(c)** If φ is as above and α, its proper Riemann integral exists, then we call α the (proper) Riemann integral, even if $\alpha = \pm\infty$ (and thus, φ is <u>not</u> Riemann integrable).
- **(c)** If ψ is as above and β, its improper integral exists, then we call β the improper Riemann integral of ψ, even if $\beta = \pm\infty$ (and thus, ψ is <u>not</u> Riemann integrable). $\square$

Definition 3.10. **(A):** Let $R \subseteq \mathbb{R}^d$ be a d dimensional rectangle, $d \in \mathbb{N}$, and $\emptyset \neq A \subseteq R$. Let $f : A \to \mathbb{R}$ be a function on A such that the function

$$\mathbf{1}_A f : R \longrightarrow \mathbb{R} \ \vec{x} \mapsto \mathbf{1}_A(\vec{x}) f(\vec{x}) \quad = \begin{cases} f(\vec{x}) & \text{if } \vec{x} \in A, \\ 0, & \text{else}\,, \end{cases} \tag{3.10}$$

possesses a Riemann integral. Then we call

$$\underset{A}{\iint \cdots \int} f(\vec{x})\,d\vec{x} := \underset{R}{\iint \cdots \int} \mathbf{1}_A(\vec{x})\, f(\vec{x})\,d\vec{x} \tag{3.11}$$

the **Riemann integral of f on** (also, **over**,) **the subset** A.

We are not yet completely done with the case $d = 1$, since we also must consider improper integrals of functions of a single variable. We do that now.

(B): Let $I \subseteq \mathbb{R}$ be an interval of infinite length, i.e., I is one of $[a, \infty[, \ ]-\infty, b], \ ]-\infty, \infty[$, for suitable $a, b \in \mathbb{R}$. Let $\emptyset \neq A \subseteq I$ and $f : A \to \mathbb{R}$ a function on A such that the function

(3.12)
$$\mathbf{1}_A f : I \longrightarrow \mathbb{R} \;\; x \mapsto \mathbf{1}_A(x) f(x) \;\; = \begin{cases} f(x) & \text{if } x \in A, \\ 0, & \text{else}\,, \end{cases}$$

possesses an improper Riemann integral. Then we call

(3.13)
$$\int_A f(x)\,dx \;:=\; \int_I \mathbf{1}_A(x)\,f(x)\,dx$$

the **Riemann integral of f on** (also, **over,**) **the subset** A. □

Theorem 3.5. *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a real–valued, nonnegative, and Riemann–integrable function on $\mathbb{R}^d$. Let*
$$\mathscr{R} \;:= \{A \subseteq \mathbb{R}^d : \mathbf{1}_A \text{ is Riemann integrable }\}\,.$$
If $\displaystyle\int_{\mathbb{R}^d} f(\vec{x})\,d\vec{x} \;=\; 1$, then the set function $\mathbb{P}(A) \;:=\; \displaystyle\int_A f(\vec{x})\,d\vec{x}$ satisfies Definition 1.2 on p.5 of a Probability measure on $\mathscr{R}$, in the following sense:

- $\mathbb{P}(\emptyset) = 0$ • $\mathbb{P}(\mathbb{R}^d) = 1$ • $0 \le \mathbb{P}(A) \le 1$, *for all $A \in \mathscr{R}$.*
- *σ–additivity: If $A_n \in \mathscr{R}$ are disjoint and* $A := \biguplus_{n\in\mathbb{N}} A_n \in \mathscr{R}$, *then* $\mathbb{P}(A) \;=\; \sum_{n\in\mathbb{N}} \mathbb{P}(A_n)$.

3.4 Series and Integrals as Tools to Compute Probabilities

3.4.1 Series and Sums

Theorem 3.6. *Let Ω be an arbitrary, nonempty, countable set. Let $p : \Omega \longrightarrow \mathbb{R}$ be a function on Ω which satisfies*

(3.14)
$$\bullet\;\; p(\omega) \ge 0 \;\text{ for all }\; \omega \in \Omega, \qquad \bullet\;\; \sum_{\omega\in\Omega} p(\omega) \;=\; 1\,.$$

Then, $\omega \mapsto p(\omega)$ defines a probability measure $\mathbb{P}$ on Ω as follows.

(3.15)
$$\mathbb{P}(\emptyset) \;:=\; 0\,; \qquad \mathbb{P}(A) \;:=\; \sum_{\omega\in A} p(\omega)$$

3.4.2 Integrals

- Throughout this section, "$\mathbb{P}$ is a probability measure on $\mathbb{R}^d$" does not imply that $A \mapsto \mathbb{P}(A)$ is defined for all $A \subseteq \mathbb{R}^d$. Rather, it suffices that $\mathbb{P}(A)$ is defined for Riemann integrable A.

4 Calculus Extensions

4.1 Extension of Lebesgue Measure to the Borel sets of $\mathbb{R}^d$

Definition 4.1. Let $A \subseteq \mathbb{R}^d$. If it exists, we call the Riemann integral of the constant function 1 over the region of integration A,

$$\lambda^d(A) := \iint\limits_A \cdots \int d\vec{x} = \iint\limits_R \cdots \int \mathbf{1}_A(\vec{x})\, d\vec{x}\,, \quad (R \text{ is a rectangle that contains } A), \tag{4.1}$$

the d **dimensional Lebesgue measure** of A. $\square$

Theorem 4.1. $\boxed{\star}$ *There exists a set of subsets of $\mathbb{R}^d$, we denote it $\mathfrak{B}^d$, and a function*

$$\lambda^d : \mathfrak{B}^d \longrightarrow \mathbb{R} \cup \{\infty\}\,; \qquad A \mapsto \lambda^d(A)\,, \tag{4.2}$$

in the abstract sense of Definition 2.18 (Function) on p.13, such that

(A) *$\mathfrak{B}^d$ satisfies the following:*

$$\text{If } \iint \cdots \int_A d\vec{x} \text{ exists, then } A \in \mathfrak{B}^d, \quad \text{and} \quad \lambda^d(A) = \iint \cdots \int_A d\vec{x}\,, \tag{4.3}$$

$$\emptyset \in \mathfrak{B}^d, \quad \text{and} \quad \mathbb{R}^d \in \mathfrak{B}^d\,, \tag{4.4}$$

$$A \in \mathfrak{B}^d \quad \Rightarrow \quad A^{\complement} \in \mathfrak{B}^d\,, \tag{4.5}$$

$$A_n \in \mathfrak{B}^d \text{ for all } n \in \mathbb{N} \quad \Rightarrow \quad \bigcup_{n\in\mathbb{N}} A_n \in \mathfrak{B}^d, \quad \text{and} \quad \bigcap_{n\in\mathbb{N}} A_n \in \mathfrak{B}^d\,. \tag{4.6}$$

(B) *λ^d satisfies the following:*

$$A \in \mathfrak{B}^d \quad \Rightarrow \quad \lambda^d(A) \geq 0, \qquad \textit{(positivity)} \tag{4.7}$$

$$\lambda^d(\emptyset) = 0\,, \tag{4.8}$$

$$A, B \in \mathfrak{B}^d \text{ and } A \subseteq B \quad \Rightarrow \quad \lambda^d(A) \leq \lambda^d(B)\,, \qquad \textit{(monotony)} \tag{4.9}$$

$$(A_n)_{n\in\mathbb{N}} \in \mathfrak{B}^d \text{ disjoint} \quad \Rightarrow \quad \lambda^d\Big(\biguplus_{n\in\mathbb{N}} A_n\Big) = \sum_{n\in\mathbb{N}} \lambda^d(A_n)\,. \qquad \textit{(σ–additivity)} \tag{4.10}$$

Definition 4.2 (Borel sets). $\boxed{\star}$ We call the elements of $\mathfrak{B}^d$ the **Borel sets** of $\mathbb{R}^d$. We also simply say that they **are Borel**. We call $B \in \mathfrak{B}^d$ **Lebesgue Null**, also, λ^d **Null**, if $\lambda^d(B) = 0$. $\square$

Theorem 4.2. $\boxed{\star}$ *All countable subsets of $\mathbb{R}^d$ are Lebesgue Null. In particular, they are Borel sets.*

Corollary 4.1. $\boxed{\star}$

(a) *All finite subsets of $\mathbb{R}^d$ are Borel. In particular, all singleton sets $\{\vec{x}\}$ ($\vec{x} \in \mathbb{R}^d$), are Borel.*
(b) *adding and/or removing countably many points to/from a Borel set results in a Borel set.*

4.2 The Lebesgue Integral

Definition 4.3 (Simple Function on $\mathbb{R}^d$)**.** Let $d, n \in \mathbb{N}$. Let $A_1, \dots, A_n$ be Borel sets of $\mathbb{R}^d$. (Thus, $\lambda^d(A_j)$ is defined for all A_j.) Further, let $c_1, c_2, \dots, c_n$ be a corresponding set of non–negative real numbers. Let

$$f : \mathbb{R}^d \longrightarrow \mathbb{R}\,; \qquad \vec{x} \mapsto f(\vec{x}) := \sum_{j=1}^{n} c_j \mathbf{1}_{A_j}(\vec{x}) \tag{4.11}$$

Then we call f a **simple function**. □

Proposition 4.1. $\boxed{\star}$

(a) *All step functions with $c_j \geq 0$ are simple functions.*
(b) *Not all simple functions are step functions.*
(c) *Not all simple functions possess a Riemann integral.*

Definition 4.4. Let $f(\vec{x}) = \sum\limits_{j=1}^{n} c_j \mathbf{1}_{A_j}(\vec{x})$ be a simple function such that $c_j \geq 0$ for all j. Then we call

$$\int f d\,\lambda^d := \int f(\vec{x}) d\,\lambda^d(\vec{x}) := \int f(\vec{x}) \lambda^d(d\vec{x}) := \sum_{j=1}^{n} c_j \lambda^d(A_j)\,. \tag{4.12}$$

the **Lebesgue integral** of the simple function f. □

Definition 4.5 (Lebesgue integral)**.** $\boxed{\star}$

(a) Either let $f : \mathbb{R}_d \to [0, \infty[$ be a nonnegative function on $\mathbb{R}_d$, such that

- there is a nondecreasing sequence of simple functions, $f_n \geq 0$, satisfying $f_n \uparrow f$;

Or let $f : \mathbb{R}_d \to]-\infty, 0]$ be a nonpositive function on $\mathbb{R}_d$, such that

- there is a nonincreasing sequence of simple functions, $f_n \leq 0$, satisfying $f_n \downarrow f$.

We define the **Lebesgue integral** of that nonnegative or nonpositive function f as

$$\int f\, d\lambda^d := \lim_{n\to\infty} \int f_n\, d\lambda^d\,. \tag{4.13}$$

(b) Let $f: \mathbb{R}_d \to \mathbb{R}$ be a function on $\mathbb{R}_d$ such that

- both f^+ and f^- are limits of nondecreasing sequences of simple functions ≥ 0;
- at least one of $\int f^+ d\lambda^d$, $\int f^- d\lambda^d$ is finite. (According to **(a)**, those integrals exist, but neither of them was guaranteed to be finite.)

Then we define the **Lebesgue integral** of the function f as the expression

$$\int f\, d\lambda^d = \int (f^+ - f^-)\, d\lambda^d := \int f^+\, d\lambda^d - \int f^-\, d\lambda^d\,. \tag{4.14}$$

(c) We call a real–valued function f **Lebesgue integrable**, if $\int f\, d\lambda^d$ exists and is finite. $\square$

Definition 4.6. $\boxed{\star}$

- We call simple functions, and real–valued functions that are limits of sequences of simple functions, **Borel measurable functions** (or simply, **Borel functions**). $\square$

Theorem 4.3. $\boxed{\star}$ *Assume that $f_1, f_2, \dots$ are Borel functions, $c_1, c_2, \dots \in \mathbb{R}$, $B \in \mathfrak{B}^d$. Each of the following also is a Borel function:*

• c_1 *(constant function)* • $c_1 f_1$ • $f_1 \pm f_2$ • $f_1 f_2$ • $\mathbf{1}_B f_1$ • f_1/f_2 *(if $f_2 \neq 0$)* • $\sum_{j=1}^{n} c_j f_j$

• $\min(f_1, f_2)$ • $\max(f_1, f_2)$ • $\min_{j=1,\dots,n} f_j$ • $\max_{j=1,\dots,n} f_j$ • $\inf_{j\in\mathbb{N}} f_j$ • $\sup_{j\in\mathbb{N}} f_j$ $\square$

If they exist (see the subsequent remark), the following also are Borel functions:

• $\lim_{j\to\infty} f_j$ • $\sum_{j=1}^{\infty} f_j$ • $\min_{j\in\mathbb{N}} f_j$ • $\max_{j\in\mathbb{N}} f_j$

Theorem 4.4. $\boxed{\star}$ *Lebesgue integrals satisfy the following. Let $B \in \mathfrak{B}^d$ and assume that f is a Borel function. Then*

(a) *If $\int f\, d\lambda^d$ exists, then $\int \mathbf{1}_B f\, d\lambda^d$ exists.*

(b) *If f is Lebesgue integrable, then $\mathbf{1}_B f$ is Lebesgue integrable.*

Definition 4.7. Let $B \in \mathfrak{B}^d$ and assume that f is a Borel function on $\mathbb{R}^d$ for which the Lebesgue integral $\int f d\lambda^d$ exists. The **Lebesgue integral of** f **on** B or **over** B is defined by the expression

$$\int_B f\,d\lambda^d \;:=\; \int_B f(\vec{x})d\,\lambda^d(\vec{x}) \;:=\; \int_B f(\vec{x})\lambda^d(d\vec{x}) \;:=\; \int \mathbf{1}_B f\,d\lambda^d\,. \tag{4.15}$$

We say that **Lebesgue integrable on** B, if $\int_B f\,d\lambda^d$ exists and <u>is finite</u>. □

Fact 4.1. *Let $D \subseteq \mathbb{R}^d$ and $f : D \to \mathbb{R}$, such that f and D are of any relevance for this course.*

- *If the Riemann integral $\displaystyle\int_D f(\vec{x})\,d\vec{x}$ exists, then the Lebesgue integral $\displaystyle\int_D f\,d\lambda^d$ exists.*
- *Further,* $\displaystyle\int_D f(\vec{x})\,d\vec{x} \;=\; \int_D f\,d\lambda^d.$
- *Accordingly, all the techniques one has learned in calculus to evaluate the Riemann integral can be used to compute the Lebesgue integral.* □

Proposition 4.2 (Integrability criterion). $\boxed{\star}$ *Let f be a Borel function and B a Borel set. Then f is integrable on B* $\;\Leftrightarrow\; \displaystyle\int_B |f|\,d\lambda^d < \infty \;\Leftrightarrow\;$ *both* $\displaystyle\int_B f^+\,d\lambda^d < \infty$ *and* $\displaystyle\int_B f^-\,d\lambda^d < \infty.$

Theorem 4.5. *Assume that $f, g, f_1, f_2, \dots$ are Borel functions, $c, c_1, c_2, \dots \in \mathbb{R}$, and B is a Borel set. Then Lebesgue integrals on B satisfy the following.*

***(a) Positivity**:* $\displaystyle\int_B 0\,d\lambda^d = 0; \qquad f \geq 0 \text{ on } B \;\Rightarrow\; \int_B f\,d\lambda^d \;\geq\; 0,$

***(b) Monotonicity**:* $\displaystyle\lambda^d\{\vec{x} \in B : f(\vec{x}) > g(\vec{x})\} = 0 \;\Rightarrow\; \int_B f\,d\lambda^d \;\leq\; \int_B g\,d\lambda^d.$

In particular, $\displaystyle f \leq g \text{ on } B \;\Rightarrow\; \int_B f\,d\lambda^d \;\leq\; \int_B g\,d\lambda^d,$

and also, $\displaystyle\lambda^d\{\vec{x} \in B : f(\vec{x}) \neq g(\vec{x})\} = 0 \;\Rightarrow\; \int_B f\,d\lambda^d \;=\; \int_B g\,d\lambda^d.$

***(c) Linearity I**: f, g integrable on B* $\displaystyle\;\Rightarrow\; \int_B (f \pm g)\,d\lambda^d \;=\; \int_B f\,d\lambda^d \pm \int_B g\,d\lambda^d$

and also, $\displaystyle\int_B (cf)\,d\lambda^d \;=\; c\int_B f\,d\lambda^d.$

***Linearity II**: $f_1 \dots, f_n$ integrable* $\displaystyle\;\Rightarrow\; \int_B \Big(\sum_{j=1}^n f_j\Big)\,d\lambda^d \;=\; \sum_{j=1}^n c_j \int_B f_j\,d\lambda^d.$

***(d) Monotone Convergence**: Assume that* $0 \leq f_1 \leq f_2 \leq \cdots,\; 0 \geq g_1 \geq g_2 \geq \cdots.$

Then $\displaystyle\int_B f_n\,d\lambda^d \uparrow \int_B \Big(\sup_{n\in\mathbb{N}} f_n\Big)\,d\lambda^d$ *and* $\displaystyle\int_B g_n\,d\lambda^d \downarrow \int_B \Big(\inf_{n\in\mathbb{N}} g_n\Big)\,d\lambda^d$ *as* $n \to \infty.$

(e) Dominated Convergence: *Assume that*

- $\lim\limits_{n\to\infty} f_n$ *exists,* • $|f_n| \leq g$ *for all* $n \in \mathbb{N}$, • $\displaystyle\int_B g \, d\lambda^d < \infty$.

$$\text{Then } \lim_{n\to\infty} \int_B f_n \, d\lambda^d = \int_B \Big(\lim_{n\to\infty} f_n \Big) \, d\lambda^d \text{ as } n \to \infty.$$

Theorem 4.6 (Fubini's theorem for Lebesgue integrals). $\boxed{\star}$ *Assume that* $f_1, f_2, \dots$ *are Borel functions, and* B_1, B_2 *are Borel sets. Then, for any rearrangement* $j_1, j_2, \dots, j_d$ *of* $1, 2, \dots, d$,

$$\begin{aligned} \int_{B_1\times B_2\times\dots\times B_d} f \, d\lambda^d &= \int_{B_1} \Big(\int_{B_2} \Big(\cdots \int_{B_d} f \, d\lambda^1 \cdots \Big) d\lambda^1 \Big) d\lambda^1 \\ &= \int_{B_{j_1}} \Big(\int_{B_{j_2}} \Big(\cdots \int_{B_{j_d}} f \, d\lambda^1 \cdots \Big) d\lambda^1 \Big) d\lambda^1 \end{aligned} \tag{4.16}$$

This formula is technically correct, but let us supply all arguments and write, [6] *e.g.,* $\lambda^1(dx_j)$ *for* $d\lambda^1$:

$$\begin{aligned} \int_{B_1\times B_2\times\dots\times B_d} f(\vec{x}) \, \lambda^d(d\vec{x}) &= \int_{B_1} \Big(\int_{B_2} \Big(\cdots \int_{B_d} f(\vec{x}) \, \lambda^1(dx_d) \cdots \Big) \lambda^1(dx_2) \Big) \lambda^1(dx_1) \\ &= \int_{B_{j_1}} \Big(\int_{B_{j_2}} \Big(\cdots \int_{B_{j_1}} f(\vec{x}) \, \lambda^1(dx_{j_d}) \cdots \Big) \lambda^1(dx_{j_2}) \Big) \lambda^1(dx_{j_1}) \,. \end{aligned} \tag{4.17}$$

In particular, assume that each B_j *is an interval* $[\alpha_j, \beta_j]$ *or* $[\alpha_j, \beta_j]$ *or* $[\alpha_j, \beta_j]$ *or* $[\alpha_j, \beta_j]$, *where* $\alpha_j \leq \beta_j$.

If we adjust the notation to that of Riemann integrals and replace $\int_{B_j}$ *with* $\int\limits_{\alpha_j}^{\beta_j}$, $\lambda^d(d\vec{x})$ *with* $d\vec{x}$, *and* $\lambda^1(dx_j)$ *with* dx_j, *then* (4.17) *matches Fubini's formula* (**??**)***(g)*** *for Riemann integrals (see p.**??**).*

Here is another version of Fubini's theorem. It features "only" two vector– valued components.

Assume that $d, d_1, d_2 \in \mathbb{N}$, *that* $d_1 + d_2 = d$, *that* $f : \mathbb{R}^d \to \mathbb{R}$ *is a nonnegative and/or* λ^d*–integrable Borel function, and that* $B_1 \in \mathfrak{B}^{d_1}$ *and* $B_2 \in \mathfrak{B}^{d_2}$. *For* $\vec{x} = (x_1, x_2 \dots, x_{d_1})$ *and* $\vec{y} = (y_1, y_2 \dots, y_{d_2})$, *let* $(\vec{x}, \vec{y}) := (x_1, \dots, x_{d_1}, y_1, \dots, y_{d_2})$. *Then*

$$\begin{aligned} \int_{B_1\times B_2} f(\vec{x}, \vec{y}) \, \lambda^d\big(d(\vec{x}, \vec{y})\big) &= \int_{B_1} \Big(\int_{B_2} f(\vec{x}, \vec{y}) \, \lambda^{d_2}(d\vec{y}) \Big) \lambda^{d_1}(d\vec{x}) \\ &= \int_{B_2} \Big(\int_{B_1} f(\vec{x}, \vec{y}) \, \lambda^{d_1}(d\vec{x}) \Big) \lambda^{d_2}(d\vec{y}) \,. \end{aligned} \tag{4.18}$$

[6] Recall that (4.4) on p.30 and (4.7) on p.32 give us a choice of notation

$$\int_B f \, d\lambda^d = \int_B f(\vec{x}) d\,\lambda^d(\vec{x}) = \int_B f(\vec{x}) \lambda^d(d\vec{x})$$

Even though there only are two integrations $\lambda^{d_1}(d\vec{x})$ *and* $\lambda^{d_2}(d\vec{y})$, (4.18) *is more general than* (4.17), *because the Borel sets* B_1, B_2, *and* $B_1 \times B_2$ *are no more cartesian products of onedimensional Borel sets.*

Theorem 4.7. *Let* $f : \mathbb{R}^d \to \mathbb{R}$ *be a real–valued, Borel–measurable function on* $\mathbb{R}^d$. *If* f *is nonnegative or Lebesgue integrable (i.e.,* $\int |f| d\lambda^d < \infty$*), then the set function*

(4.19) $$\Psi : \mathfrak{B}^d \longrightarrow [0, \infty], \qquad \Psi(A) := \int_A f \, d\lambda^d$$

is σ*–additive.*

Corollary 4.2. *Let* $f : \mathbb{R}^d \to \mathbb{R}$ *be a real–valued, nonnegative, and Borel–measurable function on* $\mathbb{R}^d$. *If* $\displaystyle\int f d\lambda^d = 1$, *then the set function*

(4.20) $$\mathbb{P} : \mathfrak{B}^d \longrightarrow [0, \infty], \qquad \mathbb{P}(A) := \int_A f \, d\lambda^d$$

defines a probability measure on $\mathbb{R}^d$.

Definition 4.8 (Support of a real–valued function). $\boxed{\star}$

Let Ω be some nonempty set and $f : \Omega \to \mathbb{R}$. We call

(4.21) $$\text{suppt}(f) := \{ \omega \in \Omega : f(\omega) \neq 0 \}$$

the **support** of the function f. $\square$

5 The Probability Model

5.1 Probability Spaces

Definition 5.1 (σ–algebra)**.** Let Ω be a nonempty set and $\mathfrak{F} \subseteq 2^\Omega$ such that

(a) $A \in \mathfrak{F} \;\Rightarrow\; A^{\complement} \in \mathfrak{F}\,.$

(b) $A_n \in \mathfrak{F}$ arbitrary $\;\Rightarrow\; \bigcup\limits_{j=1}^{\infty} A_j \in \mathfrak{F}\,.$

(c) $\emptyset \in \mathfrak{F}\,.$

Then we call $\mathfrak{F}$ a σ**–algebra** for Ω. (Also, a σ–algebra on Ω or asociated with Ω.)

$\mathfrak{F}$ is also called a σ**–field** for Ω, but that is considered old–fashioned terminology. $\square$

Proposition 5.1. *σ–algebras $\mathfrak{F}$ satisfy the following.*

(a) $\Omega \in \mathfrak{F}$.

(b) *Let $n \in \mathbb{N}$ and $A_1, \dots, A_n \in \mathfrak{F}$. Then $A_1 \cup A_2 \cup \dots \cup A_n \in \mathfrak{F}$. (finite union.)*

(c) *Let $n \in \mathbb{N}$ and $A_1, A_2, \dots \in \mathfrak{F}$. Let $A = \bigcap\limits_{k=1}^{n} A_k$ and $B = \bigcap\limits_{k=1}^{\infty} A_k$.*

Then $A \in \mathfrak{F}$ and $B \in \mathfrak{F}$. $\square$

Proposition 5.2. $\boxed{\star}$ *Assume that $(A_j)_{j \in J}$ is a countable partition of a nonempty set Ω. In other words, the sets A_j are mutually disjoint subsets of Ω, $\biguplus\left[A_j : j \in J\right] = \Omega$, and the index set J is countable. Then*

$$\mathfrak{F} := \{\,\textit{all unions involving some or all of the } A_j\,\} \tag{5.1}$$

is a σ–algebra for Ω.

Definition 5.2 (Probability measures and probability spaces)**.** Given are a nonempty set Ω with a σ–algebra $\mathfrak{F} \subseteq 2^\Omega$ and a function

$$\mathbb{P} : \mathfrak{F} \longrightarrow [0,1]\,; \quad A \mapsto \mathbb{P}(A) \qquad \text{as follows.}$$

$$\mathbb{P}(\emptyset) \;=\; 0\,, \tag{5.2}$$

$$\mathbb{P}(\Omega) \;=\; 1\,, \tag{5.3}$$

$$(A_n)_{n\in\mathbb{N}} \in \mathfrak{F} \text{ disjoint} \;\Rightarrow\; P\Big(\biguplus_{n\in\mathbb{N}} A_n\Big) \;=\; \sum_{n=1}^{\infty} \mathbb{P}(A_n) \;=\; \sum_{n\in\mathbb{N}} \mathbb{P}(A_n). \quad (\sigma\text{–additivity}) \tag{5.4}$$

- We call $\mathbb{P}$ a **probability measure** or simply a **probability**
- The triplet $(\Omega, \mathfrak{F}, \mathbb{P})$ is called a **probability space**.
- (Only) the elements of $\mathfrak{F}$ are called **events**.
- We often call disjoint events **mutually exclusive** events.
- An event A is a $\mathbb{P}$ **Null event**, also, **Null event**, if $\mathbb{P}(A) = 0$.

We suggest to reserve the term "probability" for the function value $\mathbb{P}(A)$ that belongs to a specific event A, and always refer to $\mathbb{P}$, i.e., the function $A \mapsto \mathbb{P}(A)$, as a "probability measure". $\square$

Notation 5.1 (Sample spaces and sample points)**.**

- We also call a probability space a **sample space** and an outcome a **sample point**.
- We also call Ω by itself (as opposed to the triplet $(\Omega, \mathfrak{F}, \mathbb{P})$) a probability space or sample space. Sometimes we refer to Ω as the **carrier set** or **carrier** of $(\Omega, \mathfrak{F}, \mathbb{P})$.
- We like to write Ω for the carrier set, $\mathfrak{F}$ for the σ–algebra and $\mathbb{P}$ for the probability measure of a probability space, but different notation may be used. For example, there may be a probability space $(S, \mathscr{S}, Q)$ and outcomes s or x or $\vec{y}$ (vector notation).

Definition 5.3 (Equiprobability)**.** Let $(\Omega, \mathbb{P})$ be a finite probability space, i.e., $|\Omega| < \infty$. Let $n := |\Omega|$. We say that $\mathbb{P}$ has **equiprobable** outcomes or that $\mathbb{P}$ **satisfies equiprobability**, if

$$\mathbb{P}\big(\{\omega\}\big) \;=\; \frac{1}{|\Omega|} \quad \text{(since then } \mathbb{P}\{\omega\} \text{ is constant for all } \omega \in \Omega\text{)}. \tag{5.5}$$

Synonyms for equiprobability are (discrete) [7]**uniform probability**, **Laplace probability**, $\square$

Theorem 5.1 (Continuity property of probability measures)**.** *Let $(\Omega, \mathfrak{F}, \mathbb{P})$ be a probability space. If $A_n, B_n \in \mathfrak{F}$, then the following is true:*

$$A_n \uparrow \Rightarrow \; \mathbb{P}(A_n) \uparrow \mathbb{P}\left(\bigcup_{n\in\mathbb{N}} A_n\right), \tag{5.6}$$

$$B_n \downarrow \Rightarrow \; \mathbb{P}(B_n) \downarrow \mathbb{P}\left(\bigcap_{n\in\mathbb{N}} B_n\right). \tag{5.7}$$

Definition 5.4 (Discrete probability space)**.** Assume that the probability space $(\Omega, \mathfrak{F}, \mathbb{P})$ satisfies the following:

(a) $\mathbb{P}(\{\omega\})$ is defined for all $\omega \in \Omega$. In other words, we ask that $\{\omega\} \in \mathfrak{F}$ for all $\omega \in \Omega$.
(b) There exists a countable subset A^* of Ω such that $\sum_{\omega\in A^*} \mathbb{P}\{\omega\} \;=\; 1$

Then we call $(\Omega, \mathfrak{F}, \mathbb{P})$ a **discrete probability space**. $\square$

[7] there also is the concept of uniform probability in connection with continuous random variables. See Definition 10.8 (Continuous uniform random variable) on p.78.

Theorem 5.2. *Let $(\Omega', \mathfrak{F}', \mathbb{P}')$ be a discrete probability space and $A^* \in \mathfrak{F}'$ a countable event such that $\sum\limits_{\omega' \in A^*} \mathbb{P}'\{\omega'\} = 1$. Then*

- ***(a)*** $A^* \in \mathfrak{F}'$.
- ***(b)*** $\mathbb{P}'(A^*) = 1$ *and thus,* $\mathbb{P}'\left((A^*)^{\complement}\right) = 0$.
- ***(c)*** $\mathbb{P}'(A) = \mathbb{P}'(A \cap A^*)$ *for all* $A \in \mathfrak{F}'$.
- ***(d)*** $\mathbb{P}'(A) = \sum\limits_{\omega' \in A \cap A^*} \mathbb{P}'\{\omega'\}$ *for all* $A \in \mathfrak{F}'$.
- ***(e)*** $\boxed{\star}$ *The formula* $\mathbb{P}(B) := \mathbb{P}'(B \cap A^*)$ *"extends"* $\mathbb{P}'$ *to a probability measure* $\mathbb{P}$ *on the entire power set* $2^{\Omega'}$.

Corollary 5.1.

- ***(a)*** *If* $(\Omega', \mathfrak{F}', \mathbb{P}')$ *be a discrete probability space, then* $\mathbb{P}'$ *is characterized by the probabilities* $\mathbb{P}'\{\omega'\}$ *of the outcomes* ω'.
- ***(b)*** *Let* Ω' *be some arbitrary, nonempty set. Assume that* $\left(p_j\right)_j$ *is a finite or infinite sequence of real numbers that satisfies*
  - $p_j \geq 0$ *for all* j *and* $\sum_j p_j = 1$

  Further, assume that $\left(\omega'_j\right)_j$ *is a corresponding sequence of distinct elements of* Ω'*, then* $\left(p_j\right)_j$ *defines a discrete probability space* $(\Omega', 2^{\Omega'}, \mathbb{P}')$ *as follows.*
  - $\mathbb{P}'(\emptyset) := 0, \quad \mathbb{P}'(A) := \sum\limits_{j\,:\,\omega'_j \in A} p_j$, *for* $A \neq \emptyset$. $\square$

Theorem 5.3. $\boxed{\star}$ *Let* Ω *be some arbitrary set and* $\left(\mathfrak{F}_i\right)_{i \in I}$ *a family of* σ*–algebras on* Ω*, i.e.,* $\mathfrak{F}_i \subseteq 2^{\Omega}$ *for each* $i \in I$*. No assumption is made about the index set other than* $I \neq \emptyset$*. Thus, this family may consist of finitely many* σ*–algebras or of entire sequence or even uncountably many* σ*–algebras.*

- *Let* $\mathfrak{F} := \bigcap_{i \in I} \mathfrak{F}_i$*, i.e.,* $\mathfrak{F} = \{A \subseteq \Omega : A \in \mathfrak{F}_i$ *for each index* $i\}$*. Then* $\mathfrak{F}$ *is a* σ*–algebra.*

This can also be stated as follows. *Any intersection of* σ*–algebras results in a* σ*–algebra.*

Theorem 5.4. $\boxed{\star}$ *Let* Ω *be an arbitrary set and* $\mathscr{A} \subseteq 2^{\Omega}$*. (So elements of* $\mathscr{A}$ *are subsets of* Ω*.)*

- *There exists a minimal (i.e., smallest)* σ*–algebra that contains* $\mathscr{A}$*.*
- *Further, this* σ*–algebra is uniquely determined by* $\mathscr{A}$*. This allows us to name it* $\sigma\{\mathscr{A}\}$*.*

Definition 5.5 (σ–algebra generated by a collection of sets). $\boxed{\star}$ Let Ω be a nonempty set.

(a) Let $\mathscr{A} \subseteq 2^\Omega$, i.e., the elements of $\mathscr{A}$ are subsets of Ω.
We call $\sigma\{\mathscr{A}\}$ the **σ–algebra generated by $\mathscr{A}$**. If $\mathscr{A}$ is of the form $\mathscr{A} = \{...\}$, we also write $\sigma\{...\}$ for $\sigma\{\{...\}\}$.

(b) Assume in addition that $\mathfrak{F}$ is a σ–algebra for Ω and $\mathscr{A} \subseteq \mathfrak{F}$. If $\sigma\{\mathscr{A}\} = \mathfrak{F}$, we call $\mathscr{A}$ a **generator for** $\mathfrak{F}$ a.k.a. generator of $\mathfrak{F}$, and we say that $\mathscr{A}$ **generates** $\mathfrak{F}$.

Concerning notation:

- One also can write $\sigma(\mathscr{A})$ or $\sigma[\mathscr{A}]$ for $\sigma\{\mathscr{A}\}$.
- Given a family of subsets $A_i \subseteq \Omega$, $(i \in I)$, $\sigma\{A_i : i \in I\}$ can also be written as $\sigma\{A_i : i \in I\} = \sigma\big((A_i)_{i \in I}\big) = \sigma\big[(A_i)_{i \in I}\big] = \sigma\big\{(A_i)_{i \in I}\big\}$.
  As usual, it is OK to omit the "$i \in I$" part if the meaning of I is unambiguous. $\square$

Definition 5.6 (Borel σ–algebra). $\boxed{\star}$

For $d = 1, 2, \dots,$ we define

- $\mathfrak{B}^d := \sigma\{$ d–dimensional rectangles $\}$,
- $\mathfrak{B} := \mathfrak{B}^1 = \sigma\{$ all intervals of real numbers $\}$.

$\mathfrak{B}$ and $\mathfrak{B}^d$ are the **Borel σ–algebras** and their members are the **Borel sets** of $\mathbb{R}$ and $\mathbb{R}^d$. $\square$

Fact 5.1. $\boxed{\star}$ *For the following, note that the sets $\mathfrak{I}_1, \dots, \mathfrak{I}_8$ were defined in Example* **??** *on p.***??**.

(a) *Let $\mathfrak{I}$ denote one of the collections of half-open intervals, $\mathfrak{I}_1, \mathfrak{I}_4$. Let $\mathscr{E} := \mathfrak{I} \uplus \mathbb{R}$. Then any function $\mathbb{P}_0 : \mathscr{E} \to [0,1]$ which satisfies $\mathbb{P}_0(\emptyset) = 0$, $\mathbb{P}_0(\mathbb{R}) = 1$ and σ–additivity on $\mathscr{E}$:*
$E_n \in \mathscr{E}$ *disjoint such that* $\mathbb{E} := \biguplus_{n \in \mathbb{N}} \in \mathscr{E} \Rightarrow \mathbb{P}_0(E) = \sum_{n \in \mathbb{N}} \mathbb{P}_0(E_n)$,
can be uniquely extended to a probability measure on $\mathfrak{B}$, the Borel sets of $\mathbb{R}$.

(b) *Let $\mathfrak{I}$ denote one of the collections of d–dimensional rectangles $\mathfrak{I}_5, \mathfrak{I}_8$. Let $\mathscr{E} := \mathfrak{I} \cup \{\mathbb{R}^d\}$. Then any function $\mathbb{P}_0 : \mathscr{E} \to [0,1]$ which satisfies $\mathbb{P}_0(\emptyset) = 0$, $\mathbb{P}_0(\mathbb{R}^d) = 1$ and σ–additivity on $\mathscr{E}$: $E_n \in \mathscr{E}$ disjoint such that $\mathbb{E} := \biguplus_{n \in \mathbb{N}} E_n \in \mathscr{E} \Rightarrow \mathbb{P}_0(E) = \sum_{n \in \mathbb{N}} \mathbb{P}_0(E_n)$,*
can be uniquely extended to a probability measure on $\mathfrak{B}^d$, the Borel sets of $\mathbb{R}^d$. $\square$

Notational conveniences for probabilities:

If we have a set that is written as $\{\dots\}$, i.e., with curly braces as delimiters, then we may write its probability as $\mathbb{P}\{\dots\}$ instead of $\mathbb{P}\big(\{\dots\}\big)$. Specifically for singletons $\{\omega\}$, it is OK to write $\mathbb{P}\{\omega\}$.

Theorem 5.5 (WMS Ch.02.8, Theorem 2.6). *If A, B are events in a probability space $(\Omega, \mathbb{P})$, then*

(5.8) ***Additive Law of Probability:*** $\mathbb{P}(A \cup B) \;=\; \mathbb{P}(A) \;+\; \mathbb{P}(B) \;-\mathbb{P}(A \cap B)\,.$

(5.9) ***Rule of the Complement:*** $\mathbb{P}\big[A^{\complement}\big] \;=\; 1 \;-\; \mathbb{P}[A]\,.$

Theorem 5.6 (Exclusion–Inclusion formula). $\boxed{\star}$ *If $A_1, A_2, \cdots, A_n$ are events in a probability space $(\Omega, \mathbb{P})$, then*

$$
\begin{aligned}
\mathbb{P}(A_1 \cup A_2 \cdots \cup A_n) \;&=\; \sum_i \mathbb{P}(A_i) \;-\; \sum_{i<j} \mathbb{P}(A_i \cap A_j) \\
&+\; \sum_{i<j<k} \mathbb{P}(A_i \cap A_j \cap A_k) \;-\; \cdots \;+\; (-1)^{n+1} \cdot \mathbb{P}(A_1 \cap A_2 \cdots \cap A_n)\,.
\end{aligned} \tag{5.10}
$$

Corollary 5.2 (Exclusion–Inclusion formula for 3 events). $\boxed{\star}$ *If A_1, A_2, A_3 are events in a probability space $(\Omega, \mathbb{P})$, then*

$$
\begin{aligned}
\mathbb{P}(A_1 \cup A_2 \cup A_3) \;&=\; \big[\,\mathbb{P}(A_1) + \mathbb{P}(A_2) + \mathbb{P}(A_3)\,\big] \\
&- \big[\mathbb{P}(A_1 \cap A_2) + \mathbb{P}(A_1 \cap A_3) + \mathbb{P}(A_2 \cap A_3)\,\big] \;+\; \mathbb{P}(A_1 \cap A_2 \cap A_n)\,.
\end{aligned} \tag{5.11}
$$

5.2 Conditional Probability and Independent Events

Definition 5.7 (Conditional probability). Given are a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and two events $A, B \in \mathscr{F}$. We call

$$
\mathbb{P}(A \mid B) \;:=\; \begin{cases} \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}\,, & \text{if } \mathbb{P}(B) > 0\,, \\ \textbf{undefined}\,, & \text{if } \mathbb{P}(B) = 0\,, \end{cases} \tag{5.12}
$$

(read: "probability of A given B" or "probability of A conditioned on B") the **conditional probability** of the event A, given that the event B has occurred. □

Theorem 5.7. *Given are a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and an event $B \in \mathscr{F}$ such that $\mathbb{P}(B) > 0$. Then*

$$
\mathbb{P}(\cdot \mid B) : \mathfrak{F} \longrightarrow [0,1]\,; \qquad A \mapsto \mathbb{P}(A \mid B) \tag{5.13}
$$

is another probability measure on $(\Omega, \mathscr{F})$.

In other words, $\mathbb{P}(\cdot \mid B)$ satisfies (5.2) – (5.4) of Definition 5.2 (Probability measures and probability spaces) on p.35.

Proposition 5.3. *If $(\Omega, \mathfrak{F}, \mathbb{P})$ is a probability space and $A, B, C \in \mathfrak{F}$, then*

$$\mathbb{P}(A \cap B \cap C) \;=\; \mathbb{P}(A \mid B \cap C) \cdot \mathbb{P}(B \mid C) \cdot \mathbb{P}(C)\,. \tag{5.14}$$

Proposition 5.4 (Multiplicative Law of Probability for n events)**.** *If $(\Omega, \mathfrak{F}, \mathbb{P})$ is a probability space, $n \in \mathbb{N}$ and $A_1, \dots, A_n \in \mathfrak{F}$, then*

$$\begin{aligned}\mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_n) &= \mathbb{P}(A_1 \mid A_2 \cap \dots \cap A_n) \cdot \mathbb{P}(A_2 \mid A_3 \dots \cap A_n) \cdots \\ &\quad \cdots \mathbb{P}(A_{n-2} \mid A_{n-1} \cap A_n)\, \mathbb{P}(A_{n-1} \mid A_n)\, \mathbb{P}(A_n).\end{aligned} \tag{5.15}$$

Definition 5.8 (Two independent events)**.** Given are a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and two events $A, B \in \mathscr{F}$. We say that A and B are **independent** if

$$\mathbb{P}(A \cap B) \;=\; \mathbb{P}(A) \cdot \mathbb{P}(B)\,. \;\square \tag{5.16}$$

Definition 5.9 (Three independent events)**.** Given are a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and three events $A, B, C \in \mathscr{F}$. We say that A, B and C are **independent** if

$$\begin{aligned}\mathbb{P}(A \cap B \cap C) &= \mathbb{P}(A) \cdot \mathbb{P}(B) \cdot \mathbb{P}(C)\,,\\ \mathbb{P}(A \cap B) &= \mathbb{P}(A) \cdot \mathbb{P}(B)\,,\\ \mathbb{P}(A \cap C) &= \mathbb{P}(A) \cdot \mathbb{P}(C)\,,\\ \mathbb{P}(B \cap C) &= \mathbb{P}(B) \cdot \mathbb{P}(C)\,. \;\square\end{aligned} \tag{5.17}$$

Definition 5.10 (Finitely many independent events)**.** Given are a probability space $(\Omega, \mathscr{F}, \mathbb{P})$, $n \in \mathbb{N}$ and events $A_1, A_2, \dots, A_n \in \mathscr{F}$. We say that $A_1, A_2, \dots, A_n$ are **independent** if, for ANY subselection of indices

$$1 \leq j_1 \;<\; j_2 \;<\; \cdots \;<\; j_k \;\leq\; n\,,$$

it is true that

$$\mathbb{P}(A_{j_1} \cap A_{j_1} \cap A_{j_k}) \;=\; \mathbb{P}(A_{j_1}) \cdot \mathbb{P}(A_{j_2}) \cdot \mathbb{P}(A_{j_k})\,. \;\square \tag{5.18}$$

Definition 5.11 (Sequences of independent events)**.** Given are a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and a sequence of events $A_1, A_2, \dots \in \mathscr{F}$ We say that this sequence is **independent** if, for ANY FINITE subselection of distinct indices $j_1, j_2, \dots, j_k \in \mathbb{N}$, it is true that

$$\mathbb{P}(A_{j_1} \cap A_{j_2} \cap A_{j_k}) \;=\; \mathbb{P}(A_{j_1}) \cdot \mathbb{P}(A_{j_2}) \cdot \mathbb{P}(A_{j_k})\,. \;\square \tag{5.19}$$

Definition 5.12 (Independence of arbitrarily many events). $\boxed{\star}$ Given are a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and a family $\big(A_i\big)_{i \in I}$ of events $A_i \in \mathscr{F}$. Here I denotes an arbitrary set of indices. We say that this family is **independent** if, for ANY FINITE subselection of distinct indices $i_1, i_2, \dots, i_k \in I$, it is true that

$$\mathbb{P}(A_{i_1} \cap A_{i_2} \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \cdot \mathbb{P}(A_{i_2}) \cdot \mathbb{P}(A_{i_k}). \quad \square \tag{5.20}$$

Theorem 5.8. $\boxed{\star}$ *Given are a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and a family $\big(A_i\big)_{i \in I}$ of independent events $A_i \in \mathscr{F}$. Here I denotes an arbitrary set of indices. Then we have the following:*

If some or all of the A_i are replaced by their complement $A_i^{\complement}$, then the resulting family of events also is independent.

In other words, for each $i \in I$, let B_i be either A_i or $A_i^{\complement}$. Then independence of $\big(A_i\big)_{i \in I}$ implies that of $\big(B_i\big)_{i \in I}$.

Corollary 5.3. *Given are a $(\Omega, \mathfrak{F}, \mathbb{P})$ is a probability space, $n \in \mathbb{N}$ and independent events $A_1, \dots, A_n \in \mathfrak{F}$.*
If some or all of the A_i are replaced by their complement $A_i^{\complement}$, then the resulting list of events also is independent.

In other words, for each $i = 1, 2, \dots, n$, let B_i be either A_i or $A_i^{\complement}$. Then independence of $A_1, \dots, A_n$ implies that of $B_1, \dots, B_n$.

Theorem 5.9. *Given are a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and two events $A, B \in \mathscr{F}$ such that $\mathbb{P}(B) > 0$. Then*

$$A \text{ and } B \text{ are independent} \quad \Leftrightarrow \quad \mathbb{P}(A \mid B) = \mathbb{P}(A). \tag{5.21}$$

Corollary 5.4. *If $(\Omega, \mathscr{F}, \mathbb{P})$ is a probability space and $A, B \in \mathscr{F}$ such that $\mathbb{P}(A) > 0$ and $\mathbb{P}(B) > 0$. Then*

$$A \text{ and } B \text{ are independent} \quad \Leftrightarrow \quad \mathbb{P}(A \mid B) = \mathbb{P}(A) \quad \Leftrightarrow \quad \mathbb{P}(B \mid A) = \mathbb{P}(B). \tag{5.22}$$

5.3 Random Elements and their Probability Distributions

Theorem 5.10. *Let $(\Omega, \mathbb{P})$ be a probability space, Ω' a nonempty set, and $Y : \Omega \to \Omega'$ a function. Then the formula*

(5.23) $$\mathbb{P}_Y(B) \ := \ \mathbb{P}\{Y \in B\} \ \ (B \subseteq \Omega')$$

defines a probability measure on Ω'.

Definition 5.13 (Probability Distribution)**.** Let $(\Omega, \mathbb{P})$ be a probability space, Ω' a nonempty set, and $Y : \Omega \to \Omega'$ a function. Then the probability measure $\mathbb{P}_Y$ on Ω' of Theorem 5.10, given by

(5.24) $$\mathbb{P}_Y(A') \ := \ \mathbb{P}\{Y \in A'\} \ = \ \mathbb{P}\left(Y^{-1}(A')\right) \ \ (A' \subseteq \Omega'),$$

is called the **probability distribution** or just the **distribution** of Y with respect to $\mathbb{P}$. Very often the probability space $(\Omega, \mathbb{P})$ is fixed for a long stretch. We then simply talk about the probability distribution of Y, without referring to $\mathbb{P}$. □

Definition 5.14 (Random Variables and Random Vectors)**.** Let $(\Omega, \mathbb{P})$ be a probability space and let $n \in \mathbb{N}$.
Let $B \subseteq \mathbb{R}$. A function
$$Y \, : \, \Omega \longrightarrow B\,; \quad \omega \mapsto Y(\omega)$$
is called a **random variable** (in short, **r.v.** or **rv**,) on $(\Omega, \mathfrak{F}, \mathbb{P})$. Let $B' \subseteq \mathbb{R}^n$. A function
$$\vec{X} = \left(X_1, X_2, \dots, X_n\right) \, : \, \Omega \longrightarrow B'\,; \quad \omega \mapsto \vec{X}(\omega) \ = \ \left(X_1\omega), \dots, X_n(\omega)\right)$$
is called a **random vector** on $(\Omega, \mathfrak{F}, \mathbb{P})$.
If there is a countable subset $B^* = \{y_1, y_2, \dots\}$ of B such that $\sum_j \mathbb{P}_Y\{y_j\} = 1$ (i.e., $\mathbb{P}\{Y \notin B^*\} = 0$), we call Y a **discrete random variable**. Likewise, if there is a countable subset B'^* of B' such that $\mathbb{P}\{\vec{X} \notin B'^*\} = 0$, we call $\vec{X}$ a **discrete random vector**.

Note that random variables and vectors which have a countable range are discrete. □

Definition 5.15 (σ–algebra generated by random elements)**.** $\boxed{\star}$ Let $(\Omega, \mathbb{P})$ be a probability space.

(a) Let $X : (\Omega, \mathbb{P}) \to \Omega'$ be a random element on $(\Omega, \mathbb{P})$. We call

(5.25) $$\sigma\{X\} \ := \ \sigma\{X^{-1}(A') : A' \subseteq \Omega'\}$$

the σ**–algebra generated by the random element** X.

(b) Let $X_i : (\Omega, \mathbb{P}) \to \Omega',\ i \in I$ be a family of random elements on $(\Omega, \mathbb{P})$. We call

(5.26) $$\sigma\big\{(X_i)_{i\in I}\big\} := \sigma\{X_i : i \in I\} := \sigma\{X_i^{-1}(A') : A' \subseteq \Omega',\ i \in I\}$$

the σ**–algebra generated by the family of random elements** $(X_i)_{i\in I}$.

Proposition 5.5. $\boxed{\star}$ *Let $X : (\Omega, \mathbb{P}) \to \Omega'$ be a random element and $g : \Omega' \to \mathbb{R}$.*

- *Let Z be the random variable $\omega \mapsto Z(\omega) := g\big(X(\omega)\big)$.*
- *Let $B^* \in \Omega'$ such that $\mathbb{P}_X(B^*) = 1$ and let $C^* := \{g(x) : x \in B^*\}$ be the direct image $g(B^*)$ of B^* under g. (See Definition 2.29 on p.19.)*

$$\begin{tikzcd} \Omega \arrow[r, "X"] \arrow[dr, "Z = g\circ X"'] & \Omega' \arrow[d, "g"] \\ & \mathbb{R} \end{tikzcd}$$

Then $\mathbb{P}_Z(C^*) = 1$.

Corollary 5.5. *Let $X : (\Omega, \mathbb{P}) \to \Omega'$ be a random element and $g : \Omega' \to \mathbb{R}$. Further, let Z be the random variable $g \circ X : \omega \mapsto Z(\omega) = g\big(X(\omega)\big)$. In other words, Z is the composition of g with X. Then*

(a) *If $\omega \mapsto X(\omega)$ only assumes finitely many (distinct) values $x_1, \dots, x_n$, then $\omega \mapsto Z(\omega)$ only assumes finitely many values $z_1, \dots, z_m$ (and $m \leq n$).*

(b) *If $\omega \mapsto X(\omega)$ only assumes an infinite sequence of (distinct) values (x_j), then $\omega \mapsto Z(\omega)$ assumes a countable set of function values. (This set forms a finite or infinite sequence. (See Definition 2.25 (Countable and uncountable sets) on p.15).*

(c) *If X is a discrete random element, then $Z = g(X)$ is a discrete random variable.*

5.4 Independence of Random Elements

Definition 5.16 (Independence of arbitrarily many random elements)**.** Given are a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and a family $\big(X_i\big)_{i\in I}$ of random elements on Ω. Here, I denotes an arbitrary set of indices. We say that this family is **independent** if, for ANY FINITE subselection of distinct indices $i_1, i_2, \dots, i_k \in I$ and $j = 1, 2, \dots, k$,

(5.27) $$\begin{aligned}&\mathbb{P}\{\, X_{i_1} \in A'_{i_1},\, X_{i_2} \in A'_{i_2},\, \dots,\, X_{i_k} \in A'_{i_k} \,\}\\ &\qquad = \mathbb{P}\{X_{i_1} \in A'_{i_1}\} \cdot \mathbb{P}\{X_{i_2} \in A'_{i_2}\} \cdots \mathbb{P}\{X_{i_k} \in A'_{i_k}\}, \quad \text{for all } A'_{i_j} \subseteq \Omega'.\ \square\end{aligned}$$

Fact 5.2 (Independence of discrete random elements)**.** $\boxed{\star}$ *Assume that the random elements X_i of Definition 5.16 are discrete and that $\Omega'_* \subseteq \Omega'$ is countable and satisfies $\mathbb{P}\{X_i \in \Omega'_*\} = 1$. Then* *it suffices to show that (5.27) is satisfied for events of the form* $\{X_{i_j} = \omega'\}$*, where* $\omega' \in \Omega'_*$*. In other words, it suffices to verify the following.*

- *For ANY FINITE subselection of distinct indices* $i_1, i_2, \dots, i_k \in I$ *and* $j = 1, 2, \dots, k$*,*

(5.28) $$\mathbb{P}\{X_{i_1} = \omega'_{i_1}, \dots, X_{i_k} = \omega'_{i_k}\} \;=\; P\{X_{i_1} = \omega'_{i_1}\} \cdots P\{X_{i_k} = \omega'_{i_k}\}$$

is satisfied for all $\omega'_{i_j} \in \Omega'_*$*.*

From this general case, we obtain the case $I = 1, 2$ *as follows.*

Independence of two random elements, X_1, X_2*: For all* $\omega', \tilde{\omega}' \in \Omega'_*$*,*

(5.29) $$\mathbb{P}\{X_1 = \omega', X_2 = \tilde{\omega}'\} \;=\; P\{X_1 = \omega'\} \cdot P\{X_2 = \tilde{\omega}'\}\,.$$

For $I = 1, 2, 3$*, we obtain*

Independence of three random elements, X_1, X_2, X_3*:*

(1) *For all subselections* $i_1 < i_2$ *of* $k = 2$ *elements of* $\{1, 2, 3\}$ *(there are 3 such subselections) and for all* $\omega'_{i_1}, \omega'_{i_2} \in \Omega'_*$*,*

(5.30) $$\mathbb{P}\{X_{i_1} = \omega'_{i_1}, X_{i_2} = \omega'_{i_2}\} \;=\; P\{X_{i_1} = \omega'_{i_1}\} \cdot P\{X_{i_2} = \omega'_{i_2}\}\,,$$

(2) *For* $k = 3$ *(i.e.,* $i_1 = 1, i_2 = 2, i_1 = 3$*) and for all* $\omega'_1, \omega'_2, \omega'_3 \in \Omega'_*$*,*

(5.31) $$\begin{aligned}\mathbb{P}\{X_1 = \omega'_1, X_2 =&\omega'_2, X_3 = \omega'_3\}\\ &= P\{X_1 = \omega'_1\} \cdot P\{X_2 = \omega'_2\} \cdot P\{X_3 = \omega'_3\}\,.\end{aligned}$$

For $I = 1, 2, \dots, n$*, we obtain*

Independence of n *random elements,* $X_1, X_2, \dots, X_n$*:*

For <u>EACH</u> $k = 2, 3, \dots, n-1, n$*, the following must be true: For all subselections* $i_1 < \dots < i_k$ *of* k *elements of* $\{1, \dots, n\}$ *and for all* $\omega'_{i_j} \in \Omega'_*$*,* $(1 \leq j \leq k)$*,*

(5.32) $$\mathbb{P}\{X_{i_1} = \omega'_{i_1}, \dots, X_{i_k} = \omega'_{i_k}\} \;=\; P\{X_{i_1} = \omega'_{i_1}\} \cdots P\{X_{i_k} = \omega'_{i_k}\}\,.$$

For $I = \mathbb{N}$*, we obtain*

Independence of an infinite sequence $X_1, X_2, \dots,$ *of random elements:*

For <u>EACH</u> $k = 2, 3, 4, \dots,$ *the following must be true: For all subselections* $i_1 < \dots < i_k$ *of* k *elements of* $\mathbb{N}$ *and for all* $\omega'_{i_j} \in \Omega'_*$*,* $(1 \leq j \leq k)$*,*

(5.33) $$\mathbb{P}\{X_{i_1} = \omega'_{i_1}, \dots, X_{i_k} = \omega'_{i_k}\} \;=\; P\{X_{i_1} = \omega'_{i_1}\} \cdots P\{X_{i_k} = \omega'_{i_k}\}\,.$$

Fact 5.3 (Independence of random variables). $\boxed{\star}$ *Assume that the random elements X_i of Definition 5.16 are random variables. Then it suffices to show that (5.27) is satisfied for events of the form $\{X_{i_j} \in]-\infty, \beta_{i_j}]\}$, for all $\beta_{i_j} \in \mathbb{R}$. In other words, it suffices to verify the following.*

- *For ANY FINITE subselection of distinct indices $i_1, i_2, \dots, i_k \in I$ and $j = 1, 2, \dots, k$,*

$$\mathbb{P}\{X_{i_1} \leq \beta_{i_1}, \dots, X_{i_k} \leq \beta_{i_k}\} = P\{X_{i_1} \leq \beta_{i_1}\} \cdots P\{X_{i_k} \leq \beta_{i_k}\}, \tag{5.34}$$

is satisfied for all $\beta_{i_j} \in \mathbb{R}$.

From this general case, we obtain the case $I = 1, 2$ as follows.

Independence of two random variables, Y_1, Y_2: For all $\beta_1, \beta_2 \in \mathbb{R}$,

$$\mathbb{P}\{Y_1 \leq \beta_1, Y_2 \leq \beta_2\} = P\{Y_1 \leq \beta_1\} \cdot P\{Y_2 \leq \beta_2\}. \tag{5.35}$$

For $I = 1, 2, 3$, we obtain

Independence of three random variables, Y_1, Y_2, Y_3:

(1) *For all subselections $i_1 < i_2$ of $k = 2$ elements of $\{1, 2, 3\}$ (there are 3 such subselections) and for all $\beta_{i_1}, \beta_{i_2} \in \mathbb{R}$,*

$$\mathbb{P}\{Y_{i_1} \leq \beta_{i_1}, Y_{i_2} \leq \beta_{i_2}\} = P\{Y_{i_1} \leq \beta_{i_1}\} \cdot P\{Y_{i_2} \leq \beta_{i_2}\}, \tag{5.36}$$

(2) *For $k = 3$ (i.e., $i_1 = 1, i_2 = 2, i_1 = 3$) and for all $\beta_1, \beta_2, \beta_3 \in \mathbb{R}$,*

$$\begin{aligned} &\mathbb{P}\{Y_1 \leq \beta_1,\ Y_2 \leq \beta_2,\ Y_3 \leq \beta_3\} \\ &\qquad = P\{Y_1 \leq \beta_1\} \cdot P\{Y_2 \leq \beta_2\} \cdot P\{Y_3 \leq \beta_3\}. \end{aligned} \tag{5.37}$$

For $I = 1, 2, \dots, n$, we obtain

Independence of n random variables, $Y_1, Y_2, \dots, Y_n$:
For EACH $k = 2, 3, \dots, n-1, n$, the following must be true: For all subselections $i_1 < \dots < i_k$ of k elements of $\{1, \dots, n\}$ and for all $\beta_{i_j} \in \mathbb{R}$, $(1 \leq j \leq k)$,

$$\mathbb{P}\{Y_{i_1} \leq \beta_{i_1}, \dots, Y_{i_k} \leq \beta_{i_k}\} = P\{Y_{i_1} \leq \beta_{i_1}\} \cdots P\{Y_{i_k} \leq \beta_{i_k}\}. \tag{5.38}$$

For $I = \mathbb{N}$, we obtain

Independence of an infinite sequence $Y_1, Y_2, \dots,$ of random variables:
For EACH $k = 2, 3, 4, \dots,$ the following must be true: For all subselections
$i_1 < \cdots < i_k$ *of k elements of $\mathbb{N}$ and for all $\beta_{i_j} \in \mathbb{R}$, $(1 \leq j \leq k)$,*

(5.39) $$\mathbb{P}\{Y_{i_1} \leq \beta_{i_1}, \dots, Y_{i_k} \leq \beta_{i_k}\} \;=\; P\{Y_{i_1} \leq \beta_{i_1}\} \cdots P\{Y_{i_k} \leq \beta_{i_k}\}\,.$$

Definition 5.17 (iid families)**.** Let $(X_i)_{i \in I}$ be a family of random elements $X_i : (\Omega, \mathbb{P}) \to \Omega'$. We speak of an **independent and identically distributed family**, aka **iid family** of random elements, if

(1) the X_i are independent,
(2) they all have the same distribution:

$$\mathbb{P}_{X_i}(B) \;=\; \mathbb{P}_{X_j}(B)\,, \qquad \text{for all } \; i, j \in I \; \text{ and all } \; B \subseteq \Omega'.$$

Note that this can also be written

$$\mathbb{P}\{X_i \in B\} \;=\; \mathbb{P}\{X_j \in B\}\,, \qquad \text{for all } \; i, j \in I \; \text{ and all } \; B \subseteq \Omega'.$$

In the special case of a sequence $X_1, X_2, \dots$ of iid random elements we speak of an **iid sequence** of random elements. $\square$

6 Advanced Topics – Measure and Probability ⋆

6.1 Random Variables as Measurable Functions

Definition 6.1 (Measurable functions). ⋆

(a) Let Ω be a nonempty set and $\mathfrak{F}$ a σ–algebra on Ω. We call the pair $(\Omega, \mathfrak{F})$ a **measurable space**. (This is not worthwhile remembering, but the remainder of this definition is.)

(b) Let $f : (\Omega, \mathfrak{F}) \longrightarrow (\Omega', \mathfrak{F}')$ be a function which has measurable spaces both as domain and codomain. We call this function **measurable with respect to $\mathfrak{F}$ and $\mathfrak{F}'$**, a.k.a. $(\mathfrak{F}, \mathfrak{F}')$**–measurable**, if

(6.1) $$A' \in \mathfrak{F}' \Rightarrow f^{-1}(A') \in \mathfrak{F}\,.$$

(c) If f is $\mathbb{R}^d$–valued, in particular if f is real–valued, and if we refer to f as being **$\mathfrak{F}$–measurable** or **Borel measurable**, then it is implied that $\mathfrak{F}' = \mathfrak{B}^d$, the Borel σ–algebra of $\mathbb{R}^d$. □

Definition 6.2 (Simple Function on Ω)**.** Let $(\Omega, \mathfrak{F})$ be a measurable space, $n \in \mathbb{N}$, $A_1, \dots, A_n \in \mathfrak{F}$. Further, let $c_1, c_2, \dots, c_n$ be a corresponding set of real numbers. Let

(6.2) $$f : \Omega \longrightarrow \mathbb{R}\,; \qquad \omega \mapsto f(\omega) := \sum_{j=1}^{n} c_j \mathbf{1}_{A_j}(\omega)$$

Then we call f a **simple function**. We say that f is in **standard form**, if the numhers c_j are distinct, i.e., $c_i \neq c_j$, for $i \neq j$. □

Proposition 6.1. ⋆ *Let $f = \sum_{j=1}^{n} c_j \mathbf{1}_{A_j}$ be a simple function. Then f has a representation in standard form. This standard representation is*

(6.3) $$f(\omega) = \sum_{i=1}^{k} d_i \mathbf{1}_{\{f=d_i\}}(\omega)\,, \quad \text{with distinct numbers } d_1, \dots, d_k.$$

Theorem 6.1. *Assume that $f_1, f_2, \dots$ are Borel measurable functions, $c_1, c_2, \dots \in \mathbb{R}$, $B \in \mathfrak{F}$. Then each of the following also is a Borel measurable function:*

• c_1 *(constant function)* • $c_1 f_1$ • $f_1 \pm f_2$ • $f_1 f_2$ • $\mathbf{1}_B f_1$ • f_1/f_2 *(if $f_2 \neq 0$)* • $\sum_{j=1}^{n} c_j f_j$

• $\min(f_1, f_2)$ • $\max(f_1, f_2)$ • $\min_{j=1,\dots,n} f_j$ • $\max_{j=1,\dots,n} f_j$ • $\inf_{j \in \mathbb{N}} f_j$ • $\sup_{j \in \mathbb{N}} f_j$ □

If they exist (see the subsequent remark), the following also are measurable functions:

- $\lim\limits_{j\to\infty} f_j$ • $\sum\limits_{j=1}^{\infty} f_j$ • $\min\limits_{j\in\mathbb{N}} f_j$ • $\max\limits_{j\in\mathbb{N}} f_j$

Theorem 6.2. *Let $(\Omega, \mathfrak{F})$ be a measurable space.*
Let $f : (\Omega, \mathfrak{F}) \longrightarrow [0, \infty[$ be a nonnegative, $(\mathfrak{F}, \mathfrak{B}^1)$–measurable function. Then there exists a sequence $0 \leq f_1 \leq f_2 \leq \cdots$ of simple functions such that $f_n \uparrow f$ as $n \to \infty$. In other words,

$$\lim_{n\to\infty} f_n(\omega) \;=\; f(\omega), \quad \textit{for all } \omega \in \Omega.$$

Theorem 6.3. *Let $(\Omega, \mathfrak{F}, \mathbb{P})$ be a probability space, $(\Omega', \mathfrak{F}')$ a measurable space, and*

$$X : (\Omega, \mathfrak{F}, \mathbb{P}) \longrightarrow (\Omega', \mathfrak{F}', \mathbb{P}_X)\,.$$

an $(\mathfrak{F}, \mathfrak{F}')$–measurable function. Then the formula

$$\mathbb{P}_X(A') \;:=\; \mathbb{P}\{X \in A'\} \;\; (A' \in \mathfrak{F}') \tag{6.4}$$

defines a probability measure on $\mathfrak{F}'$.

Only for the remainder of this chapter 6.1 (Advanced Topics – Measurable Functions), we modify Definitions 5.14 on p.42 and **??** on p.**??** as follows.

Definition 6.3 (Advanced level definition of random variables and random elements). $\boxed{\star}$
Given are a probability space $(\Omega, \mathfrak{F}, \mathbb{P})$, a measurable space $(\Omega', \mathfrak{F}')$, $d \in \mathbb{N}$, and an $(\mathfrak{F}, \mathfrak{F}'$–measurable function

$$X :\, (\Omega, \mathfrak{F}, \mathbb{P}) \longrightarrow (\Omega', \mathfrak{F}')\,.$$

(a) We call X a **random element**.
(b) If $(\Omega', \mathfrak{F}') = (\mathbb{R}, \mathfrak{B}))$, we also call X a **random variable**.
(c) If $(\Omega', \mathfrak{F}') = (\mathbb{R}^d, \mathfrak{B}^d))$, we also call X a **random vector**. $\square$

Theorem 6.4. $\boxed{\star}$ *Let Ω be a nonempty set and let $\mathcal{E}, \mathcal{E}_1$ and $\mathcal{E}_2$ be three collections of subsets of Ω. Then*

$$\mathcal{E}_1 \subseteq \mathcal{E}_2 \quad \Rightarrow \quad \sigma(\mathcal{E}_1) \subseteq \sigma(\mathcal{E}_2)\,, \tag{6.5}$$

$$\sigma\big(\sigma(\mathcal{E})\big) = \sigma(\mathcal{E})\,, \tag{6.6}$$

$$\sigma(\mathcal{E}_1) \supseteq \mathcal{E}_2 \;\; \textit{and} \;\; \sigma(\mathcal{E}_2) \supseteq \mathcal{E}_1 \quad \Rightarrow \quad \sigma(\mathcal{E}_1) = \sigma(\mathcal{E}_2)\,. \tag{6.7}$$

Definition 6.4 (Advanced Definition of σ–algebras generated by random elements). $\boxed{\star}$
We define for a function f and a family of functions $(f_i)i \in I$,

$$f, f_i : \Omega \longrightarrow (\Omega', \mathfrak{F}'),\; i \in I \;:$$

$$\sigma\{f\} := \sigma\{f^{-1}(A') : A' \in \mathfrak{F}'\} \tag{6.8}$$

$$\sigma\big\{(f_i)_{i\in I}\big\} := \sigma\{f_i : i \in I\} := \sigma\{f_i^{-1}(A') : A' \in \mathfrak{F}',\; i \in I\} \tag{6.9}$$

(a) We call $\sigma\{f\}$ the **σ–algebra generated by the function** f.
(b) We call $\sigma\big\{(f_i)_{i\in I}\big\}$ the **σ–algebra generated by the family of functions** $(f_i)_{i\in I}$. $\square$

Theorem 6.5. $\boxed{\star}$ *Given is a function f with measurable spaces $(\Omega, \mathfrak{F})$ as domain and $(\Omega', \mathfrak{F}')$ as codomain:*

$$f : (\Omega, \mathfrak{F}) \longrightarrow (\Omega', \mathfrak{F}')\,.$$

No assumption is made about $(\mathfrak{F}, \mathfrak{F}')$–measurability. Then

(a) *$\sigma\{f\} = \{f^{-1}(A') : A' \in \mathfrak{F}'\}$. In particular, $\{f^{-1}(A') : A' \in \mathfrak{F}'\}$is a σ–algebra for Ω.*
(b) *f is $(\mathfrak{F}, \mathfrak{F}')$–measurable $\Leftrightarrow$ $\sigma\{f\} \subseteq \mathfrak{F}$.*
(c) *We can strengthen assertion **(b)** as follows: Let $\mathcal{E}'$ be a generator of $\mathfrak{F}'$. Then*
f is $(\mathfrak{F}, \mathfrak{F}')$–measurable $\Leftrightarrow$ $\{f^{-1}(E') : E' \in \mathcal{E}'\} \subseteq \mathfrak{F}$.

Definition 6.5 (Independence of arbitrarily many random elements – advanced definition). $\boxed{\star}$ Given are a probability space $(\Omega, \mathfrak{F}, \mathbb{P})$, a measurable space $(\Omega', \mathfrak{F}')$, and a family of random elements,

$$X_i : (\Omega, \mathfrak{F}, \mathbb{P}) \longrightarrow (\Omega', \mathfrak{F}') \qquad (i \in I).$$

Here, I denotes an arbitrary set of indices. We say that this family is **independent** if, for ANY FINITE subselection of distinct indices $i_1, i_2, \dots, i_k \in I$ and $j = 1, 2, \dots, k$,

$$\begin{aligned} &\mathbb{P}\{\, X_{i_1} \in A'_{i_1},\; X_{i_2} \in A'_{i_2},\; \dots,\; X_{i_k} \in A'_{i_k} \,\} \\ &\qquad = \mathbb{P}\{X_{i_1} \in A'_{i_1}\} \cdot \mathbb{P}\{X_{i_2} \in A'_{i_2}\} \cdots \mathbb{P}\{X_{i_k} \in A'_{i_k}\}\,, \quad \text{for all } A'_{i_j} \in \mathfrak{F}'. \end{aligned} \tag{6.10}$$

$\square$

Definition 6.6 (Independence of a family of sets of measurable sets)**.** $\boxed{\star}$
Given are a probability space $(\Omega, \mathfrak{F}, \mathbb{P})$, and a family

$$\boldsymbol{\mathcal{E}}_i \subseteq \mathfrak{F} \qquad (i \in I).$$

We say that this family is **independent** if, for ANY FINITE subselection of distinct indices $i_1, i_2, \dots, i_k \in I$ and $j = 1, 2, \dots, k$, and for any choices $A_{i_j} \in \boldsymbol{\mathcal{E}}_{i_j}$,

$$\mathbb{P}\{ A_{i_1} \cap A_{i_2} \cap \cdots, \cap A_{i_k} \} = \mathbb{P}(A_{i_1}) \cdot \mathbb{P}(A_{i_2}) \cdots \mathbb{P}(A_{i_k}), \quad \text{for all } A_{i_j} \in \boldsymbol{\mathcal{E}}_{i_j}. \quad \square \tag{6.11}$$

Proposition 6.2. $\boxed{\star}$ *Given are a probability space $(\Omega, \mathfrak{F}, \mathbb{P})$, a measurable space $(\Omega', \mathfrak{F}')$, and a family of random elements,*

$$X_i : (\Omega, \mathfrak{F}, \mathbb{P}) \longrightarrow (\Omega', \mathfrak{F}') \qquad (i \in I).$$

Then,

$$\textit{The family } (X_i)_i \in I \textit{ is independent} \quad \Leftrightarrow \quad \textit{the family } \sigma\{X_i\}_{i \in I} \textit{ is independent.}$$

Definition 6.7 (Independence of arbitrarily many random elements – precise definition)**.**
REMOVED: This has already been covered in Definition 6.5 (Independence of arbitrarily many random elements – advanced definition) on p.49.

6.2 Measures

This chapter is very selective and incomplete at this point in time. Additions will be made as time allows.

Definition 6.8 (Abstract measures)**.** Let $(\Omega, \mathfrak{F})$ be a measurable space. A **measure** on $\mathfrak{F}$ is an extended real–valued function

$$\mu : \mathfrak{F} \to [0, \infty]; \quad A \mapsto \mu(A). \qquad \text{such that}$$

$$\mu(\emptyset) = 0, \tag{6.12}$$

$$A, B \in \mathfrak{F} \text{ and } A \subseteq B \quad \Rightarrow \quad \mu(A) \leq \mu(B), \qquad \text{(monotony)} \tag{6.13}$$

$$(A_n)_{n \in \mathbb{N}} \in \mathfrak{F} \text{ disjoint} \quad \Rightarrow \quad \mu\Big(\biguplus_{n \in \mathbb{N}} A_n\Big) = \sum_{n \in \mathbb{N}} \mu(A_n). \qquad (\sigma\text{–additivity}) \tag{6.14}$$

- The triplet $(\Omega, \mathfrak{F}, \mu)$ is called a **measure space**
- We call any set $N \subseteq \Omega$ with measure zero a μ **Null set**.

- We call μ a **discrete measure** if there is a countable $A^* \in \mathfrak{F}$ such that $\mu\left({A^*}^{\complement}\right) = 0$. We then call $(\Omega, \mathfrak{F}, \mu)$ a **discrete measure space**
- $\boxed{\star}$ We call μ a **finite measure** on $\mathfrak{F}$ if $\mu(\Omega) < \infty$.
- We call μ a σ**–finite measure** on $\mathfrak{F}$ if one can find a sequence $A_n \in \mathfrak{F}$ such that $\mu(A_n) < \infty$ and $\Omega = \bigcup_n A_n$.

See these footnotes concerning measurable spaces, [8] extended real numbers, [9] and μ–null sets. [10]
$\square$

Definition 6.9 (Counting measure). $\boxed{\star}$ Let $(\Omega, \mathfrak{F})$ be a measurable space, $A^* \neq \emptyset$ a countable subset of Ω

(a) We call the measure Σ_* on $\mathfrak{F}$, defined by

$$\Sigma_*(A) := |A \cap A^*|$$

the **counting measure** on $\mathfrak{F}$ with respect to A^*.

(b) In particular, if $\Omega \subseteq \mathbb{R}$ and $A^* = \mathbb{N}$, we call Σ_* the **standard counting measure** on $\mathfrak{F}$.

(c) If no reference to a σ–algebra is made, we set $\mathfrak{F} := 2^{\Omega}$. $\square$

Fact 6.1. $\boxed{\star}$

- *Let $\mathfrak{I} = \mathfrak{I}_5$ or $\mathfrak{I} = \mathfrak{I}_8$. Let the function $\mu_0 : \mathfrak{I} \to [0, \infty[$ (so $E \in \mathfrak{I} \Rightarrow \mu_0(E) < \infty$) satisfy $\mu_0(\emptyset) = 0$, $\mu_0(\mathbb{R}^d) = 1$ and σ–additivity on $\mathfrak{I}$: $E_n \in \mathfrak{I}$ disjoint such that $\mathbb{E} := \biguplus\limits_{n\in\mathbb{N}} E_n \in \mathfrak{I} \Rightarrow \mu_0(E) = \sum\limits_{n\in\mathbb{N}} \mu_0(E_n)$.*
  Then μ_0 can be uniquely extended to a measure on $\mathfrak{B}^d$, the Borel sets of $\mathbb{R}^d$.

- *One can drop the requirement that $\mu_0(A) < \infty$ for all $A \in \mathfrak{I}$, but then the extension μ is no more guaranteed to be unique.*
- *Note that for $d = 1$, the following sets are equal: $\mathfrak{I}_5 = \mathfrak{I}_1$, $\mathfrak{I}_8 = \mathfrak{I}_4$, $\mathfrak{B}^1 = \mathfrak{B}$.* $\square$

Theorem 6.6. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be nonnegative and Borel–measurable. Then the set function*

$$\mu : \mathfrak{B}^d \longrightarrow [0, \infty], \qquad \mu(A) := \int_A f \, d\lambda^d \tag{6.15}$$

defines a measure on $\mathfrak{B}^d$.

[8] See Definition 6.1 (Measurable functions) on p.47.
[9] See Definition 2.14 (Extended real numbers) on p.11.
[10] Strictly speaking any set N such that $N \subseteq A$ and $\mu(A) = 0$ is said to be μ Null. We ignore such fine points.

Theorem 6.7 (Continuity property of measures). $\boxed{\star}$ *Let* $(\Omega, \mathfrak{F}, \mu)$ *be a measure space. If* $A_n, B_n \in \mathfrak{F}$*, then the following is true:*

(6.16) $$A_n \uparrow \Rightarrow \mathbb{P}(A_n) \uparrow \mu\left(\bigcup_{n\in\mathbb{N}} A_n\right),$$

(6.17) $$B_n \downarrow \text{ and } \mu(B_1) < \infty \Rightarrow \mathbb{P}(B_n) \downarrow \mathbb{P}\left(\bigcap_{n\in\mathbb{N}} B_n\right).$$

Proposition 6.3. *Let* $(\Omega, \mathfrak{F}, \mu)$ *be a measure space and* $(\Omega', \mathfrak{F}')$ *a measurable space.*
Let $f : \Omega \to \Omega'$ *be* $(\mathfrak{F}, \mathfrak{F}')$ *measurable. Then the set function*

(6.18) $$\mu_f : \mathfrak{F}' \to [0, \infty]; A' \mapsto \mu\{f \in A'\} = \mu\{\omega \in \Omega : f(\omega) \in A'\}$$

defines a measure on $(\Omega', \mathfrak{F}')$*. Moreover, if* μ *is a probability measure on* $\mathfrak{F}$*, i.e.,* $\mu(\Omega = 1)$*, then* μ_f *is a probability measure on* $\mathfrak{F}'$*.*

Definition 6.10 (Image measure).

(1) $\boxed{\star}$ We call the measure μ_f of Proposition 10.11 the **image measure** of μ under f aka the **measure induced** by μ and f.

(2) We now switch notation from f and μ to the more customary X and $\mathbb{P}$ for the sake of clarity. In the case of a random element X on a probability space $(\Omega, \mathfrak{F}, \mathbb{P})$ with codomain $(\Omega', \mathfrak{F}')$, we call the image measure $\mathbb{P}_X$ of $\mathbb{P}$ under X which is, according to (10.47), given by

(6.19) $$\mathbb{P}_X(B) := \mathbb{P}\{X \in B\} = \mathbb{P}\{\omega \in \Omega : X(\omega) \in B\}, \quad (B \in \mathfrak{B}^1)$$

the **probability distribution** or simply the **distribution** of X. $\square$

6.3 Abstract Integrals

Definition 6.11 (Abstract integral for simple functions). Let $(\Omega, \mathfrak{F}, \mu)$ be a measure space. (see Definition 6.8 (Abstract measures) on p.50)

Let $n \in \mathbb{N}$, $A_1, \dots, A_n \in \mathfrak{F}$, $c_1, \dots, c_n \in [0, \infty[$. Let

$$f : (\Omega, \mathfrak{F}, \mu) \longrightarrow \mathbb{R}; \qquad f(\omega) = \sum_{j=1}^{n} c_j \mathbf{1}_{A_j}(\omega).$$

The **abstract integral** aka **integral** of the simple function f with respect to μ is

(6.20) $$\int f d\mu := \int f(\omega) d\mu(\omega) := \int f(\omega)\mu(d\omega) := \sum_{j=1}^{n} c_j \mu(A_j). \quad \square$$

Proposition 6.4. $\boxed{\star}$ *Let $(\Omega, \mathfrak{F}, \mu)$ be a measure space. Let $f, g_n, h_n : (\Omega, \mathfrak{F}, \mu) \longrightarrow \mathbb{R}$ be nonnegative, $(\mathfrak{F}, \mathfrak{B}^1)$measurable functions. Assume further that the functions g_n and h_n are simple. Then the following is true:*

(6.21) $$\text{If } g_n \uparrow f \text{ and } h_n \uparrow f, \text{ then } \lim_{n\to\infty} \int g_n d\mu = \lim_{n\to\infty} \int h_n d\mu .$$

Definition 6.12 (Abstract integral for measurable functions)**.**
(a) Let $(\Omega, \mathfrak{F}, \mu)$ be a measure space, $f, f_n : (\Omega, \mathfrak{F}, \mu) \longrightarrow \mathbb{R}$ $(\mathfrak{F}, \mathfrak{B}^1)$measurable, and assume that the functions f_n are simple and • **either** $0 \leq f_n \uparrow f$ • **or** $0 \geq f_n \downarrow f$. Then

(6.22) $$\int f d\mu := \lim_{n\to\infty} \int f_n d\mu$$

is called the **abstract integral** aka **integral of** f **with respect to** μ. $\square$

(b) Let $(\Omega, \mathfrak{F}, \mu)$ be a measure space and $f : (\Omega, \mathfrak{F}, \mu) \longrightarrow \mathbb{R}$ $(\mathfrak{F}, \mathfrak{B}^1)$measurable, such that

- both f^+ and f^- are limits of nondecreasing sequences of simple functions ≥ 0;
- at least one of $\int f^+ d\mu$, $\int f^- d\mu$ is finite. (According to **(a)**, those integrals exist, but neither of them was guaranteed to be finite.)

Then we define the **abstract integral** aka **integral of** f **with respect to** μ, as the expression

(6.23) $$\int f\, d\mu = \int (f^+ - f^-)\, d\mu := \int f^+\, d\mu - \int f^-\, d\mu .$$

(c) We call a real–valued function f μ**–integrable**, if $\int f\, d\mu$ exists and is finite. $\square$

Assumption 6.1. Unless explicitly stated otherwise, we assume the following for the remainder of this chapter (Chapter 6 (Advanced Topics – Measure and Probability)).

- The underlying measurable space is $(\Omega, \mathfrak{F})$.
- The underlying measure is μ.
- "measurable" means "$(\mathfrak{F}, \mathfrak{B}^1)$measurable". $\square$

Theorem 6.8. $\boxed{\star}$ *abstract integrals satisfy the following.*
Let $A \in \mathfrak{F}$ and assume that f is $(\mathfrak{F}, \mathfrak{B}^1)$measurable. Then

(a) *If $\int f\, d\mu$ exists, then $\int \mathbf{1}_A f\, d\mu$ exists.*
(b) *If f is μ–integrable, then $\mathbf{1}_A f$ is μ–integrable.*

Definition 6.13. Let $A \in \mathfrak{F}$ and assume that f is a measurable function on $(\Omega, \mathfrak{F}, \mu)$, for which the abstract integral $\int f d\mu$ exists. The **abstract integral of f on A** or **over A** is defined by the expression

$$\int_A f\,d\mu \;:=\; \int_A f(\omega) d\,\mu(\omega) \;:=\; \int_A f(\omega)\mu(d\omega) \;:=\; \int \mathbf{1}_A f\,d\mu\,. \tag{6.24}$$

We say that f is **μ–integrable on A**, if $\int_A f\,d\mu$ exists and is finite. □

Proposition 6.5 (Integrability criterion). $\boxed{\star}$ *Let f be a measurable function and $A \in \mathfrak{F}$. Then*

$$f \text{ is integrable on } A \;\Leftrightarrow\; \int_A |f|\,d\mu < \infty \;\Leftrightarrow\; \text{both} \int_A f^+\,d\mu < \infty \text{ and } \int_A f^-\,d\mu < \infty.$$

Theorem 6.9. *Assume that $f, g, f_1, f_2, \dots$ are measurable functions, $c, c_1, c_2, \dots \in \mathbb{R}$, and $A \in \mathfrak{F}$. Then μ–integrals on A satisfy the following.*

(a) Positivity: $\displaystyle\int_A 0\,d\mu = 0; \qquad f \geq 0 \text{ on } A \;\Rightarrow\; \int_A f\,d\mu \geq 0,$

(b) Monotonicity: $\displaystyle\mu\{\omega \in A : f(\omega) > g(\omega)\} = 0 \;\Rightarrow\; \int_A f\,d\mu \leq \int_A g\,d\mu.$

In particular, $\displaystyle f \leq g \text{ on } A \;\Rightarrow\; \int_A f\,d\mu \leq \int_A g\,d\mu,$

and also, $\displaystyle\mu\{\omega \in A : f(\omega) \neq g(\omega)\} = 0 \;\Rightarrow\; \int_A f\,d\mu = \int_A g\,d\mu.$

(b) Linearity I: *f, g integrable on A* $\displaystyle\Rightarrow\; \int_A (f \pm g)\,d\mu = \int_A f\,d\mu \pm \int_A g\,d\mu$

and also, $\displaystyle\int_A (cf)\,d\mu = c\int_A f\,d\mu.$

Linearity II: *$f_1 \dots, f_n$ integrable* $\displaystyle\Rightarrow\; \int_A \Big(\sum_{j=1}^n f_j\Big)\,d\mu = \sum_{j=1}^n c_j \int_A f_j\,d\mu.$

(d) Monotone Convergence: *Assume that $0 \leq f_1 \leq f_2 \leq \cdots$, $0 \geq g_1 \geq g_2 \geq \cdots$.*

Then $\displaystyle\int_A f_n\,d\mu \uparrow \int_A \Big(\sup_{n\in\mathbb{N}} f_n\Big)\,d\mu$ *and* $\displaystyle\int_A g_n\,d\mu \downarrow \int_A \Big(\inf_{n\in\mathbb{N}} g_n\Big)\,d\mu$ *as $n \to \infty$.*

(e) Dominated Convergence: *Assume that*

- $\displaystyle\lim_{n\to\infty} f_n$ *exists,*
- $f_n \leq g$ *for all $n \in \mathbb{N}$,*
- $\displaystyle\int_A g\,d\mu < \infty.$

Then $\displaystyle\lim_{n\to\infty}\int_A f_n\,d\mu = \int_A \Big(\lim_{n\to\infty} f_n\Big)\,d\mu$ *as $n \to \infty$.*

Definition 6.14 (Product measure space). $\boxed{\star}$ Let $(\Omega_1, \mathfrak{F}_1, \mu_1)$ and $(\Omega_1, \mathfrak{F}_1, \mu_1)$ be measure spaces with σ–finite measures μ_1, μ_2. Let

$$\mathfrak{F}_1 \otimes \mathfrak{F}_2 \ := \ \sigma\{A_1 \times A_2 \ : \ A_1 \in \mathfrak{F}_1,\ A_2 \in \mathfrak{F}_2\}\,. \tag{6.25}$$

Let $\mu_1 \times \mu_2 : \mathfrak{F}_1 \otimes \mathfrak{F}_2 \longrightarrow [0, \infty]$ be the measure which is uniquely determined by

$$\mu_1 \times \mu_2(A_1 \times A_2) \ = \ \mu_1(A_1) \cdot \mu_2(A_2)\,, \quad \text{for } A_1 \in \mathfrak{F}_1 \text{ and } A_2 \in \mathfrak{F}_2. \tag{6.26}$$

We call the measure space $(\Omega_1 \times \Omega_2, \mathfrak{F}_1 \otimes \mathfrak{F}_2,\ \mu_1 \times \mu_2)$ the **product measure space** aka **product space** of the factors $(\Omega_1, \mathfrak{F}_1, \mu_1)$ and $(\Omega_1, \mathfrak{F}_1, \mu_1)$, $\mathfrak{F}_1 \otimes \mathfrak{F}_2$ the **product σ–algebra** of the factors $\mathfrak{F}_1$ and $\mathfrak{F}_2$, and $\mu_1 \times \mu_2$ the **product measure** of the factors μ_1 and μ_2.

There are alternate ways to denote integrals with respect to $\mu_1 \times \mu_2$.

$$\begin{aligned} \int f \, d\, \mu_1 \times \mu_2 \ &= \ \int f(\omega_1, \omega_2)\, d\, \mu_1 \times \mu_2(\omega_1, \omega_2) \\ &= \ \int f(\omega_1, \omega_2)\, \mu_1 \times \mu_2\big(d(\omega_1, \omega_2)\big) \ = \ \int f(\omega_1, \omega_2)\, \mu_1 \times \mu_2(d\omega_1, d\omega_2) \end{aligned} \tag{6.27}$$

See (6.20) and (6.24). $\square$

Theorem 6.10 (Fubini's theorem for abstract integrals). $\boxed{\star}$ *Let $(\Omega_1, \mathfrak{F}_1, \mu_1)$ and $(\Omega_1, \mathfrak{F}_1, \mu_1)$ be measure spaces with σ–finite measures μ_1, μ_2. Let*

$$f : (\Omega_1 \times \Omega_2, \mathfrak{F}_1 \otimes \mathfrak{F}_2,\ \mu_1 \times \mu_2) \longrightarrow \mathbb{R}; \quad (\omega_1, \omega_2) \mapsto f(\omega_1, \omega_2), \quad \text{be } \mathfrak{F}_1 \otimes \mathfrak{F}_2\text{–measurable.}$$

Assume that f is nonnegative and/or $(\mu_1 \times \mu_2)$–integrable, and that $A_1 \in \mathfrak{F}_1,\ A_2 \in \mathfrak{F}_2$. Then

$$\begin{aligned} \int_{A_1 \times A_2} f \, d\mu_1 \times \mu_2 \ &= \ \int_{A_1} \left(\int_{A_2} f \, d\mu_2 \right) d\mu_1 \\ &= \ \int_{A_2} \left(\int_{A_1} f \, d\mu_1 \right) d\mu_2\,. \end{aligned} \tag{6.28}$$

When we supply the arguments, ω_1 and ω_2, (6.28) reads

$$\begin{aligned} \int_{A_1 \times A_2} f\big(\omega_1, \omega_2\big)\, \mu_1 \times \mu_2\big(d(\omega_1, \omega_2)\big) \ &= \ \int_{A_1} \left(\int_{A_2} f\big(\omega_1, \omega_2\big)\, \mu_{d_2}(d\omega_2) \right) \mu_{d_1}(d\omega_1) \\ &= \ \int_{A_2} \left(\int_{A_1} f\big(\omega_1, \omega_2\big)\, \mu_{d_1}(d\omega_1) \right) \mu_{d_2}(d\omega_2)\,. \end{aligned} \tag{6.29}$$

6.4 The ILMD Method

Theorem 6.11 (Integrals under Transforms)**.** $\boxed{\star}$ *Let $(\Omega, \mathfrak{F}, \mu)$ be a measure space and let $(\Omega', \mathfrak{F}')$ be a measurable space. Assume that $f : \Omega \to \Omega'$ is $(\mathfrak{F}, \mathfrak{F}')$–measurable. and $g : \Omega' \to \mathbb{R}$ is $(\mathfrak{F}', \mathfrak{B}^1)$–measurable. μ_f denotes the image measure of μ under f on $\mathfrak{F}'$. It was defined in Definition 6.10 on p.52 as*

$$\mu_f(A') \;=\; \mu\{f \in A'\} \;=\; \mu\big(f^{-1}(A')\big)\,.$$

If $g \geq 0$ or $g \circ f$ is integrable then

$$\int g \circ f \, d\mu \;=\; \int g \, d\mu_f, \quad \text{i.e.,} \quad \int_\Omega g\big(f(\omega)\big)\, d\mu(\omega) \;=\; \int_{\Omega'} g(\omega')\, d\mu_f(\omega')\,. \tag{6.30}$$

Theorem 6.12. $\boxed{\star}$ *Let $(\Omega, \mathfrak{F}, \mu)$ be a measure space and let f be a nonnegative, real–valued, Borel–measurable function on $(\Omega, \mathfrak{F}, \mu)$. Let ν be the measure defined by*

$$\nu(A) \;:=\; \int_A f \, d\mu \tag{6.31}$$

(see Theorem 6.6 on p.51). Further, let φ be a real–valued, Borel–measurable function on Ω, such that $\varphi \geq 0$ or φ is ν–integrable. Then

$$\int_A \varphi \, d\nu \;=\; \int_A \varphi \cdot f \, d\mu, \;\; \text{i.e.,} \;\; \int_A \varphi(\omega)\, \nu(d\omega) \;=\; \int_A \varphi(\omega)\, f(\omega)\, \mu(d\omega); \;\; A \in \mathfrak{F}\,. \tag{6.32}$$

6.5 Expectation and Variance as Probability Measure Integrals

Definition 6.15 (Expected value of a random variable)**.** Let Y be a random variable on a probability space $(\Omega, ,\mathbb{P})$.

(a) We call

$$\mathbb{E}[Y] \;:=\; \int Y \, d\mathbb{P} \tag{6.33}$$

the **expected value**, also **expectation** or **mean** of Y.

(b) We call

$$Var[Y] \;:=\; \mathbb{E}\big[(Y - \mathbb{E}[Y])^2\big] \;=\; \int (Y - \mathbb{E}[Y])^2 \, d\mathbb{P} \tag{6.34}$$

the **variance**, of Y.

(c) We call $SD[Y] \;:=\; \sigma_Y \;:=\; \sqrt{Var[Y]}$ The **standard deviation** of Y. ☐

Theorem 6.13 (LOTUS: Expectations under Transforms)**.** *Let $(\Omega, \mathfrak{F}, \mathbb{P})$ be a probability space and let $(\Omega', \mathfrak{F}')$ be a measurable space. Let X be a random element on Ω which takes values in Ω'. Moreover, let $g : \Omega' \to \mathbb{R}$; $x \mapsto g(x)$, be a random variable on $(\Omega, \mathfrak{F}', \mathbb{P}_X)$. Here, $\mathbb{P}_X$ denotes the distribution of X (this is the image measure of $\mathbb{P}$ under X on $\mathfrak{F}'$).*
If $g \geq 0$ or $g \circ X$ is integrable, then

$$\mathbb{E}[g \circ X] \;=\; \int_{\Omega} g \circ X(\omega)\, \mathbb{P}(d\omega) \;=\; \int_{\Omega'} g(x)\, \mathbb{P}_X(dx)\,. \tag{6.35}$$

In particular, if X itself is a random variable and thus, $(\Omega', \mathfrak{F}') = (\mathbb{R}, \mathfrak{B}^1)$, then

$$\mathbb{E}[X] \;=\; \int_{\Omega} X(\omega)\, \mathbb{P}(d\omega) \;=\; \int_{\mathbb{R}} x\, \mathbb{P}_X(dx)\,. \tag{6.36}$$

7 Combinatorial Analysis

7.1 The Multiplication Rule

Theorem 7.1 (Multiplication rule). ***(A)*** *Assume that two actions A and B are performed such that*

- *the first one has m outcomes, $\{a_1, a_2, \dots, a_m\}$,*
- *the second one has n outcomes $\{b_1, b_2, \dots, b_n\}$ for each outcome of the first one.*

- *Then the number of combined outcomes (a_i, b_j) is mn.*

(B) *Generalization. Assume that k actions $A_1, \dots, A_k$ are performed such that*

- *action A_1 has n_1 outcomes, $\{a_1^{(1)}, a_2^{(1)}, \dots, a_{n_1}^{(1)}\}$,*
- *action A_2 has n_2 outcomes, $\{a_1^{(2)}, a_2^{(2)}, \dots, a_{n_2}^{(2)}\}$ for each outcome of A_1,*
- *action A_3 has n_3 outcomes, $\{a_1^{(3)}, a_2^{(3)}, \dots, a_{n_3}^{(3)}\}$ for each combined outcome (x_1, x_2), where x_1 is one of the A_1–outcomes and x_2 is one of the A_2–outcomes,*
  \- -
- *action A_k has n_k outcomes, $\{a_1^{(k)}, a_2^{(k)}, \dots, a_{n_k}^{(k)}\}$ for each combined outcome (x_1, x_2, x_{k-1}), where each x_j is one of the A_j–outcomes, i.e., x_j is one of $a_1^{(j)}, \dots, a_{n_j}^{(j)}$.*

- *Then there are $n_1 \cdot n_2 \cdots n_k$ combined outcomes $(x_1, x_2, \dots, x_k)$.*
  Here, each x_j is one of the n_j outcomes $a_1^{(j)}, \dots, a_{n_j}^{(j)}$ of A_j.

7.2 Permutations

Definition 7.1 (WMS Ch.02.6, Definition 2.7 - Permutation)**.** An ordered arrangement of r distinct objects is called a **permutation** of size r. The number of ways of ordering n distinct objects taken r at a time will be designated by the symbol P_r^n. $\square$

Theorem 7.2 (WMS Ch.02.6, Theorem 2.2)**.**

$$P_r^n \;=\; n(n-1)(n-2)\cdot(n-r+1) \;=\; \frac{n!}{(n-r)!}\,. \tag{7.1}$$

Here, $n!$ ("n factorial") is defined as follows.

$$n! \;=\; \begin{cases} n(n-1)\cdots 2\cdot 1\,, & \text{if } n\in\mathbb{N}\,,\\ 1\,, & \text{if } n=0\,. \end{cases} \tag{7.2}$$

7.3 Combinations, Binomial and Multinomial Coefficients

Theorem 7.3. *Let $0 \le k \le n$. A set of size n has*

$$\frac{n!}{k!(n-k)!}\,.$$

subsets of size k.

Definition 7.2 (Number of combinations)**.** We call the number of selections of size k from a collection of n distinct items when the order in which those k items were selected is ignored, the **number of combinations of n objects taken k at a time**. We write $\binom{n}{k}$ for this number. □

Theorem 7.4. *Given are n items of which n_1 are alike, n_2 are alike, $\dots$, n_r are alike ($n_1 + \cdots + n_r = n$). Then the number of distinguishable arrangements of those n items is*

$$\binom{n}{n_1, n_2, \dots n_r} = \frac{n!}{n_1!\, n_2! \,\cdots\, n_r!}\,.$$

Definition 7.3 (Multinomial coefficients)**.** The numbers

$$\binom{n}{n_1\, n_2 \cdots n_r} = \frac{n!}{n_1! n_2! \cdots n_r!}\,. \tag{7.3}$$

that appear in Theorem 7.4 are called **multinomial coefficients.** If $r = 2$, then there is some integer $0 \le k \le n$ such that $n_1 = k$ and $n_2 = n - k$. We write

$$\binom{n}{k} := \frac{n!}{k!(n-k)!} \qquad \text{for} \qquad \binom{n}{k, n-k} \tag{7.4}$$

and speak of **binomial coefficients**. Convention: We define $\binom{n}{k} := 0$ for $k > n$. □

Theorem 7.5. *Let $r, n \in \mathbb{N}$ such $r \le n$ and $x_1, x_2, \dots x_r \in \mathbb{R}$. Then*

$$\left(x_1 + x_2 + \cdots + x_r\right)^n = \sum_{\substack{n_1,\dots,n_r \ge 0 \\ n_1 + \cdots + n_r = n}} \binom{n}{n_1, n_2, \dots n_r} x_1^{n_1}\, x_2^{n_2} \cdots x_r^{n_r}\,. \tag{7.5}$$

In particular, if $r = 2$, we obtain the ***binomial theorem****:*

$$(x_1 + x_2)^n = \sum_{j=0}^{n} \binom{n}{j} x_1^j x_2^{n-j}.$$

Theorem 7.6. *Given are n distinct items and r distinct bins of fixed sizes $n_1, n_2, \dots, n_r$ such that $n_1 + \dots + n_r = n$.*
Then the number of distinguishable placements of the n items into those r bins, when disregarding the order in which the items were placed into any one of those bins, is

$$\binom{n}{n_1, n_2, \dots n_r} = \frac{n!}{n_1!\, n_2! \cdots n_r!}.$$

Proposition 7.1. ***(A)*** *There are $\binom{n-1}{r-1}$ distinct integer–valued vectors $\vec{x} = (x_1, x_2, \dots, x_r)$ such that*

$$x_1 + x_2 + \dots + x_r = n \quad \text{and} \quad x_i > 0,\ i = 1, \dots, r.$$

(B) *There are $\binom{n+r-1}{r-1}$ distinct integer–valued vectors $\vec{y} = (y_1, y_2, \dots, y_r)$ such that*

$$y_1 + y_2 + \dots + y_r = n \quad \text{and} \quad y_i \geq 0,\ i = 1, \dots, r.$$

Proposition 7.2. ***(A)*** *There are $\binom{n-1}{r-1}$ ways to select n indistinguishable items into r distinct bins such that each bin contains at least one item.*

(B) *There are $\binom{n+r-1}{r-1}$ ways to select n indistinguishable items into r distinct bins.*

Remark 7.1. The multinomial coefficients

$$\binom{n}{n_1\, n_2 \cdots n_k} = \frac{n!}{n_1! n_2! \cdots n_k!}.$$

of Definition 7.3 appear in the following settings:

- Distinct selections of n items of which n_1 are alike, n_2 are alike, ..., n_k are alike. Example: different rearrangements of the word "BANANA".
- They are coefficients in the expansion of $\left(x_1 + x_2 + \cdots x_k\right)^n$.
- Selections of n distinct items into k distinct bins of fixed sizes $n_1, \ldots, n_k$, disregarding order within each bin. That is the WMS definition in their Theorem 2.3 of Ch.02.6. $\square$

8 More on Probability

This chapter corresponds to material found in WMS ch.2

8.1 Total Probability and Bayes Formula

Theorem 8.1 (Total Probability and Bayes Formula [11])**.** *Assume that $\{B_1, B_2, \dots\}$ is a partition of Ω and that $A \subseteq \Omega$. such that $\mathbb{P}(B_j) > 0$ for all j. Then*

$$\mathbb{P}(A) = \sum_{j=1}^{\infty} \mathbb{P}(A \mid B_j)\, \mathbb{P}(B_j)\,. \tag{8.1}$$

$$\mathbb{P}(B_j \mid A) = \frac{\mathbb{P}(A \mid B_j)\mathbb{P}(B_j)}{\sum\limits_{i=1}^{\infty} \mathbb{P}(A \mid B_i)\, \mathbb{P}(B_i)}\,. \qquad \textit{\textbf{(Bayes formula)}} \tag{8.2}$$

Note that the above also covers finite partitions $\{B_1, B_2, \dots, B_k\}$ of Ω: apply the formulas with

$$B_{k+1} := B_{k+2} := \dots := \emptyset\,.$$

8.2 Sampling and Urn Models With and Without Replacement

Definition 8.1.

(a) We call the action of picking n items $x_1, x_2, \dots, x_n$ from a collection of N items a **sampling action of size** n. Aternatively, we also use the phrases **sampling process** and **sampling procedure**. Here, $n \in \mathbb{N}$ and $N \in \mathbb{N}$ or $N = \infty$.

(b) We call the specific outcome of such a sampling action (the list $x_1, x_2, \dots, x_n$) a **realization** of that sampling action. □

(c) In yet another instance of notational abuse, both the sampling action and an outcome of this action (a realization) will be referred to as a **sample** of size n if this does not lead to any confusion. Note that we had mentioned this previously in Example **??** on p.**??**. □

Definition 8.2 (Sampling as a Random element)**.** Let $(\Omega, \mathbb{P})$ be a probability space. Let $U \neq \emptyset$ be a collection of N items ($N \in \mathbb{N}$ or $N = \infty$), which we can think of as the "population of interest". Let $n \in \mathbb{N}$ (so $n < \infty$), such that $n \leq N$.

(a) Let $\vec{X} : (\Omega, \mathbb{P}) \longrightarrow U^n$ be a random element with codomain U^n. If we interpret $\vec{X}$ as the action of picking n items

$$\vec{x} = x_1, x_2, \dots, x_n = \vec{X(\omega)} = X_1(\omega), X_2(\omega), \dots, X_n(\omega)$$

from U, then we call $\vec{X}$ a **sampling action of size** n. Aternatively, we also use the phrases **sampling process** and **sampling procedure**.

(b) We call a specific outcome (the list $\vec{x} = (x_1, x_2, \dots, x_n)$) a **realization** of that sampling action. See Example **??** on p.**??**.

(c) Both the sampling action and an outcome of this action (a realization) are called a **sample** of size n if the context makes it clear what is being discussed.

(d) If there is a specific $\vec{x}^* \in U^n$ such that $\mathbb{P}\{\vec{X} = \vec{x}^*\} = 1$, (this certainly is the case if $\vec{X}(\omega) = \vec{x}^*$ for all $\omega \in \Omega$), then we call both the sampling action $\vec{X}$ and the realization $\vec{x}^*$ a **deterministic sample**. $\square$

Definition 8.3 (Simple Random Sample)**.**

(a) We call a sampling action of size n ($n \in \mathbb{N}$) from a population of size $N < \infty$ a **simple random sampling action**, in brief, an **SRS action**, if there are no duplicates allowed (i.e., we sample without replacement) and each of the potential outcomes has equal chance of being selected.

(b) As in Definition 8.2 (Sampling as a Random element), we call both an SRS action and a realization of this action a **simple random sample of size** n. (Briefly, an **SRS**.) $\square$

Definition 8.4 (Random Sample)**.**

(a) We call a sampling action of size n ($n \in \mathbb{N}$) from a population of size $N < \infty$ a **random sampling action**, if the picks are independent of each other. See Chapter 5.4 (Independence of Random Elements) for the definition of independent random elements.

(b) As in Definition 8.2 (Sampling as a Random element), we call both a random sampling action and a realization of this action a **random sample of size** n. $\square$

Definition 8.5 (Urn models)**.**

(a) An **urn model without replacement** describes a mechanism by which a blindfolded person selects a fixed number of balls from an urn in which the balls have been well mixed. Note that the resulting realizations will contain no duplicates.

(b) An **urn model with replacement** describes a mechanism by which a blindfolded person selects a fixed number of balls from an urn as follows.
(1) The balls are well mixed.
(2) A ball is picked and the outcome is recorded.
(3) The ball is put back into the urn.
(4) Steps **(1)** through **(3)** are repeated until all n balls have been selected. $\square$

9 Discrete Random Variables and Random Elements

This chapter corresponds to material found in WMS ch.3

Assumption 9.1 (All series are absolutely convergent)**.** We assume the following for the entire remainder of these lecture notes.

- Unless explicitly stated otherwise, all sequences are either known to be absolutely convergent or assumed to be absolutely convergent.

In particular, if $p_X(x)$ is the probability mass function of a discrete random element X which takes values in a set Ω', $g : \Omega' \to \mathbb{R}$ is a real–valued function, and ω'_n is a sequence in Ω', then we assume that the series $\sum g(\omega'_n)p_X(\omega'_n)$ is absolutely convergent. $\square$

9.1 Probability Mass Function and Expectation

Proposition 9.1. *A real–valued function of a random element is a random variable.*

Definition 9.1 (Probability mass function)**.** For a discrete random element X on $(\Omega, \mathbb{P})$, define

$$p(x) \ := p_X(x) \ := \ \mathbb{P}_X\{x\} \ = \ \mathbb{P}\{X = x\}\,. \tag{9.1}$$

We call p_X the **probability mass function** (WMS: **probability function**) for X. We also write **PMF** for probability mass function. $\square$

Theorem 9.1. *If p_X is the probability mass function of a discrete random element X, then*

$$0 \ \le p_X(x) \ \le 1; \quad \textit{for all } x \tag{9.2}$$

$$\sum_{x \text{ s.t. } p_X(x)>0} p_X(x) \ = \ 1 \tag{9.3}$$

Definition 9.2 (Expected value of a discrete random variable)**.** Let Y be a discrete random variable with probability mass function $p_Y(y)$. Then

$$\mathbb{E}[Y] \ := \ \sum_y y\, p_Y(y) \ = \ \sum_y y\, \mathbb{P}\{Y = y\}\,,$$

is called the **expected value**, also **expectation** or **mean** of Y. $\square$

Proposition 9.2. $\boxed{\star}$ *Let $A_1, A_2, \dots, A_n$ a list of mutually disjoint events in a probability space $(\Omega, \mathbb{P})$. Let $y_1, y_2, \dots, y_n \in \mathbb{R}$. Then*

$$\mathbb{E}\left[\sum_{j=1}^{n} y_j \mathbf{1}_{A_j}\right] = \sum_{j=1}^{n} y_j \mathbb{P}(A_j)\,. \tag{9.4}$$

Theorem 9.2. *Let Y be a discrete random variable and $g : \mathbb{R} \to \mathbb{R};\ y \mapsto g(y)$ be a real-valued function. Then the random variable $g \circ Y : \omega \mapsto g\big(Y(\omega)\big)$ has the following expected value:*

$$\mathbb{E}[g(Y)] = \sum_{\text{all } y} g(y)\, p_Y(y) = \sum_{\text{all } y} g(y)\, \mathbb{P}\{Y = y\}\,. \tag{9.5}$$

Theorem 9.3. *Let $c \in \mathbb{R}$, Y be a discrete random variable and $g_1, g_2, g_n : \mathbb{R} \to \mathbb{R}$ be a list of n real-valued functions. Then*

$$\mathbb{E}[c] = c \qquad \textit{and} \qquad \mathbb{E}[cY] = c\mathbb{E}[Y]\,, \tag{9.6}$$

$$\mathbb{E}[cg_j(Y)] = c\mathbb{E}[g_j(Y)]\,. \tag{9.7}$$

Further, the random variable

$$\sum_{j=1}^{n} g_j \circ Y : \Omega \longrightarrow \mathbb{R}; \qquad \omega \mapsto \sum_{j=1}^{n} g_j\big(Y(\omega)\big)$$

has the following expected value:

$$\mathbb{E}\left[\sum_{j=1}^{n} g_j \circ Y\right] = \sum_{j=1}^{n} \mathbb{E}[g_j \circ Y]\,. \tag{9.8}$$

Theorem 9.4. *Let $Y_1, Y_2, \dots, Y_n : \Omega \to \mathbb{R}$ be discrete random variables which all are defined on the same probability space $(\Omega, \mathbb{P})$ ($n \in \mathbb{N}$). Then the random variable*

$$\sum_{j=1}^{n} Y_j : \Omega \longrightarrow \mathbb{R}; \qquad \omega \mapsto \sum_{j=1}^{n} Y_j(\omega)$$

has the following expected value:

$$\mathbb{E}\left[\sum_{j=1}^{n} Y_j\right] = \sum_{j=1}^{n} \mathbb{E}[Y_j]\,. \tag{9.9}$$

In other words, the expectation of the sum is the sum of the expectations.

Definition 9.3 (Variance and standard deviation of a random variable)**.** Y be a random variable. The **variance** of Y is defined as the expected value of $(Y - \mathbb{E}[Y])^2$. In other words,

$$Var[Y] := \sigma_Y^2 := \mathbb{E}\big[(Y - \mathbb{E}[Y])^2\big]\,. \tag{9.10}$$

We call $SD(Y) := \sigma_Y := \sqrt{Var[Y]}$ The **standard deviation** of Y. $\square$

Theorem 9.5. *If Y is a discrete random variable, then*

$$Var[Y] = \mathbb{E}\big[Y^2\big] - \big(\mathbb{E}[Y]\big)^2\,.$$

Theorem 9.6. *Let Y be a discrete random variable and $a, b \in \mathbb{R}$. Then*

$$Var\,[aY + b] = a^2 Var[Y]\,. \tag{9.11}$$

In other words, shifting a random variable by b, leaves its variance unchanged and multiplying it by a constant multiplies its variance by the square of that constant.

Theorem 9.7 (Bienaymé formula)**.** *Let $Y_1, Y_2, \dots, Y_n : \Omega \to \mathbb{R}$ be independent discrete random variables which all are defined on the same probability space $(\Omega, \mathbb{P})$ ($n \in \mathbb{N}$). Here we take the naive definition of independence: The outcomes of any Y_k are not influenced by the outcomes of the other Y_j. We will give a formulation of independence in terms of probabilities in a later chapter. Then*

$$Var\left[\sum_{j=1}^{n} Y_j\right] = \sum_{j=1}^{n} Var[Y_j]\,. \tag{9.12}$$

In other words, for independent random variables, the variance of the sum is the sum of the variances.

9.2 Bernoulli Variables and the Binomial Distribution

Definition 9.4 (Bernoulli trials and variables)**.** A **Bernoulli trial**. is a random element with only two outcomes, such as

⊡ S (success) or F (failure) ⊡ T (true) or F (false) ⊡ Y (Yes) or N (No) ⊡ 1 or 0

- We call $p := \mathbb{P}\{X = \text{success}\}$ the **success probability** and $q := 1 - p = \mathbb{P}\{X = \text{failure}\}$ the **failure probability** of the Bernoulli trial.
- If a Bernoulli trial X has numeric outcomes, then we call X a **Bernoulli variable**.
- If those outcomes are 1 and 0, we say that X is a **0–1 encoded Bernoulli trial**.
- A **Bernoulli sequence** is an iid sequence (Def. 5.17 on p.46) of Benoulli trials. □

Theorem 9.8 (Expected value and variance of a 0–1 encoded Bernoulli trial)**.** *Let X be a 0–1 encoded Bernoulli trial with $p := \mathbb{P}\{X = 1\}$. Then*

(9.13) $$\mathbb{E}[X] = p \qquad \textit{and} \qquad Var[X] = pq\,.$$

Definition 9.5 (Binomial Distribution)**.** Let $n \in \mathbb{N}$ and $0 \leq p \leq 1$. Let Y be a random variable with probability mass function

(9.14) $$p_Y(y) = \binom{n}{y} p^y q^{n-y}\,.$$

Then we say that Y has a **binomial distribution**. with parameters n and p or, in short, a **binom(n,p) distribution**. We also say that Y is binom(n,p). □

Theorem 9.9. *Let X_1, X_2, X_n be a Bernoulli sequence of size n with success probability p. Let Y be the number of successes in that sequence, i.e., $Y(\omega)$ = number of indices j such that $X_j(\omega) = S$.*

- *Then Y is binom(n,p).*

Theorem 9.10 (Expected value and variance of a binom(n,p) variable)**.** *Let Y be a binom(n,p) variable. Then*

(9.15) $$\mathbb{E}[Y] = np \qquad \textit{and} \qquad Var[Y] = npq\,.$$

9.3 Geometric + Negative Binomial + Hypergeometric Distributions

Definition 9.6 (Geometric distribution)**.** A random variable Y is said to have a **geometric distribution** with parameter $0 \leq p \leq 1$ or, in short, a **geom(p) distribution**, if its probability mass functions is as follows:

(9.16) $$p_Y(y) = q^{y-1}\,p\,, \qquad \text{for } y = 1, 2, 3, \dots\,.$$ □

Theorem 9.11. *Let $X_1, X_2, \cdots : (\Omega, \mathbb{P}) \to \{S, F\}$ be an infinite Bernoulli sequence with success probability $0 \le p \le 1$.*
Let $T(\Omega, \mathbb{P}) \to \mathbb{N}$ be the random variable

$$T(\omega) := \begin{cases} \text{smallest integer } k > 0 \text{ such that } X_k(\omega) = S & \text{if such a } k \text{ exists,} \\ \infty\,, & \text{else.} \end{cases}$$

- *Then T is geom(p).*

Theorem 9.12 (WMS Ch.03.5, Theorem 3.8)**.** *If Y is a geom(p) random variable, then*

$$\mathbb{E}[Y] = \frac{1}{p}\,, \qquad \text{and} \qquad Var[Y] = \frac{q}{p^2}\,.$$

Definition 9.7 (Negative binomial distribution)**.** $\boxed{\star}$ A random variable Y has a **negative binomial distribution** with parameters p and r if

$$p_Y(y) = \binom{y-1}{r-1} p^r q^{y-r}\,, \quad \text{where } r \in \mathbb{N},\ y = r, r+1, r+2, \dots,\ 0 \le p \le 1\,. \tag{9.17}$$ □

Theorem 9.13. *Let $X_1, X_2, \cdots : (\Omega, \mathbb{P}) \to \{S, F\}$ be an infinite Bernoulli sequence with success probability $0 \le p \le 1$.*
Let $t_1 < t_2 < \cdots$ be the subsequence of those indices at which a success happens. In other words,

$$X_n(\omega) = \begin{cases} S = \text{success} & \text{if } n \text{ is one of } t_1, t_2, \dots, \\ F = \text{failure}\,, & \text{else.} \end{cases}$$

Two points to note:

- *There will be different subsequences $t_1, t_2, \dots$ for different arguments $\omega \in \Omega$. In other words, we are dealing with a sequence of random variables(!)*

$$t_1 = T_1(\omega),\ t_2 = T_2(\omega),\ t_3 = T_3(\omega),\ \dots$$

- *It is possible that we are dealing with an ω for which there are only 18 successes in the entire (infinite) sequence $X_1(\omega), X_2(\omega), \dots$. In this case, we define $T_{19}(\omega) = T_{20}(\omega) = \cdots = \infty$. More generally, if $r \in \mathbb{N}$ and the sequence $X_1(\omega), X_2(\omega), \dots$ has less than r successes, we define*

$$T_r(\omega) := \infty\,.$$

Now that we have defined $T_r = T_r(\omega)$, we are ready to state the theorem.

- *The random variable T_r has a negative binomial distribution with parameters p and r.*

Theorem 9.14. $\boxed{\star}$

If the random variable Y is negative binomial with parameters p and r,

$$\mathbb{E}[Y] = \frac{r}{p} \qquad \textit{and} \qquad Var[Y] = \frac{r(1-p)}{p^2}\,.$$

Definition 9.8 (Hypergeometric distribution)**.** A random variable Y has a **hypergeometric distribution** with parameters N, R and n if its PMF is

$$p_Y(y) = \frac{\binom{R}{y}\binom{N-R}{n-y}}{\binom{N}{n}}\,, \tag{9.18}$$

where the nonnegative integers N, R, n and y are subject to the following conditions:

- $y \leq n$ • $y \leq R$ • $n - y \leq N - R$ $\square$

Theorem 9.15.

- *Given is an urn wich contains N well–mixed balls of two colors, Red and Black. We assume that R are Red and thus, the remaining $N-R$ are Black.*
- *A sample of size n is drawn without replacement from that urn, according to Definition 8.5**(a)**.*

Let the random variable Y denote the number of Red balls in that sample. Then Y is hypergeometric with parameters N, R and n. In other words, its PMF is

$$p_Y(y) = \frac{\binom{R}{y}\binom{N-R}{n-y}}{\binom{N}{n}}\,.$$

Theorem 9.16 (WMS Ch.03.7, Theorem 3.10)**.** $\boxed{\star}$

Let Y be a hypergeometric random variable with parameters N, R and n. Then

$$\mathbb{E}[Y] = \frac{nR}{N} \qquad \textit{and} \qquad Var[Y] = n\left(\frac{R}{N}\right)\left(\frac{N-R}{N}\right)\left(\frac{N-n}{N-1}\right)\,. \tag{9.19}$$

9.4 The Poisson Distribution

Proposition 9.3. *Let $\lambda > 0$. Then the function*

$$p(y) \;:=\; e^{-\lambda}\,\frac{\lambda^y}{y!}$$

defines a probability mass function on $[0,\infty[_{\mathbb{Z}} = \{0,1,2,\dots\}$.

Definition 9.9 (Poisson variable)**.** Let Y be a random variable and $\lambda > 0$. We say that Y has a **Poisson probability distribution** with parameter λ, in short, Y is **poisson(λ)**, if its probability mass function is

$$p_Y(y) \;=\; \frac{\lambda^y}{y!}\,e^{-\lambda}\,, \qquad \text{for } y = 0,1,2,\dots, \quad \square$$

Proposition 9.4. *Given is some event of interest, E.*

(1) *We define a random variable Y which counts how often E happen in a "unit". We leave it open whether this unit is a time interval (maybe a minute or a year) or a subset of d–dimensional space ($d = 1,2,3$). Let us write A for that unit.*

- *Example: Y is the number of car accidents that happen in Binghamton during a day (unit of time),*
- *Example: Y is the number of typos on a randomly picked page of these lecture notes ("page" is a twodimensional unit – square inches).*

(2) *Given $n \in \mathbb{N}$, we subdivide the unit (A) into n parts of equal size. Let*

$$\vec{X}^{(n)} \;:=\; \big(X_1^{(n)}, X_2^{(n)}, \dots, X_n^{(n)}\big)\,,$$

where $X_j^{(n)}$ = the number of times that E happens in subunit j.

- *Assume that for all big enough, FIXED n,*
  $\boxdot$ *the $X_j^{(n)}$ are independent*
  $\boxdot$ *for each j, $\mathbb{P}\{X_j^{(n)} = 0$ or $1\} = 1$: E (i.e., the event of interest) happens at most once in such a small subunit*
  $\boxdot$ $p_n := \mathbb{P}\{X_j^{(n)} = 1\}$ *is constant in j ($j = 1,2,\dots,n$)*
  $\boxdot$ $\lambda := n \cdot p_n$ *is constant in n: For large enough k, $kp_k = (k+1)p_{k+1} = (k+2)p_{k+2} = \dots = \lambda$.*

Given these assumptions, the following is true:

(a) *The random variable $Y^{(n)} \;:=\; X_1^{(n)} + X_2^{(n)} + \cdots, + X_n^{(n)}$ is binom(n, p_n) for large n.*

(b) *The binom(n, p_n) probability mass functions $p_{Y^{(n)}}$ converge to that of a poisson(λ) variable:*

$$\lim_{n\to\infty} p_{Y^{(n)}}(y) \;=\; \lim_{n\to\infty} \binom{n}{y} p_n^y (1-p_n)^{n-y} \;=\; e^{-\lambda}\cdot\frac{\lambda^y}{y!}\,, \qquad \textit{for } y = 0,1,2,\dots, \tag{9.20}$$

Theorem 9.17 (WMS Ch.03.8, Theorem 3.11)**.**

A poisson(λ) random variable has expectation and variance λ. In other words,

$$\mathbb{E}[Y] = Var[Y] = \lambda\,. \tag{9.21}$$

9.5 Moments, Central Moments and Moment Generating Functions

Definition 9.10 (kth Moment)**.** If Y is a random variable and $k \in \mathbb{N}$,

$$\mu'_k := \mathbb{E}[Y^k] \tag{9.22}$$

is called the kth **moment** of Y. μ'_k also is referred to as the kth **moment of Y about the origin**. ☐

Definition 9.11 (kth Central Moment)**.** If Y is a random variable and $k \in \mathbb{N}$,

$$\mu_k := \mathbb{E}[(Y - \mathbb{E}[Y])^k] = \mathbb{E}[(Y - \mu)^k] \tag{9.23}$$

is called the kth **central moment** of Y aka the kth **moment of Y about its mean**. ☐

Proposition 9.5 (The moments determine the distribution)**.** $\boxed{\star}$ *Under fairly slight assumptions the following is true for two random variables Y_1 and Y_2.*

$$\textit{If}\ \ \mathbb{E}[Y_1^k] = \mathbb{E}[Y_2^k] \quad \textit{for } k = 1, 2, 3, \dots, \quad \textit{then}\ \ \mathbb{P}_{Y_1} = \mathbb{P}_{Y_2}\,.$$

In other words, the distribution of a random variable is uniquely determined by its moments.

Definition 9.12 (Moment–generating function)**.** Let Y be a random variable for which one can find $\delta > 0$ (no matter how small), such that

$$m(t) := m_Y(t) := \mathbb{E}\big[e^{tY}\big] \qquad \text{is finite for } |t| < \delta\,. \tag{9.24}$$

Then we say that Y has **moment–generating function**, in short, **MGF**, $m_Y(t)$. ☐

Theorem 9.18. *Let Y be a random variable with MGF $m_Y(t)$ and $k \in \mathbb{N}$. Then its kth moment is obtained as the kth derivative of $m_Y(\cdot)$, evaluated at $t = 0$:*

$$\mu'_k = m^{(k)}(0) = \frac{d^k m(t)}{dt^k}\Big|_{t=0}\,. \tag{9.25}$$

Proposition 9.6. $\boxed{\star}$ *If Y is a poisson(λ) random variable ($\lambda > 0$), its MGF is*

(9.26) $$m_Y(t) = e^{\lambda(e^t - 1)}\,.$$

10 Continuous Random Variables

10.1 Cumulative Distribution Function of a Random Variable

Definition 10.1 (Cumulative Distribution Function)**.** Let Y denote any random variable (it need not be discrete). The **distribution function** of Y, also called its **cumulative distribution function** or **CDF (cumulative distribution function)**, is defined as follows.

(10.1) $$F(y) := F_Y(y) := \mathbb{P}\{Y \leq y\} \qquad \text{for } y \in \mathbb{R}\,. \ \square$$

Theorem 10.1 (Properties of a Cumulative Distribution Function)**.** *If $F_Y(y)$ is the cumulative distribution function of a random variable Y, then*

(1) $F_Y(-\infty) = \lim\limits_{y \to -\infty} \mathbb{P}(Y \leq y) = 0.$

(2) $F_Y(\infty) = \lim\limits_{y \to \infty} \mathbb{P}(Y \leq y) = 1.$

(3) *$F_Y(y)$ s a nondecreasing function of y. In other words, if $y_1 < y_2$, then $F_Y(y_1) \leq F_Y(y_2)$ See Definition 2.24 on p.15.*

(4) *$y \mapsto F_Y(y)$ is **right continuous** at all arguments y, i.e., $F(y) = F(y+)$ for all y.*

10.2 Continuous Random Variables and Probability Density Functions

Definition 10.2 (Continuous random variable)**.** We call a random variable Y with distribution function $F_Y(y)$ **continuous**, if $F_Y(y)$ is continuous, for all arguments y. $\square$

Proposition 10.1. *Let Y be a continuous random variable with CDF $F_Y(y)$. Then its distribution gives zero probability to all singletons $\{a\} (a \in \mathbb{R})$. Also, it gives the same probability to an interval with endpoints $-\infty < a < b < \infty$, regardless whether a and/or b do or do not belong to that interval. In other words,*

(10.2) $$a \in \mathbb{R} \Rightarrow \mathbb{P}\{Y = a\} = \mathbb{P}_Y\{a\} = 0\,,$$

(10.3) $$\begin{aligned} -\infty < a < b < \infty \Rightarrow \mathbb{P}\{a < Y < b\} &= \mathbb{P}\{a \leq Y < b\} \\ = \mathbb{P}\{a < Y \leq b\} &= \mathbb{P}\{a \leq Y \leq b\}\,. \end{aligned}$$

Assumption 10.1 (All continuous random variables have a differentiable CDF)**.** Unless explicitly stated otherwise, all continuous random variables are assumed to satisfy the following:

The first derivative $\frac{dF_Y}{dy}$ of F_Y exists and is continuous except for, at most, a finite number of points in any finite interval.

All cumulative distribution functions for continuous random variables that we deal with in this course satisfy this assumtion. $\square$

Definition 10.3 (Probability density function)**.** Let Y be a continuous random variable with CDF $F_Y(y)$. For all arguments y where the derivative $F_Y'(y) = \frac{dF_Y(y)}{dy}$ exists, we define

$$f(y) \;:=\; f_Y(y) \;:=\; \frac{dF_Y(y)}{dy}\,.$$

We call f_Y the **probability density function** or, in short, the **PDF** of the continuous random variable Y. $\square$

Theorem 10.2. *Let Y be a continuous random variable with CDF $F_Y(y)$ and PDF $f_Y(y)$.*

(1) *If $a, b \in \mathbb{R}$ and $a < b$, then*

(10.4) $$\mathbb{P}\{a < Y \leq b\} \;=\; F_Y(b) - F_Y(a) \;=\; \int_a^b f(y)dy\,.$$

(2) $f_Y(y) \;\geq\; 0$ *for* $-\infty < y < \infty$.

(3) $\int\limits_{\infty}^{\infty} f_Y(y)dy \;=\; 1.$

Theorem 10.3. *Let $\psi : \mathbb{R} \to \mathbb{R}$ satisfy the following:*

(1) *ψ is integrable: $\int\limits_a^b \psi(x)dx$ exists for $a < b$.*

(2) $\psi(x) \;\geq\; 0$ *for* $-\infty < x < \infty$.

(3) $\int\limits_{\infty}^{\infty} \psi(x)dx \;=\; 1.$

- *Then, $Q\{a < Y \leq b\} \;:=\; \int_a^b \psi(x)dx$ defines a probability measure Q on Ω.*

Definition 10.4 (pth quantile)**.** Let Y denote any random variable and $0 < p < 1$. Let ϕ_p be the number derived in the previous remark, i.e.,

(10.5) $$\phi_p \;=\; \min\{\alpha \in \mathbb{R} : F_Y(\alpha) \geq p\}$$

We call ϕ_p the pth **quantile** and also the $100p$th **percentile** of Y.

Moreover, we call $\phi_{0.25}$ the **first quartile**, $\phi_{0.5}$ the **median**, and $\phi_{0.75}$ the **third quartile**, of the random variable Y. $\square$

Proposition 10.2. *Let Y be a continuous random variable with CDF $F_Y(y)$. Then*

$$\phi_p \;=\; \min\{\alpha \in \mathbb{R} : F_Y(\alpha) = p\}\,. \tag{10.6}$$

Proposition 10.3. *Let Y be a random variable with an injective CDF $F_Y(y)$. (Note that it is not assumed that F_Y is continuous.) Then*

$$\phi\big(F_Y(y)\big) \;=\; y \qquad \text{for all } y \in \mathbb{R} \tag{10.7}$$

★ *Note that (10.7) states that ϕ is a left inverse of the injective function F_Y.*

Proposition 10.4. *Let Y be a random variable with a bijective CDF $F_Y : \mathbb{R} \xrightarrow{\sim}]0,1[$. Then $F_Y(y)$ and $\phi(p)$ are inverse to each other, i.e.,*

$$\begin{aligned} \phi\big(F_Y(y)\big) &= y\,, &\qquad &\text{for all } y \in \mathbb{R}\,,\\ F_Y\big(\phi(p)\big) &= p\,, &\qquad &\text{for all } 0 < p < 1\,. \end{aligned} \tag{10.8}$$

10.3 Expected Value, Variance and MGF of a Continuous Random Variable

Assumption 10.2 (All continuous random variables have Expectations)**. A.** Unless explicitly stated otherwise, all continuous random variables are assumed to to possess a probability density function $f_Y(y)$ that satisfies

$$\int_{-\infty}^{\infty} |y| f(y)\,dy| < \;\infty\,.$$

This technical condition guarantees the existence of $\int\limits_{-\infty}^{\infty} y f(y) dy$ which is needed to define the expected value of Y.

B. We further assume that, unless specifically stated otherwise, there is a common probability space $(\Omega, \mathbb{P})$ for all random variables. In other words, all random variables Y, be they discrete, continuous or neither, are of the form $Y : (\Omega, \mathbb{P}) \to \mathbb{R}$. $\square$

Definition 10.5 (Expected value of a continuous random variable)**.** Let Y be a continuous random variable with PDF $f_Y(y)$. We call

$$E(Y) := \int_{-\infty}^{\infty} y f_Y(y)\,dy \tag{10.9}$$

the **expected value**, also **expectation** or **mean** of Y. $\square$

Theorem 10.4. $\boxed{\star}$

Let Y be a continuous random variable with CDF F_Y and PDF f_Y.
Then

$$\mathbb{E}[Y] = \int_0^{\infty} \big(1 - F_Y(y)\big)dy - \int_0^{\infty} F_Y(-y)dy \tag{10.10}$$

$$= \int_0^{\infty} \mathbb{P}\{Y > y\}dy - \int_0^{\infty} \mathbb{P}\{Y \leq -y\}dy\,. \tag{10.11}$$

Corollary 10.1. $\boxed{\star}$

Let Y be a nonnegative, continuous random variable with CDF F_Y and PDF f_Y. Then

$$\mathbb{E}[Y] = \int_0^{\infty} \big(1 - F_Y(y)\big)dy = \int_0^{\infty} \mathbb{P}\{Y > y\}dy\,. \tag{10.12}$$

Theorem 10.5. *Let Y be a continuous random variable with PDF f_Y and $g : \mathbb{R} \to \mathbb{R};\ y \mapsto g(y)$ be a real-valued function. Then the random variable $g \circ Y : \omega \mapsto g\big(Y(\omega)\big)$ has expectation*

$$\mathbb{E}[g(Y)] = \int_{\infty}^{\infty} g(y) f_Y(y)\,dy\,. \tag{10.13}$$

Theorem 10.6. *Let $c \in \mathbb{R}$, Y be a discrete or continuous random variable and $g_1, g_2, g_n : \mathbb{R} \to \mathbb{R};\ y \mapsto g(y)$ be a list of n real-valued functions. Then*

$$\mathbb{E}[c] = c\,, \tag{10.14}$$

$$\mathbb{E}[cg_j(Y)] = c\mathbb{E}[g_j(Y)]\,. \tag{10.15}$$

Further, the random variable

$$\sum_{j=1}^{n} g_j \circ Y : \Omega \longrightarrow \mathbb{R}; \qquad \omega \mapsto \sum_{j=1}^{n} g_j\big(Y(\omega)\big)$$

has the following expected value:

$$\mathbb{E}\left[\sum_{j=1}^{n} g_j \circ Y\right] = \sum_{j=1}^{n} \mathbb{E}[g_j \circ Y]\,. \tag{10.16}$$

Definition 10.6. $\boxed{\star}$ If $Y_1, Y_2, \dots, Y_m$ is a list of discrete random variables and $Y'_1, Y'_2, \dots, Y'_n$ is a list of continuous random variables, all of which are defined on the same probability space $(\Omega, \mathbb{P})$, then we define

$$\mathbb{E}\left[\sum_{i=1}^{m} Y_i + \sum_{j=1}^{n} Y'_j\right] := \sum_{i=1}^{m} \mathbb{E}[Y_i] + \sum_{j=1}^{n} \mathbb{E}[Y'_j]\,p\,. \ \square \tag{10.17}$$

Theorem 10.7. *Let $Y_1, Y_2, \dots, Y_n : \Omega \to \mathbb{R}$ be random variables. (which all are defined on the same probability space $(\Omega, \mathbb{P})$ ($n \in \mathbb{N}$ by Assumption 10.2.**B**). Some may be continuous, others may be discrete. Then the random variable*

$$\sum_{j=1}^{n} Y_j : \Omega \longrightarrow \mathbb{R}; \qquad \omega \mapsto \sum_{j=1}^{n} Y_j(\omega)$$

has the following expected value:

$$\mathbb{E}\left[\sum_{j=1}^{n} Y_j\right] = \sum_{j=1}^{n} \mathbb{E}[Y_j]\,. \tag{10.18}$$

In other words, the expectation of the sum is the sum of the expectations.

Theorem 10.8. *Let Y be a discrete or continuous random variable. Let $Y_1, Y_2, \dots, Y_n : \Omega \to \mathbb{R}$ be independent random variables (which all are defined on the same probability space $(\Omega, \mathbb{P})$ ($n \in \mathbb{N}$ by Assumption 10.2.**B**). Some may be continuous, others may be discrete. Further, let $a, b \in \mathbb{R}$. Then*

$$Var[Y] = \mathbb{E}\big[Y^2\big] - \big(\mathbb{E}[Y]\big)^2\,, \tag{10.19}$$

$$Var\,[aY + b] = a^2 Var[Y]\,, \tag{10.20}$$

$$Var\left[\sum_{j=1}^{n} Y_j\right] = \sum_{j=1}^{n} Var[Y_j]\,. \tag{10.21}$$

Remark 10.1. Note that independence of $Y_1, \dots, Y_n$ is required for the validity of (10.21)! $\square$

Unless something different is stated, Y is a random variable $Y : (\Omega, \mathbb{P}) \to \mathbb{R}$ on some probability space $(\Omega, \mathbb{P})$. Further, $\mu = \mathbb{E}[Y]$, $\sigma^2 = Var[Y]$ and $\sigma = \sqrt{Var[Y]}$ denote expectation, variance and standard deviation of Y.

Definition 10.7. For $k \in \mathbb{N}$, we define

$$\mu'_k := \mathbb{E}[Y^k] \qquad (k\text{th } \textbf{moment of } Y \textbf{ about the origin}) \tag{10.22}$$

$$\mu_k := \mathbb{E}[(Y - \mathbb{E}[Y])^k] = \mathbb{E}[(Y-\mu)^k] \qquad (k\text{th } \textbf{central moment of } Y) \tag{10.23}$$

$$m(t) := m_Y(t) := \mathbb{E}\left[e^{tY}\right] \qquad (\textbf{moment–generating function of } Y) \tag{10.24}$$

As in the discrete case we assume that the expectations defining μ'_k and μ_k exist and that there is some $\delta > 0$ such that $m_Y(t)$ is defined (i.e., finite) for $|t| < \delta$. $\square$

Theorem 10.9. *Let Y be a random variable with MGF $m_Y(t)$ and $k \in \mathbb{N}$. Then its kth moment is obtained as the kth derivative of $m_Y(\cdot)$, evaluated at $t = 0$:*

$$\mu'_k = m^{(k)}(0) = \frac{d^k m(t)}{dt^k}\Big|_{t=0}. \tag{10.25}$$

Proposition 10.5. *Let Y be a random variable with MGF $m_Y(t)$. Let $a, b \in \mathbb{R}$, $Y' := Y + a$, $Y'' := bY$. Then*

$$m_{Y'}(t) = e^{ta} m_Y(t)\,, \tag{10.26}$$

$$m_{Y''}(t) = m_Y(bt)\,. \tag{10.27}$$

10.4 The Uniform Probability Distribution

Definition 10.8 (Continuous, uniform random variable)**.** Let Y be a random variable and $-\infty < \theta_1 < \theta_2 < \infty$. We say that Y has a **continuous uniform probability distribution** with parameters θ_1 and θ_2 — also, that Y **is uniform on** $[\theta_1, \theta_2]$ or $Y \sim$ **uniform(**θ_1, θ_2**)** — if Y has probability density function

$$f_Y(y) = \begin{cases} \dfrac{1}{\theta_2 - \theta_1}\,, & \text{if } \theta_1 \le y \le \theta_2, \\ 0\,, & \text{else.} \quad \square \end{cases} \tag{10.28}$$

Theorem 10.10 (WMS Ch.04.4, Theorem 4.6). *If $\theta_1 < \theta_2$ and Y is a uniform random variable with parameters θ_1, θ_2, then*

$$\mathbb{E}[Y] = \frac{\theta_1 + \theta_2}{2} \qquad \text{and} \qquad Var[Y] = \frac{(\theta_2 - \theta_1)^2}{12}.$$

Theorem 10.11. *Assume that Y is a continuous random variable with CDF $F_Y(y)$. Let $U := F_Y(Y)$. Then $U \sim$ uniform$(0,1)$.*

Theorem 10.12. *Given are a uniform$(0,1)$ random variable U and a continuous function $F : \mathbb{R} \to [0,1]$ that satisfies the conditions of Theorem 10.1 (Properties of a Cumulative Distribution Function) on p.73:*

- F *is nondecreasing* • $F(-\infty) := \lim_{y\to-\infty} F(y) = 0$ • $F(\infty) := \lim_{y\to\infty} F(y) = 1$

(10.29) $$\text{Let} \quad G : [0,1] \to \mathbb{R}; \quad p \mapsto G(p) := \min\{y \in \mathbb{R} : F(y) \geq p\}.$$

Let $Z := G(U)$ be the random variable $\omega \mapsto Z(\omega) := G\big(U(\omega)\big)$.
Then its CDF matches F. In other words, $F_Z(y) = F(y)$ for all $y \in \mathbb{R}$.

10.5 The Normal Probability Distribution

Definition 10.9 (Normal random variable). Let $\sigma > 0$ and $-\infty < \mu < \infty$. We say that a random variable Y has a **normal probability distribution** with mean μ and variance σ^2 if its probability density function is

(10.30) $$f_Y(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(y-\mu)^2/(2\sigma^2)}, \qquad (y \in \mathbb{R}). \ \square$$

We also express that by saying that Y is $\mathcal{N}(\mu, \sigma^2)$. Moreover, we call Y **standard normal** if Y is $\mathcal{N}(0,1)$.

Proposition 10.6. *Let the random variable Y be $\mathcal{N}(\mu, \sigma^2)$. Then*

(10.31) $$m_Y(t) = e^{\mu t + (\sigma^2 t^2)/2}.$$

Theorem 10.13 (WMS Ch.04.5, Theorem 4.7). *If Y is a normally distributed random variable with parameters μ and σ, then*

$$\mathbb{E}[Y] = \mu \quad \textit{and} \quad Var[Y] = \sigma^2 .$$

10.6 The Gamma Distribution

Definition 10.10 (Gamma random variable). Let $\sigma > 0$ and $-\infty < \mu < \infty$. We say that a random variable Y has a **gamma distribution** with **shape parameter** $\alpha > 0$ and **scale parameter** $\beta > 0$ if its probability density function is

$$f_Y(y) = \begin{cases} \dfrac{y^{\alpha-1}e^{-y/\beta}}{\beta^\alpha\Gamma(\alpha)}, & \text{if } 0 \le y < \infty, \\ 0, & \text{else}, \end{cases} \tag{10.32}$$

where $\Gamma(\alpha)$ is the **gamma function**

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1}e^{-y}\,dy . \tag{10.33}$$

We also express that by saying that Y is gamma(α, β). □

Proposition 10.7. *The gamma function satisfies the following:*

$$\Gamma(1) = 1 , \tag{10.34}$$
$$\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1) \qquad \textit{for all } \alpha > 1 , \tag{10.35}$$
$$\Gamma(n) = (n-1)! \qquad \textit{for all } n \in \mathbb{N} . \tag{10.36}$$

Proposition 10.8. *If the random variable Y is gamma(α, β) it has MGF*

$$m_Y(t) = \frac{1}{(1-\beta t)^\alpha} \qquad \textit{for } t < \frac{1}{\beta} . \tag{10.37}$$

Theorem 10.14 (WMS Ch.04.6, Theorem 4.8). *Let the random variable Y be gamma(α, β) with $\alpha, \beta > 0$. Then*

$$\mathbb{E}[Y] = \alpha\beta \quad \textit{and} \quad Var[Y] = \alpha\beta^2 .$$

Definition 10.11 (Chi–square distribution)**.** Let $\nu \in \mathbb{N}$. We say that a random variable Y has a **chi–square distribution** with ν **degrees of freedom**, in short, Y is **chi–square with** ν **df** , or $Y \sim \chi^2$(df=ν), or Y is **chi–square(**ν**)**, or Y is $\chi^2(\nu)$, if Y is gamma$(\nu/2, 2)$. In other words, Y must have a gamma distribution with shape parameter $\nu/2$ and scale parameter 2. □

Theorem 10.15 (WMS Ch.04.6, Theorem 4.9)**.** *A chi–square random variable Y with ν degrees of freedom has expectation and variance*

$$\mathbb{E}[Y] \,=\, \nu \qquad \textit{and} \qquad Var[Y] \,=\, 2\nu.$$

Definition 10.12 (Exponential distribution)**.** We say that a random variable Y has an **exponential distribution** with parameter $\beta > 0$, in short, Y is **expon(**β**)**, if $Y \sim$ gamma$(1, \beta)$; in other words, if Y has density

$$f_Y(y) \,=\, \begin{cases} \dfrac{1}{\beta}\, e^{-y/\beta}\,, & \text{for } 0 \leq y < \infty\,, \\ 0\,, & \text{else}\,. \end{cases} \tag{10.38}$$
□

Proposition 10.9. *Let Y be an exponential random variable with parameter β and $y \geq 0$. Then,*

$$\mathbb{P}\{Y > y\} \,=\, e^{-y/\beta}\,. \qquad \textit{Thus,} \quad F_Y(y) = 1 \,-\, e^{-y/\beta}\,.$$

Theorem 10.16. *An exponential random variable Y with parameter β has expectation and variance*

$$\mathbb{E}[Y] \,=\, \beta \qquad \textit{and} \qquad Var[Y] \,=\, \beta^2.$$

Proposition 10.10 (Memorylessness of the exponential distribution)**.** *Let Y be an exponential random variable. Let $t > 0$ and $h > 0$. Then*

$$\mathbb{P}\{Y \,>\, t + h \mid Y > t\} \,=\, \mathbb{P}\{Y \,>\, h\}\,. \tag{10.39}$$

Remark 10.2. The property (10.39) of an exponential distribution is referred to as the **memoryless property** of the exponential distribution. It also occurs in the geometric distribution. □

10.7 The Beta Distribution

Definition 10.13 (Beta distribution). $\boxed{\star}$ A random variable Y has a **beta probability distribution** with parameters $\alpha > 0$ and $\beta > 0$ if it has density function

$$f_Y(y) \;=\; \begin{cases} \dfrac{y^{\alpha-1}(1-y)^{\beta-1}}{B(\alpha,\beta)}\,, & \text{if } 0 \le y \le 1\,, \\ 0\,, & \text{else}\,, \end{cases} \tag{10.40}$$

where

$$B(\alpha,\beta) \;=\; \int_0^1 y^{\alpha-1}(1-y)^{\beta-1}\,dy \;=\; \frac{\Gamma(\alpha)\,\Gamma(\beta)}{\Gamma(\alpha+\beta)}\,. \tag{10.41}$$

We also express that by saying that Y is beta(α,β). $\square$

Theorem 10.17. $\boxed{\star}$ *If Y is a beta–distributed random variable with parameters $\alpha > 0$ and $\beta > 0$, then*

$$\mathbb{E}[Y] \;=\; \frac{\alpha}{\alpha+\beta} \qquad \text{and} \qquad Var[Y] \;=\; \frac{\alpha\,\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

10.8 Inequalities for Probabililities

Theorem 10.18. $\boxed{\star}$ *Let Y, Z be continuous or discrete random variables and $a > 0$. Assume further that $Y \ge 0$. Then*

$$\mathbb{P}\{Y \ge a\}) \;\le\; \frac{\mathbb{E}[Y]}{a}\,, \tag{10.42}$$

$$\mathbb{P}\{|Z| \ge a\}) \;\le\; \frac{\mathbb{E}[\,|Z|^n]}{a^n}\,. \tag{10.43}$$

(10.42) *is known as the **Markov inequality***

Theorem 10.19 (Tchebysheff Inequalities). *Let Y be a random variable with mean $\mu = \mathbb{E}[Y]$ and standard deviation σ. Let $k > 0$. Then*

$$\mathbb{P}\{|Y-\mu| \;\ge\; k\sigma\} \;\le\; \frac{1}{k^2}\,, \tag{10.44}$$

$$\mathbb{P}\{|Y-\mu| \;<\; k\sigma\} \;\ge\; 1 \;-\; \frac{1}{k^2}\,. \tag{10.45}$$

Both (10.44) and (10.45) are known as the ***Tchebysheff inequalities***

10.9 Mixed Random Variables

Definition 10.14 (Mixed random variables)**.** Let Y be a random variable on a probability space $(\Omega, \mathfrak{F}, \mathbb{P})$ as follows. There are a finite or infinite sequence $y_1 < y_2 < \cdots$ of real numbers and a function $f : \mathbb{R} \to [0, \infty[$ such that, for all Borel sets B of $\mathbb{R}$, its distribution $\mathbb{P}_Y$ satisfies

$$\begin{aligned} &\mathbb{P}_Y(B) = \mu_d(B) + \mu_c(B), \quad \text{where} \\ &\mu_d(B) = \sum_{y_j \in B} \mathbb{P}\{Y = y_j\}, \quad \text{and} \quad \mu_c(B) = \int_B f(t)\, dt. \end{aligned} \tag{10.46}$$

We say that Y is a **mixed random variable** and $\mathbb{P}_Y$ is a **mixed distribution**. We call μ_d the **PMF part** and μ_c the **PDF part** of both Y and $\mathbb{P}_Y$. $\square$

Proposition 10.11. $\boxed{\star}$ *Let μ, μ_1, μ_2 be measures on the Borel sets of $\mathbb{R}$ such that*

$$\mu = \mu_1 + \mu_2, \quad \textit{i.e.,} \quad \mu(B) = \mu_1(B) + \mu_2(B), \quad \textit{for all } B \in \mathfrak{B}^1.$$

Further, let $g : \mathbb{R} \to \mathbb{R}$ and $A \in \mathfrak{B}^1$. Then

$$\int_A g\, d\mu = \int_A g\, d\mu_1 + \int_A g\, d\mu_2. \tag{10.47}$$

Theorem 10.20 (Expectation of mixed random variables)**.** *Let Y be a mixed random variable on a probability space $(\Omega, \mathfrak{F}, \mathbb{P})$. Let $y_1 < y_2 < \cdots$ and $f : \mathbb{R} \to [0, \infty[$ be such that $\mu_d(B) = \sum_{y_j \in B} \mathbb{P}\{Y = y_j\}$ is the PMF part of Y and $\mu_c(B) = \int_B f(t)\, dt$ is the PDF part of Y. Then the expectation $\mathbb{E}[g \circ Y]$ for a function $g : \mathbb{R} \to \mathbb{R}$ is*

$$\mathbb{E}[g \circ Y] = \sum_{y_j \in \mathbb{R}} g(y_j) \cdot \mathbb{P}\{Y = y_j\} + \int_{-\infty}^{\infty} g(y) f(y)\, dy. \tag{10.48}$$

We assume again as we did in Definition 10.14 (Mixed random variables) that the y_j form a finite or infinite list of real numbers.

11 Multivariate Probability Distributions

11.1 Multivariate CDFs, PMFs and PDFs

Assumption 11.1 (Comma separation denotes intersection)**.** Separating commas are to be interpreted as "**and**" and not as "**or**". Thus, for example,

$$\begin{aligned}\{X \in B,\, Y = \alpha,\, 5 \leq Z < 8\} &= \{X \in B \textbf{ and } Y = \alpha \textbf{ and } 5 \leq Z < 8\} \\ &= \{X \in B\} \cap \{Y = \alpha\} \cap \{5 \leq Z < 8\}\,. \ \square\end{aligned}$$

Definition 11.1 (Joint cumulative distribution function)**.** Given are two random variables Y_1 and Y_2. No assumption is made whether they are discrete or continuous. We call

$$F(y_1, y_2) := F_{Y_1,Y_2}(y_1, y_2) := \mathbb{P}(Y_1 \leq y_1, Y_2 \leq y_2)\,, \qquad \text{where } y_1, y_2 \in \mathbb{R}\,, \tag{11.1}$$

the **joint cumulative distribution function** or **bivariate cumulative distribution function** or **joint CDF** or **joint distribution function** of Y_1 and Y_2. $\square$

Theorem 11.1. *Let Y_1 and Y_2 be random variables with joint CDF $F_{Y_1,Y_2}(y_1, y_2)$. Further, assume that $\vec{a} := (a_1, a_2) \in \mathbb{R}^2$ and $\vec{b} := (b_1, b_2) \in \mathbb{R}^2$ satisfy $a_1 < b_1$ and $a_2 < b_2$. Then,*

$$F_{Y_1,Y_2}(-\infty, -\infty) = F_{Y_1,Y_2}(-\infty, y_2) = F_{Y_1,Y_2}(y_1, -\infty) = 0\,. \tag{11.2}$$

$$F_{Y_1,Y_2}(\infty, \infty) = 1\,, \tag{11.3}$$

$$\begin{aligned}\mathbb{P}\{a_1 < Y_1 \leq b_1,\, a_2 < Y_2 \leq b_2\} &= F_{Y_1,Y_2}(b_1, b_2) - F_{Y_1,Y_2}(a_1, b_2) \\ &\quad - F_{Y_1,Y_2}(b_1, a_2) + F_{Y_1,Y_2}(a_1, a_2)\,,\end{aligned} \tag{11.4}$$

$$F_{Y_1,Y_2}(b_1, b_2) - F_{Y_1,Y_2}(a_1, b_2) - F_{Y_1,Y_2}(b_1, a_2) + F_{Y_1,Y_2}(a_1, a_2) \geq 0\,, \tag{11.5}$$

Definition 11.2 (Joint probability mass function)**.** Let Y_1 and Y_2 be discrete random variables. We call

$$p(y_1, y_2) := p_{Y_1,Y_2}(y_1, y_2) := \mathbb{P}\{Y_1 = y_1, Y_2 = y_2\}\,, \qquad \text{where } y_1, y_2 \in \mathbb{R}\,, \tag{11.6}$$

the **joint probability mass function** or **bivariate probability mass function** or **joint PMF** of Y_1 and Y_2. $\square$

Proposition 11.1 (WMS Ch.05.2, Theorem 5.1). *If Y_1 and Y_2 are discrete random variables with joint PMF $p_{Y_1,Y_2}(y_1,y_2)$, then*

(1) $p_{Y_1,Y_2}(y_1,y_2) \geq 0$ *for all* $y_1, y_2 \in \mathbb{R}$,

(2) $\sum\limits_{y_1,y_2} p_{Y_1,Y_2}(y_1,y_2) = 1$.

(3) $F_{Y_1,Y_2}(y_1,y_2) = \sum\limits_{u_1 \leq y_1,\, u_2 \leq y_2} p_{Y_1,Y_2}(u_1,u_2) = \sum\limits_{u_1 \leq y_1} \sum\limits_{u_2 \leq y_2} p_{Y_1,Y_2}(u_1,u_2)\,.$

Definition 11.3 (Jointly continuous random variables). Let Y_1 and Y_2 be random variables with joint CDF $F(y_1,y_2)$. We call Y_1 and Y_2 **jointly continuous** if $F(y_1,y_2)$ is a continuous function of both arguments. $\square$

Assumption 11.2 (Jointly continuous random variables have PDFs). We assume for all jointly continuous random variables Y_1 and Y_2 that $\dfrac{\partial^2 F_{Y_1,Y_2}}{\partial y_1 \partial y_2}$ exists and is continuous except for $(y_1,y_2) \in B^*$, where the set $B^* \subseteq \mathbb{R}^2$ satisfies that
$B^* \cap B$ is finite for any bounded subset $B \in \mathbb{R}^2$ (bounded sets are those contained in a circle with sufficiently large radius).

This assumption guarantees for all $y_1, y_2 \in \mathbb{R}$, when we write f_{Y_1,Y_2} for $\dfrac{\partial^2 F_{Y_1,Y_2}}{\partial y_1 \partial y_2}$, that

$$
\begin{aligned}
F_{Y_1,Y_2}(y_1,y_2) &= \int_{-\infty}^{y_1} \int_{-\infty}^{y_2} f_{Y_1,Y_2}(u_1,u_2)\, du_2\, du_1 \\
&= \int_{-\infty}^{y_2} \int_{-\infty}^{y_1} f_{Y_1,Y_2}(u_1,u_2)\, du_1\, du_2\,. \qquad (11.7)\\
&= \iint\limits_{]-\infty,y_1 \times]-\infty,y_2]} f_{Y_1,Y_2}(u_1,u_2)\, du_1\, du_2\,. \quad \square
\end{aligned}
$$

Definition 11.4 (WMS Ch.05.2, Definition 5.3). Let Y_1 and Y_2 be continuous random variables with joint distribution function $F(y_1,y_2)$ and second derivative $f_{Y_1,Y_2}(y_1,y_2) = \dfrac{\partial^2 F_{Y_1,Y_2}}{\partial y_1 \partial y_2}(y_1,y_2)$. We call $f_{Y_1,Y_2}(y_1,y_2)$ the **joint probability density function** or **joint PDF** of Y_1 and Y_2. $\square$

Theorem 11.2. *Let Y_1 and Y_2 be jointly continuous random variables with joint PDF $f_{Y_1,Y_2}(y_1, y_2)$, then*

(1) $f_{Y_1,Y_2}(y_1, y_2) \ \geq \ 0$ *for all* y_1, y_2.

(2) $\int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} f_{Y_1,Y_2}(y_1, y_2) dy_1 dy_2 = 1$.

11.2 Marginal and Conditional Probability Distributions

Definition 11.5 (Marginal distribution of two random variables)**.** Let $\vec{Y} = (Y_1, Y_2)$ be a vector of two random variables with joint distribution

$$B_1 \times B_2 \mapsto \mathbb{P}_{Y_1,Y_2}(B_1 \times B_2) \ = \ \mathbb{P}\{Y_1 \in B_1, Y_2 \in B_2\}\,, \quad \text{where } B_1, B_2 \subseteq \mathbb{R}.$$

We call the probability measures

$$Q_1 : B_1 \mapsto \mathbb{P}_{Y_1,Y_2}(B_1 \times \mathbb{R}) \quad \text{and} \quad Q_2 : B_2 \mapsto \mathbb{P}_{Y_1,Y_2}(\mathbb{R} \times B_2) \tag{11.8}$$

the **marginal distributions** of $\vec{Y} = (Y_1, Y_2)$. $\square$

Proposition 11.2. *The marginal distributions of $\vec{Y} = (Y_1, Y_2)$ are the distributions $\mathbb{P}_{Y_1}$ and $\mathbb{P}_{Y_2}$ of the coordinates*

Y_1 and Y_2. In other words, $Q_1 \ = \ \mathbb{P}_{Y_1}$ and $Q_2 \ = \ \mathbb{P}_{Y_2}$

Definition 11.6 (Marginal PMF and PDF)**. (a)** Let Y_1 and Y_2 be discrete random variables with joint PMF $p_{Y_1,Y_2}(y_1, y_2)$. We call

$$p_{Y_1}(y_1) \ = \ \sum_{\text{all } y_2} p_{Y_1,Y_2}(y_1, y_2) \quad \text{and} \quad p_{Y_2}(y_2) \ = \ \sum_{\text{all } y_1} p_{Y_1,Y_2}(y_1, y_2) \tag{11.9}$$

the **marginal probability mass functions** or **marginal PMFs** of Y_1 and Y_2.

(b) Let Y_1 and Y_2 be continuous random variables with joint PDF $f_{Y_1,Y_2}(y_1, y_2)$. We call

$$f_{Y_1}(y_1) \ = \ \int_{-\infty}^{\infty} f_{Y_1,Y_2}(y_1, y_2)\, dy_2 \quad \text{and} \quad f_{Y_2}(y_2) \ = \ \int_{-\infty}^{\infty} f_{Y_1,Y_2}(y_1, y_2)\, dy_1\,. \tag{11.10}$$

the **marginal density functions** or **marginal PDFs** of Y_1 and Y_2. $\square$

Definition 11.7 (Conditional probability mass function)**.** Let Y_1 and Y_2 be discrete random variables with joint PMF $p_{Y_1,Y_2}(y_1, y_2)$ and marginal PMFs $p_{Y_1}(y_1)$ and $p_{Y_2}(y_2)$. Then we call

$$p_{Y_1|Y_2}(y_1 \mid y_2) \ := \ \begin{cases} \mathbb{P}\{Y_1 = y_1 \ \mid \ Y_2 = y_2\}\,, & \text{if } \mathbb{P}\{Y_2 = y_2\} > 0\,, \\ \textbf{undefined}\,, & \text{if } \mathbb{P}\{Y_2 = y_2\} = 0\,, \end{cases} \tag{11.11}$$

the **conditional probability mass function** or the **conditional PMF** of Y_1 given Y_2. Likewise, we call

$$p_{Y_2|Y_1}(y_2 \mid y_1) := \begin{cases} \mathbb{P}\{Y_2 = y_2 \mid Y_1 = y_1\}, & \text{if } \mathbb{P}\{Y_1 = y_1\} > 0, \\ \textbf{undefined}, & \text{if } \mathbb{P}\{Y_1 = y_1\} = 0, \end{cases} \tag{11.12}$$

the **conditional PMF** of Y_2 given Y_1. □

Definition 11.8 (Conditional probability density function)**.** Let Y_1 and Y_2 be continuous random variables with joint PDF $f_{Y_1|Y_2}(y_1, y_2)$ and marginal densities $f_{Y_1}(y_1)$ and $f_{Y_2}(y_2)$. Then we call

$$f_{Y_1|Y_2}(y_1 \mid y_2) := \begin{cases} \dfrac{f_{Y_1,Y_2}(y_1, y_2)}{f_{Y_2}(y_2)}, & \text{if } f_{Y_2}(y_2) > 0, \\ \textbf{undefined}, & \text{if } f_{Y_2}(y_2) = 0, \end{cases} \tag{11.13}$$

the **conditional probability density function** or the **conditional PDF** of Y_1 given Y_2. Likewise we call

$$f_{Y_2|Y_1}(y_2 \mid y_1) := \begin{cases} \dfrac{f_{Y_1,Y_2}(y_1, y_2)}{f_{Y_1}(y_1)}, & \text{if } f_{Y_1}(y_1) > 0, \\ \textbf{undefined}, & \text{if } f_{Y_1}(y_1) = 0, \end{cases} \tag{11.14}$$

the **conditional PDF** of Y_2 given Y_1. □

Definition 11.9. $\boxed{\star}$ Let Y_1 and Y_2 be two jointly continuous random variables. Then,

$$F_{Y_1|Y_2}(y_1 \mid y_2) := \int_{-\infty}^{y_1} \frac{f_{Y_1,Y_2}(u_1, y_2)}{f_{Y_2}(y_2)}\, du_1 \tag{11.15}$$

defines the **conditional distribution function** or **conditional CDF** of Y_1 given $Y_2 = y_2$. □

11.3 Independence of Random Variables and Discrete Random Elements

Theorem 11.3 (CDFs of Independent random variables)**.** *Let Y_1 and Y_2 be random variables with CDFs $F_{Y_1}(y_1)$ and $F_{Y_2}(y_2)$ and with joint CDF $F_{Y_1,Y_2}(y_1, y_2)$. Then Y_1 and Y_2 are independent if and only if*

$$F_{Y_1,Y_2}(y_1, y_2) = F_{Y_1}(y_1) \cdot F_{Y_2}(y_2) \quad \textit{for all } y_1, y_2 \in \mathbb{R}. \tag{11.16}$$

We must treat discrete random elements separately since there are no CDFs.

Let X_1 and X_2 be discrete random elements with PMFs $p_{X_1}(x_1)$ and $p_{X_2}(x_2)$ and with joint PMF $p_{X_1,X_2}(x_1,x_2)$. Then X_1 and X_2 are independent if and only if

$$p_{X_1,X_2}(x_1,x_2) = p_{X_1}(x_1) \cdot p_{X_2}(x_2) \quad \text{for all } x_1, x_2 \in \mathbb{R}. \tag{11.17}$$

Theorem 11.4 (Functions of independent random variables are independent)**.**
Let $\vec{Y} = (Y_1, \dots, Y_k) : (\Omega, \mathbb{P}) \to \mathbb{R}$ be a vector of k independent random variables and $h_j : \mathbb{R} \to \mathbb{R}$.

- *Then the random variables $h_1 \circ Y_1, \dots, h_k \circ Y_k$ also are independent.*

Theorem 11.5 (WMS Ch.05.4, Theorem 5.4)**.** *If Y_1 and Y_2 are discrete random variables with joint PMF $p_{Y_1,Y_2}(y_1,y_2)$ and marginal PMFs $p_{Y_1}(y_1)$ and $p_{Y_2}(y_2)$, then*

$$Y_1, Y_2 \text{ are independent} \quad \Leftrightarrow \quad p_{Y_1,Y_2}(y_1,y_2) = p_{Y_1}(y_1) \cdot p_{Y_2}(y_2) \ \text{ for all } y_1, y_2 \in \mathbb{R}. \tag{11.18}$$

If Y_1 and Y_2 are continuous random variables with joint PDF $f_{Y_1,Y_2}(y_1,y_2)$ and marginal PDFs $f_{Y_1}(y_1)$ and $f_{Y_2}(y_2)$, then

$$Y_1, Y_2 \text{ are independent} \quad \Leftrightarrow \quad f_{Y_1,Y_2}(y_1,y_2) = f_{Y_1}(y_1) \cdot f_{Y_2}(y_2) \ \text{ for all } y_1, y_2 \in \mathbb{R}. \tag{11.19}$$

Theorem 11.6. *If Y_1 and Y_2 are independent random variables, then*

$$\mathbb{E}[Y_1 \cdot Y_2] = \mathbb{E}[Y_1] \cdot \mathbb{E}[Y_2]\,. \tag{11.20}$$

Theorem 11.7 (WMS Ch.05.4, Theorem 5.5)**.** *Let the continuous random variables Y_1 and Y_2 have a joint PDF $f_{Y_1,Y_2}(y_1,y_2)$ that is strictly positive if and only if there are constants $a < b$ and $c < d$ such that*

$$f_{Y_1,Y_2}(y_1,y_2) > 0 \quad \Leftrightarrow \quad a \le y_1 \le b \ \text{ and } \ c \le y_2 \le d\,.$$

$$\text{Then} \quad Y_1, Y_2 \text{ are independent} \quad \Leftrightarrow \quad f_{Y_1,Y_2}(y_1,y_2) = g_1(y_1) \cdot g_2(y_2) \tag{11.21}$$

for suitable nonnegative functions $g_1, g_2 : \mathbb{R} \to \mathbb{R}$ such that the only argument of g_1 is y_1 and the only argument of g_2 is y_2.

11.4 The Mulitivariate Uniform Distribution

Definition 11.10 (Multivariate continuous, uniform random variable). **(A)** Let $\vec{Y} = (Y_1, Y_2)$ be a twodimensional random vector of continuous random variables with a joint PDF $f_{\vec{Y}}(y_1, y_2)$ that satisfies the following:

- There is a constant $c > 0$ such that either $f_{\vec{Y}}(y_1, y_2) = c$ or $f_{\vec{Y}}(y_1, y_2) = 0$.

Let $C := \{(y_1, y_2) \in \mathbb{R}^2 : f_{\vec{Y}}(y_1, y_2) > 0\}$. Then we say that $\vec{Y}$ has a **continuous uniform probability distribution** on C.

(B) Let $\vec{Y} = (Y_1, Y_2, Y_3)$ be a threedimensional random vector of continuous random variables with a joint PDF $f_{\vec{Y}}(y_1, y_2, y_3)$ that satisfies the following:

- There is a constant $d > 0$ such that either $f_{\vec{Y}}(y_1, y_2, y_3) = d$ or $f_{\vec{Y}}(y_1, y_2, y_3) = 0$.

Let $D := \{(y_1, y_2, y_2, y_3) \in \mathbb{R}^3 : f_{\vec{Y}}(y_1, y_2, y_3) > 0\}$. Then we say that $\vec{Y}$ has a **continuous uniform probability distribution** on D. $\square$

11.5 The Expected Value of a Function of Several Random Variables

Notation 11.1 (Arrow notation for vectors).

- We write $\vec{x}$ as an abbreviation for a vector $(x_1, x_2, \dots, x_n)$. The explicitly stated or known from the context.
- If $f : \mathbb{R}^n \to \mathbb{R}$ is a function of n real numbers and $U = [a_1, b$ n–dimensional rectangle, we write

$$\int_A f(\vec{x})\, d\vec{x} = \int_{a_1}^{b_1} \cdots \int_{a_2}^{b_2} \int_{a_1}^{b_1} f(x_1, x_2, \dots, x_n)\, dy_1 d$$

  Note that all integrands that occur in this course are so well beh which those n integrations take place can be switched around, it in the cases $n = 2$ and $n = 3$ from multidimensional calculus
- Let $a_1 < b_1$, $a_2 < b_2, \dots, a_n < b_n$ for some $n \in \mathbb{N}$. Then $\vec{y}$ denotes the following: $\vec{y} = (y_1, y_2, \dots, y_d)$ and $a_i < y_i \le b_i$ fo

Theorem 11.8 (Expected value of $g(\vec{Y})$). ***(a)*** *Let $k \in \mathbb{N}$ and let $\vec{Y} = (Y_1, Y_2, \dots, Y_k)$ be a vector of discrete random variables on a probability space $(\Omega, \mathbb{P})$ with PMF $p_{\vec{Y}}(\vec{y})$. Further, let $g : \mathbb{R}^k \to \mathbb{R}$ be a function of k real numbers $y_1, y_2, \dots, y_k$. Then*

$$\mathbb{E}\big[g(\vec{Y})\big] = \mathbb{E}\big[g(Y_1, Y_2, \dots, Y_k)\big] := \sum_{y_1,\, y_2,\, \dots,\, y_k} \cdots \sum g(\vec{y})\, p_{\vec{Y}}(\vec{y}) \tag{11.22}$$

*is called the **expected value** or **mean** of the random variable $g(\vec{Y})$. As usual, the sum on the right is countable summation over those $\vec{y} = (y_1, y_2, \dots, y_k)$ for which $p_{\vec{Y}}(\vec{y}) \neq 0$.*

(b) *Let $k \in \mathbb{N}$ and let $\vec{Y} = \big(Y_1, Y_2, \dots, Y_k\big)$ be a vector of continuous random variables on a probability space $(\Omega, \mathbb{P})$ with PDF $f_{\vec{Y}}(\vec{y})$. Further, let $h : \mathbb{R}^k \to \mathbb{R}$ be a function of k real numbers $y_1, y_2, \dots, y_k$. Then*

$$\mathbb{E}\big[h(\vec{Y})\big] \;=\; \mathbb{E}\big[h(Y_1, Y_2, \dots, Y_k)\big] \;:=\; \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h(\vec{y})\, f_{\vec{Y}}(\vec{y}) d\vec{y} \tag{11.23}$$

*is called the **expected value** or **mean** of the random variable $g(\vec{Y})$.*

Theorem 11.9 (WMS Ch.05.6, Theorem 5.6)**.**

$$c \in \mathbb{R} \;\;\Rightarrow\;\; \mathbb{E}[c] \;=\; c\,. \tag{11.24}$$

Theorem 11.10 (WMS Ch.05.6, Theorem 5.7)**.** *Let $c \in \mathbb{R}$ and $g : \mathbb{R}^2 \to \mathbb{R}$ Then the random variable $g(Y_1, Y_2)$ satisfies*

$$\mathbb{E}[cg(Y_1, Y_2)] \;=\; c\mathbb{E}[g(Y_1, Y_2)]\,. \tag{11.25}$$

Theorem 11.11 (WMS Ch.05.6, Theorem 5.8)**.** *Let $g_1, g_2, \dots, g_k : \mathbb{R}^n \to \mathbb{R}$ and $\vec{Y} := (Y_1, \dots, Y_n)$. Then the random variables $g_j(\vec{Y})$ $(j = 1, \dots, k)$ satisfy*

$$\begin{aligned} &\mathbb{E}[g_1(\vec{Y}) + g_2(\vec{Y}) + \cdots + g_k(\vec{Y})] \\ &\quad = \mathbb{E}[g_1(\vec{Y})] \;+\; \mathbb{E}[g_2(\vec{Y})] \;+ \cdots +\; \mathbb{E}[g_k(\vec{Y})]\,. \end{aligned} \tag{11.26}$$

Theorem 11.12. *Let $g, h : \mathbb{R} \to \mathbb{R}$ be functions of a single variable and assume that the random variables Y_1 and Y_2 are independent. Then the random variables $g(Y_1)$ and $h(Y_2)$ also are independent and they satisfy*

$$\mathbb{E}[g(Y_1)\, h(Y_2)] \;=\; \mathbb{E}[g(Y_1)]\, \mathbb{E}[h(Y_2)]\,. \tag{11.27}$$

11.6 Covariance

In this entire section, we consider two random variables Y_1 and Y_2 on a probability space $(\Omega, \mathbb{P})$. As usual, we denote mean and standard deviation

$$\mu_j := \mathbb{E}[Y_j], \quad \sigma_j := \sqrt{Var[Y_j]}, \quad \text{for } j = 1, 2.$$

Definition 11.11 (Covariance)**.** The **covariance** of Y_1 and Y_2 is

$$Cov[Y_1, Y_2] = \mathbb{E}\big[(Y_1 - \mathbb{E}[Y_1])\,(Y_2 - \mathbb{E}[Y_2])\big] = \mathbb{E}\big[(Y_1 - \mu_1)\,(Y_2 - \mu_2)\big]. \;\square \tag{11.28}$$

Definition 11.12 (Correlation coefficient)**.** The **correlation coefficient**, of Y_1 and Y_2 is

$$\rho = \frac{Cov(Y_1, Y_2)}{\sigma_1 \sigma_2} \;\square \tag{11.29}$$

We say that Y_1 and Y_2 have **positive correlation** if $\rho > 0$, (i.e., if $Cov(Y_1, Y_2) > 0$), they have **negative correlation** if $\rho < 0$, (i.e., if $Cov(Y_1, Y_2) < 0$), and that they have **zero correlation** or that they are **uncorrelated** if $\rho = 0$, (i.e., if $Cov(Y_1, Y_2) = 0$).

Proposition 11.3. *The correlation coefficient satisfies the inequality*

$$-1 \le \rho \le 1 \tag{11.30}$$

Theorem 11.13.

$$Cov[Y_1, Y_2] = \mathbb{E}[(Y_1 - \mu_1)\,(Y_2 - \mu_2)] = \mathbb{E}[Y_1 Y_2] - \mathbb{E}[Y_1]\,\mathbb{E}[Y_2]. \tag{11.31}$$

Theorem 11.14. *Independent random variables are uncorrelated.*

Definition 11.13 (Linear function)**.** $\boxed{\star}$ Let $n \in \mathbb{N}$. We call a function $\varphi : \mathbb{R}^n \to \mathbb{R}$; $\vec{x} = (x_1, \dots, x_n) \mapsto \varphi(\vec{x})$, a **linear function**, of $x_1, \dots, x_n$, if there are constants $a_1, \dots, a_n \in \mathbb{R}$ such that

$$\varphi(\vec{x}) = a_1 x_1 + a_2 x_2 + \cdots + a_n x_n = \sum_{j=1}^{n} a_j x_j. \;\square \tag{11.32}$$

Theorem 11.15 (WMS Ch.05.8, Theorem 5.12). *Let $\vec{X} = X_1, \dots, X_m$ and $\vec{Y} = Y_1, \dots, Y_n$ be random variables on a probability space $(\Omega, \mathbb{P})$. For $i = 1, \dots, m$ and $j = 1, \dots, n$, let $\xi_i := E(X_i)$ and $\eta_j := E(Y_j)$. Further, let*

$$U := \sum_{i=1}^{m} a_i X_i \qquad \text{and} \qquad V := \sum_{j=1}^{n} b_j Y_j\,,$$

where $\vec{a} = (a_1, a_2, \dots, a_m)$ and $\vec{b} = (b_1, b_2, \dots, b_n)$ are two constant vectors. Then

$$\mathbb{E}[U] = \sum_{i=1}^{m} a_i \xi_i\,, \tag{11.33}$$

$$Var[U] = \sum_{i=1}^{m} a_i^2 Var[X_i] + 2 \sum\sum_{1 \le i < j \le m} a_i a_j Cov[X_i, X_j]\,. \tag{11.34}$$

$$Cov[U, V] = \sum_{i=1}^{m} \sum_{j=1}^{n} a_i b_j Cov[X_i, Y_j]\,. \tag{11.35}$$

In (11.34), $\sum\sum_{1 \le i < j \le m} \cdots$ refers to summation over all pairs (i, j) satisfying $i < j$.

Corollary 11.1 (Bienaymé formula for uncorrelated variables). $\boxed{\star}$ *Let $Y_1, Y_2, \dots, Y_n : \Omega \to \mathbb{R}$ be uncorrelated random variables (which all are defined on the same probability space $(\Omega, \mathbb{P})$ $(n \in \mathbb{N}$. Then*

$$Var\left[\sum_{j=1}^{n} Y_j\right] = \sum_{j=1}^{n} Var[Y_j]\,. \tag{11.36}$$

11.7 Conditional Expectations and Conditional Variance

11.7.1 The Conditional Expectation With Respect to an Event $\boxed{\star}$

Assumption 11.3. In all of this subsection we deal with a fixed probability space $(\Omega, \mathbb{P})$ and a fixed event $B \subseteq \Omega$ that satisfies $\mathbb{P}(B) > 0$. We further assume that $Q(\cdot)$ is the probability measure

$$A \mapsto Q(A) := \mathbb{P}(A \mid B), \quad \text{where } A \subseteq \Omega. \tag{11.37}$$

The symbols $X, X_1, X_2, \dots$ denote random elements and $Y, Y_1, Y_2, \dots$ denote random variables on Ω. We need not be specific about whether we mean $(\Omega, \mathbb{P})$ or (Ω, Q), because the definition of random element and random variable does not involve the probability measure, only the carrier space Ω. $\square$

Theorem 11.16. *If $\vec{Y} = (Y_1, Y_2, \dots, Y_k)$ is a vector of k discrete or Q–continuous random variables, then*

$$E^Q\left[\sum_{j=1}^n Y_j\right] = \sum_{j=1}^n E^Q[Y_j]\,. \tag{11.38}$$

Theorem 11.17. *If Y is a discrete or Q–continuous random variable and $\vec{Y} = (Y_1, Y_2, \dots, Y_k)$ is a vector of k Q–independent discrete or Q–continuous random variables, then*

$$Var^Q[Y] = E^Q\left[Y^2\right] - \left(E^Q[Y]\right)^2, \tag{11.39}$$

$$Var^Q\left[aY + b\right] = a^2 Var^Q[Y]\,, \tag{11.40}$$

$$Var^Q\left[\sum_{j=1}^n Y_j\right] = \sum_{j=1}^n Var^Q[Y_j]\,. \tag{11.41}$$

Theorem 11.18. *Let the events $A_1, A_2, B \subseteq \Omega$ satisfy $\mathbb{P}(A_1 \cap B > 0$, $\mathbb{P}(A_2 \cap B > 0$. (Hence, $\mathbb{P}(B > 0)$. Then*

$$\begin{aligned} & \textit{(a)} \quad \mathbb{P}(A_1 \cap A_2 \mid B) = \mathbb{P}(A_1 \mid B) \cdot \mathbb{P}(A_2 \mid B) \\ \Leftrightarrow \quad & \textit{(b)} \quad \mathbb{P}(A_1 \mid A_2 \cap B) = \mathbb{P}(A_1 \mid B) \\ \Leftrightarrow \quad & \textit{(c)} \quad \mathbb{P}(A_2 \mid A_1 \cap B) = \mathbb{P}(A_2 \mid B)\,. \end{aligned} \tag{11.42}$$

In other words, if A_i and A_j are independent with respect to "just" conditioning on B, then "further" conditioning of A_i on both A_j and B has no effect. Here, either $i = 1, j = 2$ or $i = 2, j = 1$.

11.7.2 The Conditional Expectation w.r.t a Random Variable or Random Element

Definition 11.14 (Conditional expectation)**.** Let Y_1 and Y_2 be two random variables which are either jointly discrete or jointly continuous and $g : \mathbb{R} \to \mathbb{R}$. Let

$$\mathbb{E}[g(Y_1) \mid Y_2 = y_2] := \sum_{y_1} g(y_1)\, p(y_1 \mid y_2) \quad \text{(discrete case)}, \tag{11.43}$$

$$\mathbb{E}[g(Y_1) \mid Y_2 = y_2] := \int_{-\infty}^{\infty} g(y_1)\, f(y_1 \mid y_2)\, dy_1 \quad \text{(continuous case)}. \tag{11.44}$$

We call $\mathbb{E}[g(Y_1) \mid Y_2 = y_2]$ the **conditional expectation** of $g(Y_1)$, given that $Y_2 = y_2$. $\square$

Theorem 11.19 (WMS Ch.05.11, Theorem 5.14)**.** *Let Y_1 and Y_2 be either jointly continuous or jointly discrete random variables. Then*

$$\mathbb{E}[Y_1] = \mathbb{E}\big[\mathbb{E}[Y_1 \mid Y_2]\big]\,. \tag{11.45}$$

*See Remark **??** concerning the interpretation of the right–hand side.*

Definition 11.15 (Conditional variance)**.** Let Y_1 and Y_2 be two random variables which are either jointly discrete or jointly continuous. Let

$$Var[Y_1 \mid Y_2 = y_2] := \mathbb{E}[Y_1^2 \mid Y_2 = y_2] - \big(\mathbb{E}[Y_1 \mid Y_2 = y_2]\big)^2\,. \tag{11.46}$$

We call $Var[Y_1 \mid Y_2 = y_2]$ the **conditional variance** of (Y_1), given that $Y_2 = y_2$. $\square$

Theorem 11.20. *Let Y_1 and Y_2 be jointly discrete or jointly continuous random variables. Then*

$$Var[Y_1 \mid Y_2] = \mathbb{E}\big[\,(Y_1 - \mathbb{E}[Y_1 \mid Y_2])^2 \mid Y_2\,\big]\,, \tag{11.47}$$

$$Var[Y_1] = \mathbb{E}\big[\,Var[Y_1 \mid Y_2]\,\big] + Var\big[\,\mathbb{E}[Y_1 \mid Y_2]\,\big]\,. \tag{11.48}$$

Lemma 11.1. $\boxed{\star}$

***(A)**: Let X and Y be two jointly continuous r.v.s on $(\Omega, \mathfrak{F}, \mathbb{P}$ and B a Borel set of $\mathbb{R}$. Then,*

$$\int_B \mathbb{E}[Y \mid X = x] f_X(x)\, dx = \int_B \int_{-\infty}^{\infty} y f_{(X,Y)}(x,y)\, dy\, dx\,. \tag{11.49}$$

***(B)**: Let X and Y be two jointly discrete r.v.s on $(\Omega, \mathfrak{F}, \mathbb{P}$ and $B \subseteq \mathbb{R}$. Then,*

$$\sum_{x \in B} \mathbb{E}[Y \mid X = x] p_X(x) = \sum_{x \in B, y \in \mathbb{R}} y p_{(X,Y)}(x,y)\,. \tag{11.50}$$

Theorem 11.21 (Conditional expectations preserve all partial averages).

Let X and Y be two jointly continuous or jointly discrete r.v.s on $(\Omega, \mathfrak{F}, \mathbb{P})$ and $B \subseteq \mathbb{R}$. [12] *Then,*

$$\int_{\{X \in B\}} \mathbb{E}[Y \mid X]\, d\mathbb{P} \;=\; \int_{\{X \in B\}} Y\, d\mathbb{P}\,.$$

11.7.3 Conditional Expectations as Optimal Mean Squared Distance Approximations

Theorem 11.22. *Assume that Y is a random variable and $\vec{X} = (X_1, \dots, X_k)$ is a random vector on $(\Omega, \mathbb{P})$. Then, either $\mathbb{E}\big[(Y - g \circ \vec{X})\big] \;=\; \infty$ for all real–valued functions $g : \mathbb{R}^k \to \mathbb{R}$ of k real arguments, or*

$$\mathbb{E}\left[\left(Y - \mathbb{E}[Y \mid \vec{X}]\right)^2\right] \;\leq\; \mathbb{E}\left[(Y - g \circ \vec{X})^2\right],$$

for all such functions g. Further, this is a strict inequality if $\mathbb{E}[Y \mid \vec{X}] \neq g \circ \vec{X}$.

Note that, as always, we consider equations and inequalities involving random variables to be true as long as they are satisfied on a set of probability 1.

We interpret random variables of the form $g(\vec{X})$, where $\vec{x} \mapsto g(\vec{x})$ is a (deterministic) function of $\vec{x}$, as those random variables that only use the information available to $\vec{X}$.

If we measure the quality of the approximation of a random variable Y by $g(\vec{X})$ as their mean squared distance, $\mathbb{E}\left[\left(Y - g(\vec{X})\right)^2\right]$, then

- $\mathbb{E}[Y \mid \vec{X}]$ is the best approximation of Y which is based only on information provided by $\vec{X}$.

11.8 The Multinomial Probability Distribution

Definition 11.16 (Multinomial Sequence)**.** Let $X_1, X_2, \dots$ be a finite or infinite sequence of random elements on a probability space $(\Omega, \mathbb{P})$ which take values in a set Ω'. We call this sequence a **multinomial sequence**, if the following are satisfied:

(1) The sequence is iid.

(2) There is some $k \in \mathbb{N}$ such that the outcome of each X_j is one of k distinct values $\omega'_1, \omega'_2, \dots, \omega'_k \in \Omega'$.

Since the X_j have identical distribution, there are probabilities $p_1, p_2, \dots, p_k$ such that

(3) $p_i \;:=\; \mathbb{P}\{X_j = \omega'_i\}$ is the same for all j and $p_1 + \dots + p_k = 1$.

If we consider a finite multinomial sequence $X_1, X_2, \dots, X_n$, we adopt the WMS notation and speak of a **multinomial experiment** of size n wich consists of the **trials** X_j $\square$

Definition 11.17 (Multinomial distribution)**.** Assume that $\vec{Y} = \big(Y_1, Y_2, \dots, Y_k\big)$ is a vector of random variables which possesses the joint probability mass function

$$p_{\vec{Y}}(y_1, y_2, \dots, y_k) \;=\; \binom{n}{y_1, \cdots, y_k} p_1^{y_1} p_2^{y_2} \cdots p_k^{y_k}\,, \tag{11.51}$$

subject to the following conditions:

- $p_j \geq 0$ for $j = 1, 2, \dots, k$ and $\sum\limits_{j=1}^{k} p_j = 1$.
- $y_i = 0, 1, 2, \dots, n$ for $i = 1, 2, \dots, k$ and $\sum\limits_{i=1}^{k} y_i = n$.

Then we say that the random variables Y_i have a **multinomial distribution** with parameters n and $\vec{p} = (p_1, p_2, \dots, p_k)$. $\square$

Theorem 11.23. *Let $n \in \mathbb{N}$ and $X_1, \dots, X_n$ be a multinomial sequence of size n. Let $p_j := \mathbb{P}\{X_i = \omega'_j\}$. (That probability is the same for all i, since the X_i have identical distribution.)*

Let $\vec{Y} = \big(Y_1, \dots, Y_k\big)$ be a vector of k random variables, such that each Y_j equals the number of the n trials resulting in an outcome that falls into class j. In other words,

(A) $Y_i(\omega) = y_i \;\Leftrightarrow\; X_j(\omega) = \omega'_i$ *for exactly y_i of the multinomial items X_j.*

Then $\vec{Y}$ has a multinomial distribution with parameters n and $p_{\vec{Y}}(y_1, y_2, \dots, y_k)$.

Theorem 11.24 (WMS Ch.05.9, Theorem 5.13)**.** *Assume that the random vector $\vec{Y} = (Y_1, Y_2, \dots, Y_k)$ follows a multinomial distribution with parameters n and $\vec{p} = (p_1, p_2, \dots, p_k)$. Then, for $i, i' \in [1, k]_{\mathbb{Z}}$ and $q_i = 1 - p_i$,*

(a) $\mathbb{E}[Y_i] \;=\; np_i$ ***(b)*** $Var[Y_i] = np_i q_i$ ***(c)*** *If $i \neq i'$, then* $Cov[Y_i, Y_{i'}] \;=\; -np_i p_{i'}$

11.9 Order Statistics

- We will deal in this section exclusively with continuous random variables.
- When considering a finite or infinite sequence $Y_1, Y_2, Y_3, \dots$ of such random variables, we assume that they are iid (independent and identically distributed).

Definition 11.18 (Order statistics)**.** Given n iid continuous random variables $\vec{Y} = (Y_1, Y_2, \dots, Y_n)$, we sort them in inreasing order. The resulting sequence of random variables, which we denote as $Y_{(j)}, j = 1, \dots, n$, then satisfies, for each $(\omega \in \Omega$,

$$Y_{(1)}(\omega) \;\leq\; Y_{(2)}(\omega) \;\leq\; Y_{(3)}(\omega) \;\leq \cdots \leq\; Y_{(n)}(\omega)\,. \tag{11.52}$$

We call $Y_{(j)}$ the **jth order statistic** of $\vec{Y}$.

See Example **??(b)** why we may consider strictly increasing rather than nondecreasing. □

Assumption 11.4. Unless explicitly stated otherwise,

- $\vec{Y} = (Y_1, Y_2, \dots, Y_n)$ denotes a list of n iid continuous random variables ($n \in \mathbb{N}$).
- $Y_1 \sim Y_2 \sim \cdots \sim Y_n$ implies $F_{Y_1} = F_{Y_2} = \cdots = F_{Y_n}$ and $f_{Y_1} = f_{Y_2} = \cdots = f_{Y_n}$
  We write $F(y) := F_{Y_j}(y)$ and $f(y) := f_{Y_j}(y)$ for the common CDF and PDF. □

Theorem 11.25 (CDF and PDF of the jth order statistic). *For $y \in \mathbb{R}$, the CDF of the kth order statistic ($k = 1, \dots, n$) satisfies the following:*

$$F_{Y_{(1)}(y)} = 1 - [1 - F(y)]^n , \tag{11.53}$$

$$F_{Y_{(n)}(y)} = [F(y)]^n , \tag{11.54}$$

$$F_{Y_{(k)}(y)} = 1 - \sum_{j=0}^{k-1} \binom{n}{j} [F(y)]^j \, [1 - F(y)]^{n-j} = \sum_{j=k}^{n} \binom{n}{j} [F(y)]^j \, [1 - F(y)]^{n-j} . \tag{11.55}$$

For $y \in \mathbb{R}$, the PDF of the kth order statistic ($k = 1, \dots, n$) satisfies the following:

$$f_{Y_{(1)}(y)} = n \, [1 - F(y)]^{n-1} \, f(y) , \tag{11.56}$$

$$f_{Y_{(n)}(y)} = n \, [F(y)]^{n-1} \, f(y) , \tag{11.57}$$

$$f_{Y_{(k)}(y)} = n \binom{n-1}{k-1} f(y) \cdot \big[\, F(y) \,\big]^{k-1} \cdot \big[\, 1 - F(y) \,\big]^{n-k} . \tag{11.58}$$

Theorem 11.26 (WMS Ch.06.7, Theorem 6.5). *If two indices i and j satisfy $1 \le i < j \le n$, the joint density of $Y_{(i)}$ and $Y_{(j)}$ is*

$$\begin{aligned} f_{Y_{(i)},Y_{(j)}}(y_i, y_j) &= \frac{n!}{(i-1)!(j-1-i)!(n-j)!} [F(y_i)]^{i-1} \\ &\quad \times [F(y_j) - F(y_i)]^{j-1-i} \times [1 - F(y_j)]^{n-j} \, f(y_i) f(y_j) , \quad -\infty < y_i < y_j < \infty . \end{aligned}$$

Theorem 11.27 (Joint PDF of the order statistic). ***A:*** *Let $\vec{y} \in \mathbb{R}^n$ satisfy*

$$y_1 < y_2 < \cdots < y_n . \tag{11.59}$$

For the vector $\vec{Y} = \big(Y_1, \dots, Y_n\big)$, let $\vec{Y}_{(\bullet)}$ be the vector of its associated order statistics, i.e.,

$$\vec{Y}_{(\bullet)} = \big(Y_{(1)}, \dots, Y_{(n)}\big). \tag{11.60}$$

Then its density function at $\vec{y}$ is given by

$$f_{\vec{Y}_{(\bullet)}}(\vec{y}) = n! \cdot \prod_{j=1}^{n} f(y_j) = n! \, f(y_1) \cdots f(y_n) . \tag{11.61}$$

B: *If $\vec{y}$ does not satisfy (11.59), then $f_{\vec{Y}_{(\bullet)}}(\vec{y}) = 0$.*

11.10 The Bivariate Normal Distribution

Definition 11.19 (Bivariate normal distribution). $\boxed{\star}$ We say that two continuous random variables Y_1 and Y_2 have a **bivariate normal distribution**, or that they have a **joint normal distribution**, if their joint PDF is

$$f_{Y_1,Y_2}(y_1, y_2) = \frac{e^{-Q/2}}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}, \qquad -\infty < y_1 < \infty,\ -\infty < y_2 < \infty, \tag{11.62}$$

where $Q = \dfrac{1}{1-\rho^2}\left[\dfrac{(y_1-\mu_1)^2}{\sigma_1^2} - 2\rho\,\dfrac{(y_1-\mu_1)(y_2-\mu_2)}{\sigma_1\sigma_2} + \dfrac{(y_2-\mu_2)^2}{\sigma_2^2}\right]$.

We then also write $(Y_1, Y_2) \sim \mathcal{N}(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2, \rho)$. $\square$

Theorem 11.28. *If two random variables Y_1 and Y_2 are $\mathcal{N}(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2, \rho)$, then*

(a) $Y_1 \sim \mathcal{N}(\mu_1, \sigma_1^2$ *and* $Y_1 \sim \mathcal{N}(\mu_2, \sigma_2^2$.
Thus, $\mathbb{E}[Y_1] = \mu_1$, $Var[Y_1] = \sigma_1^2$, $\mathbb{E}[Y_2] = \mu_2$, $Var[Y_2] = \sigma_2^2$.

(b) $Cov[Y_1, Y_2] = \sigma_1\,\sigma_2\,\rho$. *Thus, ρ is the correlation coefficient of Y_1 and Y_2.*

Theorem 11.29. *If two jointly normal random variables Y_1 and Y_2 are uncorrelated, then they are independent.*

11.11 Blank Page after Ch.11

This page is intentionally left blank!

12 Functions of Random Variables and their Distribution

12.1 The Method of Distribution Functions

12.2 The Method of Transformations in One Dimension

Theorem 12.1. *Given are a continuous random variable Y with density $f_Y(y)$ and a differentiable function $h(y)$ which is either strictly increasing or strictly decreasing for all $y \in suppt(f_Y)$, i.e., for all y that satisfy $f_Y(y) > 0$. Then the PDF of $U := h(Y)$ is*

$$f_U(u) \;=\; f_Y\big(h^{-1}(u)\big) \cdot |h^{-1\prime}(u)| \;=\; f_Y\big(h^{-1}(u)\big) \cdot \left| \frac{d[h^{-1}(u)]}{du} \right| . \tag{12.1}$$

12.3 The Method of Transformations in Multiple Dimension

Theorem 12.2.

- *Let $\vec{Y} = \big(Y_1, \dots, Y_n\big)$ be a vector of randomvariables with joint PDF $f_{\vec{Y}}\big(\vec{y}\big)$ and let R be a "nice" subset of $\mathbb{R}^n$ which is so big that it hosts all outcomes $\vec{Y}(\omega)$ of $\vec{Y}$.*
- *Let the function $\vec{h} : R \to \mathbb{R}^n;\;\; \vec{y} \mapsto \vec{u} = \vec{h}\big(\vec{y}\big)$ satisfy the following.*
- ⊡ *$\vec{h}$ has continuous partial derivatives $\dfrac{\partial h_i}{y_j}$ for all $1 \le i, j \le n$.*
- ⊡ *If the vector $\vec{u}$ is a function value $\vec{u} = \vec{h}\big(\vec{y}\big)$ of some argument $\vec{y}$ that satisfies $f_{\vec{Y}}\big(\vec{y}\big) > 0$, then there is no other argument $\tilde{\vec{y}}$ that satisfies all those conditions.*

Then $\vec{h}$ has an inverse $\vec{h}^{-1} = h_1^{-1},\, h_2^{-1}, \dots, h_n^{-1}$ which is defined by the relation

$$\vec{u} \;=\; \vec{h}\big(\vec{y}\big) \;\Leftrightarrow\; \vec{y} \;=\; \vec{h}^{-1}\big(\vec{u}\big)\,.$$

We can write this for the coordinate functions $h_i(\cdot)$ and $h_j^{-1}(\cdot)$ as follows:

$$u_1 \;=\; h_1\big(\vec{y}\big),\; \dots,\; u_n \;=\; h_n\big(\vec{y}\big) \quad \textit{and} \quad y_1 \;=\; h_1^{-1}\big(\vec{u}\big),\; \dots,\; y_n \;=\; h_n^{-1}\big(\vec{u}\big). \tag{12.2}$$

Also, all partial derivatives $\dfrac{\partial h_i^{-1}}{u_j}$ exist and are continuous for $1 \le i, j \le n$.

$$\textit{Let}\quad \frac{d\vec{h}}{d\vec{y}} := \begin{bmatrix} \frac{\partial h_1}{\partial y_1} & \frac{\partial h_1}{\partial y_2} & \cdots & \frac{\partial h_1}{\partial y_n} \\ \frac{\partial h_2}{\partial y_1} & \frac{\partial h_2}{\partial y_2} & \cdots & \frac{\partial h_2}{\partial y_n} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial h_n}{\partial y_1} & \frac{\partial h_n}{\partial y_2} & \cdots & \frac{\partial h_n}{\partial y_n} \end{bmatrix}, \qquad \frac{d\vec{h}^{-1}}{d\vec{u}} := \begin{bmatrix} \frac{\partial h_1^{-1}}{\partial u_1} & \frac{\partial h_1^{-1}}{\partial u_2} & \cdots & \frac{\partial h_1^{-1}}{\partial u_n} \\ \frac{\partial h_2^{-1}}{\partial u_1} & \frac{\partial h_2^{-1}}{\partial u_2} & \cdots & \frac{\partial h_2^{-1}}{\partial u_n} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial h_n^{-1}}{\partial u_1} & \frac{\partial h_n^{-1}}{\partial u_2} & \cdots & \frac{\partial h_n^{-1}}{\partial u_n} \end{bmatrix}. \tag{12.3}$$

$$\textit{Let}\quad J^{-1} \;:=\; J^{-1}(\vec{y}) \;:=\; \textit{det}\left(\frac{d\vec{h}}{d\vec{y}}\right), \qquad J \;:=\; J(\vec{u}) \;:=\; \textit{det}\left(\frac{d\vec{h}^{-1}}{d\vec{u}}\right). \tag{12.4}$$

- *We add another assumption:* $J^{-1}(\vec{y}) \neq 0$ *for all* y *that satisfy* $f_{\vec{Y}}(\vec{y}) > 0$.

(12.5) $$\text{Then} \quad J\big(h(\vec{y})\big) \neq 0 \quad \text{and} \quad J\big(h(\vec{y})\big) = 1\Big/J^{-1}\big(\vec{y}\big)\,.$$

Further, the density of the transform $\vec{U} = h\big(\vec{Y}\big)$ *is computed as*

(12.6) $$f_{\vec{U}}(\vec{u})\big) = f_{\vec{Y}}\big(h^{-1}(\vec{u})\big) \cdot |J(\vec{u})|\,.$$

Definition 12.1 (Jacobian and Jacobian matrix)**.** The matrix $\dfrac{d\vec{h}}{d\vec{y}}$ of the partial derivatives of the function $\vec{y} \mapsto \vec{h}(\vec{y})$ is called the **Jacobian matrix** of $\vec{h}(\cdot)$. We refer to its determinant, $J^{-1}(\vec{y}) = \det\left(\dfrac{d\vec{h}}{d\vec{y}}\right)$, as the **Jacobian**, sometimes also the **Jacobian determinant**, of $\vec{h}(\cdot)$. □

Notation 12.1 (Jacobian)**.**

- Stewart writes $\dfrac{\partial(u_1,\dots,u_n)}{\partial(y_1,\dots,y_n)} := \det\left(\dfrac{d\vec{h}^{-1}}{d\vec{u}}\right)$ and $\dfrac{\partial(y_1,\dots,y_n)}{\partial(u_1,\dots,u_n)} := \det\left(\dfrac{d\vec{h}^{-1}}{d\vec{u}}\right)$
- Thus, the expression $J = J(\vec{u}) = \det\left(\frac{d\vec{h}^{-1}}{d\vec{u}}\right)$, which appears in
  (12.6) $f_{\vec{U}}(\vec{u})\big) = f_{\vec{Y}}\big(h^{-1}(\vec{u})\big) \cdot |J(\vec{u})|$, is the Jacobian of $h^{-1}(\vec{u})$ and not of $h(\vec{y})$
  .
- This author follows the great majority of books on multivariable calculus in defining the the Jacobian as the determinant of $\dfrac{d\vec{h}}{d\vec{y}}$. □
- Be aware that WMS chooses instead to call $J = \det\dfrac{d\vec{h}^{-1}}{d\vec{u}}$ the Jacobian.
- The reason seems to be that most books on probability and statistics agree on using the letter J for $\det\dfrac{d\vec{h}^{-1}}{d\vec{u}}$ (without giving a name to that determinant) and WMS does not want to use the somewhat lengthy "the reciprocal of the Jacobian" in its frequent references to J

12.4 The Method of moment–generating Functions

Assumption 12.1. Unless stated otherwise, we will assume in this entire section that

(a) $\vec{Y} = (Y_1, Y_2, \dots, Y_n)$ denotes a list of n random variables ($n \in \mathbb{N}$).
• Either all Y_j are discrete, or they all are continuous random variables.

(b) $h : D \to \mathbb{R}; \quad \vec{y} \mapsto u = h(\vec{y}) = h(y_1, \dots, y_n)$
is a function with domain $D \subseteq \mathbb{R}^n$ (this covers $\mathbb{R} = \mathbb{R}^1$ for $n = 1$), such that
• there is no issue with the existence of the PMF or PDF of $U := h(\vec{Y})$.
• All MGFs, $m_{Y_j}(t) = \mathbb{E}\big[e^{tY_j}\big]$ and $m_U(t) = \mathbb{E}\big[e^{tU}\big]$ exist if $|t|$ is small enough, i.e., there is some $\delta > 0$ such that those MGFs exist for $-\delta < t < \delta$.

(c) Those assumptions also hold for differently named (vectors of) random variables and functions, e.g. $V = g\big(\vec{\widetilde{Y}}\big) = g\big(\widetilde{Y}_1, \dots, \widetilde{Y}_k\big)$. □

Theorem 12.3 (The MGF determines the distribution)**.** *Given are two random variables Y and $\widetilde{Y}$. If their moment–generating functions $m_Y(t)$ and $m_{\widetilde{Y}}(t)$ exist and coincide in a small interval that is centered at $t = 0$,*

- *Then $\mathbb{P}_Y = \mathbb{P}_{\widetilde{Y}}$, i.e., Y and $\widetilde{Y}$ have the same probability distribution.*

Theorem 12.4 (MGF of a sum of functions of independent variables)**.** *Given are n independent random variables $Y_1, Y_2, \dots, Y_n$ with MGFs $m_{Y_1}(t), m_{Y_2}(t), \dots, m_{Y_n}(t)$. and n real–valued functions $h_1(y_1), \dots, h_n(y_n)$ of real numbers $y_1, \dots, y_n$.*
Let $U := h_1(Y_1) + h_2(Y_2) + \cdots + h_n(Y_n)$. Then (under the conditions of Assumption 12.1 on 101)

$$m_U(t) = m_{h_1(Y_1) + \cdots + h_n(Y_n)} = \prod_{j=1}^{n} m_{h_j(Y_j)}(t)\,. \tag{12.7}$$

Corollary 12.1 (WMS Ch.06.5, Theorem 6.2)**.** *Let $Y_1, Y_2, \dots, Y_n$ be independent random variables with moment–generating functions $m_{Y_1}(t), m_{Y_2}(t), \dots, m_{Y_n}(t)$, respectively. Then*

$$m_{Y_1 + \cdots + Y_n}(t) = \prod_{j=1}^{n} m_{Y_j}(t) = m_{Y_1}(t) \cdot m_{Y_2}(t) \cdots m_{Y_n}(t)\,. \tag{12.8}$$

Theorem 12.5 (Linear combinations of uncorrelated normal variables are normal)**.**

Given are n uncorrelated, $\mathcal{N}(\mu_j, \sigma_j^2)$ random variables Y_j, ($j = 1, \dots, n$. In other words, each Y_j is normal with expectation μ_j and standard deviation σ_j. Let $a_1, \dots, a_n \in \mathbb{R}$. Then

$$\sum_{j=1}^{n} a_j Y_j \sim \mathcal{N}\left(\sum_{j=1}^{n} a_j \mu_j \,, \sum_{j=1}^{n} a_j^2 \sigma_j^2 \right). \tag{12.9}$$

Thus, the linear combination of uncorrelated normal random variables is normal with expectation and variance being the linear combinations of the indivicual expectations and variances.

Theorem 12.6. *Given are n independent, gamma(α_j, β) random variables Y_j, $(j = 1, \dots, n$. In other words, each Y_j is gamma with the same scale parameter β. Then*

$$\sum_{j=1}^{n} Y_j \sim \textit{gamma}\left(\sum_{j=1}^{n} \alpha_j\,, \beta\right) . \tag{12.10}$$

Thus, the sum of independent gamma random variables with the same scale parameter β is gamma with the shape parameter being the sum of the shape parameters, and scale parameter β.

Corollary 12.2. *Let $Y_1, Y_2, \dots, Y_n$ be independent χ^2 variables such that each Y_j has ν_j degrees of freedom. Then*

$$m_{Y_1 + \dots + Y_n}(t) \sim \chi^2 \left(\sum_{j=1}^{n} \nu_j \; \textit{df}\right) . \tag{12.11}$$

13 Limit Theorems

13.1 Four Kinds of Limits for Sequences of Random Variables

Definition 13.1 (Convergence of Random Variables)**.** Let Y_n ($n \in \mathbb{N}$) and Y be random variables on a probability space $(\Omega, \mathbb{P})$. We define

$$Y_n \overset{\mathbf{pw}}{\to} Y \text{ or } \text{pw} - \lim_{n\to\infty} Y_n = Y\,, \quad \text{if } \lim_{n\to\infty} Y_n(\omega) = Y(\omega), \text{ for all } \omega \in \Omega\,, \tag{13.1}$$

$$Y_n \overset{\mathbf{a.s.}}{\to} Y \text{ or } \text{a.s.} - \lim_{n\to\infty} Y_n = Y\,, \quad \text{if } \mathbb{P}\{\omega \in \Omega : \lim_{n\to\infty} Y_n(\omega) = Y(\omega)\} = 1\,, \tag{13.2}$$

$$Y_n \overset{\mathbf{P}}{\to} Y \text{ or } \text{P} - \lim_{n\to\infty} Y_n = Y\,, \quad \text{if } \forall\, \varepsilon > 0 \lim_{n\to\infty} \mathbb{P}\{\omega \in \Omega : |Y_n(\omega) - Y(\omega)| > \varepsilon\} = 0\,, \tag{13.3}$$

$$Y_n \overset{\mathbf{D}}{\to} Y, \text{ if } \lim_{n\to\infty} F_{Y_n}(y) = F_Y(y),\ \forall\, y \in \mathbb{R} \text{ where the CDF } F_Y \text{ of } Y \text{ is continuous.} \tag{13.4}$$

We also say:

If $Y_n \overset{\mathbf{pw}}{\to} Y$, Y is the **pointwise limit** of the Y_n, or: Y_n **converges pointwise** to Y.
If $Y_n \overset{\mathbf{a.s.}}{\to} Y$, Y is the **almost sure limit** of the Y_n, or: Y_n **converges almost surely** to Y.
If $Y_n \overset{\mathbf{P}}{\to} Y$, Y is the **limit in probability**; of the Y_n, or: Y_n **converges in probability** to Y.
If $Y_n \overset{\mathbf{D}}{\to} Y$, Y is the **limit in distribution** of the Y_n, or: Y_n **converges in distribution** to Y.

Theorem 13.1 (Relationship between the modes of convergence)**.**

Let Y and $Y_1, Y_2, \dots$ be random variables on a probability space $(\Omega, \mathbb{P})$. Then,

$$Y_n \overset{pw}{\to} Y \;\Rightarrow\; Y_n \overset{a.s.}{\to} Y \;\Rightarrow\; Y_n \overset{P}{\to} Y \;\Rightarrow\; Y_n \overset{D}{\to} Y\,. \tag{13.5}$$

Theorem 13.2 (Slutsky's Theorem)**.** $\boxed{\star}$ *Let $Y_1, Y_2, \dots$) and $U_1, U_2, \dots$ be two sequences of random variables. Let Y be another random variable and c a constant such that*

- $Y_n \xrightarrow{D} Y$ *(convergence in distribution)*
- $U_n \xrightarrow{P} c$ *(convergence in probability)*

Then,

$$Y_n + U_n \xrightarrow{D} Y + c\,, \tag{13.6}$$

$$Y_n \cdot U_n \xrightarrow{D} cY\,, \tag{13.7}$$

$$\frac{Y_n}{U_n} \xrightarrow{D} \frac{Y_n}{c}\,, \quad \textit{assuming that } c \neq 0. \tag{13.8}$$

Theorem 13.3 (Convergence is maintained under continuous transformations). $\boxed{\star}$
Let $Y_1, Y_2, \dots)$ and Y be random variables on some probability space $(\Omega, \mathbb{P})$. Let $f : \mathbb{R} \to \mathbb{R}$ be continuous. Then,

$$Y_n \xrightarrow{a.s.} Y \Rightarrow f \circ Y_n \xrightarrow{a.s.} f \circ Y\,.$$
$$Y_n \xrightarrow{P} Y \Rightarrow f \circ Y_n \xrightarrow{P} f \circ Y\,.$$
$$Y_n \xrightarrow{D} Y \Rightarrow f \circ Y_n \xrightarrow{D} f \circ Y\,.$$

□

13.2 Two Laws of Large Numbers

Theorem 13.4 (Weak Law of Large Numbers). *Let $Y_1, Y_2, \dots$ be an iid sequence of random variables on a probability space $(\Omega, \mathbb{P})$ with finite variance: $\sigma^2 := var[Y_n] < \infty$. Let $\mu := \mathbb{E}[Y_n]$. Then,*

$$\frac{Y_1 + Y_2 + \cdots + Y_n}{n} \text{ converges in probability to } \mu, \quad \text{i.e.,}$$

(13.9)
$$\big[\,\varepsilon > 0\,\big] \quad \Rightarrow \quad \left[\lim_{n\to\infty} \mathbb{P}\left\{\left|\frac{1}{n}\sum_{j=1}^{n} Y_j - \mu\right| > \varepsilon\right\} \;=\; 0.\right]$$

Theorem 13.5 (Strong Law of Large Numbers). *Let $Y_1, Y_2, \dots$ be an iid sequence of random variables on a probability space $(\Omega, \mathbb{P})$ and $\mu := \mathbb{E}[Y_n]$. Then,*

$$\frac{Y_1 + Y_2 + \cdots + Y_n}{n} \text{ converges almost surely to } \mu, \quad \text{i.e.,}$$

(13.10)
$$\mathbb{P}\left\{\lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} Y_j \;\neq\; \mu\right\} \;=\; 0\,.$$

13.3 Sampling Distributions

Definition 13.2 (Random samples from a distribution).
Let Y be a random variable on a probability space $(\Omega, \mathbb{P})$. Let $n \in \mathbb{N}$. We call a vector $\vec{Y} = \big(Y_1, \dots, Y_n\big)$ a **random sampling action of size n on** (or **from) the distribution of** Y, if

- the random variables $Y_1, \dots, Y_n$ are iid with distribution $\mathbb{P}_Y$.

The following are alternate names for this kind of sampling action:

- **random sampling action of size** n **on** (or **from)** Y
- "random sampling action" can be shortened to "**random sample**"
- **random sample** also refers to a realization $\vec{y} = \vec{Y}(\omega)$ of a random sampling action.

Note that the last two bulleted items are consistent with earlier definitions of sampling where we also use "sample" both for a sampling action and a realization of such an action. ☐

Definition 13.3 (Statistic). Let Y be a random variable on a probability space $(\Omega, \mathbb{P})$ and $\vec{Y} = \big(Y_1, \dots, Y_n\big)$ a random sampling action on Y. Let

$$T : \mathbb{R}^n \mapsto \mathbb{R}\,; \qquad \vec{y} \mapsto T(\vec{y})$$

be some function that can be applied to the sampling action $\vec{Y}$. We call the random variable

$$\omega \mapsto T\big(\vec{Y}(\omega)\big)$$

a **statistic** of that sampling action. We call the distribution of that random variable,

$$B \;\mapsto\; \mathbb{P}_{T \circ \vec{Y}}(B) \;=\; \mathbb{P}\{T(\vec{Y}) \in (B)\} \;=\; \mathbb{P}\{\omega \in \Omega : T\big(\vec{Y}(\omega)\big) \in B\} \tag{13.11}$$

its **sampling distribution**. Once the sampling action has been performed and a realization $\vec{y} = Y(\vec{\omega})$ has been obtained, we call $t \;=\; T\big(\vec{Y}(\omega)\big)$ the realization of the statistic. ☐

Theorem 13.6. *Let Y be a random variable on a probability space $(\Omega, \mathbb{P})$ and $\vec{Y} = \big(Y_1, \dots, Y_n\big)$ a random sampling action on Y. Let $T_1, T_2, \dots, T_k : \mathbb{R}^n \mapsto \mathbb{R}$ be statistics for that sample action. Let*

$$T^* : \mathbb{R}^k \mapsto \mathbb{R}\,; \qquad (t_1, \dots, t_k) \mapsto T^*(t_1, \dots, t_k)\,.$$

Then, setting $\vec{t} = (t_1, \dots, t_k)$ and $\vec{T} = (T_1, \dots, T_k)$, the composition

$$T^* \circ \vec{T} \circ \vec{Y} \;:\; \omega \mapsto T^*\big(\vec{T}[\vec{Y}(\omega)]\big) \;=\; T^*\big(T_1[\vec{Y}(\omega)], \dots, T_k[\vec{Y}(\omega)]\big)$$

also is a statistic of $\vec{Y}$.

A function of a function of the data is a function of the data.

Definition 13.4 (Sample variance). Let $\vec{Y} = \big(Y_1, \dots, Y_n\big)$ be a random sample action on a random variable Y.

The **sample variance** is defined as the random variable

(13.12)
$$\omega \mapsto S^2(\omega) \;:=\; \frac{1}{n-1}\sum_{j=1}^{n}\big(Y_j(\omega)-\bar{Y}(\omega)\big)^2 .$$

We further call $\omega \mapsto S(\omega) := \sqrt{S^2(\omega)}$ the The **sample standard deviation**.
We will often write s^2 and s for the realizations $S^2(\omega)$ and $S(\omega)$ that result from creating the sample.
We write S_n, S_n^2, s_n, s_n^2 for S, S^2, s, s^2, if we want to keep track of the sample size. That will be the case, e.g., if we consider the sample variance of the first n picks of a sample of infinite size. $\square$

Theorem 13.7 (WMS Ch.07.2, Theorem 7.1)**.** *Let $Y_1, Y_2, \dots, Y_n$ be a random sample of size n from a normal distribution with mean μ and variance σ^2, i.e., we sample on a random variable $Y \sim \mathcal{N}(\mu, \sigma^2)$. Then the sample mean $\bar{Y}$ follows a normal distribution with mean μ and variance σ^2/n.*

Theorem 13.8 (WMS Ch.07.2, Theorem 7.2)**.** *Let $\vec{Y} = \big(Y_1, \dots, Y_n\big)$ be a random sample on $Y \sim \mathcal{N}(\mu, \sigma^2)$. Let $Z_j = (Y_j - \mu)/\sigma$ for $j = 1, 2, \dots, n$. Then $\vec{Z} = \big(Z_1, \dots, Z_n\big)$ is a random sample on a standard normal variable. (In particular, the Z_j are iid.) Further,*

(13.13)
$$\sum_{j=1}^{n} Z_i^2 \;=\; \sum_{j=1}^{n}\left(\frac{Y_j - \mu}{\sigma}\right)^2$$

follows a χ^2 distribution with n degrees of freedom.

Proposition 13.1. $\boxed{\star}$ *Let Y_1 and Y_2 be independent standard normal random variables. Then $Y_1 + Y_2$ and $Y_1 - Y_2$ are independent and normally distributed, both with mean 0 and variance 2.*

Theorem 13.9 (Independence of sample mean and sample variance in normal populations)**.**
Let $\vec{Y} = \big(Y_1, \dots, Y_n\big)$ be a random sample on $Y \sim \mathcal{N}(\mu, \sigma^2)$. Let $Z_j = (Y_j - \mu)/\sigma$ for $j = 1, \dots, n$. Then, $\vec{Z} = (Z_1, \dots, Z_n)$ is a random sample on a standard normal variable. Moreover,

(a) $$\frac{(n-1)S^2}{\sigma^2} \;=\; \frac{1}{\sigma^2}\sum_{j=1}^{n}(Y_j - \bar{Y})^2 \;\sim\; \chi^2(df\; = n-1)$$

(b) *$\bar{Y}$ and S^2 are independent random variables.*

Definition 13.5 (Student's t–distribution)**.** Let Z and W be independent random variables such that Z is standard normal and$W$ is χ^2 with ν df. Let

$$T = \frac{Z}{\sqrt{W/\nu}} \tag{13.14}$$

Then we refer to the distribution $\mathbb{P}_T$ of T as a **t–distribution** or **Student's t–distribution** with ν df. We also write that as $T \sim t(\nu)$ or $T \sim t(\text{df} = \nu)$. ☐

The Student's t–distribution is named after the English statistician William S. Gosset (1876 – 1937). Gosset was Head Brewer of the Guinness Brewery in Dublin, Ireland and published his papers under the pseudonym "Student".

Theorem 13.10. *Let $Y \sim \mathcal{N}(\mu, \sigma^2)$ and $\vec{Y} = \big(Y_1, \dots, Y_n\big)$ be a random sample on Y. Let*

$$T := \frac{\bar{Y} - \mu}{S/\sqrt{n}}. \tag{13.15}$$

Then T follows a t–distribution with $(n-1)$ df.

Definition 13.6 (F–distribution)**.** Given are two independent random variables $W_1 \sim \chi^2(\text{df} = \nu_1)$ and $W_2 \sim \chi^2(\text{df} = \nu_2)$. with ν_1 and ν_2 df, respectively. Then we say that

$$F = \frac{W_1/\nu_1}{W_2/\nu_2}$$

follows an **F distribution** with ν_1 **numerator degrees of freedom** and ν_2 **denominator degrees of freedom**. ☐

Theorem 13.11. *Consider two random samples of sizes n_1 and n_2 from two independent populations, on random variables $Y_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $Y_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ with sample variances S_1^2 and S_2^2. Let*

$$F := \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}. \tag{13.16}$$

Then F follows an F distribution with $(n_1 - 1)$ numerator df and $(n_2 - 1)$ denominator df.

13.4 The Central Limit Theorem

Theorem 13.12 (Lévy–Cramér continuity theorem)**.** $\boxed{\star}$ *Let $Y_1, Y_2, \dots)$ be a sequence of random variables (iid is not assumed) with associated CDFs $F_{Y_1}, F_{Y_2}, \dots)$ and MGFs $m_{Y_1}(t), m_{Y_2}(t), \dots)$. Let Y be a random variable with associated CDF F_Y and MGF $m_Y(t)$. Then,*

$$\begin{aligned} &\big[\, m_{Y_n}(t) \to m_Y(t) \text{ as } n \to \infty, \ \text{ for all } t \in \mathbb{R} \,\big] \\ \Rightarrow\ &\big[\, F_{Y_n}(y) \to F_Y(y) \text{ as } n \to \infty, \ \text{ for all } y \text{ where } F_Y(\cdot) \text{ is continuous.} \big] \end{aligned} \tag{13.17}$$

Theorem 13.13 (Central Limit Theorem). ***Central Limit Theorem:***
Let $\vec{Y} = (Y_1, Y_2, \dots, Y_n)$ be a vector of iid random variables with common expectation $\mathbb{E}[Y_j] = \mu$ and finite variance $Var[Y_j] = \sigma^2$. Let Z be a standard normal variable and

$$U_n := \frac{\sum_{j=1}^{n} Y_j - n\mu}{\sigma \cdot \sqrt{n}} = \frac{\bar{Y}_n - \mu}{\sigma/\sqrt{n}}, \qquad \text{where } n \in \mathbb{N},\ \bar{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i .$$

Then, U_n converges to Z in distribution as $n \to \infty$. In other words,

$$\lim_{n\to\infty} \mathbb{P}\{U_n \le u\} = \mathbb{P}\{Z \le u\} = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt \qquad \text{for all } u .$$

Remark 13.1. Note that the CLT states the obvious if the iid sample picks Z_j are $\mathcal{N}(\mu, \sigma^2)$:

- In this case, $\bar{Y}_n \sim \mathcal{N}(\mu, \sigma^2/n)$. Hence, $U_n = \dfrac{\bar{Y}_n - \mathbb{E}[\bar{Y}_n]}{\sigma_{\bar{Y}_n}} \sim \mathcal{N}(0,1)$.
- Thus, $F_{U_n}(y) = F_Z(y)$, for all y and n. Thus, $\lim_{n\to\infty} F_{U_n}(y) = F(y)$, for all y. $\square$

Theorem 13.14 (Student t converges to normal distribution). *Let $T_1, T_2, \dots)$ be a sequence of random variables such that $T_j \sim t(df = j)$. Then T_j converges in distribution to a standard normal variable.*

Lemma 13.1. $\boxed{\star}$ *Let $\vec{y} := (y_1, \dots, y_n) \in \mathbb{R}^n, (n \in \mathbb{N})$, and $\bar{y} := \frac{1}{n}\sum_{j=1}^{n} y_j$ the arithmetic mean of $\vec{y}$. Then,*

(a)
$$\sum_{j=1}^{n}(y_j - c)^2 = \sum_{j=1}^{n}(y_j - \bar{y})^2 + \sum_{j=1}^{n}(\bar{y} - c)^2 ,$$

(b) *$\bar{y}$ minimizes the expression $\sum_{j=1}^{n}(y_j - c)^2$, where $c \in \mathbb{R}$):*

$$\sum_{j=1}^{n}(y_j - c)^2 \ge \sum_{j=1}^{n}(y_j - \bar{y})^2 \quad \text{for all } c \in \mathbb{R},$$

Corollary 13.1. $\boxed{\star}$ *The sample variance $S^2 = \dfrac{1}{n-1}\sum_{j=1}^{n}\left(Y_j - \bar{Y}\right)^2$ of any sample $\vec{Y} := (Y_1, \dots, Y_n)$, $(n \in \mathbb{N})$, satisfies*

$$(n-1)S^2 = \sum_{j=1}^{n} Y_j^2 - n\bar{Y}^2 .$$

Theorem 13.15 (Sample variance converges to population variance)**.**
Let $\vec{Y} := (Y_1, \dots, Y_n) \in \mathbb{R}^n, (n \in \mathbb{N})$*, be a random sample from the distribution of a random variable* Y *with finite variance* $\sigma^2 < \infty$*. Then the sample variance* $S_n^2 = \frac{1}{n-1}\sum_{j=1}^{n}\left(Y_j - \bar{Y}\right)^2$ *converges a.s (hence, also in probability and in distribution) to* σ^2*.*

Theorem 13.16 (CLT – Sample variance version)**.** *Let* $\vec{Y} = (Y_1, Y_2, \dots, Y_n)$ *be a vector of iid random variables with common expectation* $\mathbb{E}[Y_j] = \mu$ *and finite variance* $Var[Y_j] = \sigma^2$*. Let* Z *be a standard normal variable. For* $n \in \mathbb{N}$*, let*

$$\bar{Y}_n := \frac{1}{n}\sum_{i=1}^{n} Y_i\,, \quad S_n^2 := \frac{1}{n-1}\sum_{i=1}^{n}\left(Y_i - \bar{Y}_n\right)^2, \quad S_n := \sqrt{S^2}\,, \quad W_n := \frac{\bar{Y}_n - \mu}{S_n/\sqrt{n}}\,.$$

(Thus, $\bar{Y}_n$ *and* S_n *are sample mean and sample standard deviation of the RSA* $\vec{Y}$*).*

Then W_n *converges to* Z *in distribution as* $n \to \infty$*.*

14 Sample Problems for Exams

14.1 Practice Midterm 1 for Math 447 - Chris Haines

Here are some commented excerpts of a practice exam for the first midterm. It was written by Prof. Christopher Haines and forwarded to me by Prof. Adam Weisblat, both at Binghamton University (October 2023).

Exercise 14.1. Practice Midterm 1 (C. Haines) – # 01

SKIPPED □

Exercise 14.2. Practice Midterm 1 (C. Haines) – # 02

The Lakers and Heat are playing in the NBA Finals. The series is a best–of–seven (first team to win four games clinches the series). The Lakers will win each game with probability $3/4$.

- **(a)** Given that the Heat won game one, what is the probability the Lakers go on to win the series?
- **(b)** Given that the Heat win at least two games in the series, what is the probability the Lakers go on to win the series?

□

15 Other Appendices

15.1 Greek Letters

The following section lists all greek letters that are commonly used in mathematical texts. You do not see the entire alphabet here because there are some letters (especially upper case) which look just like our latin alphabet letters. For example: $A = \text{Alpha}$ $\quad B = \text{Beta}$. On the other hand there are some lower case letters, namely epsilon, theta, sigma and phi which come in two separate forms. This is not a mistake in the following tables!

α	alpha	θ	theta	ξ	xi	ϕ	phi
β	beta	ϑ	theta	π	pi	φ	phi
γ	gamma	ι	iota	ρ	rho	χ	chi
δ	delta	κ	kappa	ϱ	rho	ψ	psi
ϵ	epsilon	$\varkappa$	kappa	σ	sigma	ω	omega
ε	epsilon	λ	lambda	ς	sigma		
ζ	zeta	μ	mu	τ	tau		
η	eta	ν	nu	υ	upsilon		

Γ	Gamma	Λ	Lambda	Σ	Sigma	Ψ	Psi
Δ	Delta	Ξ	Xi	Υ	Upsilon	Ω	Omega
Θ	Theta	Π	Pi	Φ	Phi		

15.2 Notation

This appendix on notation has been provided because future additions to this document may use notation which has not been covered in class. It only covers a small portion but provides brief explanations for what is covered.

For a complete list check the list of symbols and the index at the end of this document.

Notation 15.1. **a)** If two subsets A and B of a space Ω are disjoint, i.e., $A \cap B = \emptyset$, then we often write $A \uplus B$ rather than $A \cup B$ or $A + B$. The complement $\Omega \setminus A$ of A is denoted $A^{\complement}$. .

b) $\mathbb{R}_{>0}$ or $\mathbb{R}^+$ denotes the interval $]0, +\infty[$, $\mathbb{R}_{\geq 0}$ or $\mathbb{R}_+$ denotes the interval $[0, +\infty[$,

c) The set $\mathbb{N} = \{1, 2, 3, \cdots\}$ of all natural numbers excludes the number zero. We write $\mathbb{N}_0$ or $\mathbb{Z}_+$ or $\mathbb{Z}_{\geq 0}$ for $\mathbb{N} \uplus \{0\}$. $\mathbb{Z}_{\geq 0}$ is the B/G notation. It is very unusual but also very intuitive. $\square$

References

List of Symbols

$A_n \downarrow A$ – nonincreasing set seq. , 15
$A_n \uparrow A$ – nondecreasing set seq. , 15
$F_Y(y)$ – CDF of random var. Y , 73
$[a, b[,\ ]a, b]$ – half-open intervals , 11
$[a, b]$ – closed interval , 11
$\binom{n}{r}$ – nbr of combinations , 59
$\mathbb{P}_r^n$ – permutation , 58
$\Rightarrow$ – implication , 7
$\mathfrak{P}(\Omega), 2^\Omega$ – power set , 9
$\emptyset$ – empty set, 6
$\exists$ – exists , 10
$\exists!$ – exists unique , 10
$\forall$ – for all , 10
$\inf\,(x_i),\ \inf\,(x_i)_{i\in I},\ \inf\limits_{i\in I} x_i$ – families , 20
$\inf\,(x_n),\ \inf\,(x_n)_{n\in\mathbb{N}},\ \inf\limits_{n\in\mathbb{N}} x_n$ – sequences , 20
$\inf(A)$ – infimun of A , 19
$\sup\,(x_n),\ \sup\,(x_n)_{n\in\mathbb{N}},\ \sup\limits_{n\in\mathbb{N}} x_n$ – sequences , 20
$\sup(A)$ – supremun of A , 19
$|x|$ – absolute value , 12
$]a, b[_{\mathbb{Q}}$ – interval of rational #s , 12
$]a, b[_{\mathbb{Z}}$ – interval of integers , 12
$]a, b[$ – open interval , 11
$x \in X$ – element of a set, 6
$x \notin X$ – not an element of a set, 6
$x_n \downarrow x$ – nonincreasing seq. , 15
$x_n \uparrow x$ – nondecreasing seq. , 15
$A^\complement$ – complement of A , 8
$\mathbb{N}_0$ – nonnegative integers, 11
$\mathbb{R}^+$ – positive real numbers, 11
$\mathbb{R}_{>0}$ – positive real numbers, 11
$\mathbb{R}_{\geq 0}$ – nonnegative real numbers, 11
$\mathbb{R}_{\neq 0}$ – non-zero real numbers, 11
$\mathbb{R}_+$ – nonnegative real numbers, 11
$\mathbb{Z}_{\geq 0}$ – nonnegative integers, 11
$\mathbb{Z}_+$ – nonnegative integers, 11

$(x_i)_{i\in I}$ – family , 16
$2^\Omega, \mathfrak{P}(\Omega)$ – power set , 9
$\binom{n}{n_1\, n_2 \cdots n_k}$ – multinom. coeff. , 59
$\binom{n}{k}$ – binomial coeff. , 59
$\mathbb{E}[g(Y_1) \mid Y_2 = y_2]$ – conditional expectation , 94
$\mathbb{E}(Y)$ – expected value , 57
$\mathbb{E}[Y]$ – expected value , 64
μ_k – kth central moment , 78
μ'_k – kth moment , 78
μ'_k – kth moment , 71
μ_k – kth central moment , 71
ϕ_p – pth quantile , 74
ρ – correlation coeff. , 91
σ_Y^2 – variance, discr. r.v. , 66
σ_Y – standard dev, discr. r.v. , 57, 66
$\text{binom}(n, p)$, 67
$Cov[Y_1, Y_2]$ – covariance , 91
$E(Y)$ – expected value , 76
$m(t)$ – MGF , 71
S, S_n – sample standard deviation , 107
s, s_n – sample standard deviation , 107
S^2, S_n^2 – sample variance , 107
s^2, s_n^2 – sample variance , 107
$SD(Y)$ – standard dev, discr. r.v. , 66
$SD[Y]$ – standard dev, discr. r.v. , 57
$Var[Y_1 \mid Y_2 = y_2]$ – conditional variance , 94
$Var[Y]$ – variance , 57
$Var[Y]$ – variance, discr. r.v. , 66
$Y_n \xrightarrow{\text{a.s.}} Y$ – almost sure limit , 104
$Y_n \xrightarrow{\text{D}} Y$ – limit in distrib. , 104
$Y_n \xrightarrow{\text{pw}} Y$ – pointwise limit , 104
$Y_n \xrightarrow{\text{P}} Y$ – limit in probab. , 104
$\Gamma(\alpha)$ – gamma function , 80
$\inf\limits_{x\in A} f(x)$ – infimum of f , 20
$\inf_A f$ – infimum of f , 20
$\mathbb{N}, \mathbb{N}_0$, 112
$\mathbb{P}(A \mid B)$ – conditional probab , 39, 86, 87
$\mathbb{R}^+, \mathbb{R}_{>0}$, 112
$\mathbb{R}_+, \mathbb{R}_{\geq 0}$, 112
$\mathbb{R}_{>0}, \mathbb{R}^+$, 112
$\mathbb{R}_{\geq 0}, \mathbb{R}_+$, 112
$\mathbb{Z}_+, \mathbb{Z}_{\geq 0}$, 112
$\overline{\mathbb{R}}$ – $[-\infty, \infty]$, 11
$\mathbf{1}_A$ – indicator function of A , 21
$\sigma\{\mathscr{A}\}$ – σ–algebra generated by $\mathscr{A}$, 38
$\sup\,(x_i),\ \sup\,(x_i)_{i\in I},\ \sup\limits_{i\in I} x_i$ – families , 20
$\sup\limits_{x\in A} f(x)$ – supremum of f , 20
$\sup_A f$ – supremum of f , 20
$\text{suppt}(f)$ – support of f , 34

$|X|$ – size of a set , 10
$\mathscr{N}(\mu, \sigma^2)$ – normal with μ, σ^2 , 79
$\mathscr{N}(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2, \rho)$ – bivariate normal , 98
$\mathfrak{B}$ – Borel σ–algebra of $\mathbb{R}$, 38
$\mathfrak{B}, \mathfrak{B}^d$ – Borel sets , 29
$\mathfrak{B}^d$ – Borel σ–algebra of $\mathbb{R}^d$, 38
$\{\}$ – empty set, 6
$A \cap B$ – A intersection B, 7
$A \setminus B$ – A minus B , 8
$A \subseteq B$ – A is subset of B , 6
$A \subsetneq B$ – A is strict subset of B, 6
$A \triangle B$ – symmetric difference of A and B , 8
$A \uplus B$ – disjoint union , 5, 7
$A^{\complement}$ – complement , 112
$B \supsetneq A$ – B is strict superset of A, 6
$B(\alpha, \beta)$, 82
$f : X \to Y$ – function, 13
$f(A)$ – direct image , 19
$f^{-1}(B)$ – indirect image, preimage , 17
$f_{Y_1|Y_2}(y_1 \mid y_2)$ – conditional PDF , 87
$(\Omega, \mathfrak{F}, \mathbb{P})$ – probability space , 35
$(\Omega, \mathfrak{F}, \mu)$ – measure space , 51
$(S, \mathscr{S}, \mathbb{P})$ – sample space , 36
χ^2(df=ν) – chi–square with ν df , 81
$\chi^2(\nu)$ – chi–square with ν df , 81
$\cup_{j \in J} A_j$ – union of all A_j w. $j \in J$, 7
$\cup_{j \in J} A_j$ – union of all A_j w. $j \in J$, 16
$\mu(\cdot)$ – measure , 50
$\sigma\{\mathscr{A}\}$ – σ–algebra generated by $\mathscr{A}$, 37
$\Sigma_*(\cdot)$ – counting measure , 51
$\uplus_{j \in J} A_j$ – union of disjoint sets , 7
$|f|, f^+, f^-$, 12
$A \cup B$ – A union B , 7
$A \supseteq B$ – A is superset of B, 6
$cap_{j \in J} A_j$ – intersection of all A_j w. $j \in J$, 7
$cap_{j \in J} A_j$ – intersection of all A_j; ($j \in J$) , 16
$f\big|_A$ – restriction of f , 14
$f \vee g, f \wedge g$ – $\max(f, g), \min(f, g)$, 13
$F_{Y_1,Y_2}(y_1, y_2)$ – joint CDF , 84
P – measure , 35
$p_{Y_1,Y_2}(y_1, y_2)$ – joint PMF , 84
$uplus_{j \in J} A_j$ – – union of disjoint sets , 16
$x \vee y$ – $\max(x, y)$, 13
$x \wedge y$ – $\min(x, y)$, 13
x^+, x^- – positive, negative parts , 12
$X_1 \times X_2 \cdots \times X_n$ – cartesian product , 21
$Y_{(j)}$ – jth order statistic, 96
beta(α, β) – beta with α, β , 82
chi–square(ν) – chi–square with ν df , 81
expon(β) – exponential with β , 81
gamma(α, β) – gamma with α, β , 80
geom(p) , 67
poisson(λ) , 70
uniform(θ_1, θ_2) – uniform distrib , 78

g.l.b.(A) – greatest lower bound of A , 19

l.u.b.(A) – least upper bound of A , 19

Index

$\chi^2(\nu)$ (chi–square distribution), 81
$\mathbb{P}$ Null event, 35
μ Null set, 51
σ–algebra, 35
  Borel σ–algebra, 38
σ–algebra generated by a collection of sets, 38
σ–algebra generated by a family of functions (advanced def., 49
σ–algebra generated by a family of random elements, 43
σ–field, 35
σ–finite measure, 51
0–1 encoded Bernoulli trial, 67

closed rectangle, 24

absolute value, 12
absolutely convergent series, 23
abstract integral, 53
abstract integral on a subset, 54
almost sure convergence, 104
almost sure limit, 104
argument, 13
assignment operator, 13

Bayes formula, 62
Bernoulli sequence, 67
Bernoulli trial, 66
  0–1 encoded, 67
  failure probability, 67
  success probability, 67
Bernoulli variable, 67
beta probability distribution, 82
beta(α, β), 82
bijective, 14
binom(n, p) distribution, 67
binomial coefficients, 59
binomial distribution, 67
binomial theorem, 60
bivariate cumulative distribution function, 84
bivariate normal distribution, 98
bivariate probability mass function, 84
Borel σ–algebra, 38
Borel function, 31
Borel measurable, 47
Borel measurable function, 31
Borel set, 29, 38
bounded, 19
bounded above, 19
bounded below, 19
box (3 dimensional rectangle), 24

carrier, 36
carrier set, 36
cartesian product, 21
CDF, 73
  conditional, 87
  joint, 84
central moment of a random variable, 71
characteristic function, 22
chi–square distribution, 81
chi–square with ν df (chi–square distribution), 81
chi–square(ν) (chi–square distribution), 81
closed interval, 11
codomain, 13
coefficient
  binomial, 59
  multinomial, 59
combination, 59
complement, 8
conditional CDF, 87
conditional distribution function, 87
conditional expectation, 94
conditional PDF, 87
conditional PMF, 87
conditional probability, 39
conditional probability density function, 87
conditional probability mass function, 87
conditional variance, 94
continuous random variable, 73
continuous unifurm probability distribution, 78, 89
convergence
  almost surely, 104
  in distribution, 104
  in probability, 104
  pointwise, 104
convergence in distribution, 104
convergence in probability, 104

correlation
- negative, 91
- positive, 91
- zero, 91

correlation coefficient, 91
countable set, 15
countably infinite set, 15
counting measure, 51
covariance, 91
cumulative distribution function, 73
- bivariate, 84
- joint, 84

De Morgan's Law, 9, 16
decreasing, 15
degrees of freedom, 81
- chi–square distribution, 81
- denominator, 108
- numerator, 108

denominator degrees of freedom, 108
density function
- marginal, 86

determinant
- Jacobian, 101

deterministic sample, 63
df = degrees of freedom, 81
direct image, 19
discrete measure, 51
discrete measure space, 51
discrete probability space, 36
discrete random variable, 42
discrete random vector, 42
disjoint, 7
distribution, 42, 52
- binomial, 67
- marginal, 86
- mixed, 83
- multinomial, 96
- uniform, 78, 89

distribution function, 73
- conditional, 87
- joint, 84

domain, 13

element of a set, 6
empty set, 6
equiprobability, 36
event, 5
- independence, 40, 41
- mutually exclusive, 35

event (precise definition), 35
exclusive events, 35
expectation
- conditional, 94

expectation - abstract integral, 57
expectation - continuous r.v., 76
expectation - discrete r.v., 64
expected value, 89, 90
expected value - abstract integral, 57
expected value - continuous r.v., 76
expected value - discrete r.v., 64
experiment
- multinomial, 95

expon(β) (exponential distribution), 81
exponential distribution, 81
extended real numbers, 11
extension of a function, 14

F distribution, 108
failure probability, 67
family, 16
- supremum, 20

finite measure, 51
finite sequence, 15
first quartile, 74
function, 13
- μ–integrable, 53
- argument, 13
- assignment operator, 13
- Borel measurable, 31
- codomain, 13
- domain, 13
- extension, 14
- function value, 13
- infimum, 20
- inverse, 13
- Lebesgue integrable, 31
- linear, 91
- maps to operator, 13
- measurable, 47
- range, 13
- restriction, 14
- simple, 47
- simple (preliminary), 30

- support, 34
- supremum, 20

function value, 13

gamma distribution, 80
gamma function, 80
gamma(α, β), 80
geom(p) distribution, 67
geometric distribution, 67
Gosset, William S., 108
graph, 13
greatest lower bound, 19
greek letters, 112

half closed rectangle, 24
half open rectangle, 24
half-open interval, 11
hypergeometric distribution, 69

iid family, 46
iid sequence, 46
image measure, 52
improper Riemann integral, 27
increasing, 15
independence
- random elements, 43, 49, 50

independent and identically distributed, 46
independent events, 40, 41
index set, 16
indexed family, 16
indicator function, 21
induced measure, 52
infimum, 19
infimum of a family, 20
infimum of a sequence, 20
infinite sequence, 15
injective, 14
integer, 10
integrable function (w.r.t. μ), 53
integrable function w.r.t. μ on a subset, 54
integral, 53
- abstract integral, 53
- Lebesgue integral, 30, 31

integral w.r.t. μ, 53
interval, 24
- closed, 11
- half-open, 11
- open, 11

inverse function, 13
irrational number, 11

Jacobian, 101
Jacobian determinant, 101
Jacobian matrix, 101
joint CDF, 84
joint cumulative distribution function, 84
joint distribution function, 84
joint normal distribution, 98
joint PDF, 85
joint PMF, 84
joint probability density function, 85
joint probability mass function, 84
jointly continuous random variables, 85

Laplace probability, 36
least upper bound, 19
Lebesgue integrable function, 31
Lebesgue integrable function on a subset, 32
Lebesgue integral, 30, 31
Lebesgue integral on a subset, 32
Lebesgue measure, 25, 29
- rectangle, 25

Lebesgue Null set, 29
limit
- almost sure, 104
- in probability, 104
- pointwise, 104

limit in probability, 104
linear function, 91
lower bound, 19

maps to operator, 13
marginal density function, 86
marginal distribution, 86
marginal PDF, 86
marginal PMF, 86
marginal probability mass function, 86
Markov inequality, 82
maximum, 13, 19
mean, 89, 90
mean - abstract integral, 57
mean - continuous r.v., 76
mean - discrete r.v., 64
measurable
- Borel measurable, 47

measurable function, 47

measurable space, 47
measure, 50
  σ–finite, 51
  counting measure, 51
  discrete, 51
  finite, 51
  induced, 52
  product, 55
measure space, 51
  discrete, 51
  product, 55
median, 74
member of a set, 6
member of the family, 16
memoryless property, 81
MGF (moment–generating function), 71
minimum, 19
mixed distribution, 83
mixed random variable, 83
  PDF part, 83
  PMF part, 83
moment about about its mean, 71
moment about the origin, 71
moment of a random variable, 71
moment–generating function, 71
multinomial coefficients, 59
multinomial distribution, 96
multinomial experiment, 95
multinomial sequence, 95
mutually disjoint, 7
mutually exclusive, 35

natural number, 10
negative binomial distribution, 68
negative correlation, 91
negative part, 12
nondecreasing, 15
nonincreasing, 15
normal distribution
  bivariate, 98
  joint, 98
normal probability distribution, 79
Null event, 35
Null set, 51
  λ^d, 29
  μ (abstract measure), 51
  Lebesgue, 29
numerator degrees of freedom, 108

open interval, 11
open rectangle, 24
or
  exclusive, 10
  inclusive, 10
order statistic, 96
outcome, 5
  probability space, 5
  sample space, 5

partition, 10, 16
partitioning, 10, 16
PDF
  conditional, 87
  joint, 85
  marginal, 86
PDF (probability density function), 74
PDF part of a mixed random variable, 83
percentile, 74
permutation, 58
PMF
  conditional, 87
  joint, 84
  marginal, 86
PMF (probability mass function), 64
PMF part of a mixed random variable, 83
pointwise convergence, 104
pointwise limit, 104
Poisson probability distribution, 70
poisson(λ), 70
positive correlation, 91
positive part, 12
power set, 9
preimage, 17
probability, 35
  conditional, 39
probability density function, 74
  conditional, 87
  joint, 85
probability distribution, 42, 52
probability function, 64
probability mass function, 64
  conditional, 87
  joint, 84
  marginal, 86

probability measure, 5, 35
probability space, 5, 35
  discrete, 36
product measure, 55
product measure space, 55
proper Riemann integral, 26

quad, 24
quantile, 74
quartile
  first, 74
  third, 74

r.v. = random variable, 42
random element, 48
  σ–algebra generated by ..., 42
  independence, 43, 49, 50
random sample, 63, 106
random sampling action, 63
  on/from a distribution, 105
  on/from a random variable, 106
random variable, 42, 48
  central moment, 71
  continuous, 73
    expectation, 76
    expected value, 76
    mean, 76
  discrete, 42
    expectation, 64
    expected value, 64
    mean, 64
    variance, 66
  distribution function, 73
  expectation - abstract integral, 57
  expected value - abstract integral, 57
  mean - abstract integral, 57
  mixed, 83
  moment, 71
  moment about its mean, 71
  moment about the origin, 71
  moment–generating function, 71
  standard deviation, 57, 66
  standard normal, 79
  uncorrelated, 91
  uniform, 78
  variance - abstract, 57
random variables
  jointly continuous, 85
random vector, 42, 48
  discrete, 42
range, 13
rational number, 10
real number, 11
realization, 62
rearrangement
  sequence, 23
  series, 23
rectangle
  d–dimensional, 24
  closed, 24
  half closed, 24
  half open, 24
  Lebesgue measure, 25
  open, 24
restriction of a function, 14
Riemann integrable, 27
Riemann integral, 25, 26
  improper, 27
  proper, 26
Riemann integral over a subset, 27
Riemann sum, 26
right continuous function, 73
rv = random variable, 42

sample, 62, 63
  deterministic, 63
  random sample, 106
  realization, 62
sample point, 36
  sample space, 5
sample space, 5, 36
sample standard deviation, 107
sample variance, 106
sampling action, 62
sampling distribution, 106
sampling procedure, 62
sampling process, 62
scale parameter, 80
sequence, 15
  finite, 15
  finite subsequence, 15
  infimum, 20
  infinite, 15
  multinomial, 95

start index, 15
subsequence, 15
supremum, 20
series
absolutely convergent, 23
set, 6
countable, 15
countably infinite, 15
difference, 8
difference set, 8
disjoint, 7
intersection, 7, 16
mutually disjoint, 7
proper subset, 6
proper superset, 6
size, 10
strict subset, 6
strict superset, 6
subset, 6
superset, 6
symmetric difference, 8
uncountable, 15
union, 7, 16
shape parameter, 80
sigma algebra, 35
generated by a collection of sets, 38
sigma algebra generated by a function (advanced def.), 49
sigma algebra generated by random elements, 42
sigma–field, 35
simple function, 47
standard form, 47
simple function (preliminary), 30
simple random sample, 63
simple random sampling action, 63
size, 10
SRS, 63
SRS action, 63
standard deviation, 57, 66
sample, 107
standard normal, 79
start index, 15
statistic, 106
step function, 25
strictly decreasing, 15
strictly increasing, 15
subsequence, 15
finite, 15
success probability, 67
support, 34
supremum, 19
supremum of a family, 20
supremum of a sequence, 20
surjective, 14

Tchebysheff inequalities, 83
third quartile, 74
triangle inequality, 13

uncorrelated random variables, 91
uncountable set, 15
uniform probability, 36
uniform probability distribution, 78
uniform random variable, 78
uniform random vector, 89
uniform(θ_1, θ_2), 78
universal set, 8
upper bound, 19
urn model with replacement, 63
urn model without replacement, 63

variance
conditional, 94
sample, 106
variance - abstract, 57
variance - discrete r.v., 66

zero correlation, 91