WuS - Complete Summary

Ruben Schenk, ruben.schenk@inf.ethz.ch

March 23, 2022

Contents

1 Introduction

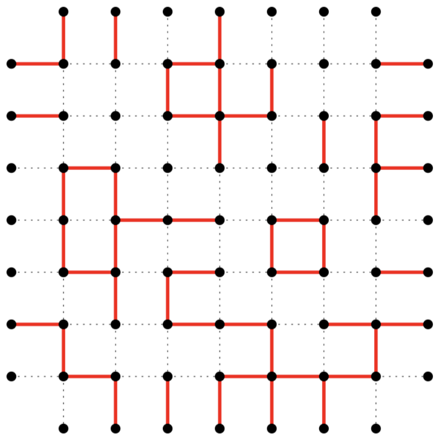
1.1 Percolation Theory

1.1.1 Overview

In physics and mathematics, **percolation theory** describes the behavior of clustered components in random networks. The common intuition is movement and filtering of fluids through porous materials, for example, filtration of water through soil and permeable rocks. In a network, let each node be a cell through which a fluid-like substance may transit to other cells. A network, i.e. a grid, then is a sponge-like substance and percolation is the determination of whether a substance introduced at one cell will reach the other side of the network (or grid).

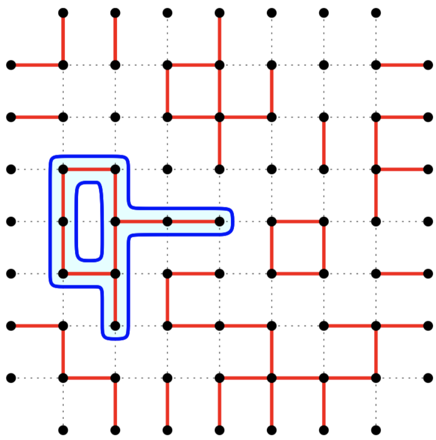
1.1.2 Percolation in a Box

Imagine a box (or grid) with vertices $V = \{-n, ..., n\}^2$ and edges $E = \{e_1, ..., e_N\}$. We introduce parameter p, with $0 \leq p \leq 1$. p denotes the probability that an edge e is *open* ($X_e = 1$). In other words, an edge e is *closed* ($X_e = 0$) with probability $1 - p$. The corresponding model could look something like this:



Note: If an edge is colored red, it means that it's open.

We denote an **open path** as a path consisting of open edges. A **cluster** is the connected component of $(V, \{e : X_e = 1\})$. The following figure shows an example of a cluster (marked in blue):



Theorem [Kesten, 1980]: For the percolation with parameter p we have:

$$\lim_{n\to\infty} \mathbb{P}[\bullet] = \begin{cases} 0, & \text{if } p < \frac{1}{2}, \\ 1, & \text{if } p > \frac{1}{2}. \end{cases}$$

where $\mathbb{P}[\bullet]$ denotes the probability that there exists an open path from the top to the bottom in an $n \times n$ box. Similarly, for the percolation with parameter p we have:

$$\mathbb{P}[\exists \text{ an infinite cluster}] = \begin{cases} 0, & \text{if } p < \frac{1}{2}, \\ 1, & \text{if } p > \frac{1}{2}. \end{cases}$$

1.2 Introduction to Probability

Probability is a mathematical language describing systems involving randomness. Probabilities are used for:

- *Describe random experiments* in the real world, such as coin flips, dice rolling, etc.
- *Express uncertainty.* For example, when a machine performs a measurement, the value is rarely exact. One may use probability theory in this context by saying that the value obtained is equal to the real value plus some small random error.
- *Decision-making.* Probability theory can be used to describe a system when only part of the information is known.
- *Randomized algorithms* in computer science. Sometimes, it is more efficient to add some randomness to perform an algorithm.
- *Simplify complex systems.* Examples include water molecules in water, cars on the highway, etc.

The **goal** of probability theory is to establish general theorems which describe the behavior of multiple random experiments. Example:

Theorem [Law of large numbers]:

$$X_i = \begin{cases} 0, & i^{th} \text{ throw is head}, \\ 1, & i^{th} \text{ throw is number}. \end{cases}$$

It holds, that:

$$\lim_{n\to\infty} \frac{X_1 + \cdots + X_n}{n} = \frac{1}{2}.$$

2 Mathematical Framework

2.1 Probability Space

2.1.1 Sample Space

Assume we want to model a random experiment. The first mathematical object needed is the set of all possible outcomes of the experiment, denoted by Ω.

The set Ω is called the **sample space.** An element $\omega \in \Omega$ is called an **outcome** (or *elementary experiment*).

Example: If we throw a die, we have the following sample space:

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

2.1.2 Events

Previously, the set of **events** was always $\mathcal{P}(\Omega)$. In this class, we will work with more general sets of events $\mathcal{F} \subset \mathcal{P}(\Omega)$, called sigma algebras.

Definition: A **sigma-algebra** is a subset $\mathcal{F} \subset \mathcal{P}(\Omega)$ satisfying the following properties:

1. $\Omega \in \mathcal{F}$
2. $A \in \mathcal{F} \implies A^C \in \mathcal{F}$
3. $A_1, A_2, ... \in \mathcal{F} \implies \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$

Example: Following are some (non-) examples of sigma-algebras for $\Omega = \{1, 2, 3, 4, 5, 6\}$:

- $\mathcal{F} = \{\emptyset, \{1, 2, 3, 4, 5, 6\}\}$ is a sigma-algebra.
- $\mathcal{F} = \{\emptyset, \{1, 2\}, \{3, 4, 5, 6\}, \{1, 2, 3, 4, 5, 6\}\}$ is a sigma-algebra.
- $\mathcal{F} = \{\{1, 2, 3, 4, 5, 6\}\}$ is not a sigma-algebra because $P2$ is not satisfied.
- $\mathcal{F} = \{\emptyset, \{1, 2, 3\}, \{4, 5, 6\}, \{1\}, \{2, 3, 4, 5, 6\}, \Omega\}$ is not a sigma-algebra because $P3$ is not satisfied.

2.1.3 Probability Measure

Definition: Let Ω be a sample space, let $\mathcal{F}$ be a sigma-algebra. A **probability measure** on $(\Omega, \mathcal{F})$ is a map

$$
\begin{aligned}
\mathbb{P} : &\mathcal{F} \to [0, 1] \\
&A \to \mathbb{P}[A]
\end{aligned}
$$

that satisfies the following two properties:

- **P1.** $\mathbb{P}[\Omega] = 1$.
- **P2. (countable additivity)** $\mathbb{P}[A] = \sum_{i=1}^{\infty} \mathbb{P}[A_i]$ if $A = \bigcup_{i=1}^{\infty} A_i$ (*disjoint union*).

2.1.4 Notion of Probability Space

Definition: Let Ω be a sample space, $\mathcal{F}$ a sigma-algebra, and $\mathbb{P}$ a probability measure. The triple $(\Omega, \mathcal{F}, \mathbb{P})$ is called a **probability space.**

2.2 Examples of Probability Space

2.2.1 Example with Ω Finite

We discuss a particular type of probability spaces where the sample space Ω is an arbitrary **finite** set, and all the outcomes have the **same** probability $p_\omega = \frac{1}{|\Omega|}$.

Definition: Let Ω be a finite sample space. The **Laplace model** on Ω is the triple $(\Omega, \mathcal{F}, \mathbb{P})$, where:

- $\mathcal{F} = \mathcal{P}(\Omega)$,
- $\mathbb{P} : \mathcal{F} \to [0, 1]$ is defined by

$$\forall A \in \mathcal{F} \quad \mathbb{P}[A] = \frac{|A|}{|\Omega|}$$

Example: We consider $n \geq 3$ points on a circle, from which we select 2 at random. What is the probability that these two points selected are neighbors? We consider the Laplace model one

$$\Omega = \{E \subset \{1, 2, ..., n\} : |E| = 2\}.$$

The event "the two points of E are neighbors" is given by

$$A = \{\{1, 2\}, \{2, 3\}, ..., \{n-1, n\}, \{n, 1\}\}$$

and we have

$$\mathbb{P}[A] = \frac{|A|}{|\Omega|} = \frac{n}{\binom{n}{2}} = \frac{2}{n-1}.$$

2.2.2 Example with Ω Infinite Countable

Example: We throw a biased coin multiple times, at each throw, the coin falls on head with probability p, and it falls on tail with probability $1-p$ (p is a fixed parameter in $[0, 1]$). We stop at the first time we see a tail. The probability that we stop exactly at time k is given by

$$p_k = p^{k-1}(1-p).$$

For this experiment, one possible probability space is given by:

- $\Omega = \mathbb{N} \setminus \{0\} = \{1, 2, 3, ...\}$
- $\mathcal{F} = \mathcal{P}(\Omega)$
- for $A \in \mathcal{F}$, $\mathbb{P}[A] = \sum_{k \in A} p_k$

2.3 Properties of Events

2.3.1 Operations on Events and Interpretation

The following propositions asserts that the different well-known set operations are allowed.

Proposition (Consequences of the definition): Let $\mathcal{F}$ be a sigma-algebra on Ω. We have:

- **P4.** $\emptyset \in \mathcal{F}$
- **P5.** $A_1, A_2, ... \in \mathcal{F} \implies \bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$
- **P6.** $A, B \in \mathcal{F} \implies A \cup B \in \mathcal{F}$
- **P7.** $A, B \in \mathcal{F} \implies A \cap B \in \mathcal{F}$

A short summary of the common set-operations is given below:

- A^C : A does not occur.
- $A \cap B$: A and B occur.
- $A \cup B$: A or B occurs
- $A \Delta B$: one and only one of A or B occurs
- $A \subset B$: If A occurs, then B occurs
- $A \cap B = \emptyset$: A and B cannot occur at the same time
- $\Omega = A_1 \cup A_2 \cup A_3$ with A_1, A_2, A_3 pairwise disjoint: for each outcome ω, one and only one of the events A_1, A_2, A_3 is satisfied.

2.4 Properties of Probability Measures

2.4.1 Direct Consequences of the Definition

Proposition: Let $\mathbb{P}$ be an arbitrary measure on $(\Omega, \mathcal{F})$. We have:

- **P3.** $\mathbb{P}[\emptyset] = 0$.
- **P4. (additivity)** Let $k \geq 1$. let $A_1, ..., A_k$ be k pairwise disjoint events, then $\mathbb{P}[A_1 \cup \cdots \cup A_k] = \mathbb{P}[A_1] + \cdots + \mathbb{P}[A_k]$.
- **P5.** Let A be an event, then $\mathbb{P}[A^C] = 1 - \mathbb{P}[A]$.
- **P6.** If A and B are two events (not necessarily disjoin), then $\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A \cap B]$.

2.4.2 Useful Inequalities

Proposition (Monotonicity): Let A, $B \in \mathcal{F}$, then

$$A \subset B \implies \mathbb{P}[A] \leq \mathbb{P}[B].$$

Proposition (Union bound): Let A_1, $A_2, ...$ be a sequence of events (not necessarily disjoint), then we have

$$\mathbb{P}[\bigcup_{i=1}^{\infty} A_i] \leq \sum_{i=1}^{\infty} \mathbb{P}[A_i].$$

Remark: The union bound also applies to a *finite* collection of events.

2.4.3 Continuity Properties of Probability Measures

Proposition: Let (A_n) be an increasing sequence of events (i.e. $A_n \subset A_{n+1}$ for every n). then

$$\lim_{n\to\infty} P[A_n] = \mathbb{P}[\bigcup_{n=1}^{\infty} A_n]. \quad \text{(increasing limit)}$$

Let (B_n) be a decreasing sequence of events (i.e. $B_n \supset B_{n+1}$ for every n). Then

$$\lim_{n\to\infty} P[B_n] = \mathbb{P}[\bigcap_{n=1}^{\infty} B_n]. \quad \text{(decreasing limit)}$$

Remark: By monotonicity, we have $\mathbb{P}[A_n] \leq \mathbb{P}[A_{n+1}]$ and $\mathbb{P}[B_n] \geq \mathbb{P}[B_{n+1}]$ for every n. Hence the limits in the proposition are well defined as monotone limits.

2.5 Conditional Probabilities

Definition (Conditional probability): Let $(\Omega, \mathcal{F}, \mathbb{P})$ be some probability space. Let A, B be two events with $\mathbb{P}[B] > 0$. The **conditional probability of** A **given** B is defined by

$$\mathbb{P}[A \mid B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}.$$

Remark: $\mathbb{P}[B \mid B] = 1$.

Proposition: Let Ω, $\mathcal{F}$, $\mathbb{P}$ be some probability space. Let B be an event with positive probability. Then $\mathbb{P}[\,.\mid B]$ is a probability measure on Ω.

Proposition (Formula of total probability): Let $B_1, ..., B_n$ be a partition of the sample space Ω with $\mathbb{P}[B_i] > 0$ for every $1 \leq i \leq n$. Then, one has

$$\forall A \in \mathcal{F} : \mathbb{P}[A] = \sum_{i=1}^{n} \mathbb{P}[A \mid B_i]\mathbb{P}[B_i].$$

Here, a *partition* B_i is such that $\Omega = B_1 \cup \cdots \cup B_n$ and the events are pariwise disjoint.

Proposition (Bayes formula): Let $B_1, ..., B_n \in \mathcal{F}$ be a partition of Ω with $\mathbb{P}[B_i] > 0$ for every i. For every event A with $\mathbb{P}[A] > 0$, we have

$$\forall i = 1, ..., n : \mathbb{P}[B_i \mid A] = \frac{\mathbb{P}[A \mid B_i] \cdot \mathbb{P}[B_i]}{\sum_{j=1}^{n} \mathbb{P}[A \mid B_j] \cdot \mathbb{P}[B_j]}.$$

2.6 Independence

2.6.1 Independence of Events

Definition (Independence of two events): Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Two events A and B are said to be **independent** If

$$\mathbb{P}[A \cap B] = \mathbb{P}[A] \cdot \mathbb{P}[B].$$

Remark: If $\mathbb{P}[A] \in \{0, 1\}$, then A is independent of every event, i.e. $\forall B \in \mathcal{F} : \mathbb{P}[A \cap B] = \mathbb{P}[A] \cdot \mathbb{P}[B]$. Furthermore we might also state, that A is independent of B if and only if A is independent of B^C.

Proposition: Let A, $B \in \mathcal{F}$ be two events with $\mathbb{P}[A]$, $\mathbb{P}[B] > 0$. Then the following are equivalent:

- $\mathbb{P}[A \cap B] = \mathbb{P}[A] \cdot \mathbb{P}[B]$ (A and B are independent)
- $\mathbb{P}[A \mid B] = \mathbb{P}[A]$ (the occurrence of B has no influence on A)
- $\mathbb{P}[B \mid A] = \mathbb{P}[B]$ (the occurrence of A has no influence on B)

Definition: Let I be an arbitrary set of indices. A collection of events $(A_i)_{i \in I}$ is said to be **independent** if

$$\forall J \subset I \text{ infinite} : \mathbb{P}[\bigcap_{j \in J} A_j] = \prod_{j \in J} \mathbb{P}[A_j].$$

3 Random Variables and Distribution Functions

3.1 Abstract Definition

Definition: Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A **random variable (r.v.)** is a map $X : \Omega \to \mathbb{R}$ such that for all $a \in \mathbb{R}$,

$$\{\omega \in \Omega : X(\omega) \le a\} \in \mathcal{F}.$$

The condition $\{\omega \in \Omega : X(\omega) \le a\} \in \mathcal{F}$ is needed for $\mathbb{P}[\{w \in \Omega : X(\omega) \le a\}]$ to be well-defined.

Example (Indicator function of an event): Let $A \in \mathcal{F}$. Consider the **indicator function** $\mathbb{1}_A$ of A, defined by

$$\forall \omega \in \Omega : X(\omega) = \begin{cases} 0 & \text{if } \omega \notin A, \\ 1 & \text{if } \omega \in A. \end{cases}$$

Then $\mathbb{1}_A$ is a random variable. Indeed, we have

$$\{\omega : \mathbb{1}_A(\omega) \le a\} = \begin{cases} \emptyset & \text{if } a < 0, \\ A^C & \text{if } 0 \le a \le 1, \\ \Omega & \text{if } a \ge 1, \end{cases}$$

and $\emptyset$, A^C, and Ω are three elements of $\mathcal{F}$.

Notation: When events are defined in terms of random variables, we will *omit the dependence in* ω. For example, for $a \le b$ we write:

$$\begin{aligned} \{X \le a\} &= \{\omega \in \Omega : X(\omega) \le a\}, \\ \{a < X \le b\} &= \{\omega \in \Omega : aX(\omega) < b\}, \\ \{X \in \mathbb{Z}\} &= \{\omega \in \Omega : X(\omega) \in \mathbb{Z}\} \end{aligned}$$

When considering the probability of the events above, we omit the brackets and, for example, simply write:

$$\mathbb{P}[X \le a] = \mathbb{P}[\{X \le a\}] = \mathbb{P}[\{\omega \in \Omega : X(\omega) \le a\}].$$

3.2 Distribution Function

Definition: Let X be a random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The **distribution function** of X is the function $F_X : \mathbb{R} \to [0, 1]$ defined by

$$\forall a \in \mathbb{R} : F_X(a) = \mathbb{P}[X \le a]$$

The idead is that the distribution function F_X encodes the probabilistic properties of the random variable X.

Proposition (Basic identity): Let $a < b$ be two real numbers. Then

$$\mathbb{P}[a < X \le b] = F(b) - F(a)$$

Theorem (Properties of distribution functions): Let X be a random variable on some probability space $(\Omega, \mathcal{F}), \mathbb{P}$. The distribution function $F = F_X : \mathbb{R} \to [0, 1]$ of X satisfies the following properties:

1. F is nondecreasing.
2. F is right continuous, i.e. $F(a) = \lim_{h \downarrow 0} F(a + h)$ for every $a \in \mathbb{R}$.
3. $\lim_{a \to -\infty} F(a) = 0$ and $\lim_{a \to \infty} F(a) = 1$.

3.3 Independence

3.3.1 Independence of Random Variables

Definition: Let $X_1, ..., X_n$ be n random variables on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We say that $X_1, ..., X_n$ are **independent** if

$$\forall x_1, ..., x_n \in \mathbb{R} : \mathbb{P}[X_1 \leq x_1, ..., X_n \leq x_n] = \mathbb{P}[X_1 \leq x_1] \cdots \mathbb{P}[X_n \leq x_n].$$

Definition: An infinite sequence $X_1, X_2, ...$ of random variables is said to be:

- **independent** ix $X_1, ..., X_n$ are independent, for every n.
- **independent and identically distributed (iid)** if they are independent and have the same distribution function, i.e. $\forall i, y : F_{X_i} = F_{X_j}$.

3.4 Transformation of Random Variables

Once we have some random variables $X_1, X_2, ...$ on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we can create and consider many new random variables on the same probability space by using operations. For example, one can consider $Z_1 = X_1 + X_2$. However, one should not forget that random variables are maps $\Omega \to \mathbb{R}$. For example, the random variable Z_1 corresponds to the map, defined for every $\omega \in \Omega$, $Z_1(\omega) = X_1(\omega) + X_2(\omega)$.

Formally, we introduce the following notation, which allows us to work with random variables as if they were just real numbers. If X is the random variable, and $\phi : \mathbb{R} \to \mathbb{R}$, then we write

$$\phi(X) := \phi \circ X.$$

This way, $\phi(X)$ is a new mapping $\Omega \to \mathbb{R}$ as show in the following diagram:

$$\Omega \to^X \mathbb{R} \to^\phi \mathbb{R}$$
$$\omega \to X(\omega) \to \phi(X(\omega)).$$

3.5 Construction of Random Variables

The goal of this section is to construct general random variables. Our approach will rely on the abstract theorem of Kolmogorov, that guarantees existences of iid sequences. The construction proceeds in 4 steps:

Step 1: Komogorov theorem and iid sequence of Bernoulli random variables Our construction starts with Bernoulli random variables, that we define now.

Definition: Let $p \in [0, 1]$. A random variable X is said to be a **Bernoulli random variable with parameter** p if

$$\mathbb{P}[X = 0] = 1 - p \text{ and } \mathbb{P}[X = 1] = p.$$

In this case, we write $X \sim \text{Ber}(p)$.

> **Theorem (Existence theorem of Kolmogorov):** There exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and an infinite sequence of random variables $X_1, X_2, ...$ (on this probability space) that is an iid sequence of Bernoulli random variables with parameter $\frac{1}{2}$.

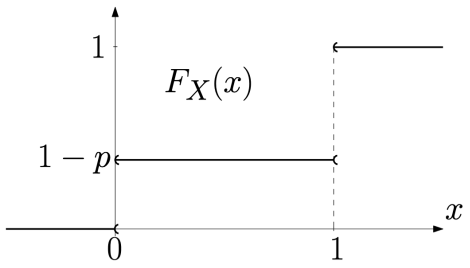
Step 2: Construction of a uniform random variable in $[0, 1]$ Here we use Bernoulli random variables to construct a uniform random variable in $[0, 1]$. Intuitively, one can imagine a droplet of water falling in the interval $[0, 1]$. A uniform random variable in $[0, 1]$ represents the position at which such a droplet falls.

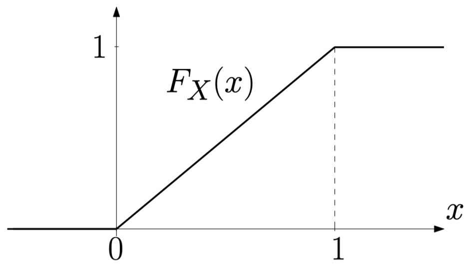
Definition: A random variable U is said to be a **uniform random variable in** $[0, 1]$ if its distribution function is equal to

$$F_U(x) = \begin{cases} 0, & x < 0, \\ x, & 0 \leq x \leq 1, \\ 1, & x > 1. \end{cases}$$

In this case, we write $U \sim \mathcal{U}([0, 1])$.

The figure below shows the distribution function of a Bernoulli r.v. with parameter p (left) and the distribution function of a uniform random variable in $[0, 1]$ (right).





Let $X_1, X_2, \ldots$ be a sequence of independent Bernoulli random variables with parameter $\frac{1}{2}$. For every fixed ω, we have $X_1(\omega), X_2(\omega), \ldots \in \{0, 1\}$. Hence the infinite series

$$Y(\omega) = \sum_{n=1}^{\infty} 2^{-n} X_n(\omega)$$

is absolutely convergent, and we have $Y(\omega) \in [0, 1]$.

Proposition: The mapping $Y : \Omega \to [0, 1]$ defined by the equation above is a uniform random variable in $[0, 1]$.

Step 3: Construction of a random variable with an arbitrary distribution F Let $F : \mathbb{R} \to [0, 1]$ satisfying item $(1) - (3)$ at the beginning of the section. If F is strictly increasing and continuous then F is one-to-one and one can define its inverse F^{-1}. For every $\alpha \in [0, 1]$, $F^{-1}(\alpha)$ is the unique real number x such that $F(x) = \alpha$. In such a case, the defines the inverse distribution function. More generally, we can define a generalized inverse for F.

Definition (Generalized inverse): The generalized inverse of F is the mapping $F^{-1} : (0, 1) \to \mathbb{R}$ defined by

$$\forall \alpha \in (0, 1) : F^{-1}(\alpha) = \inf\{x \in \mathbb{R} : F(x) \geq \alpha\}.$$

By definition of the infimum and using right continuity of F, we have for every $x \in \mathbb{R}$ and $\alpha \in (0, 1)$

$$(F^{-1}(\alpha) \leq x) \iff (\alpha \leq F(x)).$$

Theorem (inverse transform sampling): Let $F : \mathbb{R} \to [0, 1]$ satisfying items $(1) - (3)$ at the beginning of the section. Let U be a uniform random variable in $[0, 1]$. Then the random variable

$$X = F^{-1}(U)$$

has distribution $F_X = F$.

Step 4: General sequence of independent random variables Finally, we introduce the following theorem:

Let $F_1, F_2, ...$ be a sequence of functions $\mathbb{R} \rightarrow [0, 1]$ satisfying items $(1) - (3)$ at the beginning of the section. Then there exists a probability space $(\Omega, \mathcal{F}. \mathbb{P})$ and a sequence of independent random variables $X_1, X_2, ...$ on this probability space such that

- for every i X_i has a distribution function F_i (i.e. $\forall x \mathbb{P}[X_i \leq x] = F_i(x)$), and
- $X_1, X_2, ...$ are independent.

4 Discrete and Continuous Random Variables

4.1 Discontinuity & Continuity Points of F

We have seen that the distribution function $F = F_X$ of a random variable X is always *rigt continuous*. What about left continuous?

Example: For a Bernoulli random variable $X \sim \text{Ber}(p)$ with $p < 1$, we have $F_X(-h) = 0$ for every $h > 0$, but $F_X(0) = 1 - p \neq 0$. Therefore, F_X is not left continuous at 0, i.e.

$$\lim_{h \downarrow 0} F_X(-h) = 0 \neq F_X(0).$$

The following proposition gives an interpretation of the limit

$$F(a-) := \lim_{h \downarrow 0} F(a - h)$$

at a given point a for a general distribution function.

Proposition (probability of a given value): Let $X : \Omega \to \mathbb{R}$ be a random variable with distribution function F. Then for every a in $\mathbb{R}$ we have

$$\mathbb{P}[X = a] = F(a) - F(a-).$$

We give the following interpretation of the above introduce proposition. Fix some $a \in \mathbb{R}$. Then:

- If F is not continuous at a point $a \in \mathbb{R}$, then the "jump size" $F(a) - F(a-)$ is equal to the probability that $X = a$.
- If F is continuous at a point $a \in \mathbb{R}$, then $\mathbb{P}[X = a] = 0$.

4.2 Almost Sure Events

Definition: Let $A \in \mathcal{F}$ be an event. We say that A occurs **almost surely (a.s.)** if

$$\mathbb{P}[A] = 1.$$

Remark: This notion can be extended to any set $A \subset \Omega$: We say that A occurs almost surely if there exists an event $A' \in \mathcal{F}$ such that $A' \subset A$ and $\mathbb{P}[A'] = 1$.

4.3 Discrete Random Variables

Definition (Discrete Random Variables): A random variable $X : \Omega \to \mathbb{R}$ is said to be **discrete** if there exists some set $W \subset \mathbb{R}$ finite or countable such that

$$X \in W \quad a.s.$$

Remark: If the sample space Ω is finite or countable, then every random variable $X : \Omega \to \mathbb{R}$ is discrete.

Definition: Let X be a discrete random variable taking some values in some finite or countable set $W \subset \mathbb{R}$. The **distribution of** X is the sequence of numbers $(p(x))_{x \in W}$ defined by

$$\forall x \in W : p(x) := \mathbb{P}[X = x].$$

Proposition: The distribution $(p(x))_{w \in W}$ of a discrete random variable satisfies

$$\sum_{x \in W} p(x) = 1.$$

Example: Consider the random variable defined by

$$\forall \omega \in \Omega : X(\omega) := \begin{cases} -1, & \text{if } \omega = 1, 2, 3, \\ 0, & \text{if } \omega = 4, \\ 2, & \text{if } \omega = 5, 6. \end{cases}$$

Then X takes values in $W = \{-1, 0, 2\}$ almost surely and its distribution is given by

$$p(-1) = \frac{1}{2}, \quad p(0) = \frac{1}{6}, \quad p(2) = \frac{1}{3}.$$

Remark: Conversely, if we are given a sequence of numbers $(p(x))_{x\in W}$ with values in $[0, 1]$ and such that $\sum_{x\in W} p(x) = 1$, then there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a random variable X with associated distribution $(p(x))$. This observation is important in practice, it allows us to write: "*Let X be a discrete random variable with distribution $(p(x))_{x\in W}$.*"

4.3.1 From p to F_X

Proposition: Let X be a discrete random variable with values in a finite or countable set W almost surely, and distribution p. Then the distribution function of X is given by

$$\forall x \in \mathbb{R} : F_X(x) = \sum_{y\leq x,\, y\in W} p(y).$$

4.3.2 From F_X to p

Given a discrete random variable X. A random variable with a piecewise cosntant function F is discrete and W and p are given by:

- $W = \{\text{positions of the jumps of } F_X\}$
- $p(x) =$ "height of the jump" at $x \in W$

4.4 Examples of Discrete Random Variables

4.4.1 Bernoulli Distribution

Definition (Bernoulli): Let $0 \leq p \leq 1$. A random variable X is said to be a **Bernoulli random variable with parameter** p if it takes values in $W = \{0, 1\}$ and

$$\mathbb{P}[X = 0] = 1 - p \quad \text{and} \quad \mathbb{P}[X = 1] = p.$$

In that case, we write $X \sim \text{Ber}(p)$.

4.4.2 Binomial Distribution

Definition (Binomial): Let $0 \leq p \leq 1$, let $n \in \mathbb{N}$. A random variable X is said to be a **binomial random variable with parameters** n **and** p if it takes values in $W = \{0, ..., n$ and

$$\forall k \in \{0, ..., n\} : \mathbb{P}[X = k] = \binom{n}{k} p^k (1 - p)^{n-k}.$$

In that case we write $X \sim \text{Bin}(n, p)$. This appears in applications when we consider the number of successes in a repetition of Bernoulli experiments.

Proposition (Sum of independent Bernoulli and binomial): Let $0 \leq p \leq 1$, let $n \in \mathbb{N}$. Let $X_1, ..., X_n$ be independent Bernoulli random variables with parameter p. Then

$$S_n := X_1 + \cdots + X_n$$

is a binomial random variable with parameter n and p.

Remark: In particular, the distribution Bin(1, p) is the same as the distribution Ber(p). One can also check that if $X \sim \text{Bin}(m, p)$ and $Y \sim \text{Bin}(n, p)$ and X, Y are independent, then $X + Y = \text{Bin}(m+n, p)$.

4.4.3 Geometric Distribution

Definition (Geometric): Let $0 \leq p \leq 1$. A random variable X is said to be a **geometric random variable with parameter** p if it takes values in $W = \mathbb{N} \setminus \{0\}$ and

$$\forall k \in \mathbb{N} \setminus \{0\} : \mathbb{P}[X = k] = (1-p)^{k-1} \cdot p.$$

In that case, we write $X \sim \text{Geom}(p)$.

The geometric random variable appears naturally as the first success in an infinite sequence of Bernoulli experiments with parameter p. This is formalized by the following proposition.

Proposition: Let $X_1, X_2, ...$ be a sequence of infinitely many independent Bernoulli r.v.'s with parameter p. Then

$$T := \min\{n \geq 1 : X_n = 1\}$$

is a geometric random variable with parameter p.

Proposition: Let $T \sim \text{Geom}(p)$ for some $0 < p < 1$. Then

$$\forall n \geq 0, \forall k \geq 1 : \mathbb{P}[T \geq n + k \,|\, T > n] = \mathbb{P}[T \geq k].$$

4.4.4 Poisson Distribution

Definition: Let $\lambda > 0$ be a positive real number. A random variable X is said to be a **Poisson random variable with parameter** λ if it takes values in $W = \mathbb{N}$ and

$$\forall k \in \mathbb{N} : \mathbb{P}[X = k] = \frac{\lambda^k}{k!} e^{-\lambda}.$$

In this case, we write $X \sim \text{Poisson}(\lambda)$.

The Poisson distribution appears naturally as an approximation of a binomial distribution when the parameter n is large and the parameter p is small, as stated formally in the following proposition.

Proposition (Poisson approximation of the binomial): Let $\lambda > 0$. For every $n \geq 1$, consider a random variable $X_n \sim \text{Bin}(n, \frac{\lambda}{n})$. Then

$$\forall k \in \mathbb{N} : \lim_{n \to \infty} \mathbb{P}[X_n = k] = \mathbb{P}[N = k],$$

where N is a Poisson random variable with parameter λ.

4.5 Continuous Random Variables

Definition (Continuous Random Variables): A random variable $X : \Omega \to \mathbb{R}$ is said to be **continuous** if its distribution function F_X can be written as

$$F_X(a) = \int_{-\infty}^{a} f(x)\, dx \quad \text{for all } a \in \mathbb{R}$$

for some nonnegative function $f : \mathbb{R} \to \mathbb{R}_+$, called the **density** of X.

Intuition: $f(x)\,dx$ represents the probability that X takes a value in the infinitesimal interval $[x,\, x+dx]$.

Proposition: The density f of a random variable satisfies

$$\int_{-\infty}^{+\infty} f(x)\,dx = 1.$$

4.5.1 From f to F_X

Let X be a continuous random variable with density f. By definition, the distribution function F_X can be calculated as the integral

$$F_X(x) = \int_{-\infty}^{x} f(y)\,dy.$$

4.5.2 From F_X to f

Since one goes from f to F_X by integrating, it is natural to expect that the reverse operation is to take the derivative. This is in general the case, provided F_X is regular enough. The following theorem will be useful in applications to calculate densities.

Theorem: Let X be a random variable. Assume that the distribution function F_X is continuous and piecewise $\mathcal{C}^1$, i.e. that there exists $x_0 = -\infty < x_1 < \cdots < x_{n-1} < x_n = +\infty$ such that F_X is $\mathcal{C}^1$ on every interval $(x_i,\, x_{i+1})$. Then X is a continuous random variable and a density f can be constructed by defining

$$\forall x \in (x_i,\, x_{i+1}) : f(x) = F_X'(x)$$

and setting arbitrary values at $x_1, ..., x_{n-1}$.