A Gentle Introduction to Probability Theory

Alexander Ritz

University of Göttingen

8. Juni 2021

Overview

1. Probability Spaces
2. Random Variables
3. Distributions
4. Moments
5. Convergence of Random Variables

Section 1

Probability Spaces

Sample Space

Let Ω denote the set of all potential outcomes or results ω of a random experiment.

Then Ω is called sample space or possibility space.

Events

Any subset $A \subset \Omega$ is called an event.

σ -Algebra

A system $\mathscr{A}$ of subsets of Ω is called σ-Algebra in Ω, if it has the following properties:

σ -Algebra

A system $\mathscr{A}$ of subsets of Ω is called σ-Algebra in Ω, if it has the following properties:

- $\Omega \in \mathscr{A}$

σ -Algebra

A system $\mathscr{A}$ of subsets of Ω is called σ-Algebra in Ω, if it has the following properties:

- $\Omega \in \mathscr{A}$
- $A \in \mathscr{A} \implies A^{\mathrm{C}} \in \mathscr{A}$

σ -Algebra

A system $\mathscr{A}$ of subsets of Ω is called σ-Algebra in Ω, if it has the following properties:

- $\Omega \in \mathscr{A}$
- $A \in \mathscr{A} \implies A^{\mathrm{C}} \in \mathscr{A}$
- $A_1, A_2, A_3, \ldots \in \mathscr{A} \implies \bigcup_{I \in \mathbb{N}} A_I \in \mathscr{A}$

Power Set

The power set of a set A is given by

$$\mathcal{P}(A) := \{M \mid M \subset A\},$$

being the set of all subsets M of A. It holds that $\mid \mathcal{P}(A) \mid = 2^{|A|}$.

Generated σ-Algebra

Given an arbitrary family of subsets F of a sample space Ω, the σ-algebra

$$\sigma(F) := \bigcap_{F \subset \mathscr{A}} \mathscr{A}, \quad \text{with } \mathscr{A} \text{ being } \sigma\text{-algebras on } \Omega$$

denotes the smallest σ-algebra on Ω, which contains F.

Borel σ-Algebra

Given $\Omega = \mathbb{R}$ and $I := \{(s,t) : -\infty \leq s \leq t \leq \infty\}$, the σ-algebra generated by the open intervals I are usually denoted

$$\mathscr{B}(\mathbb{R}) := \sigma(I)$$

This is called the Borel σ algebra on $\mathbb{R}$. By definition of σ-algebras, $\mathscr{B}$ contains all open, closed and half open intervals.

Measurable Space

A tuple $(\Omega, \mathscr{A})$, consisting of a sample space Ω and a σ-algebra $A \subset \mathcal{P}(\Omega)$ is called measurable space.

Measures

Let $\mathscr{A}$ be a σ-algebra. A function $\mu : \mathscr{A} \to [0, \infty]$ is called a measure on $\mathscr{A}$, if

Measures

Let $\mathscr{A}$ be a σ-algebra. A function $\mu : \mathscr{A} \to [0, \infty]$ is called a measure on $\mathscr{A}$, if

- $\mu(\emptyset) = 0$

Measures

Let $\mathscr{A}$ be a σ-algebra. A function $\mu : \mathscr{A} \to [0, \infty]$ is called a measure on $\mathscr{A}$, if

- $\mu(\emptyset) = 0$
- μ is σ-additive, i.e. for all sequences of pairwise disjoint sets $A_1, A_2, \ldots$ holds:

$$\mu\left(\dot{\bigcup_{i\in\mathbb{N}}} A_i\right) = \sum_{i\in\mathbb{N}} \mu(A_i)$$

Probability Measures

Let $\mathscr{A}$ be a σ-algebra on the sample space Ω. A function $P : \mathscr{A} \to [0, 1]$ is called probability measure on $\mathscr{A}$, if

Probability Measures

Let $\mathscr{A}$ be a σ-algebra on the sample space Ω. A function $P : \mathscr{A} \to [0, 1]$ is called probability measure on $\mathscr{A}$, if

- $P(\Omega) = 1$

Probability Measures

Let $\mathscr{A}$ be a σ-algebra on the sample space Ω. A function $P : \mathscr{A} \to [0, 1]$ is called probability measure on $\mathscr{A}$, if

- $P(\Omega) = 1$
- P is σ-additive, that is, for all sequences of pairwise disjoint sets $A_1, A_2, \ldots$ holds:

$$P\left(\dot{\bigcup_{i \in \mathbb{N}}} A_i\right) = \sum_{i \in \mathbb{N}} P(A_i)$$

Probability Space

The triple $(\Omega, \mathscr{A}, P)$, consisting of a sample space Ω, a σ-algebra $\mathscr{A}$ on Ω and a probability measure P defined on $\mathscr{A}$, is called probability space.

Random Quantities

A function on the sample space Ω

$$X : \Omega \to \mathbb{R}^d \quad \text{with} \quad \omega \mapsto X(\omega)$$

is called random quantity.

Random Quantities

A function on the sample space Ω

$$X : \Omega \to \mathbb{R}^d \quad \text{with} \quad \omega \mapsto X(\omega)$$

is called random quantity.
In case of $d = 1$, X is called a random variable.

Random Quantities

A function on the sample space Ω

$$X : \Omega \to \mathbb{R}^d \quad \text{with} \quad \omega \mapsto X(\omega)$$

is called random quantity.
In case of $d = 1$, X is called a random variable.
In case of $d > 1$ X is a vector of random variables, that is $X = (X_1, ..., X_d)$. In that case, X is called d-dimensional random vector.

Random Quantities

A function on the sample space Ω

$$X : \Omega \to \mathbb{R}^d \quad \text{with} \quad \omega \mapsto X(\omega)$$

is called random quantity.
In case of $d = 1$, X is called a random variable.
In case of $d > 1$ X is a vector of random variables, that is $X = (X_1, ..., X_d)$. In that case, X is called d-dimensional random vector.
The function does not have to be real valued, but the concept generalises trivially!

Measurability

In order to facilitate the notation of the following concepts, we define:

$$X^{-1}(]-\infty, z]) = \{\omega \in \Omega \mid -\infty < X(\omega) \leq z\} =: \{X \leq z\}$$

Measurability

A random quantity $X : \Omega \to \mathbb{R}^d$ is called measurable with respect to the σ-algebra $\mathscr{A}$, if:

$$\forall z \in \mathbb{R}^d : \{X \leq z\} \in \mathscr{A}$$

with $\{X \leq z\} \in \mathscr{A}$ being shorthand for the set of outcomes $\omega \in \Omega$, for which the function $X = (X_1, ..., X_d)$ results in values below $z = (z_1, ..., z_d)$, i.e.:

$$\{X \leq z\} = \{\omega \in \Omega \mid X(\omega) \leq z\} = \{\omega \in \Omega \mid X_1(\omega) \leq z_1, ..., X_d(\omega) \leq z_d\}$$

Distributions of Random Variables

Let $X : \Omega \to \mathbb{R}^d$ be a measurable random variable on the measure space $(\Omega, \mathscr{A})$. The probabilities of *all* events belonging to X as a whole are called *probability distribution* or *cumulative distribution* of the random variable X. The entire distribution is determined by the function:

$$F_X : \mathbb{R}^d \to [0, 1] \tag{1}$$

$$z \mapsto F_X(z) := P(\{X \leq z\}) = P(\{X_1 \leq z_1, \dots, X_d \leq z_d\}), \tag{2}$$

leading to F_X being called *distribution function* of X

Distribution Function

Using *Riemann-Stieltjes integrals*, one can write:

$$P(\{a < X \leq b\}) = F_X(b) - F_X(a) = \int_a^b \mathrm{d}F_X(z)$$

, given that F_X is continuous.

Distribution Function

In case F_X is differentiable, the probabilities can be given as:

$$\int_a^b \mathrm{d}F_X(z) = \int_a^b F_X'(z)\,\mathrm{d}z$$

Under the further restriction that F_X' be continuous, the following holds for infinitesimal changes $\mathrm{d}z$:

$$P(\{z < X \leq z + \mathrm{d}z\}) = F_X(z + \mathrm{d}z) - F_X(z) = F_X'(z)\,\mathrm{d}z = \mathrm{d}F_X(z)$$

Density Function

If a function $f : \mathbb{R} \to \mathbb{R}$ exists with

$$F_X(z) = \int_{-\infty}^{z} f(s)ds$$

, it is called density function of X. In case f is continuous, $F_X' = f$ follows.

Expected Value

The expected value $\mathbb{E}(X)$ of a random variable X is defined as its mean function value

$$\mathbb{E}(X) = \int_{\mathbb{R}^d} z \,\mathrm{d}F_X(z)$$

Variance and Standard Deviation

The variance $\mathrm{Var}(X)$ of a random variable X is defined as the quadratic deviation of X from its expected value, i.e.

$$\mathrm{Var}(X) = \mathbb{E}(\mid X - \mathbb{E}(X) \mid^2)$$

A better intuition of the strength of dispersion can be had by examining the *standard deviation* $s(X)$

$$s(X) := \sqrt{\mathrm{Var}(X)}$$

Deterministic Convergence

Recall that convergence towards a limit L for a deterministic sequence S_n, with $n \in \mathbb{N}$, can be expressed by

$$\forall \epsilon > 0 \quad \exists N : \quad \forall n \geq N : \quad |S_n - L| < \epsilon$$

Convergence in Probability

A sequence of random variables $X_1, X_2, ...$, denoted by X_n, is said to converge in probability towards X, written as $X_n \xrightarrow{P} X$ or $\text{plim}_{n\to\infty} X_n = X$, if

$$\forall \epsilon > 0 : \quad P(|X_n - X| > \epsilon) \to 0 \quad \text{as} \quad n \to \infty$$

Convergence in Distribution

X_n is said to *converge in distribution* towards X, written as $X_n \xrightarrow{D} X$, if

$$\forall x \in \mathbb{R}: \quad P(X_n \leq x) \to P(X \leq x) \quad \text{as} \quad n \to \infty$$

, given that $P(X \leq x)$ is continuous at x.

Convergence in rth Mean

X_n converges *in rth mean* (for $r \geq 1$), written as $X_n \xrightarrow{r} X$, if

$$\forall n \in \mathbb{N} : \quad \mathbb{E}(|X_n|^r) < \infty$$

and

$$\mathbb{E}(|X_n - X|^r) \to 0 \quad \text{as} \quad n \to \infty$$