Probability Theory Lecture Slides

Mátyás Barczy and Gyula Pap

University of Szeged
Bolyai Institute
Szeged

Academic year 2023/2024, Semester II

Selective bibliography

- R. B. Ash: *Probability and Measure Theory, 2nd edition.*
  2000, Academic Press.
- K. B. Athreya, S. N. Lahiri: *Measure Theory and Probability Theory.*
  2006, Springer.
- P. Billingsley: *Probability and Measure, 3rd edition.*
  1995, Wiley-Interscience Publication.
- A. Klenke: *Probability Theory, 2nd edition.*
  2014, Springer.
- A. N. Shiryaev: *Probability-1, Probability-2, 3rd editions.*
  2016 and 2019, Springer.

Notations

- $\mathbb{Z}$: set of integers,
- $\mathbb{N}$: set of positive integers,
- $\mathbb{Z}_+$: set of nonnegative integers,
- $\mathbb{R}$: set of real numbers,
- $\mathbb{R}_+$: set of nonnegative real numbers,
- $\mathbb{C}$: set of complex numbers.

If Ω is a nonempty set and A is a subset of Ω, then we will denote it by $A \subset \Omega$ (where $\subset$ is not necessarily for strict inclusion, i.e., if $A \subset \Omega$, then $A = \Omega$ can occur as well).

Required knowledge of measure theory

Algebra, σ-algebra

Let $\Omega \neq \emptyset$ be a non-empty set. A set $\mathcal{H} \subset 2^{\Omega}$ consisting of certain subsets of Ω is called an **algebra** if

(i) $\Omega \in \mathcal{H}$,

(ii) closed under the union of pairs of sets, i.e., for any $A, B \in \mathcal{H}$, we have $A \cup B \in \mathcal{H}$,

(iii) closed under the complements of individual sets, i.e., for any $A \in \mathcal{H}$, we have $\overline{A} := \Omega \setminus A \in \mathcal{H}$.

An algebra $\mathcal{A} \subset 2^{\Omega}$ is called a **σ-algebra** if the following stricter version of (ii) holds:

(ii') closed under countable unions, i.e., for any $A_1, A_2, \dots \in \mathcal{A}$, we have $\displaystyle\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$.

Then the pair $(\Omega, \mathcal{A})$ is called a **measurable space**.

Required knowledge of measure theory

Measure

Let $\Omega \neq \emptyset$ be a nonempty set and $\mathcal{H} \subset 2^{\Omega}$ be an algebra.
A function $\mu : \mathcal{H} \to [0, \infty]$ is called

- **finitely additive**, if for any disjoint sets $A, B \in \mathcal{H}$, we have $\mu(A \cup B) = \mu(A) + \mu(B)$.
- a **measure**, if $\mu(\emptyset) = 0$ and it is $\boldsymbol{\sigma}$**-additive**, i.e.,
$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n)$$
for any pairwise disjoint sets $A_1, A_2, \dots \in \mathcal{H}$ satisfying $\displaystyle\bigcup_{n=1}^{\infty} A_n \in \mathcal{H}$.

If $\mu : \mathcal{H} \to [0, \infty]$ is finitely additive, then, by induction, one can show that for any $n \in \mathbb{N}$ and any pairwise disjoint sets $\{A_k\}_{k=1}^{n} \subset \mathcal{H}$, we have $\mu(\bigcup_{k=1}^{n} A_k) = \sum_{k=1}^{n} \mu(A_k)$.

Required knowledge of measure theory

Measure

Let $\Omega \neq \emptyset$ be a nonempty set, and $\mathcal{H} \subset 2^{\Omega}$ be an algebra.
A measure $\mu : \mathcal{H} \to [0, \infty]$ is called

- **finite**, if $\mu(\Omega) < \infty$.
- a **probability measure**, if $\mu(\Omega) = 1$.
- **σ-finite**, if there exist sets $\Omega_1, \Omega_2, \dots \in \mathcal{H}$ such that $\Omega = \bigcup_{k=1}^{\infty} \Omega_k$, and $\mu(\Omega_k) < \infty$, $k \in \mathbb{N}$.

A function $\mu : \mathcal{H} \to [-\infty, \infty]$ is called

- a **signed measure**, if it can be written in the form $\mu = \mu_1 - \mu_2$, where μ_1, μ_2 are measures, and at least one of them is finite.

Let Ω be a nonempty set. For each $n \in \mathbb{N}$, let $A_n \subset \Omega$.

If $A_1 \subset A_2 \subset \dots$ and $A := \bigcup_{n=1}^{\infty} A_n$, then we write that $A_n \uparrow A$.

If $A_1 \supset A_2 \supset \dots$ and $A := \bigcap_{n=1}^{\infty} A_n$, then we write that $A_n \downarrow A$.

Properties of a measure (e.g., continuity of a measure)

Let $\Omega \neq \emptyset$ be a nonempty set, and $\mathcal{H} \subset 2^{\Omega}$ be an algebra.
Let $\mathrm{P} : \mathcal{H} \to [0, \infty]$ be a finitely additive function such that $\mathrm{P}(\Omega) = 1$.
Then

1. $\mathrm{P}(\emptyset) = 0$;
2. for each $A \in \mathcal{H}$, we have $0 \leqslant \mathrm{P}(A) \leqslant 1$;
3. P is **monotone**, i.e., for each $A, B \in \mathcal{H}$, $A \subset B$, we have $\mathrm{P}(A) \leqslant \mathrm{P}(B)$, and we also have $\mathrm{P}(B \setminus A) = \mathrm{P}(B) - \mathrm{P}(A)$;
4. for each $A \in \mathcal{H}$, we have $\mathrm{P}(\overline{A}) = 1 - \mathrm{P}(A)$;
5. the following assertions are equivalent:
   - (a) P is σ-additive.
   - (b) P is **continuous from below**, i.e., for each $A_1, A_2, \dots \in \mathcal{H}$, $A_n \uparrow A$ and $A \in \mathcal{H}$, we have $\lim_{n\to\infty} \mathrm{P}(A_n) = \mathrm{P}(A)$.
   - (c) P is **continuous from above**, i.e., for each $A_1, A_2, \dots \in \mathcal{H}$, $A_n \downarrow A$ and $A \in \mathcal{H}$, we have $\lim_{n\to\infty} \mathrm{P}(A_n) = \mathrm{P}(A)$.
   - (d) P is **„continuous from above on the emptyset"**, i.e., for each $A_1, A_2, \dots \in \mathcal{H}$ and $A_n \downarrow \emptyset$, we have $\lim_{n\to\infty} \mathrm{P}(A_n) = 0$.

Required knowledge of measure theory

Additivity, subadditivity of a measure

Let $(\Omega, \mathcal{A})$ be a measurable space and $\mathrm{P} : \mathcal{A} \to [0, 1]$ be a probability measure. Then

1. P is finitely additive;
2. P is **σ-subadditive**, i.e., for each $A_1, A_2, \dots \in \mathcal{A}$, we have
$$\mathrm{P}\left(\bigcup_{n=1}^{\infty} A_n\right) \leqslant \sum_{n=1}^{\infty} \mathrm{P}(A_n).$$

One can check that the intersection of any nonempty family of σ-algebras is a σ-algebra.

Generated σ-algebra by a family of sets

Let $\Omega \neq \emptyset$ be a nonempty set, and $\mathcal{H} \subset 2^{\Omega}$ be an algebra. Let $\Gamma \neq \emptyset$ and for each $\gamma \in \Gamma$, let $A_\gamma \in \mathcal{H}$. The intersection of all the σ-algebras containing the sets A_γ, $\gamma \in \Gamma$, is called **the σ-algebra generated by the family of sets A_γ, $\gamma \in \Gamma$.** In notation: $\sigma(A_\gamma : \gamma \in \Gamma)$.

The definition of a generated σ-algebra can be also given for an arbitrary family of sets $A_\gamma \subset \Omega$, $\gamma \in \Gamma$ (not necessarily belonging to an algebra).

Required knowledge of measure theory

In fact, $\sigma(A_\gamma : \gamma \in \Gamma)$ is the smallest σ-algebra, which contains the sets A_γ, $\gamma \in \Gamma$.

Carathéodory extension theorem

Let $\Omega \neq \emptyset$ be a nonempty set, and $\mathcal{H} \subset 2^\Omega$ be an algebra.
Let $\mu : \mathcal{H} \to [0, \infty]$ be a σ-finite measure.
Then there exists a uniquely determined σ-finite measure $\nu : \sigma(\mathcal{H}) \to [0, \infty]$ such that for each $A \in \mathcal{H}$, we have $\nu(A) = \mu(A)$.

Probability space

By a **probability space**, we mean a triplet $(\Omega, \mathcal{A}, \mathrm{P})$, where $(\Omega, \mathcal{A})$ is a measurable space, and $\mathrm{P} : \mathcal{A} \to [0, 1]$ is a probability measure.

The elements of Ω are called **elementary (atomic) events**, and the elements of $\mathcal{A}$ are called **events.** The set Ω is called the **sure (certain) event**, and the emptyset $\emptyset$ is called the **impossible event**.

Required knowledge of measure theory

Random variable, and its distribution

Let $(\Omega, \mathcal{A}, \mathrm{P})$ be a probability space, $(X, \mathcal{X})$ be a measurable space. A function $\xi : \Omega \to X$ is called a **random variable**, if it is measurable, i.e., for each $B \in \mathcal{X}$, we have

$$\xi^{-1}(B) := \{\xi \in B\} := \{\omega \in \Omega : \xi(\omega) \in B\} \in \mathcal{A}.$$

The **distribution** of a random variable $\xi : \Omega \to X$ is the function $\mathrm{P}_\xi : \mathcal{X} \to \mathbb{R}$,

$$\mathrm{P}_\xi(B) := \mathrm{P}(\xi \in B) = \mathrm{P}(\xi^{-1}(B)), \qquad B \in \mathcal{X},$$

which is a probability measure on the measurable space $(X, \mathcal{X})$ (can be checked easily).

Required knowledge of measure theory

Discrete and simple random vectors

A random variable $\xi : \Omega \to X$ is called **discrete**, if its range, the set $\xi(\Omega)$, is countable. A random variable $\xi : \Omega \to X$ is called **simple**, if its range is a finite set.
If $\xi : \Omega \to X$ and $\eta : \Omega \to X$ are random variables and $\mathrm{P}(\xi = \eta) = 1$, then we write that $\xi = \eta$ P-a.s. (equality P–almost surely).

If $X = \mathbb{R}$, or $X = \mathbb{R}^d$, then we always choose $\mathcal{X} := \mathcal{B}(\mathbb{R})$, and $\mathcal{X} := \mathcal{B}(\mathbb{R}^d)$, respectively. So in this lecture a function $f : \mathbb{R} \to \mathbb{R}$ is called measurable if $f^{-1}(B) \in \mathcal{B}(\mathbb{R})$ for each $B \in \mathcal{B}(\mathbb{R})$ (in measure theory it is calld Borel measurability). If $\xi : \Omega \to \mathbb{R}^d$ is a random variable, then we call it a random vector as well.

If (E, ϱ) is a metric space, then we always furnish it with the Borel–σ–algebra $\mathcal{B}(E)$ (i.e., with the σ–algebra generated by the open sets).

Required knowledge of measure theory

Simple random vector

If $\xi : \Omega \to \mathbb{R}^d$ is a simple random vector and its range $\xi(\Omega) = \{x_1, \dots, x_k\}$, where $x_1, \dots, x_k \in \mathbb{R}^d$ are pairwise distinct, then

$$\xi = \sum_{j=1}^{k} x_j \mathbb{1}_{A_j},$$

where $A_j := \{\omega \in \Omega : \xi(\omega) = x_j\} \in \mathcal{A}$, $j = 1, \dots, k$, are pairwise disjoint events and $\bigcup_{j=1}^{k} A_j = \Omega$, i.e., $A_1, \dots, A_k$ is a so-called partition of Ω.

Required knowledge of measure theory

Generated σ–algebra

Let Γ be a nonempty set, and for each $\gamma \in \Gamma$ let $(X_\gamma, \mathcal{X}_\gamma)$ be a measurable space, and let $\xi_\gamma : \Omega \to X_\gamma$ be a random variable.
The **σ–algebra generated** by the random variables $\{\xi_\gamma : \gamma \in \Gamma\}$:

$$\sigma(\xi_\gamma : \gamma \in \Gamma) := \sigma(\xi_\gamma^{-1}(B) : \gamma \in \Gamma,\ B \in \mathcal{X}_\gamma).$$

The σ–algebra generated by the random variables $\{\xi_\gamma : \gamma \in \Gamma\}$ is the smallest σ–algebra with respect to all the random variables $\{\xi_\gamma : \gamma \in \Gamma\}$ are measurable.

σ–algebra generated by a single random variable

The σ–algebra generated by the (single) random variable $\xi : \Omega \to X$:

$$\sigma(\xi) = \xi^{-1}(\mathcal{X}) := \{\xi^{-1}(B) : B \in \mathcal{X}\}.$$

This σ–algebra consists of those events A which can be decided whether they occured or not ($\omega \in A$ holds or not) by observing ξ (in the knowledge of $\xi(\omega)$).

Required knowledge of measure theory

Note that if $\sigma(\xi) = \sigma(\eta)$, then in general it does not hold that $\mathsf{P}(\xi = \eta) = 1$. For example, if $\eta := \xi + 1$, then $\sigma(\xi) = \sigma(\eta)$, but $\mathsf{P}(\xi = \eta) = \mathsf{P}(\xi = \xi + 1) = 0$.

The definition of a generated σ-algebra can be given in case of a set of not necessarily mesaurable functions as well.

For example, the generated σ-algebra by an arbitrary function $g : \Omega \to \mathbb{R}^d$:

$$\sigma(g) := \sigma(g^{-1}(B) : B \in \mathcal{B}(\mathbb{R}^d)) = g^{-1}(\mathcal{B}(\mathbb{R}^d)) = \{g^{-1}(B) : B \in \mathcal{B}(\mathbb{R}^d)\},$$

and $\sigma(g)$ is the smallest σ-algebra with respect to g is measurable.

Measurability with respect to a sub-σ-algebra

Let $(\Omega, \mathcal{A}, \mathsf{P})$ be a probability space, $(X, \mathcal{X})$ be a measurable space, $\xi : \Omega \to X$ be a random variable and $\mathcal{F} \subseteq \mathcal{A}$ be a sub-σ-algebra. We say that ξ is $\mathcal{F}$-**measurable**, if $\xi^{-1}(B) \in \mathcal{F}$, $\forall\, B \in \mathcal{X}$, i.e., $\sigma(\xi) \subset \mathcal{F}$.

Required knowledge of measure theory

Separable metric space

A metric space is called **separable**, if it contains a countable, dense subset. A subset A of a metric space is called separable, if it is separable as a metric space by restricting the domain of the original metric to $A \times A$.

Approximation by simple random variables

Let (E, ϱ) be a separable metric space. For an arbitrary random variable $\xi : \Omega \to E$, there exist simple random variables $\{\xi_n\}_{n=1}^{\infty}$ such that for all $\omega \in \Omega$, we have $\lim_{n \to \infty} \xi_n(\omega) = \xi(\omega)$. If $(E, \|\,.\,\|)$ is a separable normed space, then ξ_n can be chosen such that $\|\xi_n\| \leqslant \|\xi\|, \forall\, n \in \mathbb{N}$.

Approximation by simple random variables

For an arbitrary nonnegative random variable $\eta : \Omega \to \mathbb{R}$, there exists a sequence $\{\eta_n\}_{n=1}^{\infty}$ of nonnegative simple random variables such that for each $\omega \in \Omega$, we have $\eta_n(\omega) \uparrow \eta(\omega)$ as $n \to \infty$.

Required knowledge of measure theory

For example, one can choose the following sequences:

$$\eta_n = \sum_{j=1}^{n2^n} (j-1)2^{-n} \mathbb{1}_{\{(j-1)2^{-n} \leqslant \eta < j2^{-n}\}}, \qquad n \in \mathbb{N},$$

$$\eta_n = \sum_{j=1}^{n2^n} (j-1)2^{-n} \mathbb{1}_{\{(j-1)2^{-n} \leqslant \eta < j2^{-n}\}} + n\mathbb{1}_{\{\eta \geqslant n\}}, \qquad n \in \mathbb{N}.$$

Measurable function of a random variable

Let $(X, \mathcal{X})$ be a measurable space, $\xi : \Omega \to X$ be a random variable.

1. If $(Y, \mathcal{Y})$ is a measurable space, $g : X \to Y$ is a measurable function, then the composite function $g \circ \xi : \Omega \to Y$ is a $\sigma(\xi)$–measurable random variable, i.e., $\sigma(g \circ \xi) \subset \sigma(\xi)$.
2. If $\eta : \Omega \to \mathbb{R}^d$ is a $\sigma(\xi)$–measurable random variable, then there exists a measurable function $g : X \to \mathbb{R}^d$ such that $\eta = g \circ \xi$.

Required knowledge of measure theory

„Good sets” principle

Let $(\Omega, \mathcal{A})$ and $(X, \mathcal{X})$ be measurable spaces, $\mathcal{E} \subset \mathcal{X}$, and $\xi : \Omega \to X$ be a mapping. Then $\sigma(\xi^{-1}(\mathcal{E})) = \xi^{-1}(\sigma(\mathcal{E}))$. Further, supposing that $\sigma(\mathcal{E}) = \mathcal{X}$, the mapping ξ is a random variable if and only if $\xi^{-1}(\mathcal{E}) \subset \mathcal{A}$.

Measurability of vector-valued mappings

Let $(\Omega, \mathcal{A})$ be a measurable space. Then a mapping $\xi : \Omega \to \mathbb{R}^d$ is a random vector if and only if $\{\omega \in \Omega : \xi(\omega) < x\} \in \mathcal{A}$ for all $x \in \mathbb{R}^d$.
For a mapping $\xi : \Omega \to \mathbb{R}^d$, the σ-algebra $\sigma(\xi)$ is the smallest sub-σ-algebra with respect to ξ is measurable.

Let $(\Omega, \mathcal{A})$ be a measurable space, $d \in \mathbb{N}$, and $\xi_1, \ldots, \xi_d : \Omega \to \mathbb{R}$ be mappings. Let $\xi : \Omega \to \mathbb{R}^d$, $\xi(\omega) := (\xi_1(\omega), \ldots, \xi_d(\omega))$, $\omega \in \Omega$.
Then ξ is a $\mathbb{R}^d$-valued random vector if and only if ξ_i, $i = 1, \ldots, d$, are real-valued random variables.

Required knowledge of measure theory

Distribution function of a random vector

By the distribution function of a random variable $\xi : \Omega \to \mathbb{R}^d$, $\xi = (\xi_1, \dots, \xi_d)$, we mean the function $F_\xi : \mathbb{R}^d \to [0, 1]$,

$$F_\xi(x) := \mathrm{P}(\xi < x) = \mathrm{P}(\xi_1 < x_1, \dots, \xi_d < x_d), \quad x = (x_1, \dots, x_d)^\top \in \mathbb{R}^d.$$

Let $g : \mathbb{R}^d \to \mathbb{R}$, $a_j, b_j \in \mathbb{R}$, $a_j < b_j$, $j \in \{1, \dots, d\}$, and $\Delta^{(j)}_{[a_j,b_j)} g : \mathbb{R}^d \to \mathbb{R}$,

$$\begin{aligned}(\Delta^{(j)}_{[a_j,b_j)} g)(x) := g(x_1, \dots, x_{j-1}, b_j, x_{j+1}, \dots, x_d)& \\ - g(x_1, \dots, x_{j-1}, a_j, x_{j+1}, \dots, x_d),& \qquad x \in \mathbb{R}^d.\end{aligned}$$

Then for each $x \in \mathbb{R}^d$, we have

$$\Delta^{(1)}_{[a_1,b_1)} \dots \Delta^{(d)}_{[a_d,b_d)} g(x) = \sum_{(\varepsilon_1,\dots,\varepsilon_d) \in \{0,1\}^d} (-1)^{\sum_{k=1}^d \varepsilon_k} g(c_1, \dots, c_d),$$

where $c_k := \varepsilon_k a_k + (1 - \varepsilon_k) b_k$, $k = 1, \dots, d$.
Hence $\Delta^{(1)}_{[a_1,b_1)} \dots \Delta^{(d)}_{[a_d,b_d)} g$ is a constant function.

Required knowledge of measure theory

If $a, b \in \mathbb{R}^d$, then $a \leqslant b$, and $a < b$ means that for each $j = 1, \ldots, d$, we have $a_j \leqslant b_j$, and $a_j < b_j$, respectively, and let $[a, b) := \{x \in \mathbb{R}^d : a \leqslant x < b\}$.

Characterisation of a multidimensional distribution function

A function $F : \mathbb{R}^d \to \mathbb{R}$ is a distribution function of some random variable $\xi : \Omega \to \mathbb{R}^d$ if and only if

(1) F is monotone increasing in all its variables,

(2) F is left-continuous in all its variables,

(3) $\lim\limits_{\min\{x_1,\ldots,x_d\}\to\infty} F(x) = 1$, and

$$\lim_{x_i \to -\infty} F(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_d) = 0$$

for all $i \in \{1, \ldots, d\}$ and $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_d \in \mathbb{R}$,

(4) for each $a, b \in \mathbb{R}^d$, $a < b$, we have $\Delta^{(1)}_{[a_1,b_1)} \ldots \Delta^{(d)}_{[a_d,b_d)} F \geqslant 0$.

Required knowledge of measure theory

If $d = 1$, then condition (4) is implied by condition (1).

Probability of belonging to a rectangle

If $\xi : \Omega \to \mathbb{R}^d$ is a random variable, then for each $a, b \in \mathbb{R}^d$, $a < b$, we have

$$\mathrm{P}_\xi([a,b)) = \mathrm{P}(\xi \in [a,b)) = \Delta^{(1)}_{[a_1,b_1)} \dots \Delta^{(d)}_{[a_d,b_d)} F_\xi \geqslant 0,$$

where F_ξ denotes the distribution function of ξ. Hence P_ξ is nothing else but the Lebesgue-Stieltjes measure corresponding to the distribution function F_ξ.

Equality of one-dimensional distribution functions

Let $F : \mathbb{R} \to [0,1]$ and $G : \mathbb{R} \to [0,1]$ be one-dimensional distribution functions. If $F(x) = G(x)$ for all the common continuity points $x \in \mathbb{R}$ of F and G, then $F = G$. More generally, if $S \subset \mathbb{R}$ is a dense subset of $\mathbb{R}$ such that $F(x) = G(x)$ for all $x \in S$, then $F = G$.

Independence, Kolmogorov 0–1 law

Independence of σ-algebras, events and random vectors

Let $(\Omega, \mathcal{A}, \mathrm{P})$ be a probability space, $\Gamma \neq \emptyset$ be a nonempty set.

- For each $\gamma \in \Gamma$, let $\mathcal{F}_\gamma \subset \mathcal{A}$ be a sub-σ-algebra. We say that the sub-σ-algebras $\{\mathcal{F}_\gamma : \gamma \in \Gamma\}$ are **independent**, if for each finite subset $\{\gamma_1, \ldots, \gamma_n\}$ consisting of distinct elements of Γ and for each $A_{\gamma_1} \in \mathcal{F}_{\gamma_1}, \ldots, A_{\gamma_n} \in \mathcal{F}_{\gamma_n}$, we have
$$\mathrm{P}(A_{\gamma_1} \cap \ldots \cap A_{\gamma_n}) = \mathrm{P}(A_{\gamma_1}) \cdots \mathrm{P}(A_{\gamma_n}).$$
- For each $\gamma \in \Gamma$, let $A_\gamma \in \mathcal{A}$. We say that the events $\{A_\gamma : \gamma \in \Gamma\}$ are **independent**, if the corresponding sub-σ-algebras $\big\{\{\emptyset, A_\gamma, \Omega \setminus A_\gamma, \Omega\} : \gamma \in \Gamma\big\}$ are independent.
- For each $\gamma \in \Gamma$, let $(X_\gamma, \mathcal{X}_\gamma)$ be a measurable space and $\xi_\gamma : \Omega \to X_\gamma$ be a random variable. We say that the random variables $\{\xi_\gamma : \gamma \in \Gamma\}$ are **independent**, if the corresponding (generated) σ–algebras $\big\{\sigma(\xi_\gamma) : \gamma \in \Gamma\big\}$ are independent.

Independence, Kolmogorov 0–1 law

The random variables $\xi : \Omega \to \mathbb{R}$ and $\eta : \Omega \to \mathbb{R}$ are independent if and only if $F_{\xi,\eta}(x, y) = F_\xi(x)F_\eta(y)$, $x, y \in \mathbb{R}$, where $F_{\xi,\eta}$, F_ξ and F_η denotes the distribution function of (ξ, η), ξ, and η, respectively.

Functions of independent random vectors are independent

Let $(\Omega, \mathcal{A}, \mathrm{P})$ be a probability space. If the random vectors $\xi : \Omega \to \mathbb{R}^k$ and $\eta : \Omega \to \mathbb{R}^\ell$ are independent, than for all measurable functions $g : \mathbb{R}^k \to \mathbb{R}^r$ and $h : \mathbb{R}^\ell \to \mathbb{R}^p$, we have the random vectors $g \circ \xi : \Omega \to \mathbb{R}^r$ and $h \circ \eta : \Omega \to \mathbb{R}^p$ are independent as well.

σ-algebras generated by independent algebras are independent

Let $(\Omega, \mathcal{A}, \mathrm{P})$ be a probability space. If the sub-algebras $\mathcal{F}_0 \subset \mathcal{A}$ and $\mathcal{G}_0 \subset \mathcal{A}$ are independent in the sense that for each $A \in \mathcal{F}_0$ and $B \in \mathcal{G}_0$, we have

$$\mathrm{P}(A \cap B) = \mathrm{P}(A)\mathrm{P}(B),$$

then the generated sub–σ–algebras $\mathcal{F} := \sigma(\mathcal{F}_0)$ and $\mathcal{G} := \sigma(\mathcal{G}_0)$ are independent as well.

Independence, Kolmogorov 0–1 law

Notation for σ-algebra generated by sub-σ-algebras

Let $(\Omega, \mathcal{A}, \mathrm{P})$ be a probability space, $\Gamma \neq \emptyset$ be a nonempty set. For each $\gamma \in \Gamma$, let $\mathcal{F}_\gamma$ be a sub–σ–algebra of $\mathcal{A}$.
Let $\mathcal{F}_\emptyset := \{\emptyset, \Omega\}$ (i.e., the trivial σ–algebra).
If $\Lambda \subset \Gamma$, $\Lambda \neq \emptyset$, then let

$$\mathcal{F}_\Lambda := \bigvee_{\gamma \in \Lambda} \mathcal{F}_\gamma := \sigma\left(\mathcal{F}_\gamma : \gamma \in \Lambda\right) := \sigma\left(\bigcup_{\gamma \in \Lambda} \mathcal{F}_\gamma\right).$$

σ-algebras generated by independent σ-algebras are independent

Let $(\Omega, \mathcal{A}, \mathrm{P})$ be a probability space, $\Gamma \neq \emptyset$ be a nonempty set. If $\{\mathcal{F}_\gamma : \gamma \in \Gamma\}$ are independent sub–σ–algebras of $\mathcal{A}$, and F_1, F_2 are finite, disjoint subsets of Γ, then $\mathcal{F}_{F_1}$ and $\mathcal{F}_{F_2}$ are independent.

Let $(\Omega, \mathcal{A}, \mathrm{P})$ be a probability space, $\Gamma \neq \emptyset$ be a nonempty set. If $\{\mathcal{F}_\gamma : \gamma \in \Gamma\}$ are independent sub–σ–algebras of $\mathcal{A}$, and F_1, F_2 are disjoint subsets of Γ, then $\mathcal{F}_{F_1}$ and $\mathcal{F}_{F_2}$ are independent.

Independence, Kolmogorov 0–1 law

Tail–σ–algebra

Let $(\Omega, \mathcal{A})$ be a measurable space, $\Gamma \neq \emptyset$ be a nonempty set. For each $\gamma \in \Gamma$, let $\mathcal{F}_\gamma$ be a sub–σ–algebra of $\mathcal{A}$. The **tail–σ–algebra** corresponding to the σ–algebras $\{\mathcal{F}_\gamma : \gamma \in \Gamma\}$ is defined by

$$\mathcal{T} := \bigcap_{\{F \,:\, F \subset \Gamma,\, F \text{ finite}\}} \mathcal{F}_{\Gamma \setminus F}.$$

1. If Γ is finite, then $\mathcal{T} = \{\emptyset, \Omega\}$, and hence $\mathrm{P}(A) \in \{0, 1\}$, $A \in \mathcal{T}$.
2. For a sequence of sub–σ–algebras $\{\mathcal{F}_n\}_{n=1}^\infty$, the tail–$\sigma$–algebra is
$$\mathcal{T} = \bigcap_{n=1}^\infty \sigma(\mathcal{F}_k : k \geqslant n),$$
where $\sigma(\mathcal{F}_k : k \geqslant n) \downarrow \mathcal{T}$ as $n \to \infty$.

Independence, Kolmogorov 0–1 law

3. If $(\Omega, \mathcal{A}, \mathrm{P})$ is a probability space, ξ_n, $n \in \mathbb{N}$, are random variables, then the following events belong to the tail–σ–algebra corresponding to the sub–σ–algebras $\sigma(\xi_n)$, $n \in \mathbb{N}$:

$$\left\{\omega \in \Omega : \lim_{n\to\infty} \xi_n(\omega) \text{ exists}\right\},$$

$$\left\{\omega \in \Omega : \limsup_{n\to\infty} \xi_n(\omega) \leqslant x\right\}, \qquad x \in \mathbb{R},$$

$$\left\{\omega \in \Omega : \lim_{n\to\infty} \xi_n(\omega) \text{ exists and } \lim_{n\to\infty} \xi_n(\omega) \leqslant x\right\}, \qquad x \in \mathbb{R},$$

$$\left\{\omega \in \Omega : \lim_{n\to\infty} \frac{\xi_1(\omega) + \cdots + \xi_n(\omega)}{n} \text{ exists}\right\}.$$

An event belongs to the tail–σ–algebra in question if and only if its occurrence does not depend on changing the values of finite number of ξ_n. Indeed, for each $N \in \mathbb{N}$,

$$\mathcal{T} = \bigcap_{n=1}^{\infty} \sigma(\xi_n, \xi_{n+1}, \ldots) = \bigcap_{n=N}^{\infty} \sigma(\xi_n, \xi_{n+1}, \ldots).$$

Independence, Kolmogorov 0–1 law

However, the event

$$\Big\{\omega \in \Omega : \xi_n(\omega) = 0,\ \forall\, n \in \mathbb{N}\Big\}$$

does not belong to the tail–σ–algebra corresponding to the sub–σ–algebras $\sigma(\xi_n)$, $n \in \mathbb{N}$:

Tail–σ–algebra for countably infinite Γ

Let Γ be a countably infinite set. For each $\gamma \in \Gamma$, let $\mathcal{F}_\gamma$ be a sub–σ–algebra of $\mathcal{A}$. Further, let $F_n \subset \Gamma$, $n \in \mathbb{N}$, be finite subsets of Γ such that $F_n \uparrow \Gamma$ as $n \to \infty$. Then the tail–σ–algebra corresponding to the σ–algebras $\{\mathcal{F}_\gamma : \gamma \in \Gamma\}$ takes the form

$$\mathcal{T} = \bigcap_{n=1}^{\infty} \mathcal{F}_{\Gamma \setminus F_n}.$$

In particular, in case of $\Gamma = \mathbb{N}$, we have $\mathcal{T} = \bigcap_{n=1}^{\infty} \sigma(\mathcal{F}_k : k \geqslant n)$ (as we already saw).

Independence, Kolmogorov 0–1 law

Kolmogorov 0–1 law

Let $(\Omega, \mathcal{A}, \mathrm{P})$ be a probability space, $\Gamma \neq \emptyset$ be a nonempty set. For each $\gamma \in \Gamma$, let $\mathcal{F}_\gamma$ be a sub–σ–algebra of $\mathcal{A}$, and denote by $\mathcal{T}$ the corresponding tail–σ–algebra.
If the sub–σ–algebras $\{\mathcal{F}_\gamma : \gamma \in \Gamma\}$ are independent, then for each $A \in \mathcal{T}$, we have $\mathrm{P}(A) = 0$ or $\mathrm{P}(A) = 1$.

Kolmogorov 0–1 law

Let $\{\xi_n\}_{n=1}^\infty$ be independent random variables and let $\mathcal{T} = \bigcap_{n=1}^\infty \sigma(\xi_k : k \geqslant n)$ denote the tail–σ–algebra corresponding to the sub–σ–algebras $\{\sigma(\xi_n)\}_{n=1}^\infty$.
Then for each $A \in \mathcal{T}$, we have $\mathrm{P}(A) = 0$ or $\mathrm{P}(A) = 1$.

Independence, Kolmogorov 0–1 law

Example: If $\xi_1, \xi_2, \dots$ are independent random variables and

$$\overline{S}_n := \frac{\xi_1 + \cdots + \xi_n}{n}, \qquad n \in \mathbb{N},$$

then

$$\mathrm{P}\left(\{\overline{S}_n\}_{n=1}^{\infty} \text{ converges} \right) \in \{0, 1\},$$

and there exist $-\infty \leqslant a \leqslant b \leqslant \infty$ such that

$$\mathrm{P}\left(\liminf_{n\to\infty} \overline{S}_n = a \right) = 1, \qquad \mathrm{P}\left(\limsup_{n\to\infty} \overline{S}_n = b \right) = 1.$$

So, if $\mathrm{P}\left(\{\overline{S}_n\}_{n=1}^{\infty} \text{ converges} \right) = 1$, then there exits $c \in \mathbb{R}$ such that

$$\mathrm{P}\left(\lim_{n\to\infty} \overline{S}_n = c \right) = 1.$$

Independence, Kolmogorov 0–1 law

$\limsup$ and $\liminf$ of countably many sets

If $\Omega \neq \emptyset$ is a nonempty set, and for each $n \in \mathbb{N}$, $A_n \subset \Omega$, then let

$$\limsup_{n\to\infty} A_n := \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k = \{\omega \in \Omega : \omega \in A_n \text{ for infinitely many } n \in \mathbb{N}\},$$

$$\liminf_{n\to\infty} A_n := \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k = \{\omega \in \Omega : \omega \in A_n \text{ except finitely many } n \in \mathbb{N}\}.$$

Let $\{A_n\}_{n=1}^{\infty}$ be events in a probability space $(\Omega, \mathcal{A}, \mathrm{P})$. Then the events $\limsup_{n\to\infty} A_n$ and $\liminf_{n\to\infty} A_n$ are in the tail–σ–algebra corresponding to the σ–algebras $\{\emptyset, A_n, \Omega \setminus A_n, \Omega\}$, $n \in \mathbb{N}$.

If A_n, $n \in \mathbb{N}$, are independent as well, then, by Kolmogorov 0–1 law, $\mathrm{P}(\limsup_{n\to\infty} A_n) \in \{0, 1\}$, i.e., either infinitely many of these events occur with probability 1 or at most finitely many of them occur with probability 1.

Borel–Cantelli lemmas (1909, 1917)

Let $(\Omega, \mathcal{A}, \mathrm{P})$ be a probability space, and $A_1, A_2, \dots \in \mathcal{A}$ be events.

1. If $\displaystyle\sum_{n=1}^{\infty} \mathrm{P}(A_n) < \infty$, then $\mathrm{P}\left(\limsup_{n\to\infty} A_n\right) = 0$ (i.e., at most finitely many of these events occur with probability 1).
2. If the events $\{A_n\}_{n=1}^{\infty}$ are independent and $\displaystyle\sum_{n=1}^{\infty} \mathrm{P}(A_n) = \infty$, then $\mathrm{P}\left(\limsup_{n\to\infty} A_n\right) = 1$ (i.e., infinitely many of these events occur with probability 1).

Independence, Kolmogorov 0–1 law

For each $\omega \in \Omega$, let $\mathcal{N}(\omega)$ be the number of events A_n, $n \in \mathbb{N}$, for which $\omega \in A_n$ holds.

Then $\mathcal{N}(\omega) \in \{0, 1, 2, \ldots\} \cup \{\infty\}$, $\mathcal{N} = \sum_{n=1}^{\infty} \mathbb{1}_{A_n}$, $\mathcal{N}$ is an (extended real valued) random variable, and using the properties of expectation (presented later on), we have

$$\mathrm{E}(\mathcal{N}) = \mathrm{E}\left(\sum_{n=1}^{\infty} \mathbb{1}_{A_n}\right) = \sum_{n=1}^{\infty} \mathrm{P}(A_n).$$

Part 1 of Borel-Cantelli lemma states that if the expectation of the number of events occuring is finite, then the number of events occuring is finite with probability one.

Further, since $\limsup_{n\to\infty} A_n = \{\mathcal{N} = \infty\}$, by part 2 of Borel-Cantelli lemma, in case of independent events, if the expectation of the number of events occuring is infinite, then $\mathcal{N}$, the number of events occuring, is infinite with probability 1.

Expectation (expected value)

Expectation of simple random variables

Let $\xi : \Omega \to \mathbb{R}$ be a simple random variable, and $\xi(\Omega) = \{x_1, \ldots, x_\ell\}$, where $x_1, \ldots, x_\ell \in \mathbb{R}$ are pairwise distinct. Then the quantity

$$\mathrm{E}(\xi) := \int_\Omega \xi(\omega)\, \mathrm{P}(\mathrm{d}\omega) := \sum_{j=1}^{\ell} x_j \mathrm{P}(\xi = x_j)$$

is called the **expectation** of ξ.

One can check that the expectation is finitely additive and monotone on the set of simple random variables.

Let $\xi : \Omega \to \mathbb{R}$ be a nonnegative random variable.

1. If ζ and $\{\eta_n\}_{n=1}^\infty$ are nonnegative simple random variables, and for each $\omega \in \Omega$, we have $\eta_n(\omega) \uparrow \xi(\omega) \geqslant \zeta(\omega)$, then $\lim_{n\to\infty} \mathrm{E}(\eta_n) \geqslant \mathrm{E}(\zeta)$.
2. If $\{\eta_n\}_{n=1}^\infty$ and $\{\zeta_n\}_{n=1}^\infty$ are nonnegative simple random variables, and for each $\omega \in \Omega$, we have $\eta_n(\omega) \uparrow \xi(\omega)$ and $\zeta_n(\omega) \uparrow \xi(\omega)$, then $\lim_{n\to\infty} \mathrm{E}(\eta_n) = \lim_{n\to\infty} \mathrm{E}(\zeta_n)$.

Expectation (expected value)

Expectation of nonnegative random variables

Let $\xi : \Omega \to \mathbb{R}$ be a nonnegative random variable. Let $\{\xi_n\}_{n=1}^{\infty}$ be a sequence of nonnegative simple random variables such that for each $\omega \in \Omega$, we have $\xi_n(\omega) \uparrow \xi(\omega)$ as $n \to \infty$.
Then the quantity

$$\mathsf{E}(\xi) := \int_\Omega \xi(\omega)\, \mathsf{P}(\mathrm{d}\omega) := \lim_{n\to\infty} \mathsf{E}(\xi_n)$$

is called the **expectation** of ξ.

The expectation $\mathsf{E}(\xi) \in [0, \infty]$ of a nonnegative random variable ξ is uniquely defined. Further,

$$\mathsf{E}(\xi) = \sup \{\mathsf{E}(\eta) : \eta \text{ is a simple random variable such that } 0 \leqslant \eta \leqslant \xi\}.$$

Expectation (expected value)

Decomposition of a r. v. by positive and negative parts

If $\xi : \Omega \to \mathbb{R}$ is a random variable, then $\xi^+ := \max\{\xi, 0\}$ (positive part of ξ) and $\xi^- := -\min\{\xi, 0\}$ (negative part of ξ) are nonnegative random variables as well, and $\xi = \xi^+ - \xi^-$, $|\xi| = \xi^+ + \xi^-$.

Expectation of a random variable

We say that **there exists the expectation (integral)** of a random variable $\xi : \Omega \to \mathbb{R}$, if the at least one of the expectations $\mathrm{E}(\xi^+)$ and $\mathrm{E}(\xi^-)$ is finite, and then

$$\mathrm{E}(\xi) := \int_\Omega \xi(\omega)\, \mathrm{P}(\mathrm{d}\omega) := \mathrm{E}(\xi^+) - \mathrm{E}(\xi^-).$$

We say that the **expectation of ξ is finite (ξ is integrable)**, if the expectations $\mathrm{E}(\xi^+)$ and $\mathrm{E}(\xi^-)$ are finite.

If $\xi : \Omega \to \mathbb{R}$ is a random variable and its expectation exists, then $\mathrm{E}(\xi) \in [-\infty, \infty]$.

Expectation (expected value)

Let ξ, η, $(\xi_n)_{n\in\mathbb{N}}$ be random variables on the prob. space $(\Omega, \mathcal{A}, \mathsf{P})$.

Properties of expectation

1. ξ is integrable if and only if $|\xi|$ is integrable.
2. If $\exists\, \mathsf{E}(\xi)$ and $c \in \mathbb{R}$, then $\exists\, \mathsf{E}(c\xi)$, and $\mathsf{E}(c\xi) = c\,\mathsf{E}(\xi)$.
3. If $\exists\, \mathsf{E}(\xi) > -\infty$ and $\xi \leqslant \eta$ P-a.s., then $\exists\, \mathsf{E}(\eta)$ and $\mathsf{E}(\xi) \leqslant \mathsf{E}(\eta)$.
4. If $\exists\, \mathsf{E}(\xi)$, then $|\mathsf{E}(\xi)| \leqslant \mathsf{E}(|\xi|)$.
5. If $\exists\, \mathsf{E}(\xi)$, then for all $A \in \mathcal{A}$, we have $\exists\, \mathsf{E}(\xi \mathbb{1}_A)$; if ξ is integrable, then for all $A \in \mathcal{A}$, we have $\xi \mathbb{1}_A$ is integrable as well.
6. If $\exists\, \mathsf{E}(\xi), \mathsf{E}(\eta)$ and the expression $\mathsf{E}(\xi) + \mathsf{E}(\eta)$ is meaningful (i.e., it is not of the form $\infty - \infty$ or $-\infty + \infty$), then $\exists\, \mathsf{E}(\xi + \eta)$ and $\mathsf{E}(\xi + \eta) = \mathsf{E}(\xi) + \mathsf{E}(\eta)$.
7. If $\xi = 0$ P-a.s., then $\mathsf{E}(\xi) = 0$.
8. If $\exists\, \mathsf{E}(\xi)$ and $\xi = \eta$ P-a.s., then $\exists\, \mathsf{E}(\eta)$ and $\mathsf{E}(\xi) = \mathsf{E}(\eta)$.
9. If $\xi \geqslant 0$ P-a.s. and $\mathsf{E}(\xi) = 0$, then $\xi = 0$ P-a.s.

Expectation (expected value)

Properties of expectation

10. **Monotone convergence theorem:** If for each $n \in \mathbb{N}$, we have $\xi_n \geqslant \eta$ P-a.s., $\mathrm{E}(\eta) > -\infty$, and $\xi_n \uparrow \xi$ P-a.s., then $\mathrm{E}(\xi_n) \uparrow \mathrm{E}(\xi)$ as $n \to \infty$.

11. If $\{\xi_n\}_{n=1}^{\infty}$ are nonnegative, then $\mathrm{E}\left(\sum_{n=1}^{\infty} \xi_n\right) = \sum_{n=1}^{\infty} \mathrm{E}(\xi_n)$.

12. **Fatou-lemma:**
    (a) If for each $n \in \mathbb{N}$, we have $\xi_n \geqslant \eta$ P-a.s. and $\mathrm{E}(\eta) > -\infty$, then $\mathrm{E}\left(\liminf_{n\to\infty} \xi_n\right) \leqslant \liminf_{n\to\infty} \mathrm{E}(\xi_n)$.
    (b) If for each $n \in \mathbb{N}$, we have $\xi_n \leqslant \eta$ P-a.s. and $\mathrm{E}(\eta) < \infty$, then $\limsup_{n\to\infty} \mathrm{E}(\xi_n) \leqslant \mathrm{E}\left(\limsup_{n\to\infty} \xi_n\right)$.
    (c) If for each $n \in \mathbb{N}$, we have $|\xi_n| \leqslant \eta$ P-a.s. and $\mathrm{E}(\eta) < \infty$, then

$$\mathrm{E}\left(\liminf_{n\to\infty} \xi_n\right) \leqslant \liminf_{n\to\infty} \mathrm{E}(\xi_n) \leqslant \limsup_{n\to\infty} \mathrm{E}(\xi_n) \leqslant \mathrm{E}\left(\limsup_{n\to\infty} \xi_n\right).$$

Expectation (expected value)

Properties of expectation

13. **Dominated convergence theorem:** If for each $n \in \mathbb{N}$, we have $|\xi_n| \leqslant \eta$ P-a.s., $\mathrm{E}(\eta) < \infty$, and $\xi_n \to \xi$ P-a.s., then $\mathrm{E}(|\xi|) < \infty$, $\mathrm{E}(\xi_n) \to \mathrm{E}(\xi)$ and $\mathrm{E}(|\xi_n - \xi|) \to 0$ as $n \to \infty$.
14. **Generalized dominated convergence theorem:**
    (a) If for each $n \in \mathbb{N}$, we have $|\xi_n| \leqslant \eta_n$ P-a.s., $\mathrm{E}(\eta_n) < \infty$, $\xi_n \to \xi$ P-a.s., $\eta_n \to \eta$ P-a.s., and $\mathrm{E}(\eta_n) \to \mathrm{E}(\eta)$ as $n \to \infty$, where $\mathrm{E}(\eta) < \infty$, then $\mathrm{E}(|\xi|) < \infty$ and $\mathrm{E}(\xi_n) \to \mathrm{E}(\xi)$ as $n \to \infty$.

    (b) If for each $n \in \mathbb{N}$, we have $|\xi_n| \leqslant \eta$ P-a.s., $\mathrm{E}(\eta) < \infty$, and ξ_n converges in probability to ξ as $n \to \infty$, then $\mathrm{E}(|\xi|) < \infty$, $\mathrm{E}(\xi_n) \to \mathrm{E}(\xi)$ and $\mathrm{E}(|\xi_n - \xi|) \to 0$ as $n \to \infty$.
15. **Cauchy–Schwarz inequality:** If $\mathrm{E}(\xi^2), \mathrm{E}(\eta^2) < \infty$, then $\mathrm{E}(|\xi\eta|) \leqslant \sqrt{\mathrm{E}(\xi^2)\,\mathrm{E}(\eta^2)}$.

Expectation (expected value)

Properties of expectation

16. **Jensen inequality:**
    (a) If $\mathsf{E}(|\xi|) < \infty$, $I \subset \mathbb{R}$ is an open (not necessarily bounded) interval such that $\mathsf{P}(\xi \in I) = 1$, and $g : I \to \mathbb{R}$ is convex, then $\mathsf{E}(\xi) \in I$ and $g(\mathsf{E}(\xi)) \leqslant \mathsf{E}(g(\xi))$.
    (b) Let $C \subset \mathbb{R}$ be a nonempty, Borel measurable, convex set, $g : C \to \mathbb{R}$ be a convex function, $\xi : \Omega \to C$ be a random variable such that $\mathsf{E}(|\xi|) < \infty$ and $g \circ \xi : \Omega \to \mathbb{R}$ is a random variable as well. Then $\mathsf{E}(\xi) \in C$, the expectation $\mathsf{E}(g(\xi))$ exists and $\mathsf{E}(g(\xi)) \in (-\infty, +\infty]$, further $g(\mathsf{E}(\xi)) \leqslant \mathsf{E}(g(\xi))$.
17. **Lyapunov inequality:** If $0 < s < t$, then $(\mathsf{E}(|\xi|^s))^{1/s} \leqslant \big(\mathsf{E}(|\xi|^t)\big)^{1/t}$.
18. **Hölder inequality:** Let $p, q \in (1, \infty)$ be such that $p^{-1} + q^{-1} = 1$. If $\mathsf{E}(|\xi|^p) < \infty$ and $\mathsf{E}(|\eta|^q) < \infty$, then $\mathsf{E}(|\xi\eta|) \leqslant (\mathsf{E}(|\xi|^p))^{1/p} (\mathsf{E}(|\eta|^q))^{1/q}$.
19. **Minkowski inequality:** If $p \in [1, \infty)$, $\mathsf{E}(|\xi|^p) < \infty$ and $\mathsf{E}(|\eta|^p) < \infty$, then $(\mathsf{E}\,|\xi + \eta|^p)^{1/p} \leqslant (\mathsf{E}(|\xi|^p))^{1/p} + (\mathsf{E}(|\eta|^p))^{1/p}$.

Expectation (expected value)

Properties of expectation

20. **Markov inequality:** If $\xi \geqslant 0$ P-a.s., then $\mathsf{P}(\xi \geqslant c) \leqslant \frac{\mathsf{E}(\xi)}{c}$ for all $c > 0$.
21. **Chebyshev inequality:** If $\mathsf{E}(\xi^2) < \infty$, then $\mathsf{P}(|\xi - \mathsf{E}(\xi)| \geqslant c) \leqslant \frac{\mathsf{Var}(\xi)}{c^2}$ for all $c > 0$.
22. If $\mathsf{E}(\xi)$ exists, then

$$\mathsf{E}(\xi) = \int_0^\infty \mathsf{P}(\xi \geqslant x)\,dx - \int_{-\infty}^0 \mathsf{P}(\xi < x)\,dx = \int_0^\infty (1 - F_\xi(x))\,dx - \int_{-\infty}^0 F_\xi(x)\,dx.$$

If $\xi \geqslant 0$ P-a.s., then $\mathsf{E}(\xi) = \int_0^\infty \mathsf{P}(\xi \geqslant x)\,dx = \int_0^\infty (1 - F_\xi(x))\,dx$.
In particular, if $\mathsf{P}(\xi \in \mathbb{Z}_+) = 1$, then $\mathsf{E}(\xi) = \sum_{n=1}^\infty \mathsf{P}(\xi \geqslant n)$.

23. If $\mathsf{E}(|\xi|) < \infty$, $\mathsf{E}(|\eta|) < \infty$ and ξ, η are independent, then $\mathsf{E}(|\xi\eta|) < \infty$ and $\mathsf{E}(\xi\eta) = \mathsf{E}(\xi)\,\mathsf{E}(\eta)$. If ξ and η are nonnegative and independent, then $\mathsf{E}(\xi\eta) = \mathsf{E}(\xi)\,\mathsf{E}(\eta)$.

Expectation (expected value)

Transformation theorem

If $\xi : \Omega \to \mathbb{R}^d$ is a random vector and $g : \mathbb{R}^d \to \mathbb{R}$ is a measurable function, then

$$\mathrm{E}(g(\xi)) = \int_\Omega g(\xi(\omega))\,\mathrm{P}(\mathrm{d}\omega) = \int_{\mathbb{R}^d} g(x)\,\mathrm{P}_\xi(\mathrm{d}x) = \int_{\mathbb{R}^d} g(x)\,\mathrm{d}F_\xi(x)$$

in the sense that, the integrals exist at the same time, and if they exist, then they are equal.

Expectation (expected value)

Expectation of a function of a nonnegative random variable

Let ξ be a nonnegative random variable with distribution function F_ξ, and let $g : \mathbb{R}_+ \to \mathbb{R}$ be a monotone and absolute continuous function (i.e., for each $\varepsilon > 0$, there exists $\delta > 0$ such that if $k \in \mathbb{N}$, $0 \leqslant a_1 < b_1 \leqslant a_2 < b_2 \leqslant \ldots \leqslant a_k < b_k$ and $\sum_{j=1}^k (b_j - a_j) < \delta$, then $\sum_{j=1}^k |g(b_j) - g(a_j)| < \varepsilon$). Then

$$\mathrm{E}(g(\xi)) = g(0) + \int_0^\infty g'(x)(1 - F_\xi(x))\,\mathrm{d}x,$$

which is understood in the sense that if one of the two sides is finite, then the other side is finite as well, and the two sides coincide.

Expectation (expected value)

Moments and moment generating function of a nonnegative random variable

Let ξ be a nonnegative random variable with distribution function F_ξ.

(i) For each $\alpha > 0$, we have

$$\mathrm{E}(\xi^\alpha) = \alpha \int_0^\infty x^{\alpha-1}(1 - F_\xi(x))\,\mathrm{d}x.$$

Further, if $\mathrm{E}(\xi^\alpha) < \infty$ with some $\alpha > 0$ (i.e., if ξ has a finite moment of order $\alpha > 0$), then

$$\lim_{x\to\infty} x^{\alpha-1}\mathrm{P}(\xi \geqslant x) = \lim_{x\to\infty} x^{\alpha-1}(1 - F_\xi(x)) = 0.$$

In particular, if $n \in \mathbb{N}$, then a necessary condition for the finiteness of the n^{th}-moment of ξ is that the tail probabilities $\mathrm{P}(\xi \geqslant x)$, $x \geqslant 0$, tend to zero at least of order x^{n-1} (polynomially) at infinity.

Expectation (expected value)s

(ii) For each $r \in \mathbb{R}$, we have

$$\mathrm{E}(\mathrm{e}^{r\xi}) = 1 + r \int_0^\infty \mathrm{e}^{rx}(1 - F_\xi(x))\,\mathrm{d}x.$$

Further, if $\mathrm{E}(\mathrm{e}^{r\xi}) < \infty$ with some $r \in \mathbb{R}$ (i.e., if the moment generating function of ξ exists at some point $r \in \mathbb{R}$), then

$$\lim_{x\to\infty} \mathrm{e}^{rx}\mathrm{P}(\xi \geqslant x) = \lim_{x\to\infty} \mathrm{e}^{rx}(1 - F_\xi(x)) = 0.$$

In particular, if $r > 0$, then a necessary condition for the finiteness of the moment generating function of ξ at the point r is that the tail probabilities of ξ tend to zero at least of order e^{rx} (exponentially) at infinity.

Expectation (expected value)

Absolute continuity

Let $(X, \mathcal{X})$ be a measurable space. We say that a mapping $\mu : \mathcal{X} \to [-\infty, \infty]$ is **absolutely continuous** with respect to the mapping $\nu : \mathcal{X} \to [-\infty, \infty]$, if for each $B \in \mathcal{X}$, $\nu(B) = 0$, we have $\mu(B) = 0$. In notation: $\mu \ll \nu$.

Density theorem

Let $(X, \mathcal{X})$ be a measurable space, $\nu : \mathcal{X} \to [0, \infty]$ be a measure, $g : X \to \mathbb{R}_+$ be a nonnegative measurable function. Then the mapping $\mu : \mathcal{X} \to [0, \infty]$,

$$\mu(B) := \int_B g(x)\, \nu(\mathrm{d}x), \qquad B \in \mathcal{X},$$

is a measure, which is finite if and only if g is integrable. Further, $\mu \ll \nu$, and for each measurable function $h : X \to \mathbb{R}$, we have

$$\int_X h(x) \mu(\mathrm{d}x) = \int_X h(x) g(x)\, \nu(\mathrm{d}x)$$

in the sense that the integrals exist at the same time, and if they exist, then they are equal.

Expectation (expected value)

Radon–Nikodym theorem

Let $(X, \mathcal{X})$ be a measurable space and $\nu : \mathcal{X} \to [0, \infty]$ be a σ–finite measure. A signed measure $\mu : \mathcal{X} \to [-\infty, \infty]$ is absolutely continuous with respect to the measure ν if and only if there exists a measurable function $g : X \to [-\infty, \infty]$ such that for each $B \in \mathcal{X}$, we have

$$\mu(B) = \int_B g(x)\, \nu(dx).$$

The function g is ν–a.s. uniquely determined, i.e., if $h : X \to [-\infty, \infty]$ is a measurable function such that

$$\mu(B) = \int_B h(x)\, \nu(dx)$$

for each $B \in \mathcal{X}$, then $\nu(\{x \in X : g(x) \neq h(x)\}) = 0$.

The (ν-a.s. uniquely determined) function g in the Radon–Nikodym theorem is called the Radon–Nikodym derivative of the mesure μ with respect to the measure ν. In notation: $\frac{\mathrm{d}\mu}{\mathrm{d}\nu}$.

Expectation (expected value)

Absolutely continuous random variable

Let $(X, \mathcal{X})$ be a measurable space, and $\nu : \mathcal{X} \to [0, \infty]$ be a σ–finite measure. We say that a random variable $\xi : \Omega \to X$ is **absolutely continuous with respect to the measure** ν, if $\mathrm{P}_\xi \ll \nu$. We say that a random vector $\xi : \Omega \to \mathbb{R}^d$ is **absolutely continuous**, if it is absolutely continuous with respect to the *d*-dimensional Lebesgue measure λ_d (more precisely, with respect to the restriction of λ_d to $\mathcal{B}(\mathbb{R}^d)$), and then its Radon–Nikodym derivative $f_\xi := \frac{\mathrm{d}\mathrm{P}_\xi}{\mathrm{d}\lambda_d}$ is called the **density function** of ξ.

Absolutely continuous random variable

A random variable $\xi : \Omega \to \mathbb{R}$ is absolutely continuous if and only if its distribution function F_ξ is absolutely continuous, i.e., $\forall\, \varepsilon > 0$ there exists $\delta > 0$ such that if $k \in \mathbb{N}$, $a_1 < b_1 \leqslant a_2 < b_2 \leqslant \ldots \leqslant a_k < b_k$ and $\sum_{j=1}^{k}(b_j - a_j) < \delta$, then $\sum_{j=1}^{k}(F_\xi(b_j) - F_\xi(a_j)) < \varepsilon$.

Expectation (expected value)

Characterization of density function

A function $f : \mathbb{R}^d \to \mathbb{R}$ is a density function of some d-dimensional random variable if and only if it is (Borel) measurable, nonnegative Lebesgue almost everywhere and $\int_{\mathbb{R}^d} f(x)\,\mathrm{d}x = 1$.

Connection between density function and distribution function

If a random vector $X : \Omega \to \mathbb{R}^d$ is absolutely continuous, then $f_X(x) = \partial_1 \dots \partial_d F_X(x)$ λ_d-a.e. $x \in \mathbb{R}^d$.

Expectation of a function of an absolutely continuous random vector

If $\xi : \Omega \to \mathbb{R}^d$ is an absolutely continuous random vector and $g : \mathbb{R}^d \to \mathbb{R}$ is a measurable function, then

$$\mathrm{E}(g(\xi)) = \int_{\mathbb{R}^d} g(x) f_\xi(x)\,\mathrm{d}x$$

in the sense that the integrals exist at the same time, and if they exist, then they are equal. (It is a consequence of Transformation and Density theorems.)

Expectation (expected value)

Injective function of an absolutely continuous random variable

Let $\xi : \Omega \to \mathbb{R}$ be an absolutely continuous random variable with density function f_ξ. Let $D \subset \mathbb{R}$ be an open set such that $\mathrm{P}(\xi \in D) = 1$. Let $g : D \to \mathbb{R}$ be a continuously differentiable function, which is injective on D, and its derivative is not zero at any point. (It is known that in this case $g(D) \subset \mathbb{R}$ is open, and the inverse function $h : g(D) \to D$ is continuously differentiable with nonzero derivative.) Then the random variable $g(\xi)$ is absolutely continuous as well, and its density function

$$f_{g(\xi)}(y) = \begin{cases} f_\xi(h(y))|h'(y)| = \frac{f_\xi(g^{-1}(y))}{|g'(g^{-1}(y))|}, & \text{if } y \in g(D), \\ 0, & \text{otherwise.} \end{cases}$$

Expectation (expected value)

Sum, product and ratio of independent absolutely continuous random variables

Let ξ and η be independent, absolutely continuous random variables with density functions f_ξ and f_η, respectively. Then

(i) the random variable $\xi+\eta$ is absolutely continuous, and

$$f_{\xi+\eta}(z)=\int_{-\infty}^{\infty} f_\xi(x) f_\eta(z-x)\,\mathrm{d}x=\int_{-\infty}^{\infty} f_\xi(z-y) f_\eta(y)\,\mathrm{d}y, \quad \lambda_1\text{-a.e. } z\in\mathbb{R}.$$

This formula is called a **convolution formula** as well.

(ii) the random variable $\xi\eta$ is absolutely continuous, and

$$f_{\xi\eta}(z)=\int_{-\infty}^{\infty} f_\xi(x) f_\eta\left(\frac{z}{x}\right)\frac{\mathrm{d}x}{|x|}=\int_{-\infty}^{\infty} f_\xi\left(\frac{z}{y}\right) f_\eta(y)\,\frac{\mathrm{d}y}{|y|}, \quad \lambda_1\text{-a.e. } z\in\mathbb{R},$$

(iii) the random variable $\frac{\xi}{\eta}$ is absolutely continuous, and

$$f_{\frac{\xi}{\eta}}(z)=\frac{1}{z^2}\int_{-\infty}^{\infty} f_\xi(x) f_\eta\left(\frac{x}{z}\right)|x|\,\mathrm{d}x=\int_{-\infty}^{\infty} f_\xi(zy) f_\eta(y)|y|\,\mathrm{d}y, \quad \lambda_1\text{-a.e. } z\in\mathbb{R}.$$

Expectation (expected value)

Sum, product and ratio of jointly absolutely continuous random variables

If ξ and η are jointly absolutely continuous random variables with density function $f_{\xi,\eta}$, then

(i) the random variable $\xi+\eta$ is absolutely continuous, and

$$f_{\xi+\eta}(z) = \int_{-\infty}^{\infty} f_{\xi,\eta}(x, z-x)\,\mathrm{d}x = \int_{-\infty}^{\infty} f_{\xi,\eta}(z-y, y)\,\mathrm{d}y, \quad \lambda_1\text{-a.e. } z \in \mathbb{R}.$$

(ii) the random variable $\xi\eta$ is absolutely continuous, and

$$f_{\xi\eta}(z) = \int_{-\infty}^{\infty} f_{\xi,\eta}\left(x, \frac{z}{x}\right) \frac{\mathrm{d}x}{|x|} = \int_{-\infty}^{\infty} f_{\xi,\eta}\left(\frac{z}{y}, y\right) \frac{\mathrm{d}y}{|y|}, \quad \lambda_1\text{-a.e. } z \in \mathbb{R},$$

(iii) the random variable $\frac{\xi}{\eta}$ is absolutely continuous, and

$$f_{\frac{\xi}{\eta}}(z) = \frac{1}{z^2} \int_{-\infty}^{\infty} f_{\xi,\eta}\left(x, \frac{x}{z}\right) |x|\,\mathrm{d}x = \int_{-\infty}^{\infty} f_{\xi,\eta}(zy, y)|y|\,\mathrm{d}y, \quad \lambda_1\text{-a.e. } z \in \mathbb{R}.$$

Expectation (expected value)

Concentration of a mesaure into a subset

Let $(X, \mathcal{X})$ be a measurable space, $\mu : \mathcal{X} \to [-\infty, \infty]$ be a signed measure. We say that the signed measure μ is **concentrated into a set** $\boldsymbol{B} \in \boldsymbol{\mathcal{X}}$**,** if $\mu(X \setminus B) = 0$.

Support of discrete distribution

Let $\xi : \Omega \to \mathbb{R}^d$ be a **discrete** random vector (i.e., $\xi(\Omega)$ is a countable set). We say that ξ is concentrated into a set $B \in \mathcal{B}(\mathbb{R}^d)$, if P_ξ is concentrated into B, equivalently $\mathsf{P}_\xi(B) = \mathsf{P}(\xi \in B) = 1$.
The intersection of all the sets with this property (i.e., the smallest set with this property) is called the **support** of the measure P_ξ.
In notation: $\operatorname{supp}(\xi)$.

Then

$$\operatorname{supp}(\xi) = \left\{ x \in \mathbb{R}^d : \mathsf{P}_\xi(\{x\}) > 0 \right\} = \left\{ x \in \mathbb{R}^d : \mathsf{P}(\xi = x) > 0 \right\},$$

of which the elements are called the **atoms** of the measure P_ξ.

Distribution and distribution function of a discrete random vector

The distribution of a discrete random variable $\xi : \Omega \to \mathbb{R}^d$ is given by

$$\mathrm{P}_\xi = \sum_{x \in \mathrm{supp}(\xi)} \mathrm{P}(\xi = x)\delta_x,$$

where for each $x \in \mathbb{R}^d$, δ_x denotes the Dirac mesaure concentrated on the point x, i.e., $\delta_x(A) = 1$, if $x \in A$, and $\delta_x(A) = 0$, if $x \notin A$. The distribution function of ξ is given by

$$F_\xi(x) = \sum_{\{y \in \mathrm{supp}(\xi)\,:\, y < x\}} \mathrm{P}(\xi = y), \qquad x \in \mathbb{R}^d.$$

Expectation (expected value)

Expectation of a function of a discrete random vector

Let $\xi : \Omega \to \mathbb{R}^d$ be a discrete random vector and $g : \mathbb{R}^d \to \mathbb{R}$ be a measurable function. The random variable $g(\xi)$ is integrable if and only if

$$\mathrm{E}(|g(\xi)|) = \sum_{x \in \mathrm{supp}(\xi)} |g(x)| \mathrm{P}(\xi = x) < \infty,$$

and then

$$\mathrm{E}(g(\xi)) = \sum_{x \in \mathrm{supp}(\xi)} g(x) \mathrm{P}(\xi = x).$$

(This statement is a special case of the Transformation theorem.)

Expectation (expected value)

Singularity

Let $(X, \mathcal{X})$ be a measurable space. The measures $\mu : \mathcal{X} \to [0, \infty]$ and $\nu : \mathcal{X} \to [0, \infty]$ are called **singular with respect to each other,** if there exist disjoint sets $A, B \in \mathcal{X}$ such that μ and ν are concentrated in the set A and in the set B, respectively.
In notation: $\mu \perp \nu$.

Singular random vectors

A random vector $\xi : \Omega \to \mathbb{R}^d$ is called **singular,** if $\mathrm{P}_\xi \perp \lambda_d$, where λ_d denotes the d-dimensional Lebesgue measure, equivalently, $\exists\, B \in \mathcal{B}(\mathbb{R}^d)$ such that $\lambda_d(B) = 0$ and $\mathrm{P}(\xi \in B) = 1$.

A discrete random vector is singular (can be checked easily).

Singular random variable

A random variable $\xi : \Omega \to \mathbb{R}$ is singular if and only if $F'_\xi(x) = 0$ λ_1-a.e. $x \in \mathbb{R}$.

Expectation (expected value)

Lebesgue decomposition theorem

Let $(X, \mathcal{X})$ be a measurable space, μ and ν be σ-finite measures on $\mathcal{X}$. Then there exist a measurable function $f : X \to [0, \infty]$ and a measure ν_s on $\mathcal{X}$ such that $\mu \perp \nu_s$ and

$$\nu(A) = \int_A f \, \mathrm{d}\mu + \nu_s(A), \qquad A \in \mathcal{X}.$$

Such a function f is uniquely determined μ-almost everywhere, i.e., if $g : X \to [0, \infty]$ is a measurable function such that $\nu(A) = \int_A g \, \mathrm{d}\mu + \nu_s(A)$, $A \in \mathcal{X}$, then

$$\mu(\{x \in X : f(x) \neq g(x)\}) = 0.$$

The above decomposition is called the **Lebesgue decomposition** of ν with respect to μ.

Expectation (expected value)

Decomposition theorem of distribution functions

Any distribution function $F : \mathbb{R} \to [0,1]$ can be uniquely decomposed in the form

$$F = p_1 F_{\mathrm{d}} + p_2 F_{\mathrm{af}} + p_3 F_{\mathrm{fs}},$$

where $p_1, p_2, p_3 \geqslant 0$, $p_1 + p_2 + p_3 = 1$, F_{d} is a discrete, F_{af} is an absolutely continuous and F_{fs} is a continuous singular distribution function.

Moments

Let $\xi : \Omega \to \mathbb{R}$ be a random variable.

- Let $\alpha \in \mathbb{R}_+$. The α^{th} **absolute moment** of ξ: $\mathrm{E}(|\xi|^\alpha)$.
- If $k \in \mathbb{N}$ and the k^{th} absolute moment of ξ is finite, then the $\boldsymbol{k^{th}}$ **moment** of ξ: $\mathrm{E}(\xi^k) \in \mathbb{R}$,
  the $\boldsymbol{k^{th}}$ **central moment** of ξ: $\mathrm{E}\left((\xi - \mathrm{E}(\xi))^k\right) \in \mathbb{R}$.
- If ξ has a finite second (absolute) moment, then the second central moment of ξ is called the **variance (squared deviation)** of ξ. In notation: $\mathrm{Var}(\xi) := \mathrm{D}^2(\xi) := \mathrm{E}\left[(\xi - \mathrm{E}(\xi))^2\right]$.

Expectation (expected value)

Expectation vector of a random vector

Let $\xi = (\xi_1, \dots, \xi_d) : \Omega \to \mathbb{R}^d$ be a random vector.
If $\mathsf{E}(|\xi_1|) < \infty, \dots, \mathsf{E}(|\xi_d|) < \infty$, then the **expectation vector** of ξ is

$$\mathsf{E}(\xi) := (\mathsf{E}(\xi_1), \dots, \mathsf{E}(\xi_d))^\top \in \mathbb{R}^d.$$

Multidimensional Jensen inequality

Let $\xi : \Omega \to \mathbb{R}^d$ be a random vector such that $\mathsf{E}(\|\xi\|) < \infty$.

1. If $K \subset \mathbb{R}^d$ is nonempty, convex, closed and $\mathsf{P}(\xi \in K) = 1$, then $\mathsf{E}(\xi) \in K$.
2. If $g : \mathbb{R}^d \to \mathbb{R}$ is convex and $\mathsf{E}(|g(\xi)|) < \infty$, then $g(\mathsf{E}(\xi)) \leqslant \mathsf{E}(g(\xi))$.

Expectation (expected value)

Covariance matrix (variance matrix) of a random vector

Let $\xi = (\xi_1, \dots, \xi_d) : \Omega \to \mathbb{R}^d$ be a random vector. If $\mathrm{E}(\|\xi\|^2) < \infty$, i.e., $\mathrm{E}(\xi_1^2) < \infty, \dots, \mathrm{E}(\xi_d^2) < \infty$, then the **covariance matrix** of ξ is

$$\mathrm{Cov}(\xi) := \mathrm{E}\left[(\xi - \mathrm{E}(\xi))(\xi - \mathrm{E}(\xi))^\top\right] \in \mathbb{R}^{d \times d},$$

of which the entries are $\mathrm{E}\left[(\xi_i - \mathrm{E}(\xi_i))(\xi_j - \mathrm{E}(\xi_j))\right] =: \mathrm{Cov}(\xi_i, \xi_j)$.

Properties of covariance matrices

Let $\xi = (\xi_1, \dots, \xi_d) : \Omega \to \mathbb{R}^d$ be a random vector with $\mathrm{E}(\|\xi\|^2) < \infty$.

- $\mathrm{Cov}(\xi)$ is symmetric: $\mathrm{Cov}(\xi)^\top = \mathrm{Cov}(\xi)$.
- $\mathrm{Cov}(\xi)$ is positive semidefinite, i.e., for all $x \in \mathbb{R}^d$ we have
$$x^\top \mathrm{Cov}(\xi) x = \langle \mathrm{Cov}(\xi) x, x \rangle = \sum_{i=1}^d \sum_{j=1}^d \mathrm{Cov}(\xi_i, \xi_j) x_i x_j \geqslant 0.$$
- If $A \in \mathbb{R}^{r \times d}$ and $b \in \mathbb{R}^r$, then $\mathrm{E}(A\xi + b) = A\,\mathrm{E}(\xi) + b$ and $\mathrm{Cov}(A\xi + b) = A\,\mathrm{Cov}(\xi) A^\top$.

Expectation (expected value)

Expectation of a complex valued random variable

We say that a complex valued random variable $\xi = \operatorname{Re}\xi + \mathrm{i}\operatorname{Im}\xi : \Omega \to \mathbb{C}$ has a **finite expectation (it is integrable)**, if the expectations $\mathsf{E}(\operatorname{Re}\xi)$ and $\mathsf{E}(\operatorname{Im}\xi)$ are finite, and then $\mathsf{E}(\xi) := \mathsf{E}(\operatorname{Re}\xi) + \mathrm{i}\,\mathsf{E}(\operatorname{Im}\xi)$.

Expectation of a complex valued random variable

Let $\xi : \Omega \to \mathbb{C}$ be a complex valued random variable.

- ξ has a finite expectation if and only if $\mathsf{E}(|\xi|) < \infty$.
- If $\mathsf{E}(|\xi|) < \infty$, then $|\,\mathsf{E}(\xi)| \leqslant \mathsf{E}(|\xi|)$.

Expectation (expected value)

Independence of complex valued random variables

Let $\Gamma \neq \emptyset$ be an (index) set, and for each $\gamma \in \Gamma$ let $\xi_\gamma : \Omega \to \mathbb{C}$ be a random variable. We say that the random variables $\{\xi_\gamma : \gamma \in \Gamma\}$ are **independent**, if the random variables $\{(\operatorname{Re} \xi_\gamma, \operatorname{Im} \xi_\gamma) : \gamma \in \Gamma\}$ are independent.

Independence of complex valued random variables

If $\xi_1, \ldots, \xi_n : \Omega \to \mathbb{C}$ are independent random variables such that $\mathsf{E}(|\xi_i|) < \infty$, $i = 1 \ldots, n$, then $\mathsf{E}(|\xi_1 \cdots \xi_n|) < \infty$ and

$$\mathsf{E}(\xi_1 \cdots \xi_n) = \mathsf{E}(\xi_1) \cdots \mathsf{E}(\xi_n).$$

Characteristic function

Characteristic function

The **characteristic function** $\varphi_X : \mathbb{R}^d \to \mathbb{C}$ of a random vector $X : \Omega \to \mathbb{R}^d$ is defined by

$$\varphi_X(t) := \mathsf{E}(\mathrm{e}^{\mathrm{i}\langle t,X\rangle}) = \int_{\mathbb{R}^d} \mathrm{e}^{\mathrm{i}\langle t,x\rangle} \, F_X(\mathrm{d}x) = \mathsf{E}(\cos(\langle t,X\rangle)) + \mathrm{i}\,\mathsf{E}(\sin(\langle t,X\rangle)),$$

where $t \in \mathbb{R}^d$.

If X is a discrete random vector with values $\{x_k, k \in \mathbb{N}\}$ and with distribution $\{p_k, k \in \mathbb{N}\}$, then

$$\varphi_X(t) = \sum_{k=1}^{\infty} \mathrm{e}^{\mathrm{i}\langle t,x_k\rangle} p_k, \qquad t \in \mathbb{R}^d,$$

and if X is absolutely continuous with density function f_X, then

$$\varphi_X(t) = \int_{\mathbb{R}^d} \mathrm{e}^{\mathrm{i}\langle t,x\rangle} f_X(x) \, \mathrm{d}x, \qquad t \in \mathbb{R}^d.$$

Characteristic function

Properties of a characteristic function

1. $|\varphi_X| \leqslant 1$, and $\varphi_X(0) = 1$.
2. φ_X is uniformly continuous.
3. For each $t \in \mathbb{R}^d$, we have $\varphi_X(-t) = \overline{\varphi_X(t)}$, i.e., φ_X is Hermite symmetric.
4. **Bochner theorem:** A function $\varphi : \mathbb{R}^d \to \mathbb{C}$ is the characteristic function of some random vector if and only if $\varphi(0) = 1$, it is continuous and **positive semidefinite**, i.e., for each $n \in \mathbb{N}$ and $t_1, \ldots, t_n \in \mathbb{R}^d$, we have that the matrix $\big(\varphi(t_j - t_\ell)\big)_{j,\ell=1,\ldots,n} \in \mathbb{C}^{n\times n}$ is positive semidefinite, i.e., for each $z_1, \ldots, z_n \in \mathbb{C}$, we have

$$\sum_{j=1}^{n} \sum_{\ell=1}^{n} \varphi(t_j - t_\ell) z_j \overline{z_\ell} \geqslant 0.$$

Characteristic function

Properties of a characteristic function

5. For each $A \in \mathbb{R}^{r \times d}$, $b \in \mathbb{R}^r$ and $t \in \mathbb{R}^r$, we have
$$\varphi_{AX+b}(t) = \mathrm{e}^{\mathrm{i}\langle t, b\rangle} \varphi_X(A^\top t).$$
6. **Uniqueness theorem:** $\mathrm{P}_X = \mathrm{P}_Y$ if and only if $\varphi_X = \varphi_Y$.
7. $X_1 : \Omega \to \mathbb{R}^{d_1}, \ldots, X_\ell : \Omega \to \mathbb{R}^{d_\ell}$ are independent if and only if for each $t_1 \in \mathbb{R}^{d_1}, \ldots, t_\ell \in \mathbb{R}^{d_\ell}$, we have
$$\varphi_{X_1,\ldots,X_\ell}(t_1, \ldots, t_\ell) = \prod_{j=1}^{\ell} \varphi_{X_j}(t_j).$$
8. If $X_1, \ldots, X_\ell : \Omega \to \mathbb{R}^d$ are independent, then for each $t \in \mathbb{R}^d$, we have
$$\varphi_{X_1+\cdots+X_\ell}(t) = \prod_{j=1}^{\ell} \varphi_{X_j}(t).$$

Characteristic function

Properties of a characteristic function

9. If $X = (X_1, \dots, X_d) : \Omega \to \mathbb{R}^d$ is a random vector and $\mathsf{E}(\|X\|^n) < \infty$ for some $n \in \mathbb{N}$, then φ_X is n times continuously differentiable, and for any nonnegative integers $r_1, \dots, r_d$ with $r_1 + \dots + r_d \leqslant n$, we have

$$\partial_1^{r_1} \dots \partial_d^{r_d} \varphi_X(t) = \mathrm{i}^{r_1+\dots+r_d} \, \mathsf{E}(X_1^{r_1} \cdots X_d^{r_d} \, \mathrm{e}^{\mathrm{i}\langle t, X\rangle}), \qquad t \in \mathbb{R}^d,$$

$$\mathsf{E}(X_1^{r_1} \cdots X_d^{r_d}) = \frac{\partial_1^{r_1} \dots \partial_d^{r_d} \, \varphi_X(0)}{\mathrm{i}^{r_1+\dots+r_d}},$$

moreover,

$$\varphi_X(t) = \sum_{\substack{r_1+\dots+r_d \leqslant n, \\ r_1,\dots,r_d \in \mathbb{Z}_+}} \frac{\mathrm{i}^{r_1+\dots+r_d} \, t_1^{r_1} \cdots t_d^{r_d}}{r_1! \cdots r_d!} \, \mathsf{E}(X_1^{r_1} \cdots X_d^{r_d}) + R_n(t), \qquad t \in \mathbb{R}^d,$$

where $R_n(t) = \mathrm{O}(\|t\|^n)$, $t \in \mathbb{R}^d$, and $R_n(t) = \mathrm{o}(\|t\|^n)$ as $t \to 0$, in a way that $|R_n(t)| \leqslant 3 \frac{\|t\|^n}{n!} \mathsf{E}(\|X\|^n)$, and $\lim_{t \to 0} \frac{R_n(t)}{\|t\|^n} = 0$.

Properties of a characteristic function

10. If $X : \Omega \to \mathbb{R}$ is a random variable and $\varphi_X^{(2n)}(0)$ exists and finite for some $n \in \mathbb{N}$, i.e., $\varphi_X^{(2n)}(0) \in \mathbb{R}$, then $\mathsf{E}(X^{2n}) < \infty$.
11. If for each $n \in \mathbb{N}$, we have $\mathsf{E}(\|X\|^n) < \infty$, and

$$R := \frac{1}{\limsup\limits_{n\to\infty} \sqrt[n]{\mathsf{E}(\|X\|^n)/n!}} \in (0, \infty],$$

then for each $t \in \mathbb{R}^d$, $\|t\| < R$, we have

$$\varphi_X(t) = \sum_{r_1=0}^{\infty} \cdots \sum_{r_d=0}^{\infty} \frac{\mathrm{i}^{r_1+\cdots+r_d}\,\mathsf{E}(X_1^{r_1}\cdots X_d^{r_d})}{r_1!\cdots r_d!} t_1^{r_1}\cdots t_d^{r_d}.$$

12. **Inversion formula:** If $\varphi_X \in L^1(\mathbb{R}^d)$, i.e., $\int_{\mathbb{R}^d} |\varphi_X(t)|\,\mathrm{d}t < \infty$, then X is absolutely continuous, and its density function

$$f_X(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \mathrm{e}^{-\mathrm{i}\langle t,x\rangle} \varphi_X(t)\,\mathrm{d}t, \qquad x \in \mathbb{R}^d.$$

Then f_X is bounded and continuous.

Characteristic function

Properties of a characteristic function

13. Let $d = 1$. Then $\varphi_X(t) \in \mathbb{R}$, $t \in \mathbb{R}$, if and only if X is symmetric, i.e., $X \stackrel{\mathcal{D}}{=} -X$.

Pólya theorem

If $\varphi : \mathbb{R} \to [0, \infty)$ is a function such that it is continuous, even, $\varphi(0) = 1$, $\lim_{t\to\infty} \varphi(t) = 0$, and $\varphi|_{[0,\infty)}$ is convex, then φ is the characteristic function of some random variable $X : \Omega \to \mathbb{R}$.

Using Pólya theorem one can easily give examples for characteristic functions which coincide on a finite interval, but the distribution functions corresponding uniquely to them do not coincide.

Characteristic function of a standard normally distributed random variable

If $X \sim \mathcal{N}(0, 1)$, then $\varphi_X(t) = \mathrm{e}^{-\frac{t^2}{2}}$, $t \in \mathbb{R}$.

Characteristic function

Convergence in distribution of random vectors

Let $X_n : \Omega \to \mathbb{R}^d$, $n \in \mathbb{N}$, and $X : \Omega \to \mathbb{R}^d$ be random vectors.
We say that the **sequence** $\boldsymbol{(X_n)_{n\geqslant 1}}$ **converges in distribution to** $\boldsymbol{X}$, if $F_{X_n}(x) \to F_X(x)$ at every continuity point x of F_X.
In notation: $X_n \xrightarrow{\mathcal{D}} X$.

Continuity theorem (Paul Lévy)

Let $X_n : \Omega \to \mathbb{R}^d$, $n \in \mathbb{N}$, be random vectors.

1. If there exists a random vector $X : \Omega \to \mathbb{R}^d$ such that $X_n \xrightarrow{\mathcal{D}} X$ as $n \to \infty$, then $\varphi_{X_n} \to \varphi_X$ as $n \to \infty$, uniformly on each bounded interval.
2. If for each $t \in \mathbb{R}^d$, there exists $\lim_{n\to\infty} \varphi_{X_n}(t) =: \varphi(t)$, and φ is continuous at the point $0 \in \mathbb{R}^d$, then there exists a random vector $X : \Omega \to \mathbb{R}^d$ such that $\varphi_X = \varphi$, and $X_n \xrightarrow{\mathcal{D}} X$ as $n \to \infty$.

Characteristic function

Generating function

If the coordinates of the random vector $X : \Omega \to \mathbb{R}^d$ are nonnegative integers, i.e., X is concentrated in the set $\mathbb{Z}_+^d$, i.e., $\mathsf{P}(X \in \mathbb{Z}_+^d) = 1$, then **the generating function of $\boldsymbol{X = (X_1, \ldots, X_d)}$** is the d-variable complex power series (where it exists):

$$G_X(z) := G_{X_1,\ldots,X_d}(z_1, \ldots, z_d) := \mathsf{E}(z^X) := \mathsf{E}(z_1^{X_1} \cdots z_d^{X_d})$$
$$= \sum_{k_1=0}^{\infty} \cdots \sum_{k_d=0}^{\infty} \mathsf{P}(X_1 = k_1, \ldots, X_d = k_d)\, z_1^{k_1} \cdots z_d^{k_d}.$$

This power series is absolutely convergent on the set

$$\{(z_1, \ldots, z_d) \in \mathbb{C}^d : |z_1| \leqslant 1, \ldots, |z_d| \leqslant 1\},$$

and the characteristic function of X is the periodic function

$$\varphi_X(t) = \varphi_X(t_1, \ldots, t_d) = G_X(\mathrm{e}^{\mathrm{i}t_1}, \ldots, \mathrm{e}^{\mathrm{i}t_d}), \qquad t = (t_1, \ldots, t_d) \in \mathbb{R}^d.$$

Properties of a generating function

1. $G_X(1, \ldots, 1) = 1$.
2. G_X is analytical on the set $\{(z_1, \ldots, z_d) \in \mathbb{C}^d : |z_1| < 1, \ldots, |z_d| < 1\}$.
3. For each $k_1, \ldots, k_d \in \mathbb{Z}_+$, we have

$$\mathrm{P}(X_1 = k_1, \ldots, X_d = k_d) = \frac{\partial_1^{k_1} \ldots \partial_d^{k_d} G_X(0, \ldots, 0)}{k_1! \cdots k_d!}.$$

4. Uniqueness theorem for generating functions: $\mathrm{P}_X = \mathrm{P}_Y \iff \forall\, x \in [-1, 1]^d$ for all $G_X(x) = G_Y(x)$.
5. If X and Y are independent, then $G_{X+Y}(z) = G_X(z)G_Y(z)$ on the set $\{(z_1, \ldots, z_d) \in \mathbb{C}^d : |z_1| \leqslant 1, \ldots, |z_d| \leqslant 1\}$.

Characteristic function

Properties of a generating function

6. For each $r_1, \ldots, r_d \in \mathbb{Z}_+$, we have

$$\mathrm{E}(X_1^{r_1} \cdots X_d^{r_d}) < \infty \quad \Longleftrightarrow \quad \partial_1^{r_1} \ldots \partial_d^{r_d} G_X(1-, \ldots, 1-) < \infty,$$

and

$$\partial_1^{r_1} \ldots \partial_d^{r_d} G_X(1-, \ldots, 1-)$$
$$= \mathrm{E}(X_1(X_1 - 1) \cdots (X_1 - r_1 + 1) \cdots X_d(X_d - 1) \cdots (X_d - r_d + 1)).$$

Continuity theorem for generating functions

Let $X : \Omega \to \mathbb{R}^d$ and $X_n : \Omega \to \mathbb{R}^d$, $n \in \mathbb{N}$, be random vectors such that $\mathrm{P}(X \in \mathbb{Z}_+^d) = 1$ and $\mathrm{P}(X_n \in \mathbb{Z}_+^d) = 1$, $n \in \mathbb{N}$.
Then the following assertions are equivalent:

- $X_n \xrightarrow{\mathcal{D}} X$ as $n \to \infty$.
- $\mathrm{P}(X_n = k) \to \mathrm{P}(X = k)$ as $n \to \infty$ for all $k \in \mathbb{Z}_+^d$.
- $G_{X_n}(x) \to G_X(x)$ as $n \to \infty$ for each $x \in [-1, 1]^d$.

Characteristic function

Laplace transform

If the coordinates of the random vector $X = (X_1, \dots, X_d) : \Omega \to \mathbb{R}^d$ are nonnegative, i.e., X is concentrated in the set $\mathbb{R}_+^d$, i.e., $\mathsf{P}(X \in \mathbb{R}_+^d) = 1$, then the **Laplace transform** $\psi_X : \mathbb{R}_+^d \to \mathbb{R}$ of X is defined by

$$\begin{aligned}\psi_X(s) &:= \psi_{X_1,\dots,X_d}(s_1, \dots, s_d) := \mathsf{E}(\mathrm{e}^{-\langle s, X\rangle}) \\ &= \int_0^\infty \cdots \int_0^\infty \mathrm{e}^{-s_1x_1-\cdots-s_dx_d} \,\mathrm{d}F_{X_1,\dots,X_d}(x_1, \dots, x_d),\end{aligned}$$

where $s \in \mathbb{R}_+^d$.

If $\mathsf{P}(X \in \mathbb{Z}_+^d) = 1$, then

$$\psi_X(s_1, \dots, s_d) = G_X(\mathrm{e}^{-s_1}, \dots, \mathrm{e}^{-s_d}), \qquad (s_1, \dots, s_d) \in \mathbb{R}_+^d,$$

$$G_X(x_1, \dots, x_d) = \psi_X(-\log x_1, \dots, -\log x_d), \qquad (x_1, \dots, x_d) \in (0, 1)^d.$$

Characteristic function

Properties of Laplace transform

1. $0 \leqslant \psi_X \leqslant 1$, and $\psi_X(0) = 1$.
2. ψ_X is analitic on the set $(0, \infty)^d$.
3. Uniqueness theorem for Laplace transforms:
   $\mathrm{P}_X = \mathrm{P}_Y$ if and only if $\psi_X = \psi_Y$.
4. If X and Y are independent, then $\psi_{X+Y} = \psi_X \psi_Y$.
5. For each $r_1, \ldots, r_d \in \mathbb{Z}_+$, we have
   $$\mathrm{E}(X_1^{r_1} \cdots X_d^{r_d}) < \infty \qquad \Longleftrightarrow \qquad \partial_1^{r_1} \ldots \partial_d^{r_d} \psi_X(0+, \ldots, 0+) < \infty,$$
   and $\partial_1^{r_1} \ldots \partial_d^{r_d} \psi_X(0+, \ldots, 0+) = (-1)^{r_1 + \cdots + r_d} \mathrm{E}(X_1^{r_1} \cdots X_d^{r_d})$.

Continuity theorem for Laplace transforms

Let $X : \Omega \to \mathbb{R}^d$ and $X_n : \Omega \to \mathbb{R}^d$, $n \in \mathbb{N}$, be random vectors such that $\mathrm{P}(X \in \mathbb{R}_+^d) = 1$ and $\mathrm{P}(X_n \in \mathbb{R}_+^d) = 1$, $n \in \mathbb{N}$.
Then the following statements are equivalent:

- $X_n \xrightarrow{\mathcal{D}} X$ as $n \to \infty$,
- $\psi_{X_n}(s) \to \psi_X(s)$ as $n \to \infty$ for all $s \in \mathbb{R}_+^d$.

Notable distributions

Bernoulli distribution with parameter p

Let $p \in [0,1]$. A discrete random variable X is called **Bernoulli distributed with parameter** $\boldsymbol{p}$, if it can have values: 0 and 1, and its distribution is

$$\mathrm{P}(X=1)=p, \qquad \mathrm{P}(X=0)=1-p.$$

If $A \in \mathcal{A}$ is an event, then the r. v. $\mathbb{1}_A := \begin{cases} 1 & \text{if } A \text{ occurs,} \\ 0 & \text{if } A \text{ does not occur,} \end{cases}$

is Bernoulli distributed with parameter $\mathrm{P}(A)$.

Generating function

$$G_X(z) = 1-p+pz = 1+p(z-1), \qquad z \in \mathbb{C}.$$

Laplace transform

$$\psi_X(s) = 1-p+p\,\mathrm{e}^{-s} = 1-p(1-\mathrm{e}^{-s}), \qquad s \in \mathbb{R}_+.$$

Characteristic function

$$\varphi_X(t) = 1-p+p\,\mathrm{e}^{\mathrm{i}t} = 1+p(\mathrm{e}^{\mathrm{i}t}-1), \qquad t \in \mathbb{R}.$$

Notable distributions

Binomial distribution with parameter (n, p)

Let $n \in \mathbb{N}$ and $p \in [0, 1]$. A discrete random variable X is called **binomial distributed with parameter (n, p)**, if it can have values: 0, 1, . . . , n, and its distribution is

$$\mathrm{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \qquad k \in \{0, 1, \dots, n\}.$$

If we carry out n independent experiments related to an event $A \in \mathcal{A}$ and

$$X_i := \begin{cases} 1 & \text{if } A \text{ occurs at the } i^{th} \text{ repetition,} \\ 0 & \text{otherwise,} \end{cases} \qquad i = 1, \dots, n,$$

then the random variable $X = X_1 + \cdots + X_n$ is binomial distributed with parameter $(n, \mathrm{P}(A))$, and $X_1, \dots, X_n$ are independent, Bernoulli distributed with parameter $\mathrm{P}(A)$.

Notable distributions

Let X be a binomial distributed random variable with parameter (n, p), where $n \in \mathbb{N}$ and $p \in [0, 1]$.

Generating function

$$G_X(z) = (1 - p + pz)^n = (1 + p(z - 1))^n, \qquad z \in \mathbb{C}.$$

Laplace transform

$$\psi_X(s) = (1 - p + p\,\mathrm{e}^{-s})^n = (1 - p(1 - \mathrm{e}^{-s}))^n, \qquad s \in \mathbb{R}_+.$$

Characteristic function

$$\varphi_X(t) = (1 - p + p\,\mathrm{e}^{\mathrm{i}t})^n = (1 + p(\mathrm{e}^{\mathrm{i}t} - 1))^n, \qquad t \in \mathbb{R}.$$

Hipergeometric distribution with parameter $(n, M, N-M)$

Let $n, N, M \in \mathbb{N}$ be such that $M \leqslant N$. A discrete random variable X is called **hipergeometric distributed with parameter $(n, M, N-M)$**, if it can have values those integers k for which $0 \leqslant k \leqslant n$, $k \leqslant M$ and $n-k \leqslant N-M$, and its distribution is

$$\mathrm{P}(X = k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}.$$

If there are M red and $N-M$ black balls in an urn, and we choose n balls without replacement, and X denotes the number of red balls chosen, then X is a hipergeometric distributed random variable with parameter $(n, M, N-M)$.

Notable distributions

Negative binomial distribution with parameters p and r

Let $r \in \mathbb{N}$ and $p \in (0,1]$. A discrete random variable X is called **negative binomial distributed with parameters p and r**, if it can have values: 0, 1, ..., and its distribution is

$$P(X = k) = \binom{k+r-1}{r-1} p^r (1-p)^k, \qquad k \in \{0, 1, \dots\}.$$

A negative binomial distribution with parameters p and 1, is called **a geometric distribution with parameter p** as well.

If we carry out independent experiments related to an event $A \in \mathcal{A}$ and $r + X$ denotes the number of repetitions needed for the r^{th} occurence of A, then the random variable X is negative binomial distributed with parameters $P(A)$ and r.

Convolution of geometric distributions

If the random variables $X_1, \dots, X_r$ are independent and have geometric distribution with parameter p, then the random variable $X_1 + \dots + X_r$ is negative binomial distributed with parameters p and r.

Notable distributions

Let X be a negative binomial distributed random variable with parameters p and r, where $r \in \mathbb{N}$ and $p \in (0,1]$.

Generating function

$$G_X(z) = \left(\frac{p}{1-(1-p)z} \right)^r, \qquad z \in \mathbb{C}, \quad |z| < \frac{1}{1-p},$$

where in case of $p = 1$, we define $\frac{1}{1-p} := \infty$.

Characteristic function

$$\varphi_X(t) = \left(\frac{p}{1-(1-p)\mathrm{e}^{\mathrm{i}t}} \right)^r, \qquad t \in \mathbb{R}.$$

Memorylessness property of geometric distribition

If X is a random variable having geometric distribution with paramater p, then

$$\mathrm{P}(X \geqslant k+\ell \,|\, X \geqslant k) = \mathrm{P}(X \geqslant \ell), \qquad k, \ell \in \{0, 1, \dots\}.$$

Notable distributions

Poisson distribution with parameter λ

Let $\lambda \in \mathbb{R}_+$. A discrete random variable X is called **Poisson distributed with parameter λ**, if it can have values: 0, 1, ..., and its distribution is

$$\mathrm{P}(X = k) = \frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}, \qquad k \in \{0, 1, \dots\}.$$

Generating function

$$G_X(z) = \mathrm{e}^{\lambda(z-1)}, \qquad z \in \mathbb{C}.$$

Characteristic function

$$\varphi_X(t) = \mathrm{e}^{\lambda(\mathrm{e}^{\mathrm{i}t}-1)}, \qquad t \in \mathbb{R}.$$

Approximation of binomial distribution by Poisson distribution

If X_n, $n \in \mathbb{N}$, are binomial distributed random variables with parameter (n, p_n), and $np_n \to \lambda \in (0, \infty)$ as $n \to \infty$, then $X_n \xrightarrow{\mathcal{D}} X$ as $n \to \infty$, where the random variable X is Poisson distributed with parameter λ.

Notable distributions

Uniform distribution on the set $\{0, 1, \dots, N-1\}$

A discrete random variable X is called **uniformly distributed on the set $\{0, 1, \dots, N-1\}$**, if

$$\mathrm{P}(X = k) = \frac{1}{N}, \qquad k \in \{0, 1, \dots, N-1\}.$$

Generating function

$$G_X(z) = \frac{1}{N}(1 + z + \dots + z^{N-1}) = \begin{cases} \frac{1}{N}\frac{z^N - 1}{z-1} & \text{if } z \in \mathbb{C} \setminus \{1\}, \\ 1 & \text{if } z = 1. \end{cases}$$

Characteristic function

$$\varphi_X(t) = \frac{1}{N}(1 + \mathrm{e}^{\mathrm{i}t} + \dots + \mathrm{e}^{\mathrm{i}t(N-1)}) = \begin{cases} \frac{1}{N}\frac{\mathrm{e}^{\mathrm{i}tN} - 1}{\mathrm{e}^{\mathrm{i}t}-1} & \text{if } \mathrm{e}^{\mathrm{i}t} \in \mathbb{C} \setminus \{1\}, \\ 1 & \text{if } \mathrm{e}^{\mathrm{i}t} = 1, \end{cases}$$

where $t \in \mathbb{R}$.

Notable distributions

Uniform distribution on the interval (a, b)

Let $a, b \in \mathbb{R}$ such that $a < b$. An absolutely continuous random variable X is called **uniformly distributed on the interval (a, b)**, if its density function is

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & x \in (a, b), \\ 0, & \text{otherwise.} \end{cases}$$

Characteristic function

$$\varphi_X(t) = \begin{cases} \frac{\mathrm{e}^{\mathrm{i}bt} - \mathrm{e}^{\mathrm{i}at}}{\mathrm{i}(b-a)t}, & t \neq 0, \\ 1, & t = 0. \end{cases}$$

Approximation of continuous uniform distribution

If X_n, $n \in \mathbb{N}$, are uniformly distributed random variables on the sets $\{0, 1, \ldots, n-1\}$, $n \in \mathbb{N}$, then $\frac{X_n}{n} \xrightarrow{\mathcal{D}} X$ as $n \to \infty$, where the random variable X is uniformly distributed on the interval $(0, 1)$.

Notable distributions

Exponential distribution with parameter λ

Let $\lambda > 0$. An absolutely continuous random variable X is called **exponentially distributed with parameter λ**, if its density function is

$$f_X(x) = \begin{cases} \lambda \mathrm{e}^{-\lambda x}, & x > 0, \\ 0, & \text{otherwise.} \end{cases}$$

Memorylessness property of exponential distribution

If the random variable X is exponentially distributed with parameter λ, then

$$\mathrm{P}(X \geqslant t + h \mid X \geqslant t) = \mathrm{P}(X \geqslant h), \qquad t, h \geqslant 0.$$

Laplace transform

$$\psi_X(s) = \frac{\lambda}{s + \lambda}, \qquad s \in \mathbb{R}_+.$$

Characteristic function

$$\varphi_X(t) = \left(1 - \mathrm{i}\frac{t}{\lambda}\right)^{-1}, \qquad t \in \mathbb{R}.$$

Notable distributions

Normal distribution with parameter (m, σ^2)

Let $m \in \mathbb{R}$ and $\sigma > 0$. An absolutely continuous random variable X is called **normally distributed with parameter** $(\boldsymbol{m}, \boldsymbol{\sigma^2})$, if its density function is

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{(x-m)^2}{2\sigma^2}}$$

Characteristic function

$$\varphi_X(t) = \mathrm{e}^{\mathrm{i}mt - \frac{\sigma^2 t^2}{2}}, \qquad t \in \mathbb{R}$$

de Moivre CLT, approximation of binomial distribution by normal distribution

If X_n, $n \in \mathbb{N}$, are binomially distributed random variables with parameter (n, p), where $p \in (0, 1)$, then $\frac{X_n - np}{\sqrt{np(1-p)}} \xrightarrow{\mathcal{D}} X$ as $n \to \infty$, where the random variable X is normally distributed with parameter $(0, 1)$.

Multidimensional normal distribution

- A random vector $Y : \Omega \to \mathbb{R}^d$ is called **standard normally distributed**, if $Y = (Y_1, \ldots, Y_d)$, where $Y_1, \ldots, Y_d : \Omega \to \mathbb{R}$ are independent, standard normally distributed random variables.
- A random vector $X : \Omega \to \mathbb{R}^d$ is called **normally distributed**, if the distribution of X coincides with the distribution of $AY + m$, where $Y : \Omega \to \mathbb{R}^d$ is standard normally distributed, $A \in \mathbb{R}^{d\times d}$ and $m \in \mathbb{R}^d$.

Multidimensional normal distribution

Characteristic function, density function

- A random vector $X : \Omega \to \mathbb{R}^d$ is normally distributed if and only if its characteristic function has the form
$$\varphi_X(t) = \exp\left\{ \mathrm{i}\langle m, t\rangle - \frac{1}{2}\langle Dt, t\rangle \right\}, \qquad t \in \mathbb{R}^d,$$
where $m \in \mathbb{R}^d$, and $D \in \mathbb{R}^{d\times d}$ is a symmetric, positive semidefinite matrix, i.e., $D^\top = D$, and for each $t \in \mathbb{R}^d$ we have $\langle Dt, t\rangle \geqslant 0$. Further, $m = \mathrm{E}(X)$, $D = \mathrm{Cov}(X)$.
- If D is invertible, then X is absolutely continuous and its density function is
$$f_X(x) = \frac{1}{\sqrt{(2\pi)^d \det(D)}} \exp\left\{ -\frac{1}{2}\langle D^{-1}(x-m), x-m\rangle \right\}, \quad x \in \mathbb{R}^d.$$

A random vector $X : \Omega \to \mathbb{R}^d$ is called **normally distributed with parameters** $(\boldsymbol{m}, \boldsymbol{D})$, if the characteristic function of X has the form given in the theorem above. In notation: $X \stackrel{\mathcal{D}}{=} \mathcal{N}(m, D)$.

Multidimensional normal distribution

Linear transform of multidimensional normal distribution

If $X \stackrel{\mathcal{D}}{=} \mathcal{N}(m, D)$ is a d-dimensional normally distributed random vector, and $a \in \mathbb{R}^\ell$, $B \in \mathbb{R}^{\ell \times d}$, then $a + BX \stackrel{\mathcal{D}}{=} \mathcal{N}(a + Bm, BDB^\top)$ is an ℓ-dimensional normally distributed random vector.

Characterisation of multidimensional normal distribution

A random vector $X : \Omega \to \mathbb{R}^d$ is normally distributed if and only if for each $c \in \mathbb{R}^d$, the random variable $c^\top X$ is normally distributed.

Multidimensional normal distribution

Independence of coordinates of multidimensional normal distribution

Let $(X_1, \ldots, X_k, Y_1, \ldots, Y_\ell)$ be a $k+\ell$-dimensional normally distributed random vector, and let us suppose that for each $i \in \{1, \ldots, k\}$ and $j \in \{1, \ldots, \ell\}$, we have $\mathrm{Cov}(X_i, Y_j) = 0$. Then the random vectors $(X_1, \ldots, X_k)$ and $(Y_1, \ldots, Y_\ell)$ are independent.

Independence of linear combinations

Let $X_1, \ldots, X_d$ be independent, standard normally distributed random variables. The linear combinations $a_1X_1 + \cdots + a_dX_d$ and $b_1X_1 + \cdots + b_dX_d$ are independent if and only if the vectors $(a_1, \ldots, a_d)$ and $(b_1, \ldots, b_d)$ are orthogonal.

Convergence of random vectors

Let $X : \Omega \to \mathbb{R}^d$ and $X_n : \Omega \to \mathbb{R}^d$, $n \in \mathbb{N}$, be random vectors.
We say that **the sequence $X_1, X_2, \ldots$ converges to X**

- **almost surely** (in notation $X_n \xrightarrow{\text{a.s.}} X$ or $X_n \to X$ P-a.s.), if
$$\mathrm{P}\left(\lim_{n\to\infty} X_n = X\right) = 1;$$
- **stochastically** (in notation $X_n \xrightarrow{\mathrm{P}} X$), if for each $\varepsilon > 0$, we have
$$\lim_{n\to\infty} \mathrm{P}(\|X_n - X\| \geqslant \varepsilon) = 0;$$
- **in distribution** (in notation $X_n \xrightarrow{\mathcal{D}} X$), if
$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$
for each point $x \in \mathbb{R}^d$, where F_X is continuous;
- **in r^{th} mean**, where $r > 0$ (in notation $X_n \xrightarrow{\|\cdot\|_r} X$ or $X_n \xrightarrow{L_r} X$), if $\mathrm{E}(\|X\|^r) < \infty$, $\mathrm{E}(\|X_n\|^r) < \infty$, $n \in \mathbb{N}$, and $\lim_{n\to\infty} \mathrm{E}\left(\|X_n - X\|^r\right) = 0.$

Convergence of random vectors

Connection between modes of convergences

Let $X : \Omega \to \mathbb{R}^d$ and $X_n : \Omega \to \mathbb{R}^d$, $n \in \mathbb{N}$, be random vectors.

- If $X_n \xrightarrow{\text{a.s.}} X$, or $X_n \xrightarrow{\|\cdot\|_r} X$ for some $r > 0$, then $X_n \xrightarrow{\text{P}} X$.
- If $X_n \xrightarrow{\|\cdot\|_r} X$ for some $r > 0$, then for each $s \in (0, r)$, we have $X_n \xrightarrow{\|\cdot\|_s} X$.

Limit of stochastic convergence is uniquely determined

If $X : \Omega \to \mathbb{R}^d$, $Y : \Omega \to \mathbb{R}^d$, $X_n : \Omega \to \mathbb{R}^d$ and $Y_n : \Omega \to \mathbb{R}^d$, $n \in \mathbb{N}$, are random vectors such that $X_n \xrightarrow{\text{P}} X$, $Y_n \xrightarrow{\text{P}} Y$, and $X_n = Y_n$ P-a.s. for each $n \in \mathbb{N}$, then $X = Y$ P-a.s. In particular, if $X_n \xrightarrow{\text{a.s.}} X$, $Y_n \xrightarrow{\text{a.s.}} Y$ and $X_n = Y_n$ P-a.s. for each $n \in \mathbb{N}$, then $X = Y$ P-a.s.

Convergence of random vectors

An equivalent formulation of convergence in probability

Let $X_n : \Omega \to \mathbb{R}^d$, $n \in \mathbb{N}$, be random vectors. Then X_n converges in probability to some random vector $X : \Omega \to \mathbb{R}^d$ as $n \to \infty$, if and only if for all $\varepsilon > 0$, we have

$$\lim_{n\to\infty} \sup_{\{m\in\mathbb{N}:\ m>n\}} \mathrm{P}(\|X_m - X_n\| > \varepsilon) = 0.$$

Montone decreasing sequence converging in probability to 0

Let $X_n : \Omega \to \mathbb{R}$, $n \in \mathbb{N}$, be random variables. If $X_n \xrightarrow{\mathrm{P}} 0$ as $n \to \infty$, and $\mathrm{P}(0 \leqslant X_{n+1} \leqslant X_n) = 1$ for each $n \in \mathbb{N}$, then $X_n \xrightarrow{\text{a.s.}} 0$ as $n \to \infty$.

Convergence of random vectors

Let $X : \Omega \to \mathbb{R}^d$ and $X_n : \Omega \to \mathbb{R}^d$, $n \in \mathbb{N}$, be random vectors.

- The following statements are equivalent:
  - (a) $X_n \xrightarrow{\text{a.s.}} X$ as $n \to \infty$,
  - (b) $\sup_{\{k\in\mathbb{N}:\ k\geqslant n\}} \|X_k - X\| \xrightarrow{\mathrm{P}} 0$ as $n \to \infty$, i.e.,

$$\lim_{n\to\infty} \mathrm{P}\left(\sup_{\{k\in\mathbb{N}:\ k\geqslant n\}} \|X_k - X\| > \varepsilon \right) = 0, \quad \forall\, \varepsilon > 0,$$

  - (c) $\sup_{\{k\in\mathbb{N}:\ k\geqslant n\}} \|X_k - X\| \xrightarrow{\text{a.s.}} 0$ as $n \to \infty$.

Convergence of random vectors

- The following statements are equivalent:
  (a) $(X_n)_{n\in\mathbb{N}}$ converges almost surely to some d-dimensional random vector,
  (b) $\sup\limits_{\{k\in\mathbb{N}:\, k\geqslant n\}} \|X_k - X_n\| \xrightarrow{\mathrm{P}} 0$ as $n \to \infty$, i.e.,
  $$\lim_{n\to\infty} \mathrm{P}\left(\sup_{\{k\in\mathbb{N}:\, k\geqslant n\}} \|X_k - X_n\| > \varepsilon \right) = 0, \quad \forall\, \varepsilon > 0,$$
  (c) $\sup\limits_{\{k\in\mathbb{N}:\, k\geqslant n\}} \|X_k - X_n\| \xrightarrow{\text{a.s.}} 0$ as $n \to \infty$.
- $\sum\limits_{k=1}^{\infty} \mathrm{P}(\|X_k - X\| \geqslant \varepsilon) < \infty$ for all $\varepsilon > 0 \quad \Longrightarrow \quad X_n \xrightarrow{\text{a.s.}} X$.
- $X_n \xrightarrow{\mathrm{P}} X$ as $n \to \infty \quad \Longleftrightarrow \quad$ for each sequence of positive integers $n_1 < n_2 < \ldots$ there exists a subsequence $n_{k_1} < n_{k_2} < \ldots$ such that $X_{n_{k_i}} \xrightarrow{\text{a.s.}} X$ as $i \to \infty$.

Convergence of random vectors

Convergence of continuous functions of random vectors

Let $X : \Omega \to \mathbb{R}^d$, $Y : \Omega \to \mathbb{R}^d$, $X_n : \Omega \to \mathbb{R}^d$, and $Y_n : \Omega \to \mathbb{R}^d$, $n \in \mathbb{N}$, be random vectors and $g : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^r$ is a continuous function.

- If $X_n \xrightarrow{\text{a.s.}} X$ and $Y_n \xrightarrow{\text{a.s.}} Y$, then $g(X_n, Y_n) \xrightarrow{\text{a.s.}} g(X, Y)$.
- If $X_n \xrightarrow{\mathsf{P}} X$ and $Y_n \xrightarrow{\mathsf{P}} Y$, then $g(X_n, Y_n) \xrightarrow{\mathsf{P}} g(X, Y)$.

Connection between modes of convergences and operations

Let $X : \Omega \to \mathbb{R}^d$, $Y : \Omega \to \mathbb{R}^d$, $X_n : \Omega \to \mathbb{R}^d$, and $Y_n : \Omega \to \mathbb{R}^d$, $n \in \mathbb{N}$, be random vectors.

- If $X_n \xrightarrow{\text{a.s.}} X$ and $Y_n \xrightarrow{\text{a.s.}} Y$, then $X_n + Y_n \xrightarrow{\text{a.s.}} X + Y$ and $\langle X_n, Y_n \rangle \xrightarrow{\text{a.s.}} \langle X, Y \rangle$.
- If $X_n \xrightarrow{\mathsf{P}} X$ and $Y_n \xrightarrow{\mathsf{P}} Y$, then $X_n + Y_n \xrightarrow{\mathsf{P}} X + Y$ and $\langle X_n, Y_n \rangle \xrightarrow{\mathsf{P}} \langle X, Y \rangle$.
- If $X_n \xrightarrow{\|\cdot\|_r} X$ and $Y_n \xrightarrow{\|\cdot\|_r} Y$ for some $r > 0$, then $X_n + Y_n \xrightarrow{\|\cdot\|_r} X + Y$.

Convergence of random vectors

Uniform integrability of random vectors

Let $(\Omega, \mathcal{A}, \mathrm{P})$ be a probability space, $\Gamma \neq \emptyset$ be a nonempty set, and for each $\gamma \in \Gamma$, let $X_\gamma : \Omega \to \mathbb{R}^d$ be a random vector. The family $\{X_\gamma : \gamma \in \Gamma\}$ is called **uniformly integrable**, if

$$\lim_{K \to \infty} \sup_{\gamma \in \Gamma} \mathrm{E}\left(\|X_\gamma\| \mathbb{1}_{\{\|X_\gamma\| > K\}}\right) = 0.$$

If $\Gamma \neq \emptyset$ is a nonempty *finite* set, then the uniform integrability of the random vectors $\{X_\gamma : \gamma \in \Gamma\}$ is equivalent to $\sup_{\gamma \in \Gamma} \mathrm{E}(\|X_\gamma\|) < \infty$.

Especially, if X_n, $n \in \mathbb{N}$, is a sequence of identically distributed, integrable random vectors, then $\{X_n : n \in \mathbb{N}\}$ is uniformly integrable.

In case of an infinite set Γ, the next theorem gives a set of necessary and sufficient conditions.

Convergence of random vectors

Uniform integrability

Let $(\Omega, \mathcal{A}, \mathrm{P})$ be a probability space, $\Gamma \neq \emptyset$ be a nonempty set, and for each $\gamma \in \Gamma$, let $X_\gamma : \Omega \to \mathbb{R}^d$ be a random vector. The family $\{X_\gamma : \gamma \in \Gamma\}$ is uniformly integrable if and only if

$$\sup_{\gamma \in \Gamma} \mathrm{E}(\|X_\gamma\|) < \infty$$

and

$$\lim_{\mathrm{P}(A) \to 0} \sup_{\gamma \in \Gamma} \mathrm{E}\left(\|X_\gamma\| \mathbb{1}_A\right) = 0,$$

which is understood in a way that $\forall \varepsilon > 0$ there exists $\delta > 0$ such that $\mathrm{E}\left(\|X_\gamma\| \mathbb{1}_A\right) < \varepsilon$ for all $\gamma \in \Gamma$ and for all events $A \in \mathcal{A}$ satisfying $\mathrm{P}(A) < \delta$.

Convergence of random vectors

Uniform integrability

Let $(\Omega, \mathcal{A}, \mathrm{P})$ be a probability space, $\Gamma \neq \emptyset$ be a nonempty set, and for each $\gamma \in \Gamma$, let $X_\gamma : \Omega \to \mathbb{R}^d$, $Y_\gamma : \Omega \to \mathbb{R}^d$ be random vectors.

- If there exists $r > 1$ such that $\sup_{\gamma \in \Gamma} \mathrm{E}(\|X_\gamma\|^r) < \infty$, then the random vectors $\{X_\gamma : \gamma \in \Gamma\}$ are uniformly integrable.
- If the random vectors $\{X_\gamma : \gamma \in \Gamma\}$ and $\{Y_\gamma : \gamma \in \Gamma\}$ are uniformly integrable, then the random vectors $\{X_\gamma + Y_\gamma : \gamma \in \Gamma\}$ are uniformly integrable as well.
- If the random vectors $\{Y_\gamma : \gamma \in \Gamma\}$ are uniformly integrable and for each $\gamma \in \Gamma$, we have $\|X_\gamma\| \leqslant \|Y_\gamma\|$ P-a.s., then the random vectors $\{X_\gamma : \gamma \in \Gamma\}$ are uniformly integrable as well.

Momentum convergence theorem (Vitali)

Let $X, X_1, X_2, \ldots$ be d-dimensional random vectors, and $r > 0$.

The convergence $X_n \xrightarrow{\|\cdot\|_r} X$ is equivalent to that $X_n \xrightarrow{\mathrm{P}} X$ and the uniform integrability of the random vectors $\{\|X_n\|^r : n \in \mathbb{N}\}$.

Convergence of random vectors

Weak convergence of probability measures

Let μ_n, $n \in \mathbb{N}$, and μ be probability measures on the measurable space $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$.
We say that the sequence μ_n, $n \in \mathbb{N}$, **converges weakly** to μ (in notation: $\mu_n \Rightarrow \mu$), if $\lim_{n\to\infty} \mu_n(A) = \mu(A)$ for each $A \in \mathcal{B}(\mathbb{R}^d)$ such that $\mu(\partial A) = 0$, where $\partial A = A^- \setminus A^\circ$ denotes the boundary of the set A.

Convergence of random vectors

Portmanteau theorem

Let μ_n, $n \in \mathbb{N}$, and μ be probability measures on the measurable space $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. The following assertions are equivalent:

1. $\lim\limits_{n\to\infty} \int_{\mathbb{R}^d} g(y)\,\mu_n(\mathrm{d}y) = \int_{\mathbb{R}^d} g(y)\,\mu(\mathrm{d}y)$ for all bounded and continuous functions $g : \mathbb{R}^d \to \mathbb{R}$.
2. $\lim\limits_{n\to\infty} \int_{\mathbb{R}^d} g(y)\,\mu_n(\mathrm{d}y) = \int_{\mathbb{R}^d} g(y)\,\mu(\mathrm{d}y)$ for all bounded and uniformly continuous functions $g : \mathbb{R}^d \to \mathbb{R}$.
3. $\limsup\limits_{n\to\infty} \mu_n(F) \leqslant \mu(F)$ for all closed sets $F \in \mathcal{B}(\mathbb{R}^d)$.
4. $\liminf\limits_{n\to\infty} \mu_n(G) \geqslant \mu(G)$ for all open sets $G \in \mathcal{B}(\mathbb{R}^d)$.
5. $\lim\limits_{n\to\infty} \mu_n(A) = \mu(A)$ for all $A \in \mathcal{B}(\mathbb{R}^d)$ such that $\mu(\partial A) = 0$.

The word ”portmanteau” originally means a big travel suitcase. Nowadays, in linguistics it means blend of words: a new word is formed by combining two existing words that relate to a singular concept (for example: breakfast + lunch -> brunch or Hungarian + English -> Hunglish).

Connection between weak convergence and convergence in distribution

Let $X_n : \Omega \to \mathbb{R}^d$, $n \in \mathbb{N}$, and $X : \Omega \to \mathbb{R}^d$ be random vectors. The following assertions are equivalent:

1. $X_n \xrightarrow{\mathcal{D}} X$.
2. $\mathrm{P}_{X_n} \Rightarrow \mathrm{P}_X$.
3. $\lim\limits_{n\to\infty} \mathrm{E}(g(X_n)) = \mathrm{E}(g(X))$ for all bounded and continuous functions $g : \mathbb{R}^d \to \mathbb{R}$.
4. $\lim\limits_{n\to\infty} \mathrm{E}(g(X_n)) = \mathrm{E}(g(X))$ for all bounded and uniformly continuous functions $g : \mathbb{R}^d \to \mathbb{R}$.
5. $\limsup\limits_{n\to\infty} \mathrm{P}(X_n \in F) \leqslant \mathrm{P}(X \in F)$ for all closed sets $F \in \mathcal{B}(\mathbb{R}^d)$.
6. $\liminf\limits_{n\to\infty} \mathrm{P}(X_n \in G) \geqslant \mathrm{P}(X \in G)$ for all open sets $G \in \mathcal{B}(\mathbb{R}^d)$.
7. $\lim\limits_{n\to\infty} \mathrm{P}(X_n \in A) = \mathrm{P}(X \in A)$ for all Borel sets $A \in \mathcal{B}(\mathbb{R}^d)$ such that $\mathrm{P}(X \in \partial A) = 0$.

Convergence of random vectors

For a measurable function $h : \mathbb{R}^d \to \mathbb{R}^\ell$, let D_h be the set of discontinuity points of h, i.e.,

$$D_h := \Big\{ x \in \mathbb{R}^d : \text{there exists a sequence } (x_n)_{n\in\mathbb{N}} \text{ in } \mathbb{R}^d \text{ such that } x_n \to x, \text{ but } h(x_n) \nrightarrow h(x) \Big\}.$$

From measure theory it is known that $D_h \in \mathcal{B}(\mathbb{R}^d)$.

Mapping theorem

Let $X : \Omega \to \mathbb{R}^d$, $X_n : \Omega \to \mathbb{R}^d$, $n \in \mathbb{N}$, be random vectors, and $h : \mathbb{R}^d \to \mathbb{R}^\ell$ be a measurable function.
If $X_n \xrightarrow{\mathcal{D}} X$ and $\mathrm{P}(X \in D_h) = 0$, then $h(X_n) \xrightarrow{\mathcal{D}} h(X)$.

If $X_n \xrightarrow{\mathcal{D}} X$ and h is continuous, then $D_h = \emptyset$, and $h(X_n) \xrightarrow{\mathcal{D}} h(X)$, and in this case the mapping theorem is called continuous mapping theorem as well.

Convergence of random vectors

Cramér–Slutsky lemma

Let $X : \Omega \to \mathbb{R}^d$, $X_n : \Omega \to \mathbb{R}^d$, $n \in \mathbb{N}$, and $Y_n : \Omega \to \mathbb{R}^d$, $n \in \mathbb{N}$, be random vectors.
If $X_n \xrightarrow{\mathcal{D}} X$ and $X_n - Y_n \xrightarrow{\mathrm{P}} 0$, then $Y_n \xrightarrow{\mathcal{D}} X$.

Joint convergence in distribution

Let $X : \Omega \to \mathbb{R}^d$, $X_n : \Omega \to \mathbb{R}^d$, $Y_n : \Omega \to \mathbb{R}^d$, $n \in \mathbb{N}$, be random vectors, and $c \in \mathbb{R}^d$. If $X_n \xrightarrow{\mathcal{D}} X$ and $Y_n \xrightarrow{\mathrm{P}} c$, then $(X_n, Y_n) \xrightarrow{\mathcal{D}} (X, c)$.

Convergence of random vectors

Convergence of random vectors

Let $X : \Omega \to \mathbb{R}^d$, $X_n : \Omega \to \mathbb{R}^d$, $Y_n : \Omega \to \mathbb{R}$ and $Z_n : \Omega \to \mathbb{R}^d$, $n \in \mathbb{N}$, be random vectors, and $a \in \mathbb{R}^d$, $b \in \mathbb{R}$.

- If $X_n \xrightarrow{\mathrm{P}} X$, then $X_n \xrightarrow{\mathcal{D}} X$.
- **(Cramér–Slutsky)** If $X_n \xrightarrow{\mathcal{D}} X$, $Y_n \xrightarrow{\mathrm{P}} b$ and $Z_n \xrightarrow{\mathrm{P}} a$, then $Y_n X_n + Z_n \xrightarrow{\mathcal{D}} bX + a$. Especially, if $X_n \xrightarrow{\mathcal{D}} X$, and $a, a_n \in \mathbb{R}^d$, $n \in \mathbb{N}$, $b, b_n \in \mathbb{R}$, $n \in \mathbb{N}$, such that $a_n \to a$ and $b_n \to b$, then $b_n X_n + a_n \xrightarrow{\mathcal{D}} bX + a$.
- $X_n \xrightarrow{\mathrm{P}} a$ if and only if $X_n \xrightarrow{\mathcal{D}} a$.

Mapping theorem (for stochastic convergence)

Let $X_n : \Omega \to \mathbb{R}^d$, $n \in \mathbb{N}$, be random vectors, $h : \mathbb{R}^d \to \mathbb{R}^\ell$ be a measurable function, and $x \in \mathbb{R}^d$. If $X_n \xrightarrow{\mathrm{P}} x$ and $x \notin D_h$, then $h(X_n) \xrightarrow{\mathrm{P}} h(x)$.

Convergence of random vectors

Continuous mapping theorem (for stochastic convergence)

Let $X_n : \Omega \to \mathbb{R}^d$, $n \in \mathbb{N}$, and $X : \Omega \to \mathbb{R}^d$ be random vectors, and $h : \mathbb{R}^d \to \mathbb{R}^\ell$ be a continous function. If $X_n \xrightarrow{\mathrm{P}} X$, then $h(X_n) \xrightarrow{\mathrm{P}} h(X)$.

Mapping theorem (for expectation)

Let $X_n : \Omega \to \mathbb{R}$, $n \in \mathbb{N}$, be random vectors and $h : \mathbb{R} \to \mathbb{R}$ be a bounded and measurable function such that $\mathrm{P}(X \in D_h) = 0$.
If $X_n \xrightarrow{\mathcal{D}} X$, then $\mathrm{E}(h(X_n)) \to \mathrm{E}(h(X))$.

Convergence of random vectors

Let $X : \Omega \to \mathbb{R}$ and $X_n : \Omega \to \mathbb{R}$, $n \in \mathbb{N}$, be random variables.
If $X_n \xrightarrow{\mathcal{D}} X$, then $\mathsf{E}(|X|) \leqslant \liminf_{n\to\infty} \mathsf{E}(|X_n|)$.

Convergence in distribution and uniform integrability, I

Let $X : \Omega \to \mathbb{R}$ and $X_n : \Omega \to \mathbb{R}$, $n \in \mathbb{N}$, be random variables.
If $X_n \xrightarrow{\mathcal{D}} X$ and $\{X_n : n \in \mathbb{N}\}$ is uniformly intregrable, then
$\mathsf{E}(|X|) < \infty$ and $\mathsf{E}(X_n) \to \mathsf{E}(X)$.

Convergence in distribution and uniform integrability, II

Let $X : \Omega \to \mathbb{R}$ and $X_n : \Omega \to \mathbb{R}$, $n \in \mathbb{N}$, be random variables.
If $X_n \geqslant 0$, $n \in \mathbb{N}$, $X \geqslant 0$, $\mathsf{E}(X_n) < \infty$, $n \in \mathbb{N}$, $\mathsf{E}(X) < \infty$, $X_n \xrightarrow{\mathcal{D}} X$
and $\mathsf{E}(X_n) \to \mathsf{E}(X)$, then $\{X_n : n \in \mathbb{N}\}$ is uniformly integrable.

Conditional probability, conditional expectation

Let $(\Omega, \mathcal{A}, \mathrm{P})$ be a probability space.

Conditional relative frequency

If we carry out n independent experiments, then the **conditional relative frequency of an event** $A \in \mathcal{A}$ **given that the event** $B \in \mathcal{A}$ **occured** is

$$r_n(A \,|\, B) := \frac{k_n(A \cap B)}{k_n(B)} = \frac{r_n(A \cap B)}{r_n(B)},$$

where $k_n(A \cap B)$ and $k_n(B)$ denotes the frequency of the event $A \cap B$, and B, respectively, and $r_n(A \cap B)$, and $r_n(B)$ denotes their relative frequencies.

Conditional probability

Let $B \in \mathcal{A}$ be an event such that $\mathrm{P}(B) > 0$. The **conditional probability** of an event $A \in \mathcal{A}$ given the event $B \in \mathcal{A}$ (i.e., if we know that the event B occured) is

$$\mathrm{P}(A \,|\, B) := \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)}.$$

Conditional probability, conditional expectation

Conditional probability

Let $(\Omega, \mathcal{A}, \mathrm{P})$ be a probability space, and $B \in \mathcal{A}$ be such that $\mathrm{P}(B) > 0$. Then the mapping $\mathrm{Q}_B : \mathcal{A} \to [0,1]$, $\mathrm{Q}_B(A) := \mathrm{P}(A \,|\, B)$, $A \in \mathcal{A}$, is a probability measure on the measurable space $(\Omega, \mathcal{A})$, i.e., $(\Omega, \mathcal{A}, \mathrm{Q}_B)$ is a probability space.

Conditional probability

Let $(\Omega, \mathcal{A}, \mathrm{P})$ be a probability space, and $B \in \mathcal{A}$ be such that $\mathrm{P}(B) > 0$. Further, let $\mathcal{A}_B := \{A \cap B : A \in \mathcal{A}\}$. Then $\mathcal{A}_B$ is a σ-algebra and the mapping $\mathrm{Q}_B : \mathcal{A}_B \to [0,1]$, $\mathrm{Q}_B(A) := \mathrm{P}(A \,|\, B)$, $A \in \mathcal{A}_B$, is a probability measure on the measurable space $(B, \mathcal{A}_B)$, i.e., $(B, \mathcal{A}_B, \mathrm{Q}_B)$ is a probability space.

Conditional distribution, conditional expectation, conditional variance of a discrete random variable

Let B be an event having positive probability. If X is a discrete random variable with distribution $\mathrm{P}(X = x_k)$, $k = 1, 2, \ldots$, then the **conditional distribution** of X given B is

$$\mathrm{P}(X = x_k \,|\, B) = \mathrm{Q}_B(X = x_k), \qquad k = 1, 2, \ldots,$$

the **conditional expectation** of X given B is

$$\mathrm{E}(X|B) := \sum_k x_k \cdot \mathrm{P}(X = x_k \,|\, B) = \sum_k x_k \cdot \mathrm{Q}_B(X = x_k),$$

provided that this series is absolutely convergent, i.e., $\sum_k |x_k| \cdot \mathrm{P}(X = x_k \,|\, B) < \infty$, and the **conditional variance** is

$$\mathrm{Var}(X|B) := \mathrm{E}\left[(X - \mathrm{E}(X|B))^2 | B\right] = \mathrm{E}(X^2|B) - \left[\,\mathrm{E}(X|B)\right]^2$$

$$= \sum_k x_k^2 \cdot \mathrm{P}(X = x_k \,|\, B) - \left(\sum_k x_k \cdot \mathrm{P}(X = x_k \,|\, B)\right)^2,$$

provided that the series $\sum_k x_k^2 \cdot \mathrm{P}(X = x_k \,|\, B)$ is convergent.

Conditional probability, conditional expectation

- If X is a discrete random variable, then the sequence $\mathrm{P}(X = x_k \mid B)$, $k \in \mathbb{N}$, is a probability distribution, since these numbers are nonnegative and their sum is 1:

$$\sum_k \mathrm{P}(X = x_k \mid B) = \frac{1}{\mathrm{P}(B)} \sum_k \mathrm{P}(\{X = x_k\} \cap B)$$

$$= \frac{1}{\mathrm{P}(B)} \mathrm{P}\left(\bigcup_k (\{X = x_k\} \cap B) \right) = \frac{1}{\mathrm{P}(B)} \mathrm{P}\left(\left(\bigcup_k \{X = x_k\} \right) \cap B \right)$$

$$= \frac{1}{\mathrm{P}(B)} \mathrm{P}(\Omega \cap B) = 1.$$

- Especially, if B is an event such that $\mathrm{P}(B) = 1$ (e.g., $B = \Omega$), then the conditional distribution, expectation and variance of X given B coincides with the distribution, expectation and variance of X.
- If $\mathrm{E}(|X|) < \infty$, then for each event B having positive probability, we have $\mathrm{E}(|X| \mid B) < \infty$.

Conditional probability, conditional expectation

Let us roll two fair dices. What is the conditional distribution of the difference of the numbers shown on the dices given that their sum is ℓ?

Denote by X and Y the two numbers. Then $\ell \in \{2, 3, \dots, 12\}$ and

$$\mathrm{P}(X + Y = \ell) = \begin{cases} \frac{\ell - 1}{36} & \text{if } 2 \leqslant \ell \leqslant 7, \\ \frac{13 - \ell}{36} & \text{if } 7 \leqslant \ell \leqslant 12. \end{cases}$$

Further, $|X - Y|$ can have values: 0,1,2,3,4,5, and for $\ell \in \{2, \dots, 6\}$, the conditional probabilities in question are:

$$\mathrm{P}(|X - Y| = 0 \,|\, X + Y = 2) = 1, \qquad \mathrm{P}(|X - Y| = 1 \,|\, X + Y = 3) = 1,$$

$$\mathrm{P}(|X - Y| = 0 \,|\, X + Y = 4) = \frac{1}{3}, \qquad \mathrm{P}(|X - Y| = 2 \,|\, X + Y = 4) = \frac{2}{3},$$

$$\mathrm{P}(|X - Y| = 1 \,|\, X + Y = 5) = \frac{1}{2}, \qquad \mathrm{P}(|X - Y| = 3 \,|\, X + Y = 5) = \frac{1}{2},$$

$$\mathrm{P}(|X - Y| = 0 \,|\, X + Y = 6) = \frac{1}{5}, \qquad \mathrm{P}(|X - Y| = 2 \,|\, X + Y = 6) = \frac{2}{5},$$

$$\mathrm{P}(|X - Y| = 4 \,|\, X + Y = 6) = \frac{2}{5}.$$

Conditional distribution and conditional expectation of an absolutely continuous random variable

The **conditional distribution function** of a real-valued random variable X given an event B having positive probability is $F_{X|B} : \mathbb{R} \to [0, 1]$,

$$F_{X|B}(x) := \mathrm{P}(X < x \mid B) = \mathrm{Q}_B(X < x), \qquad x \in \mathbb{R}.$$

If there exists a Borel mesaurable function $f_{X|B} : \mathbb{R} \to \mathbb{R}$ such that

$$F_{X|B}(x) = \int_{-\infty}^{x} f_{X|B}(u)\,\mathrm{d}u$$

for all $x \in \mathbb{R}$, then the function $f_{X|B}$ is called a **conditional density function** of X given B.

The conditional distribution function $F_{X|B}$ is nothing else but the distribution function of the probability measure Q_B. The conditional density function $f_{X|B}$, provided that it exists, is Borel measurable, nonnegative Lebesgue almost everywhere, and $\int_{-\infty}^{\infty} f_{X|B}(u)\,\mathrm{d}u = 1$, and hence it is (usual) density function.

Conditional probability, conditional expectation

Conditional variance of an absolutely continuous r. v.

If there exists a conditional density function $f_{X|B}$, then the **conditional expectation** of X given B is

$$\mathrm{E}(X|B) := \int_{-\infty}^{\infty} x \cdot f_{X|B}(x)\,\mathrm{d}x$$

provided that this improper integral is absolutely convergent, i.e., $\int_{-\infty}^{\infty} |x| \cdot f_{X|B}(x)\,\mathrm{d}x < \infty$; and **the conditional variance** is of X given B is

$$\begin{aligned}\mathrm{Var}(X|B) &:= \mathrm{E}\left[(X - \mathrm{E}(X|B))^2 | B\right] = \mathrm{E}(X^2|B) - \left[\,\mathrm{E}(X|B)\right]^2 \\ &= \int_{-\infty}^{\infty} x^2 \cdot f_{X|B}(x)\,\mathrm{d}x - \left(\int_{-\infty}^{\infty} x \cdot f_{X|B}(x)\,\mathrm{d}x\right)^2,\end{aligned}$$

provided that $\int_{-\infty}^{\infty} x^2 \cdot f_{X|B}(x)\,\mathrm{d}x < \infty$.

If there exists a conditional density function $f_{X|B}$ and $\mathrm{E}(|X|) < \infty$, then for each event B having positive probability, we have $\mathrm{E}(|X| \mid B) < \infty$.

Example: Let X be a standard normally distributed random variable, and $B := \{X \geqslant 0\}$. Then $\mathrm{P}(B) = 1/2$, and

$$F_{X|B}(x) = \frac{\mathrm{P}(0 \leqslant X < x)}{\mathrm{P}(X \geqslant 0)} = \begin{cases} 0 & \text{if } x \leqslant 0, \\ 2\mathrm{P}(0 \leqslant X < x) & \text{if } x > 0. \end{cases}$$

If $x > 0$, then

$$F_{X|B}(x) = 2(\Phi(x) - \Phi(0)) = \sqrt{\frac{2}{\pi}} \int_0^x \mathrm{e}^{-u^2/2}\,\mathrm{d}u.$$

Hence the conditional density function of X given B is

$$f_{X|B}(x) = \begin{cases} \sqrt{\frac{2}{\pi}}\mathrm{e}^{-x^2/2} & \text{if } x > 0, \\ 0 & \text{if } x \leqslant 0. \end{cases}$$

Conditional probability, conditional expectation

Consequently, the conditional expectation of X given B is

$$\begin{aligned}
\mathrm{E}(X|B) &= \int_{-\infty}^{\infty} x \cdot f_{X|B}(x)\,\mathrm{d}x = \int_0^{\infty} x\sqrt{\frac{2}{\pi}}\mathrm{e}^{-x^2/2}\,\mathrm{d}x \\
&= \sqrt{\frac{2}{\pi}}\Big[-\mathrm{e}^{-x^2/2}\Big]_0^{\infty} = \sqrt{\frac{2}{\pi}}.
\end{aligned}$$

Further, if $Y := |X|$, then

$$\begin{aligned}
F_Y(x) = \mathrm{P}(|X| < x) &= \begin{cases} 0 & \text{if } x \leqslant 0, \\ \mathrm{P}(-x < X < x) & \text{if } x > 0, \end{cases} \\
&= \begin{cases} 0 & \text{if } x \leqslant 0, \\ 2\mathrm{P}(0 \leqslant X < x) & \text{if } x > 0, \end{cases} = F_{X|B}(x), \qquad x \in \mathbb{R},
\end{aligned}$$

i.e., the conditional distribution of X given B coincides with the distribution of $|X|$.

Conditional density function and conditional expectation given an absolutely continuous random variable

Let (X, Y) be an absolutely continuous random vector with density function $f_{X,Y}$. Then **the conditional density function of X given $Y = y$** is defined by

$$f_{X|Y}(x|y) := \begin{cases} \dfrac{f_{X,Y}(x,y)}{f_Y(y)} & \text{if } f_Y(y) \neq 0, \\ h(x) & \text{if } f_Y(y) = 0, \end{cases} \qquad x \in \mathbb{R},$$

where f_Y is the density function of Y and h is an arbitrary density function.

the conditional distribution function of X given $Y = y$ is

$$\mathsf{P}(X < x \mid Y = y) := \int_{-\infty}^{x} f_{X|Y}(u|y)\,\mathrm{d}u, \qquad x \in \mathbb{R}.$$

the conditional expectation of X given $Y = y$ is

$$\mathsf{E}(X|Y = y) := \int_{-\infty}^{\infty} x \cdot f_{X|Y}(x|y)\,\mathrm{d}x,$$

Conditional variance and regression curve given an absolutely continuous random variable

The conditional variance of X given $Y = y$ is

$$\begin{aligned}\operatorname{Var}(X|Y=y) &:= \mathrm{E}\left[(X - \mathrm{E}(X|Y=y))^2 | Y = y\right] \\ &= \mathrm{E}(X^2|Y=y) - \left[\mathrm{E}(X|Y=y)\right]^2 \\ &= \int_{-\infty}^{\infty} x^2 \cdot f_{X|Y}(x|y)\,\mathrm{d}x - \left(\int_{-\infty}^{\infty} x \cdot f_{X|Y}(x|y)\,\mathrm{d}x\right)^2,\end{aligned}$$

provided that $\int_{-\infty}^{\infty} x^2 \cdot f_{X|Y}(x|y)\,\mathrm{d}x < \infty$.

The regression curve of X given Y is

$$\text{the function } \mathbb{R} \ni y \mapsto \mathrm{E}(X|Y=y).$$

This minimizes the quantity $\mathrm{E}\left[(X - f(Y))^2\right]$, i.e., if $\mathrm{E}(X^2) < \infty$ and $f : \mathbb{R} \to \mathbb{R}$ is a measurable function such that $\mathrm{E}\left[f(Y)^2\right] < \infty$, then

$$\mathrm{E}\left[(X - \mathrm{E}(X|Y))^2\right] \leqslant \mathrm{E}\left[(X - f(Y))^2\right].$$

Theorem of total expectation given a partition of Ω

If $B_1, B_2, \ldots$ is a partition of Ω such that $\mathrm{P}(B_i) > 0, i \in \mathbb{N}$, X is a random variable and $\mathrm{E}(|X|) < \infty$, then

$$\mathrm{E}(X) = \sum_k \mathrm{E}(X \mid B_k) \cdot \mathrm{P}(B_k).$$

Proof. Let X be a discrete random variable having possible values $x_1, x_2, \ldots$. Then

$$\sum_k \mathrm{E}(X \mid B_k) \cdot \mathrm{P}(B_k) = \sum_k \sum_j x_j \mathrm{P}(X = x_j \mid B_k) \cdot \mathrm{P}(B_k)$$

$$= \sum_k \sum_j x_j \mathrm{P}(\{X = x_j\} \cap B_k) = \sum_j x_j \sum_k \mathrm{P}(\{X = x_j\} \cap B_k)$$

$$= \sum_j x_j \mathrm{P}(X = x_j) = \mathrm{E}(X),$$

where we used the condition $\mathrm{E}(|X|) < \infty$ for interchanging the sums. The case of an absolutely coninuous random variable X can be handled similarly.

Conditional probability, conditional expectation

Conditional expectation of a discrete random variable given a partition

Let X be a discrete random variable such that $\mathrm{E}(|X|) < \infty$, and let $\mathcal{G} := \{B_1, B_2, \ldots\}$ be a partition Ω such that $\mathrm{P}(B_k > 0)$, $k \in \mathbb{N}$. Then **the conditional expectation of X given $\mathcal{G}$** is the discrete random variable

$$\mathrm{E}(X \mid \mathcal{G}) := \sum_k \mathrm{E}(X \mid B_k) \mathbb{1}_{B_k}.$$

The random variable $\mathrm{E}(X \mid \mathcal{G})$ takes the value $\mathrm{E}(X \mid B_k)$ on the event B_k.

Example:

We roll a fair dice until we see 6 as the result.
Let X be the number of times we have to roll.
Then X is geometrically distributed with parameter $\frac{1}{6}$, so $\mathrm{E}(X) = 6$.

In what follows we determine $\mathrm{E}(X)$ using the theorem of total expectation as well.
Then $B_k := \{\text{the first roll is } k\}$, $k = 1, \ldots, 6$, is a partition of Ω consisting of events having positive (1/6) probability.
By the theorem of total expectation, since $\mathrm{E}(|X|) < \infty$, we have

$$\mathrm{E}(X) = \sum_{k=1}^{6} \mathrm{E}(X \mid B_k) \cdot \mathrm{P}(B_k)$$

We show that

$$\mathrm{E}(X \mid B_k) = \begin{cases} 1 + \mathrm{E}(X) & \text{if } 1 \leqslant k \leqslant 5, \\ 1 & \text{if } k = 6. \end{cases}$$

Then $\mathrm{P}(X = 1 \,|\, B_6) = 1$ and so $\mathrm{E}(X \,|\, B_6) = 1 \cdot 1 = 1$.
If $1 \leqslant k \leqslant 5$, then

$$\mathrm{E}(X \,|\, B_k) = \mathrm{E}(1 + X - 1 \,|\, B_k) = 1 + \mathrm{E}(X - 1 \,|\, B_k) = 1 + \mathrm{E}(X),$$

where at the last step we used that the conditional distribution of $X - 1$ given B_k $(k = 1, 2, 3, 4, 5)$ coincides with the distribution of X, since

$$\mathrm{P}(X-1 = n \,|\, B_k) = \frac{\mathrm{P}(X - 1 = n, B_k)}{\mathrm{P}(B_k)} = \frac{\frac{1}{6}\mathrm{P}(X = n)}{\frac{1}{6}} = \mathrm{P}(X = n), \quad n \in \mathbb{N}.$$

Hence

$$\mathrm{E}(X) = \frac{1}{6}\big(1 + 5(1 + \mathrm{E}(X))\big),$$

yielding

$$\mathrm{E}(X) = 6.$$

Conditional probability, conditional expectation

The conditional expectations $\mathrm{E}(X \mid B_k)$, $1 \leqslant k \leqslant 5$, can be (also) calculated directly (by definition).

If $1 \leqslant k \leqslant 5$, then $\mathrm{P}(X = 1 \mid B_k) = 0$ and

$$\mathrm{P}(X = n \mid B_k) = \frac{5^{n-2}}{6^{n-1}}, \quad n = 2, 3, \dots.$$

So for $k = 1, \dots, 5$, we have

$$\begin{aligned}
\mathrm{E}(X \mid B_k) &= \sum_{n=2}^{\infty} n \frac{5^{n-2}}{6^{n-1}} = \frac{1}{5} \sum_{n=2}^{\infty} n \left(\frac{5}{6}\right)^{n-1} = \frac{1}{5} \sum_{n=2}^{\infty} (x^n)' \Big|_{x=5/6} \\
&= \left(\sum_{n=2}^{\infty} x^n\right)' \Big|_{x=5/6} = \frac{1}{5} \left(\frac{x^2}{1-x}\right)' \Big|_{x=5/6} \\
&= \frac{1}{5} \frac{2x - x^2}{(1-x)^2} \Big|_{x=5/6} = 7.
\end{aligned}$$

Hence $\mathrm{E}(X) = \frac{1}{6}(1 + 5 \cdot 7) = 6.$

Conditional probability, conditional expectation

Further, the conditional expectation of X given the partition $\mathcal{G} := \{B_1, \dots, B_6\}$ consisting of events with positive probability is the discrete random variable

$$\mathrm{E}(X \mid \mathcal{G}) = 7(\mathbb{1}_{B_1} + \cdots + \mathbb{1}_{B_5}) + 1 \cdot \mathbb{1}_{B_6} = 7 \cdot \mathbb{1}_{\Omega \setminus B_6} + 1 \cdot \mathbb{1}_{B_6}.$$

That is

$$\mathrm{E}(X \mid \mathcal{G})(\omega) = \begin{cases} 7 & \text{if } \omega \notin B_6, \\ 1 & \text{if } \omega \in B_6. \end{cases}$$

Another property of expectation

Let $(\Omega, \mathcal{A}, \mathrm{P})$ be a probability space and $\mathcal{F} \subset \mathcal{A}$ be a sub-σ-algebra.

1. If $\zeta : \Omega \to \mathbb{R}$ is an $\mathcal{F}$-measurable random variable such that $\mathrm{E}(|\zeta|) < \infty$ and $\mathrm{E}(\zeta \mathbb{1}_A) \geqslant 0$ for each $A \in \mathcal{F}$, then $\zeta \geqslant 0$ P–a.s.
2. If $\xi : \Omega \to \mathbb{R}$ and $\eta : \Omega \to \mathbb{R}$ are $\mathcal{F}$-measurable random variables such that $\mathrm{E}(|\xi|) < \infty$, $\mathrm{E}(|\eta|) < \infty$, and $\mathrm{E}(\xi \mathbb{1}_A) \leqslant \mathrm{E}(\eta \mathbb{1}_A)$ for each $A \in \mathcal{F}$, then $\xi \leqslant \eta$ P–a.s.
3. If $\xi : \Omega \to \mathbb{R}$ and $\eta : \Omega \to \mathbb{R}$ are $\mathcal{F}$-measurable random variables such that $\mathrm{E}(|\xi|) < \infty$, $\mathrm{E}(|\eta|) < \infty$, and $\mathrm{E}(\xi \mathbb{1}_A) = \mathrm{E}(\eta \mathbb{1}_A)$ for each $A \in \mathcal{F}$, then $\xi = \eta$ P–a.s.

Conditional probability, conditional expectation

Conditional expectation given a σ-algebra

Let $(\Omega, \mathcal{A}, \mathrm{P})$ be a probability space, $\mathcal{F} \subset \mathcal{A}$ be a sub-σ-algebra, and $X : \Omega \to \mathbb{R}$ is a random variable such that $\mathrm{E}(|X|) < \infty$.
A random variable $X_{\mathcal{F}} : \Omega \to \mathbb{R}$ is called **a conditional expectation of *X* given $\mathcal{F}$**, if

1. $X_{\mathcal{F}}$ is $\mathcal{F}$-measurable (i.e., $\sigma(X_{\mathcal{F}}) \subset \mathcal{F}$) and $\mathrm{E}(|X_{\mathcal{F}}|) < \infty$,
2. for each $A \in \mathcal{F}$, we have $\mathrm{E}(X_{\mathcal{F}} \mathbb{1}_A) = \mathrm{E}(X \mathbb{1}_A)$.

Conditional expectation given a σ-algebra

Let $(\Omega, \mathcal{A}, \mathrm{P})$ be a probability space, $\mathcal{F} \subset \mathcal{A}$ be a sub-σ-algebra, and $X : \Omega \to \mathbb{R}$ be a random variable such that $\mathrm{E}(|X|) < \infty$.
Then there exists a conditional expectation $X_{\mathcal{F}} : \Omega \to \mathbb{R}$, which is uniquely determined P-a.s.

In notation: $\mathrm{E}(X \mid \mathcal{F})$ denotes the equivalence class of the random variable $X_{\mathcal{F}}$ with respect to P, and its arbitraty representative as well.

Conditional probability, conditional expectation

Properties of conditional expectation

Let $(\Omega, \mathcal{A}, \mathrm{P})$ be a probability space, $\mathcal{F} \subset \mathcal{A}$ be a sub-σ-algebra.

1. If $\mathrm{E}(|X|) < \infty$, $\mathrm{E}(|Y|) < \infty$ and $X \leqslant Y$, then $\mathrm{E}(X \,|\, \mathcal{F}) \leqslant \mathrm{E}(Y \,|\, \mathcal{F})$.
2. If $\mathrm{E}(|X|) < \infty$, then $|\mathrm{E}(X \,|\, \mathcal{F})| \leqslant \mathrm{E}(|X| \,|\, \mathcal{F})$.
3. If $\mathrm{E}(|X|) < \infty$, then $\mathrm{E}(X \,|\, \mathcal{A}) = X$.
4. If X is $\mathcal{F}$-measurable and $\mathrm{E}(|X|) < \infty$, then $\mathrm{E}(X \,|\, \mathcal{F}) = X$.
5. If $\mathrm{E}(|X|) < \infty$, then $\mathrm{E}[\mathrm{E}(X \,|\, \mathcal{F})] = \mathrm{E}(X)$.
6. If $\mathrm{E}(|X|) < \infty$ and X is independent of $\mathcal{F}$, then $\mathrm{E}(X \,|\, \mathcal{F}) = \mathrm{E}(X)$.
7. **Tower rule:** if $\mathrm{E}(|X|) < \infty$ and $\mathcal{G} \subset \mathcal{F}$ is a sub-σ-algebra, then $\mathrm{E}[\mathrm{E}(X \,|\, \mathcal{F}) \,|\, \mathcal{G}] = \mathrm{E}[\mathrm{E}(X \,|\, \mathcal{G}) \,|\, \mathcal{F}] = \mathrm{E}(X \,|\, \mathcal{G})$.
8. If $\mathrm{E}(|X|) < \infty$ and $\mathrm{E}(|Y|) < \infty$, then for each $a, b \in \mathbb{R}$, we have $\mathrm{E}(aX + bY \,|\, \mathcal{F}) = a\,\mathrm{E}(X \,|\, \mathcal{F}) + b\,\mathrm{E}(Y \,|\, \mathcal{F})$.
9. If $\mathrm{E}(|X|) < \infty$, $\mathrm{E}(|XY|) < \infty$ and Y is $\mathcal{F}$–measurable, then $\mathrm{E}(XY \,|\, \mathcal{F}) = Y\,\mathrm{E}(X \,|\, \mathcal{F})$.

Properties of conditional expectation

10. If $X_1, X_2, \ldots$ are P-integrable, $X_n \uparrow X$ P-a.s. and X is P-integrable as well, further there exists a random variable Y such that for each $n \in \mathbb{N}$, we have $X_n \geqslant Y$ P-a.s. and $\mathsf{E}(|Y|) < \infty$, then $\mathsf{E}(X_n \,|\, \mathcal{F}) \uparrow \mathsf{E}(X \,|\, \mathcal{F})$ P-a.s.
11. If $X_1, X_2, \ldots$ are P-integrable, for each $n \in \mathbb{N}$, we have $X_n \geqslant Y$ P-a.s., where Y is a random variable such that $\mathsf{E}(|Y|) < \infty$, and $\mathsf{E}\left(\left|\liminf_{n\to\infty} X_n\right|\right) < \infty$, then $\mathsf{E}\left(\liminf_{n\to\infty} X_n \,|\, \mathcal{F}\right) \leqslant \liminf_{n\to\infty} \mathsf{E}(X_n \,|\, \mathcal{F})$.
12. If $X_n \xrightarrow{\text{a.s.}} X$, and there exists a P-integrable random variable Y such that for each $n \in \mathbb{N}$, we have $|X_n| \leqslant Y$ P-a.s., then $\mathsf{E}(X_n \,|\, \mathcal{F}) \xrightarrow{\text{a.s.}} \mathsf{E}(X \,|\, \mathcal{F})$, and $\mathsf{E}(|X_n - X| \,|\, \mathcal{F}) \xrightarrow{\text{a.s.}} 0$.
13. If $X_1, X_2, \ldots$ are P-integrable, for each $n \in \mathbb{N}$, we have $X_n \geqslant 0$ P-a.s., and $\sum_{n=1}^{\infty} X_n$ is P-integrable as well, then $$\mathsf{E}\left(\sum_{n=1}^{\infty} X_n \,\Big|\, \mathcal{F}\right) = \sum_{n=1}^{\infty} \mathsf{E}(X_n \,|\, \mathcal{F}).$$

Conditional probability, conditional expectation

Multidimensional conditional Jensen inequality

Let $(\Omega, \mathcal{A}, \mathrm{P})$ be a probability space, $\mathcal{F} \subset \mathcal{A}$ be a sub-σ-algebra, and $X = (X_1, \ldots, X_d) : \Omega \to \mathbb{R}^d$ be a random vector such that $\mathrm{E}(\|X\|) < \infty$.

1. If $K \subset \mathbb{R}^d$ is nonempty, convex, closed and $X \in K$ P-a.s., then $\mathrm{E}(X \,|\, \mathcal{F}) := (\mathrm{E}(X_1 \,|\, \mathcal{F}), \ldots, \mathrm{E}(X_d \,|\, \mathcal{F})) \in K$ P-a.s.
2. If $g : \mathbb{R}^d \to \mathbb{R}$ is convex and $\mathrm{E}(|g(X)|) < \infty$, then $g(\mathrm{E}(X \,|\, \mathcal{F})) \leqslant \mathrm{E}(g(X) \,|\, \mathcal{F})$.

Conditional probability given a σ-algebra

Let $(\Omega, \mathcal{A}, \mathrm{P})$ be a probability space and $\mathcal{F} \subset \mathcal{A}$ be a sub-σ-algebra. The **conditional probability of an event $A \in \mathcal{A}$ given $\mathcal{F}$** is given by $\mathrm{P}(A \,|\, \mathcal{F}) := \mathrm{E}(\mathbb{1}_A \,|\, \mathcal{F})$.

Conditional expectation given a random vector

Let $(\Omega, \mathcal{A}, \mathrm{P})$ be a probability space, $X : \Omega \to \mathbb{R}$ be a random variable such that $\mathrm{E}(|X|) < \infty$, and $Y : \Omega \to \mathbb{R}^d$ be a random vector. Then the **conditional expectation of X given Y** is $\mathrm{E}(X \,|\, Y) := \mathrm{E}(X \,|\, \sigma(Y))$.

Conditional probability, conditional expectation

Conditional expectation given a random vector

There exists a measurable function $f : \mathbb{R}^d \to \mathbb{R}$ such that $\mathrm{E}(X \mid Y) = f(Y)$.
This is the P_Y-a.s. uniquely determined measurable function $f : \mathbb{R}^d \to \mathbb{R}$ such that for each $B \in \mathcal{B}(\mathbb{R}^d)$, we have

$$\int_B f(y)\,\mathrm{P}_Y(\mathrm{d}y) = \mathrm{E}(X \mathbb{1}_{Y^{-1}(B)}),$$

where P_Y denotes the distribution of Y, i.e., $\mathrm{P}_Y(B) := \mathrm{P}(Y \in B)$ for all $B \in \mathcal{B}(\mathbb{R}^d)$.

In notation: $f(y) = \mathrm{E}(X \mid Y = y)$, $y \in \mathbb{R}^d$.

Here f is nothing else but the Radon-Nikodym derivative of the finite, signed (i.e., not necessarily nonnegative) measure $\mathbb{Q}(B) := \mathrm{E}(X \mathbb{1}_{Y^{-1}(B)})$, $B \in \mathcal{B}(\mathbb{R}^d)$, with respect to P_Y on the mesurable space $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$.

Monotone class

A family $\mathcal{C}$ of subsets of a nonempty set Ω is called a monotone class, if $A_n \in \mathcal{C}$, $n \in \mathbb{N}$ and $A_n \uparrow A$ as $n \to \infty$ yield $A \in \mathcal{C}$.

Monotone class theorem

Let $\Omega \neq \emptyset$, $\mathcal{H}$ be an algebra consisting of some subsets of Ω, and $\mathcal{C}$ be a monotone class of some subsets of Ω such that $\mathcal{H} \subset \mathcal{C}$. Then $\sigma(\mathcal{H}) \subset \mathcal{C}$.

Conditional probability, conditional expectation

Properties of conditional expectation given a random vector

Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a probability space, $X : \Omega \to \mathbb{R}$ be a random variable such that $\mathrm{E}(|X|) < \infty$, and $Y : \Omega \to \mathbb{R}^d$ be a random vector.

1. If $g : \mathbb{R}^d \to \mathbb{R}$ is a measurable function such that $\mathrm{E}(|Xg(Y)|) < \infty$, then $\mathrm{E}(Xg(Y) \mid Y = y) = g(y)\,\mathrm{E}(X \mid Y = y)$.
2. If X and Y are independent, and $g : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}$ is a measurable function such that $\mathrm{E}(|g(X, Y)|) < \infty$, then $\mathrm{E}(g(X, Y) \mid Y = y) = \mathrm{E}(g(X, y) \mid Y = y) = \mathrm{E}(g(X, y))$ and $\mathrm{E}(g(X, Y) \mid Y) = \mathrm{E}(g(X, y))\big|_{y=Y}$.
3. If $g : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}$ is a measurable function such that $\mathrm{E}(|g(X, Y)|) < \infty$, then $\mathrm{E}(g(X, Y) \mid Y) = \mathrm{E}(g(X, y) \mid Y)\big|_{y=Y}$.

Conditional probability, conditional expectation

Conditional probability given a random vector

Let $(\Omega, \mathcal{A}, \mathrm{P})$ be a probability space, and $Y : \Omega \to \mathbb{R}^d$ be a random vector. **The conditional probability of an event $A \in \mathcal{A}$ given Y** is

$$\mathrm{P}(A \,|\, Y) := \mathrm{P}(A \,|\, \sigma(Y)) := \mathrm{E}(\mathbb{1}_A \,|\, \sigma(Y)).$$

As we saw earlier, there exists a P_Y-a.s. uniquely determined measurable function $f : \mathbb{R}^d \to \mathbb{R}$ such that $\mathrm{P}(A \,|\, Y) = f(Y)$.
The equivalence class of this function f with respect to P_Y, and its arbitrary representative as well, is denoted by
$\mathrm{P}(A \,|\, Y = y) = \mathrm{E}(\mathbb{1}_A \,|\, Y = y)$.

Properties of a conditional density function

Let $(\Omega, \mathcal{A}, \mathrm{P})$ be a probability space, and $(X, Y) : \Omega \to \mathbb{R}^2$ be an absolutely continuous random vector. Denote by $f_{X,Y}$ the density function of (X, Y). Let us define the function $f_{X|Y} : \mathbb{R}^2 \to [0, \infty)$,

$$f_{X|Y}(x|y) := \begin{cases} \dfrac{f_{X,Y}(x,y)}{f_Y(y)} & \text{if } f_Y(y) \neq 0, \\ h(x) & \text{if } f_Y(y) = 0, \end{cases}$$

where f_Y is the density function of Y, and $h : \mathbb{R} \to [0, \infty)$ is an arbitrary density function. Then the following assertions hold:

1. For each $y \in \mathbb{R}$, the function $\mathbb{R} \ni x \mapsto f_{X|Y}(x|y)$ is a density function.
2. For each $A \in \mathcal{B}(\mathbb{R})$, we have $\mathrm{P}(X \in A \mid Y = y) = \int_A f_{X|Y}(x|y)\,\mathrm{d}x$.
3. If $g : \mathbb{R} \to \mathbb{R}$ is a measurable function such that $\mathrm{E}(|g(X)|) < \infty$, then $\mathrm{E}(g(X) \mid Y = y) = \int_{-\infty}^{\infty} g(x) f_{X|Y}(x|y)\,\mathrm{d}x$.

The 2nd and 3rd statements hold for P_Y-a.e. $y \in \mathbb{R}$.

Conditional probability, conditional expectation

Conditional density function

Let $(\Omega, \mathcal{A}, \mathrm{P})$ be a probability space, and $(X, Y) : \Omega \to \mathbb{R}^2$ be an absolutely continuous random variable. The function $f_{X|Y}$ defined above is called a **conditional density function** of X given Y.

Theorems of total probability and total expectation

Let $(\Omega, \mathcal{A}, \mathrm{P})$ be a probability space, and $Y : \Omega \to \mathbb{R}$ be a random variable.

1. Then for each event $A \in \mathcal{A}$, we have

$$\mathrm{P}(A) = \int_{-\infty}^{\infty} \mathrm{P}(A \,|\, Y = y)\, \mathrm{P}_Y(\mathrm{d}y).$$

2. If $X : \Omega \to \mathbb{R}$ is a random variable such that $\mathrm{E}(|X|) < \infty$, then

$$\mathrm{E}(X) = \int_{-\infty}^{\infty} \mathrm{E}(X \,|\, Y = y)\, \mathrm{P}_Y(\mathrm{d}y).$$

Continuous version of Bayes theorem

Let $(\Omega, \mathcal{A}, \mathrm{P})$ be a probability space, and $(X, Y) : \Omega \to \mathbb{R}^2$ be an absolutely continuous random vector. Then for each Borel set $A \in \mathcal{B}(\mathbb{R})$, we have

$$\mathrm{P}(X \in A \mid Y = y) = \frac{\displaystyle\int_A f_{Y|X}(y|x) f_X(x)\,\mathrm{d}x}{\displaystyle\int_{-\infty}^{\infty} f_{Y|X}(y|x) f_X(x)\,\mathrm{d}x} \qquad \mathrm{P}_Y\text{-a.e. } y \in \mathbb{R}.$$

Conditional probability, conditional expectation

Best mean squared, $\mathcal{F}$-measurable prediction

Let $(\Omega, \mathcal{A}, \mathrm{P})$ be a probability space, $\mathcal{F} \subset \mathcal{A}$ be a sub-σ-algebra, and $X : \Omega \to \mathbb{R}$ be a square P-integrable random variable. A random variable $Y : \Omega \to \mathbb{R}$ is called **a best mean squared, $\mathcal{F}$-measurable prediction of X**, if

1. Y is $\mathcal{F}$-measurable and square P-integrable,
2. for each $\mathcal{F}$-measurable square P-integrable random variable $Z : \Omega \to \mathbb{R}$, we have $\mathrm{E}((X - Y)^2) \leqslant \mathrm{E}((X - Z)^2)$.

In fact, given the vector $X \in L^2(\Omega, \mathcal{A}, \mathrm{P})$ we search for a vector $Y \in L^2(\Omega, \mathcal{F}, \mathrm{P})$ such that $\|X - Y\|_{L^2} \leqslant \|X - Z\|_{L^2}$ holds for all $Z \in L^2(\Omega, \mathcal{F}, \mathrm{P})$, and this is of course the orthogonal projection of X onto the closed, linear subspace $L^2(\Omega, \mathcal{F}, \mathrm{P})$.

Best mean squared, $\mathcal{F}$-measurable prediction

There exists a best mean squared, $\mathcal{F}$-measurable prediction of X, namely, $\mathrm{E}(X \mid \mathcal{F})$ (which is square integrable).

Best mean squared linear prediction

Let $(\Omega, \mathcal{A}, \mathrm{P})$ be a probability space, and $X, Y_1, \ldots, Y_n : \Omega \to \mathbb{R}$ be square P–integrable random variables. A random variable $Y : \Omega \to \mathbb{R}$ is called **a best mean square linear prediction of $\boldsymbol{X}$ given $\boldsymbol{Y_1, \ldots, Y_n}$,** if

1. Y is an element of the closed, linear subspace $L^2(Y_1, \ldots, Y_n)$ of the Hilbert space $L^2(\Omega, \mathcal{A}, \mathrm{P})$ which consists of the linear combinations of $Y_1, \ldots, Y_n$,
2. for each $Z \in L^2(Y_1, \ldots, Y_n)$, we have $\mathrm{E}((X - Y)^2) \leqslant \mathrm{E}((X - Z)^2)$.

In fact, given the vector $X \in L^2(\Omega, \mathcal{A}, \mathrm{P})$, we search for a vector $Y \in L^2(Y_1, \ldots, Y_n)$ such that $\|X - Y\|_{L^2} \leqslant \|X - Z\|_{L^2}$ for all $Z \in L^2(Y_1, \ldots, Y_n)$; and this is of course the orthogonal projection of X onto the closed, linear subspace $L^2(Y_1, \ldots, Y_n)$. Since $L^2(Y_1, \ldots, Y_n)$ is contained in $L^2(\Omega, \sigma(Y_1, \ldots, Y_n), \mathrm{P})$, a best mean squared linear prediction given $Y_1, \ldots, Y_n$ is in general "worse" than a best mean squared, $\sigma(Y_1, \ldots, Y_n)$-measurable prediction, which has the form $f(Y_1, \ldots, Y_n)$ with some $\mathrm{P}_{Y_1,\ldots,Y_n}$-a.e. uniquely determined measurable function $f : \mathbb{R}^n \to \mathbb{R}$.

Conditional probability, conditional expectation

Best mean squared linear prediction

Let $(X, Y_1, \ldots, Y_n)$ be a $n+1$–dimensional normally distributed random variable, and let us suppose that $\mathrm{E}(X) = \mathrm{E}(Y_1) = \ldots = \mathrm{E}(Y_n) = 0$. Then the best mean squared linear prediction of X given $Y_1, \ldots, Y_n$ coincides with the best mean squared, $\sigma(Y_1, \ldots, Y_n)$–measurable prediction, so it is $\mathrm{E}(X \mid Y_1, \ldots Y_n)$ as well.

Conditional probability, conditional expectation

Example: Let (X, Y) be a normally distributed random vector such that $\mathrm{D}^2(Y) > 0$. Then

$$\mathrm{E}(X \mid Y = y) = \mathrm{E}(X) + \frac{\mathrm{Cov}(X, Y)}{\mathrm{D}^2(Y)}(y - \mathrm{E}(Y)),$$

i.e., the regression curve is a line.

Further, if the covariance matrix of (X, Y) is invertible, i.e., $\mathrm{D}^2(X)\mathrm{D}^2(Y) - (\mathrm{Cov}(X, Y))^2 > 0$, then the conditional distribution of X given $Y = y$ is normal distribution such that

$$\mathcal{N}\left(\mathrm{E}(X \mid Y = y), \mathrm{D}^2(X) - \frac{(\mathrm{Cov}(X, Y))^2}{\mathrm{D}^2(Y)}\right).$$

Hence

$$\mathrm{D}^2(X \mid Y = y) = \mathrm{D}^2(X) - \frac{(\mathrm{Cov}(X, Y))^2}{\mathrm{D}^2(Y)},$$

which does not depend on y.

Weak laws of large numbers

Let $X_1, X_2, \dots$ be random variables, and let

$$S_n := X_1 + \cdots + X_n, \qquad \overline{X}_n := \frac{X_1 + \cdots + X_n}{n}.$$

L_2-convergence of arithmetic mean

If $\mathrm{E}(X_n^2) < \infty$ for each $n \in \mathbb{N}$ and $\mathrm{E}(X_k X_\ell) = 0$ for $k \neq \ell$, then for all $\varepsilon > 0$ and $n \in \mathbb{N}$ we have

$$\mathrm{P}(|\overline{X}_n| \geqslant \varepsilon) \leqslant \frac{1}{\varepsilon^2} \mathrm{E}(\overline{X}_n^2) \leqslant \frac{1}{n\varepsilon^2} \sup_{\ell \geqslant 1} \mathrm{E}(X_\ell^2).$$

Especially, if $\sup_{\ell \geqslant 1} \mathrm{E}(X_\ell^2) < \infty$, then $\overline{X}_n \xrightarrow{\|\cdot\|_2} 0$, and hence $\overline{X}_n \xrightarrow{\mathrm{P}} 0$.

Weak laws of large numbers

Chebyshev theorem

If $X_1, X_2, \ldots$ are pairwise uncorrelated such that $\sup\limits_{\ell \geqslant 1} \mathrm{Var}(X_\ell) < \infty$ and $\mathrm{E}(X_n) = m$ for each $n \in \mathbb{N}$, where $m \in \mathbb{R}$, then $\overline{X}_n \xrightarrow{\|\cdot\|_2} m$, and hence $\overline{X}_n \xrightarrow{\mathrm{P}} m$.

Markov theorem

If $X_1, X_2, \ldots$ are pairwise uncorrelated such that $\sup\limits_{\ell \geqslant 1} \mathrm{Var}(X_\ell) < \infty$ and $\exists \lim\limits_{n \to \infty} \mathrm{E}(\overline{X}_n) =: m \in \mathbb{R}$, then $\overline{X}_n \xrightarrow{\|\cdot\|_2} m$, and hence $\overline{X}_n \xrightarrow{\mathrm{P}} m$.

Khinchin theorem (1929)

If $X_1, X_2, \ldots$ are pairwise independent, identically distributed random variables and $\mathrm{E}(|X_1|) < \infty$, then $\overline{X}_n \xrightarrow{\mathrm{P}} \mathrm{E}(X_1)$.

Weak laws of large numbers

L_1-convergence of arithmetic mean

If $X_1, X_2, \ldots$ are uniformly integrable, (totally) independent random variables, then

$$\overline{X}_n - \mathrm{E}(\overline{X}_n) \xrightarrow{\|\cdot\|_1} 0, \qquad \text{and hence} \qquad \overline{X}_n - \mathrm{E}(\overline{X}_n) \xrightarrow{\mathrm{P}} 0.$$

Especially, if $X_1, X_2, \ldots$ are independent, identically distributed random variables and $\mathrm{E}(|X_1|) < \infty$, then $\overline{X}_n \xrightarrow{\|\cdot\|_1} \mathrm{E}(X_1)$, and hence $\overline{X}_n \xrightarrow{\mathrm{P}} \mathrm{E}(X_1)$.

Strong laws of large numbers

L_4- and P-a.s. convergence of arithmetic mean

If $X_1, X_2, \ldots$ are independent random variables and $\mathrm{E}(X_n) = 0$ for each $n \in \mathbb{N}$, then for all $\varepsilon > 0$ and $n \in \mathbb{N}$, we have

$$\mathrm{P}(|\overline{X}_n| \geqslant \varepsilon) \leqslant \frac{\mathrm{E}(\overline{X}_n^4)}{\varepsilon^4} \leqslant \frac{3}{n^2 \varepsilon^4} \sup_{\ell \geqslant 1} \mathrm{E}(X_\ell^4).$$

Especially, if $\sup_{\ell \geqslant 1} \mathrm{E}(X_\ell^4) < \infty$, then $\overline{X}_n \xrightarrow{\|\cdot\|_4} 0$ and $\overline{X}_n \xrightarrow{\text{a.s.}} 0$.

A strong law under second order moment assumption

If $X_1, X_2, \ldots$ are pairwise independent, identically distributed random variables and $\mathrm{E}(X_1^2) < \infty$, then $\overline{X}_n \xrightarrow{\text{a.s.}} \mathrm{E}(X_1)$.

Kolmogorov inequality

If $X_1, \ldots, X_n$ are independent random variables and $\mathrm{E}(X_k^2) < \infty$ for each $k \in \{1, \ldots, n\}$, then for all $\varepsilon > 0$, we have

$$\mathrm{P}\left(\max_{1 \leqslant k \leqslant n} |S_k - \mathrm{E}(S_k)| \geqslant \varepsilon \right) \leqslant \frac{\mathrm{Var}(S_n)}{\varepsilon^2}.$$

Strong laws of large numbers

Kolmogorov one series theorem

If $X_1, X_2, \ldots$ are independent random variables and $\sum\limits_{n=1}^{\infty} \mathrm{Var}(X_n) < \infty$, then

$$\mathrm{P}\left(\sum_{n=1}^{\infty} (X_n - \mathrm{E}(X_n)) \text{ is convergent} \right) = 1.$$

Kolmogorov two series theorem

If $X_1, X_2, \ldots$ are independent random variables such that $\sum\limits_{n=1}^{\infty} \mathrm{E}(X_n)$ is convergent and $\sum\limits_{n=1}^{\infty} \mathrm{Var}(X_n) < \infty$, then

$$\mathrm{P}\left(\sum_{n=1}^{\infty} X_n \text{ is convergent} \right) = 1.$$

Strong laws of large numbers

Kolmogorov three series theorem

If $X_1, X_2, \ldots$ are independent random variables and there exists $c > 0$ such that

(1) $\sum_{n=1}^{\infty} \mathrm{E}(X_n^{(c)})$ is convergent,

(2) $\sum_{n=1}^{\infty} \mathrm{Var}(X_n^{(c)}) < \infty$,

(3) $\sum_{n=1}^{\infty} \mathrm{P}(|X_n| \geqslant c) < \infty$,

where

$$X_n^{(c)} := X_n \mathbb{1}_{\{|X_n| < c\}} = \begin{cases} X_n, & \text{if } |X_n| < c, \\ 0, & \text{if } |X_n| \geqslant c, \end{cases}$$

then

$$\mathrm{P}\left(\sum_{n=1}^{\infty} X_n \text{ is convergent} \right) = 1.$$

Kronecker lemma

Let $b_1, b_2, \ldots$ be a sequence of positive numbers such that $b_n \uparrow \infty$, and for each $n \in \mathbb{N}$ let $\beta_n := b_n - b_{n-1}$, where $b_0 := 0$.

1. If $s_1, s_2, \ldots$ is a real sequence and $s_n \to s \in \mathbb{R}$, then $\frac{1}{b_n}\sum_{\ell=1}^{n} \beta_\ell s_\ell \to s$. Especially, for a convergent sequence, the sequence of its arithmetic means converges to the same limit.
2. If $x_1, x_2, \ldots$ is a real sequence and $\sum_{n=1}^{\infty} \frac{x_n}{b_n}$ is convergent, then $\frac{1}{b_n}\sum_{\ell=1}^{n} x_\ell \to 0$.
3. **(Discrete L'Hôspital rule)** Let us suppose that $\beta_n > 0$, $n \in \mathbb{N}$, and let $(x_n)_{n\in\mathbb{N}}$ be a real sequence such that $\frac{x_n}{\beta_n} \to c \in \mathbb{R}$. Then

$$\frac{1}{b_n}\sum_{\ell=1}^{n} x_\ell = \frac{\sum_{\ell=1}^{n} x_\ell}{\sum_{\ell=1}^{n} \beta_\ell} \to c.$$

The reason for calling it as discrete L'Hôspital rule is that the condition $\frac{x_n}{\beta_n} \to c \in \mathbb{R}$ can also be written in the form

$$\frac{\Delta\left(\sum_{\ell=1}^{n} x_\ell\right)}{\Delta\left(\sum_{\ell=1}^{n} \beta_\ell\right)} \to c \in \mathbb{R},$$

where $\Delta x_n := x_n - x_{n-1}$, $n \in \mathbb{N}$, with $x_0 := 0$.

Strong laws of large numbers

Kolmogorov theorem (1929)

Let $X_1, X_2, \ldots$ be independent random variables. Let $b_1, b_2, \ldots$ be a sequence of positive numbers such that $b_n \uparrow \infty$. If $\sum_{n=1}^{\infty} \frac{\mathrm{Var}(X_n)}{b_n^2} < \infty$, then

$$\frac{1}{b_n} \sum_{\ell=1}^{n} (X_\ell - \mathrm{E}(X_\ell)) \xrightarrow{\text{a.s.}} 0.$$

Especially, if $\sum_{n=1}^{\infty} \frac{\mathrm{Var}(X_n)}{n^2} < \infty$, then $\overline{X}_n - \mathrm{E}(\overline{X}_n) \xrightarrow{\text{a.s.}} 0$.

Kolmogorov theorem (1933)

Let $X_1, X_2, \ldots$ be independent, identically distributed random variables.

1. If $\mathrm{E}(|X_1|) < \infty$, then $\overline{X}_n \xrightarrow{\text{a.s.}} \mathrm{E}(X_1)$.
2. If $\mathrm{P}\left((\overline{X}_n)_{n\geqslant 1} \text{ converges}\right) > 0$, then $\mathrm{E}(|X_1|) < \infty$.

Strong laws of large numbers

Etemadi (1981)

Let $X_1, X_2, \ldots$ be pairwise independent, identically distributed random variables such that $\mathrm{E}(|X_1|) < \infty$. Then $\overline{X}_n \xrightarrow{\text{a.s.}} \mathrm{E}(X_1)$.

Chandra and Goswami (1992)

Let $X_1, X_2, \ldots$ be pairwise independent random variables such that

$$\int_0^\infty \sup_{n\in\mathbb{N}} \mathrm{P}(|X_n| > t)\,\mathrm{d}t < \infty.$$

Then $\frac{S_n - \mathrm{E}(S_n)}{n} \xrightarrow{\text{a.s.}} 0$.

Degenerate random variable

Let $(\Omega, \mathcal{A}, \mathrm{P})$ be a probability space. A random vector $X : \Omega \to \mathbb{R}^d$ is called **degenerate**, if there exists $x_0 \in \mathbb{R}^d$ such that $\mathrm{P}(X = x_0) = 1$.

For each $n \in \mathbb{N}$, let $X_{n,1}, \dots, X_{n,k_n}$ be independent (real-valued) random variables such that not all of them are degenerate and $\mathrm{E}(X_{n,j}^2) < \infty$, $j = 1, \dots, k_n$. For each $n \in \mathbb{N}$ and $j = 1, \dots, k_n$, let

- $\sigma_{n,j} := \sqrt{\mathrm{Var}(X_{n,j})}$,
- $S_n := X_{n,1} + \cdots + X_{n,k_n}$,
- $D_n := \sqrt{\mathrm{Var}(S_n)} = \sqrt{\sum_{j=1}^{k_n} \sigma_{n,j}^2} > 0$,
- $\widehat{S}_n := (S_n - \mathrm{E}(S_n))/D_n$. Then $\mathrm{E}(\widehat{S}_n) = 0$ and $\mathrm{Var}(\widehat{S}_n) = 1$.
- $r_n := \frac{1}{D_n} \max_{1 \leqslant j \leqslant k_n} \sigma_{n,j}$,
- $L_n(\varepsilon) := \frac{1}{D_n^2} \sum_{j=1}^{k_n} \mathrm{E}\left[(X_{n,j} - \mathrm{E}(X_{n,j}))^2 \mathbb{1}_{\{|X_{n,j} - \mathrm{E}(X_{n,j})| \geqslant \varepsilon D_n\}}\right]$, $\varepsilon > 0$.

Let $Y \sim \mathcal{N}(0, 1)$.

Central limit theorems

Lindeberg theorem

For each $n \in \mathbb{N}$, let $X_{n,1}, \dots, X_{n,k_n}$ be independent random variables such that not all of them are degenerate and $\mathrm{E}(X_{n,j}^2) < \infty$, $j = 1, \dots, k_n$. Let $g : \mathbb{R} \to \mathbb{R}$ be a three times continuously differentiable function.

1. For each $n \in \mathbb{N}$ and $\varepsilon > 0$, we have

$$\left| \mathrm{E}[g(\widehat{S}_n)] - \mathrm{E}[g(Y)] \right| \leqslant \left(\frac{\varepsilon}{6} + \frac{r_n}{2} \right) \|g'''\|_\infty + L_n(\varepsilon) \|g''\|_\infty,$$

   where $\|h\|_\infty := \sup_{x \in \mathbb{R}} |h(x)|$ for any $h : \mathbb{R} \to \mathbb{R}$.

2. If $\|g''\|_\infty < \infty$, $\|g'''\|_\infty < \infty$, and the so called **Lindeberg condition** holds, i.e., $\lim_{n\to\infty} L_n(\varepsilon) = 0$ for each $\varepsilon > 0$, then $\lim_{n\to\infty} \mathrm{E}[g(\widehat{S}_n)] = \mathrm{E}[g(Y)]$.

Lindeberg central limit theorem for triangular arrays

For each $n \in \mathbb{N}$, let $X_{n,1}, \ldots, X_{n,k_n}$ be independent random variables such that not all of them are degenerate and $\mathrm{E}(X_{n,j}^2) < \infty$, $j = 1, \ldots, k_n$. If $\lim_{n\to\infty} L_n(\varepsilon) = 0$ for each $\varepsilon > 0$, and $g : \mathbb{R} \to \mathbb{C}$ is a continuous function such that

$$\sup_{x\in\mathbb{R}} \frac{|g(x)|}{1 + x^2} < \infty,$$

then

$$\lim_{n\to\infty} \mathrm{E}[g(\widehat{S}_n)] = \mathrm{E}[g(Y)].$$

Especially, $\widehat{S}_n \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1)$.

Central limit theorems

Lindeberg central limit theorem for triangular arrays

For each $n \in \mathbb{N}$, let $X_{n,1}, \ldots, X_{n,k_n}$ be indendent random variables such that not all of them are degenerate and $\mathrm{E}(X_{n,j}^2) < \infty$, $j = 1, \ldots, k_n$. If $\lim_{n\to\infty} L_n(\varepsilon) = 0$ for each $\varepsilon > 0$, and $g_n : \mathbb{R} \to \mathbb{C}$, $n \in \mathbb{N}$, are continuous functions such that

$$\sup_{n\in\mathbb{N}} \sup_{x\in\mathbb{R}} \frac{|g_n(x)|}{1+x^2} < \infty,$$

and g_n converges uniformly on compact sets to some continuous function $g : \mathbb{R} \to \mathbb{C}$ as $n \to \infty$ (i.e., for each compact set $K \subset \mathbb{R}$, we have $g_n|_K$ converges uniformly to $g|_K$ as $n \to \infty$, i.e., for each compact set $K \subset \mathbb{R}$, we have $\lim_{n\to\infty} \sup_{x\in K} |g_n(x) - g(x)| = 0$), then

$$\lim_{n\to\infty} \mathrm{E}[g_n(\widehat{S}_n)] = \mathrm{E}[g(Y)].$$

Especially, $\widehat{S}_n \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$.

Central limit theorems

Lyapunov central limit theorem for triangular arrays

For each $n \in \mathbb{N}$, let $X_{n,1}, \dots, X_{n,k_n}$ be independent random variables such that not all of them are degenerate. If for some $\delta > 0$, we have $\mathrm{E}(|X_{n,j}|^{2+\delta}) < \infty$, $n \in \mathbb{N}$, $j = 1, \dots, k_n$, and

$$\frac{1}{D_n^{2+\delta}} \sum_{j=1}^{k_n} \mathrm{E}\Big[\big|X_{n,j} - \mathrm{E}(X_{n,j})\big|^{2+\delta}\Big] \to 0, \qquad \text{as } n \to \infty,$$

then $\widehat{S}_n = \frac{S_n - \mathrm{E}(S_n)}{D_n} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$. (Here $D_n = \sqrt{\mathrm{Var}(S_n)}$, $n \in \mathbb{N}$.)

Lévy central limit theorem: independent, identically distributed case

Let $X_1, X_2, \dots$ be independent, identically distributed random variables, and let $S_n := X_1 + \cdots + X_n$.

If $\mathrm{E}(X_1^2) < \infty$ and $\mathrm{Var}(X_1) > 0$, then $\frac{S_n - \mathrm{E}(S_n)}{\sqrt{\mathrm{Var}\, S_n}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$.

Further,

$$\sup_{x \in \mathbb{R}} \left| \mathrm{P}\left(\frac{S_n - \mathrm{E}(S_n)}{\sqrt{\mathrm{Var}\, S_n}} < x \right) - \Phi(x) \right| \to 0, \qquad \text{as } n \to \infty.$$

Central limit theorems

For each $n \in \mathbb{N}$, let $X_{n,1}, \ldots, X_{n,k_n}$ be independent random variables such that not all of them are degenerate, and $\mathsf{E}(X_{n,j}^2) < \infty$, $j = 1, \ldots, k_n$. If the Lindeberg condition holds, i.e., $L_n(\varepsilon) \to 0$ for each $\varepsilon > 0$, then the so called **uniformly asymptotically negligible condition** holds, i.e., for each $\varepsilon > 0$, we have

$$\max_{1 \leqslant j \leqslant k_n} \mathsf{P}\left(\left|\frac{X_{n,j} - \mathsf{E}(X_{n,j})}{\sqrt{\mathrm{Var}(S_n)}}\right| \geqslant \varepsilon\right) \to 0.$$

The uniformly asymptotically negligible condition is called infinitesimality condition as well.

Feller theorem (1935)

For each $n \in \mathbb{N}$, let $X_{n,1}, \ldots, X_{n,k_n}$ be independent random variables such that not all of them are degenerate, $\mathsf{E}(X_{n,j}^2) < \infty$, $j = 1, \ldots, k_n$, and let $S_n = X_{n,1} + \cdots + X_{n,k_n}$. If $\widehat{S}_n = \frac{S_n - \mathsf{E}(S_n)}{\sqrt{\mathrm{Var}(S_n)}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$ and the uniformly asymptotically negligible condition holds, then the Lindeberg condition holds.

Corollary

For each $n \in \mathbb{N}$, let $X_{n,1}, \dots, X_{n,k_n}$ be independent random variables such that not all of them are degenerate, $\mathrm{E}(X_{n,j}^2) < \infty$, $j = 1, \dots, k_n$, and let $S_n = X_{n,1} + \dots + X_{n,k_n}$.

(i) If $r_n = \frac{1}{D_n} \max_{1 \leqslant j \leqslant k_n} \sigma_{n,j} \to 0$, as $n \to \infty$, then the uniformly asymptotically negligible condition holds.

(ii) If the uniformly asymptotically negligible condition holds, then $\widehat{S}_n \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$, as $n \to \infty$ holds if and only if the Lindeberg condition holds.

Lindeberg multidimensional central limit theorem for triangular arrays

For each $n \in \mathbb{N}$, let $X_{n,1}, \ldots, X_{n,k_n}$ be independent d-dimensional random vectors, and $\mathrm{E}(\|X_{n,j}\|^2) < \infty$, $j = 1, \ldots, k_n$.
If

1. $\sum_{j=1}^{k_n} \mathrm{Var}(X_{n,j}) \to \Sigma$ as $n \to \infty$, where $\Sigma \in \mathbb{R}^{d \times d}$ is invertible,
2. for each $\varepsilon > 0$, we have

$$\sum_{j=1}^{k_n} \mathrm{E}\left[\|X_{n,j} - \mathrm{E}(X_{n,j})\|^2 \mathbb{1}_{\{\|X_{n,j}-\mathrm{E}(X_{n,j})\| \geqslant \varepsilon\}}\right] \to 0,$$

then $S_n - \mathrm{E}(S_n) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma)$, where $\mathcal{N}(0, \Sigma)$ denotes a d-dimensional normal distribution with mean vector $0 \in \mathbb{R}^d$ and covariance matrix $\Sigma \in \mathbb{R}^{d \times d}$.

Multidimensional central limit theorem: IID case

Let X_n, $n \in \mathbb{N}$, be independent, identically distributed d-dimensional random variables, and let $S_n = X_1 + \cdots + X_n$, $n \in \mathbb{N}$, denote the partial sums. If $\mathrm{E}(\|X_1\|^2) < \infty$ and $\mathrm{Var}(X_1) \in \mathbb{R}^{d \times d}$ is invertible, then

$$\frac{1}{\sqrt{n}}(S_n - \mathrm{E}(S_n)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mathrm{Var}(X_1)) \qquad \text{as } n \to \infty,$$

where $\mathcal{N}(0, \mathrm{Var}(X_1))$ denotes a d-dimensional normal distribution with mean vector $0 \in \mathbb{R}^d$ and covariance matrix $\mathrm{Var}(X_1)$.

Central limit theorems

Poisson convergence theorem

For each $n \in \mathbb{N}$, let $X_{n,1}, \ldots, X_{n,k_n}$ be independent random variables such that $\mathrm{P}(X_{n,j} = 1) = p_{n,j} = 1 - \mathrm{P}(X_{n,j} = 0)$, $j = 1, \ldots, k_n$, and let $S_n := X_{n,1} + \cdots + X_{n,k_n}$. If $\sum_{j=1}^{k_n} p_{n,j} \to \lambda \in \mathbb{R}_+$ and $\max_{1 \leqslant j \leqslant k_n} p_{n,j} \to 0$, then $S_n \xrightarrow{\mathcal{D}} \mathrm{Poisson}(\lambda)$.

An auxiliary lemma for estimation of difference of products

If $m \in \mathbb{N}$ and $a_1, \ldots, a_m, b_1, \ldots, b_m \in [-1, 1]$, then

$$\left| \prod_{j=1}^{m} a_j - \prod_{j=1}^{m} b_j \right| \leqslant \sum_{j=1}^{m} |a_j - b_j|.$$

Stochastic processes

Stochastic process

Let $(\Omega, \mathcal{A}, \mathrm{P})$ be a probability space, T be an arbitrary nonempty set, and for each $t \in T$, let $\xi_t : \Omega \to \mathbb{R}$ be a random variable. Then the family $\{\xi_t : t \in T\}$ is called a **stochastic process**. We say that T is the **parameter set** (or index set) of the process, and $\mathbb{R}$ is its **phase space** (or state space).

We say that a stochastic process $\{\xi_t : t \in T\}$ is in the state $x \in \mathbb{R}$ at the parameter $t \in T$, if for a realized outcome $\omega \in \Omega$, we have $\xi_t(\omega) = x$. For denoting the value of the process, we will use $\xi(t)(\omega)$, and $\xi(t, \omega)$, $t \in T, \omega \in \Omega$ as well (since a process can be naturally considered as a single mapping $\xi : T \times \Omega \to \mathbb{R}$: $\xi(t, \omega) := \xi_t(\omega)$).

Trajectory (realization, sample function)

For a fixed $\omega \in \Omega$, the mapping $T \ni t \mapsto \xi_t(\omega) \in \mathbb{R}$ is called a **trajectory** (realization, sample function) of the process.

Stochastic processes

Discrete and continuous time processes

Let $T \subset \mathbb{R}_+$ and $\{\xi_t : t \in T\}$ be a real valued stochastic process. We say that the process is of discrete time, if T is a countable set. Then usually $T = \mathbb{Z}_+$, so the process is a sequence of random variables. The process is called a continuous time process, if T is a finite or infinite subinterval of the nonnegative real line. Then for example $T = \mathbb{R}_+$ or $T = [0, 1]$.

Finite dimensional distributions

Let $T \subset \mathbb{R}$. By the finite dimensional distributions of a stochastic process $\{\xi_t : t \in T\}$, we mean the distributions of the random vectors:

$$\{(\xi_{t_1}, \ldots, \xi_{t_k}) : k \in \mathbb{N},\ t_1, \ldots, t_k \in T\} .$$

Modification, indistinguishability

Let T be a nonempty set. The stochastic processes $\{\xi_t : t \in T\}$ and $\{\eta_t : t \in T\}$ are called

1. **equivalent in the wide sense**, if their finite dimensional distributions coincide.
2. **equivalent**, if they are defined on the same probability space and $\mathrm{P}(\xi_t = \eta_t) = 1$ holds for all $t \in T$. The equivalent processes are also called **modifications** of each other.
3. **indistinguishable**, if they are defined on the same probability space and $\mathrm{P}(\xi_t = \eta_t, \ \forall\, t \in T) = 1$.

Stochastic processes

1. If the stochastic processes $\{\xi_t : t \in T\}$ and $\{\eta_t : t \in T\}$ are equivalent (i.e., modifications of each other), then they are equivalent in the wide sense as well (i.e., their finite dimensional distributions coincide).
2. If the stochastic processes $\{\xi_t : t \in T\}$ and $\{\eta_t : t \in T\}$ are indistinguishable, then they are equivalent as well (i.e., modifications of each other).

Independent, stationary increments

A stochastic process $\{\xi_t : t \geqslant 0\}$ is said to have **independent increments**, if $\mathrm{P}(\xi_0 = 0) = 1$, and for any $k \in \mathbb{N}$ and any time points $0 \leqslant t_1 < t_2 < \ldots < t_k$, the increments $\xi_{t_1}, \xi_{t_2} - \xi_{t_1}, \ldots, \xi_{t_k} - \xi_{t_{k-1}}$ are (totally) independent. A stochastic process $\{\xi_t : t \geqslant 0\}$ is said to have **independent, stationary increments**, if it has independent increments, and the distribution of the increments is invariant with respect to time translation, i.e., for any time points $t, h \geqslant 0$, the distribution of $\xi_{t+h} - \xi_t$ does not depend on t (and consequently it coincides with the distribution of ξ_h).

Convolution of distribution functions

Let X and Y be independent (real valued) random variables with distribution functions F and G, respectively. Let H denote the distribution function of $X + Y$, which is called the **the convolution of the distribution functions** F and G, and it is denoted by $F * G$. Then

$$H(z) = \int_{-\infty}^{\infty} F(z - y)\,\mathrm{d}G(y), \quad z \in \mathbb{R}.$$

Finite dimensional distributions of processes with independent increments

The finite dimensional distributions of a stochastic process $\{\xi_t : t \geqslant 0\}$ with independent increments is uniquely determined by the distributions of the increments $\xi_t - \xi_s$, $0 \leqslant s < t$.

Finite dimensional distributions of processes with independent and stationary increments

The finite dimensional distributions of a stochastic process $\{\xi_t : t \geqslant 0\}$ with independent and stationary increments is uniquely determined by the distributions of the random variables ξ_t, $t \geqslant 0$ (i.e., by the one-dimensional distributions).

Further, for the family $\{F_{\xi_t} : t \geqslant 0\}$ of distribution functions, it holds that $F_{\xi_{s+t}} = F_{\xi_s} * F_{\xi_t}$ for all $s, t \geqslant 0$ (where $*$ denotes the convolution of distribution functions).

One-parameter convolutional semigroup of distribution functions

A family $\{F_t : t \geqslant 0\}$ of (one-dimensional) distribution functions is called a **one-parameter convolutional semigroup**, if $F_{s+t} = F_s * F_t$ for all $s, t \geqslant 0$, and $F_0 = \mathbb{1}_{(0,\infty)}$.

Stochastic processes

Expectation and covariance function

Let $T \neq \emptyset$ and $\{\xi_t : t \in T\}$ be a real valued stochastic process such that $\mathsf{E}(|\xi_t|) < \infty$, $t \in T$. Then the function $m : T \to \mathbb{R}$, $m(t) := \mathsf{E}(\xi_t)$, $t \in T$, is called the expectation function of the process. Further, if $\mathsf{E}(\xi_t^2) < \infty$, $t \in T$, then the function $K : T \times T \to \mathbb{R}$,

$$K(s, t) := \operatorname{Cov}(\xi_s, \xi_t), \qquad (s, t) \in T \times T,$$

is called the covariance function of the process.

Let $T \neq \emptyset$ and $\{\xi_t : t \in T\}$ be a real valued stochastic process such that $\mathsf{E}(\xi_t^2) < \infty$, $t \in T$. Then

1. $K(s, t) = K(t, s)$, $s, t \in T$ (i.e., K is symmetric),
2. $\forall\, k \in \mathbb{N}$, $\forall\, t_1, \ldots, t_k \in T$, $\forall\, \lambda_1, \ldots, \lambda_k \in \mathbb{C}$, we have
$$\sum_{i,j=1}^{k} \lambda_i \overline{\lambda_j} K(t_i, t_j) \geqslant 0.$$
Especially, for each $k \in \mathbb{N}$ and $t_1, \ldots, t_k \in T$, the matrix $(K(t_j, t_l))_{j,l=1,\ldots,k}$ is positive semidefinite.

Kolmogorov consistency and existence theorem

Let $\xi := \{\xi_t : t \in T\}$ be a real valued stochastic process, where T is a nonempty index set.

Let $\mathbb{R}^T := \{x \,|\, x : T \to \mathbb{R}\}$.

The stochastic process ξ can be also considered as a function which is defined on the sample space Ω, and it can take values in the space $\mathbb{R}^T$, namely

$$\xi : \Omega \to \mathbb{R}^T, \; \Omega \ni \omega \mapsto \xi(\omega),$$

where $\xi(\omega) : T \to \mathbb{R}$, $T \ni t \mapsto \xi(\omega)(t) := \xi_t(\omega)$.

It were convenient if ξ would be a random element of the space $\mathbb{R}^T$, i.e., if the function $\xi : \Omega \to \mathbb{R}^T$ would be measurable with respect to some appropriately defined measurable structure.

We furnish the space $\mathbb{R}^T$ with a σ-algebra denoted by $\sigma(\mathcal{C})$, with a σ-algebra generated by the so called cylinder sets.

For this, first we introduce the so called (finite dimensional) projections.

Kolmogorov consistency and existence theorem

Projections

Let T be a nonempty index set, $n \in \mathbb{N}$ and $S = \{s_1, \ldots, s_n\} \subset T$. A mapping $p_S : \mathbb{R}^T \to \mathbb{R}^S$,

$$(p_S(x))(s_i) := x(s_i), \qquad i = 1, \ldots, n, \qquad x \in \mathbb{R}^T,$$

is called the **projection** onto $\mathbb{R}^S$.
(The mapping $p_S(x)$ can be written in an abbreviated form $(x_{s_1}, \ldots, x_{s_n})$ as well, where x_{s_i} denotes the value $x(s_i)$ of the mapping x at the point s_i in an abbreviated form, where $i = 1, \ldots, n$.)

Product of measurable spaces

Let $(X_1, \mathcal{A}_1)$ and $(X_2, \mathcal{A}_2)$ be mesaurable spaces. The elements of the set

$$\mathcal{T} := \{A_1 \times A_2 : A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2\}$$

are called measurable rectangles, and the measurable space $(X_1 \times X_2, \sigma(\mathcal{T}))$ is called the product of measurable spaces $(X_1, \mathcal{A}_1)$ and $(X_2, \mathcal{A}_2)$. The σ-algebra $\sigma(\mathcal{T})$ is usually denoted by $\mathcal{A}_1 \times \mathcal{A}_2$ (or $\mathcal{A}_1 \otimes \mathcal{A}_2$).

Kolmogorov consistency and existence theorem

Product of measurable spaces, cylinder sets

The previous definition can be extended to the product of finitely many measurable spaces $(X_i, \mathcal{A}_i)$, $i = 1, \ldots, n$, as well in an obvious way.

If $(X_\alpha, \mathcal{A}_\alpha)_{\alpha \in T}$ are infinitely many measurable spaces, where T is an arbitrary (not necessarily finite) index set, then by their product we mean the measurable space $(X, \sigma(\mathcal{C}))$, where $X := \prod_{\alpha \in T} X_\alpha$ and $\sigma(\mathcal{C})$ is the σ-algebra generated by the so called cylinder sets.

By a **cylinder set**, we mean a set $C \subset X$ for which there exist $n \in \mathbb{N}$, $\alpha_1, \ldots, \alpha_n \in T$, $\alpha_i \neq \alpha_j$, if $i \neq j$, $i, j \in \{1, \ldots, n\}$, and $B \in \prod_{k=1}^n \mathcal{A}_{\alpha_k} := \mathcal{A}_{\alpha_1} \times \cdots \times \mathcal{A}_{\alpha_n}$ such that

$$C = \{x \in X : (x_{\alpha_1}, \ldots, x_{\alpha_n}) \in B\}.$$

The indices $\alpha_1, \ldots, \alpha_n$ are called the base points (coordinates) of C, and the set B is called a base set (a base) of C.
The collection of cylinder sets is denoted by $\mathcal{C}$.

Kolmogorov consistency and existence theorem

Example for a cylinder set

Let $T := \{1,2,3\}$ and $X_i := \mathbb{R}$, $i = 1,2,3$. Then for each $r > 0$, the set $\left\{(x_1, x_2, x_3) \in \mathbb{R}^3 : x_1^2 + x_2^2 \leqslant r^2\right\}$ is a cylinder set, since

$$\left\{(x_1, x_2, x_3) \in \mathbb{R}^3 : x_1^2 + x_2^2 \leqslant r^2\right\} = \left\{(x_1, x_2, x_3) \in \mathbb{R}^3 : (x_1, x_2) \in B\right\},$$

where B denotes the disk in the plane (x_1, x_2) having center as the origin and with radius r (including its boundary as well).

This cylinder set is nothing else but the cylinder which is rotation invariant with respect to the coordinate ax x_3 and has radius r.

This cylinder set can be also given in the form

$$\left\{(x_1, x_2, x_3) \in \mathbb{R}^3 : x_1^2 + x_2^2 \leqslant r^2\right\} = \left\{(x_1, x_2, x_3) \in \mathbb{R}^3 : (x_1, x_2, x_3) \in B \times \mathbb{R}\right\},$$

so, we can see that a cylinder set can be given in different forms.

Kolmogorov consistency and existence theorem

The collection of cylinder sets, denoted by $\mathcal{C}$, is an algebra.

Product measurability

A set $B \subset X$ is measurable, i.e., $B \in \sigma(\mathcal{C})$ holds if and only if there exist $(\alpha_k)_{k=1}^{\infty}$ and $\widetilde{B} \in \prod_{k=1}^{\infty} \mathcal{A}_{\alpha_k}$ such that $p^{-1}_{(\alpha_k)_{k=1}^{\infty}}(\widetilde{B}) = B$, where $p_{(\alpha_k)_{k=1}^{\infty}} : \prod_{\alpha \in T} X_\alpha \to \prod_{k=1}^{\infty} X_{\alpha_k}$, $p_{(\alpha_k)_{k=1}^{\infty}}(x) := (x_{\alpha_k})_{k=1}^{\infty}$, $x \in \prod_{\alpha \in T} X_\alpha$ (i.e., the set B depends „only on countably many coordinates").

In case of $X_\alpha := \mathbb{R}$, $\alpha \in T$, the previous result means picturesquely that a set $B \subset \mathbb{R}^T$ is $\sigma(\mathcal{C})$-measurable if and only if the functions belonging to B are defined in a way that their values are commonly given at countably many points, while they can take arbitrary values at other points.

After this it is meaningful to ask whether a mapping $\xi : \Omega \to \mathbb{R}^T$ is measurable or not.

Kolmogorov consistency and existence theorem

Measurability of a stochastic process

Let $\xi := \{\xi_t : t \in T\}$ be a real valued stochastic process, where T is a nonempty index set. Then $\xi : \Omega \to \mathbb{R}^T$ is measurable with respect to the measurable spaces $(\Omega, \mathcal{A})$ and $(\mathbb{R}^T, \sigma(\mathcal{C}))$.

Distribution of a stochastic process

By the distribution of a (real valued) stochastic process $\xi := \{\xi_t : t \in T\}$ (where T is a nonempty index set), we mean the following probability measure defined on the space $(\mathbb{R}^T, \sigma(\mathcal{C}))$:

$$\mathrm{P}_\xi(M) := \mathrm{P}(\xi \in M) = \mathrm{P}(\xi^{-1}(M)), \qquad M \in \sigma(\mathcal{C}).$$

Kolmogorov consistency and existence theorem

Connection between the distribution and finite dimensional distributions of a stochastic process

Let $\xi := \{\xi_t : t \in T\}$ be a (real valued) stochastic process, where T is a nonempty index set.

(i) Then the distribution of ξ uniquely determines the finite dimensional distributions of ξ.

(ii) If $\eta := \{\eta_t : t \in T\}$ is a (real valued) stochastic process such that its finite dimensional distributions coincide with those of ξ, then the distributions of ξ and η coincide, i.e., $\mathrm{P}_\xi = \mathrm{P}_\eta$.
Hence the finite dimensional distributions of a stochastic process uniquely determines its distribution on the space $(\mathbb{R}^T, \sigma(\mathcal{C}))$.

In what follows we investigate the question raised earlier: what is a minimal condition under which a family of probability distributions coincides with the family of the finite dimensional distributions of some stochastic process.

Kolmogorov consistency and existence theorem

Consistent family of probability measures

Let $T \neq \emptyset$ be an index set. Let

$$T^* := \Big\{(t_1, \dots, t_n) \in T^n : \ n \in \mathbb{N}, \ t_i \neq t_j, \ \text{ if } \ i \neq j, \ i, j \in \{1, \dots, n\}\Big\},$$

and for each $(t_1, \dots, t_n) \in T^*$, let us given a probability measure $P_{t_1,\dots,t_n}$ on the measurable space $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$.
The family $\{P_{t_1,\dots,t_n} : (t_1, \dots, t_n) \in T^*, \ n \in \mathbb{N}\}$ is called **consistent** if it satisfies the following two conditions:

(a) **permutation invariance**: if π is a permutation of $(1, 2, \dots, n)$, then for all Borel measurable sets $A_i \in \mathcal{B}(\mathbb{R}), i = 1, \dots, n,$ the probability measures $P_{t_1,\dots,t_n}$ and $P_{t_{\pi(1)},\dots,t_{\pi(n)}}$ satisfy the equation

$$P_{t_1,\dots,t_n}(A_1 \times A_2 \times \cdots \times A_n) = P_{t_{\pi(1)},\dots,t_{\pi(n)}}(A_{\pi(1)} \times A_{\pi(2)} \times \cdots \times A_{\pi(n)}),$$

(b) **compatibility condition**: for each $n \in \mathbb{N}$, $(t_1, \ldots, t_n, t_{n+1}) \in T^*$ and for each $A \in \mathcal{B}(\mathbb{R}^n)$, we have

$$P_{t_1,\ldots,t_n}(A) = P_{t_1,\ldots,t_n,t_{n+1}}(A \times \mathbb{R}).$$

The condition (a) means picturesquely that the measure of a rectangular cuboid does not depend on the order of its coordinates.

The condition (b) is a generalization of the principle „the volume of a prism is the product of the area of the base and the height". One can call it compatibility condition, since it is about a connection between probability measures on Euclidean spaces with different dimensions.

Example for a consistent family of probability measures

Let $T \neq \emptyset$, $f : \mathbb{R} \to [0, \infty)$ be a density function, and for each $(t_1, \ldots, t_n) \in T^*$, $n \in \mathbb{N}$, and $A \in \mathcal{B}(\mathbb{R}^n)$, let

$$P_{t_1,\ldots,t_n}(A) := \int_A f(x_1) \cdots f(x_n) \, \mathrm{d}x_1 \mathrm{d}x_2 \ldots \mathrm{d}x_n.$$

Kolmogorov consistency and existence theorem

Let $T \neq \emptyset$ be an index set. Let P be a probability measure on the product space $(\mathbb{R}^T, \sigma(\mathcal{C}))$, where $\sigma(\mathcal{C})$ denotes the σ-algebra generated by the cylinder sets. For each $(t_1, \ldots, t_n) \in T^*$, $n \in \mathbb{N}$, let

$$P_{t_1,\ldots,t_n}(A) := \mathrm{P}\left(\left\{x \in \mathbb{R}^T : (x_{t_1}, \ldots, x_{t_n}) \in A\right\}\right), \quad A \in \mathcal{B}(\mathbb{R}^n).$$

Then the family $\{P_{t_1,\ldots,t_n} : (t_1, \ldots, t_n) \in T^*,\ n \in \mathbb{N}\}$ consisting of probability measures is consistent.

Kolmogorov consistency theorem

Let $T \neq \emptyset$ be an index set, and for each $(t_1, \ldots, t_n) \in T^*$, $n \in \mathbb{N}$ let $P_{t_1,\ldots,t_n}$ be a probability measure on the measurable space $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. Let us suppose that the family $\{P_{t_1,\ldots,t_n} : (t_1, \ldots, t_n) \in T^*,\ n \in \mathbb{N}\}$ is consistent. Then there exists a unique probability measure P on the measurable space $(\mathbb{R}^T, \sigma(\mathcal{C}))$ (where $\sigma(\mathcal{C})$ is the σ-algebra generated by cylinder sets) such that for each $(t_1, \ldots, t_n) \in T^*$, $n \in \mathbb{N}$, we have $P_{t_1,\ldots,t_n}(A) = \mathrm{P}\left(\left\{x \in \mathbb{R}^T : (x_{t_1}, \ldots, x_{t_n}) \in A\right\}\right)$, $\forall\ A \in \mathcal{B}(\mathbb{R}^n)$.

Kolmogorov consistency and existence theorem

In the proof of Kolmogorov consistency theorem, the following results from measure theory play important roles.

Inner regularity of a measure

Let $X \neq \emptyset$, $\mathcal{A}$ be a set of some subsets of X, and $\mu : \mathcal{A} \to [0, \infty]$ be a function. We say that μ is **inner regular** with respect to a system $\mathcal{K} \subset \mathcal{A}$, if for each $A \in \mathcal{A}$, we have $\mu(A) = \sup\{\mu(K) : K \subset A, K \in \mathcal{K}\}$.

σ-compact family

A family $\mathcal{K}$ consisting of some subsets of $X \neq \emptyset$ is called σ**-compact**, if for each sequence $K_n \in \mathcal{K}$, $n \in \mathbb{N}$, satisfying $\bigcap_{n=1}^{\infty} K_n = \emptyset$, we can find an $N \in \mathbb{N}$ such that $\bigcap_{n=1}^{N} K_n = \emptyset$.

Let $X \neq \emptyset$, $\mathcal{A}$ be an algebra of some subsets of X, $\mu : \mathcal{A} \to [0, \infty)$ be a finitely additive function having finite values, and $\mathcal{K} \subset \mathcal{A}$ be a σ-compact family. If μ is inner regular with respect to the system $\mathcal{K} \subset \mathcal{A}$, then μ is σ-additive on the algebra $\mathcal{A}$.

Kolmogorov consistency and existence theorem

Kolmogorov existence theorem

Let $T \subset [0, \infty)$ be a nonempty set. For each $k \in \mathbb{N}$, $t_1, \dots, t_k \in T$, $t_1 < \cdots < t_k$, let $F_{t_1, t_2, \dots, t_k} : \mathbb{R}^k \to [0, 1]$ be a k-dimensional distribution function. Let us suppose that the family

$$\{F_{t_1, t_2, \dots, t_k} : k \in \mathbb{N},\ t_1, t_2, \dots, t_k \in T,\ t_1 < t_2 < \cdots < t_k\}$$

is compatible, i.e., for any $k \in \mathbb{N}$, $t_1, \dots, t_k \in T$, $\ell \in \{1, \dots, k\}$, and integers $1 \leqslant i_1 < i_2 < \ldots < i_\ell \leqslant k$, we have

$$\lim_{x_j \to \infty,\, j \notin \{i_1, \dots, i_\ell\}} F_{t_1, \dots, t_k}(x_1, \dots, x_k) = F_{t_{i_1}, \dots, t_{i_\ell}}(x_{i_1}, \dots, x_{i_\ell}), \quad \forall\ x_{i_1}, \dots, x_{i_\ell} \in \mathbb{R}.$$

Then there exist a probability space $(\Omega, \mathcal{A}, \mathrm{P})$ and a real valued stochastic process $\{\xi_t : t \in T\}$ on it such that for any $k \in \mathbb{N}$, $t_1, \dots, t_k \in T$, $t_1 < \cdots < t_k$, we have the distribution function of $\xi_{t_1}, \dots, \xi_{t_k}$ is $F_{t_1, \dots, t_k}$.

Existence of a stochastic process with independent and stationary increments corresponding to a given one-parameter convolution semigroup

Let $\{F_t : t \geqslant 0\}$ be a one-parameter convolution semigroup of distribution functions. Then there exist a probability space $(\Omega, \mathcal{A}, \mathrm{P})$ and a stochastic process $\{\xi_t : t \geqslant 0\}$ with independent and stationary increments on it such that $F_{\xi_t} = F_t$ for all $t \geqslant 0$.

Then, as we saw earlier, the finite dimensional distributions of $\{\xi_t : t \geqslant 0\}$ are uniquely determined by the family $\{F_t : t \geqslant 0\}$.