Probability

1. | Probabilistic models

σ-algebras

Definition 1 (Algebra). Let Ω be a set and $\mathcal{A} \subset \mathcal{P}(\Omega)$. We say that $\mathcal{A}$ is an *algebra* over Ω if:

1. $\Omega \in \mathcal{A}$.
2. If $A \in \mathcal{A}$, then $A^c \in \mathcal{A}$.
3. If $A, B \in \mathcal{A}$, then $A \cup B \in \mathcal{A}$.

Proposition 2. Let $\mathcal{A}$ be an algebra over a set Ω. Then:

1. $\varnothing \in \mathcal{A}$.
2. If $A, B \in \mathcal{A}$, then $A \cap B \in \mathcal{A}$.
3. For all $n \in \mathbb{N}$, if $A_1, \ldots, A_n \in \mathcal{A}$, then:
$$\bigcup_{i=1}^{n} A_i \in \mathcal{A} \quad \text{and} \quad \bigcap_{i=1}^{n} A_i \in \mathcal{A}$$

Definition 3 (σ-algebra). Let Ω be a set and $\mathcal{A} \subset \mathcal{P}(\Omega)$. We say that $\mathcal{A}$ is a *σ-algebra* over Ω if:

1. $\Omega \in \mathcal{A}$.
2. If $A \in \mathcal{A}$, then $A^c \in \mathcal{A}$.
3. If $A_1, A_2, \ldots \in \mathcal{A}$, then:
$$\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$$

Proposition 4. Let Ω be a set, I be an index set and $\{\mathcal{A}_i : i \in I\}$ be a collection of σ-algebras. Then, $\bigcap_{i\in I} \mathcal{A}_i$ is a σ-algebra.

Proposition 5. Let $\mathcal{A}$ be an σ-algebra over a set Ω. Then:

1. $\varnothing \in \mathcal{A}$.
2. If $A_1, A_2, \ldots \in \mathcal{A}$, then:
$$\bigcap_{n=1}^{\infty} A_n \in \mathcal{A}$$
3. For all $n \in \mathbb{N}$, if $A_1, \ldots, A_n \in \mathcal{A}$, then:
$$\bigcup_{i=1}^{n} A_i \in \mathcal{A} \quad \text{and} \quad \bigcap_{i=1}^{n} A_i \in \mathcal{A}$$

Definition 6. Let Ω be a set. The *trivial σ-algebra* is the smallest σ-algebra over Ω, that is, $\{\varnothing, \Omega\}$.

Definition 7. Let Ω be a set. The *discrete σ-algebra* is the largest σ-algebra over Ω, that is, $\mathcal{P}(\Omega)$.

Definition 8. Let Ω be a set and $A \subseteq \Omega$ be a subset. The *σ-algebra generated* by A, $\sigma(A)$, is the smallest σ-algebra over Ω containing A, that is:
$$\sigma(A) = \{\varnothing, \Omega, A, A^c\}$$

Definition 9. Let Ω be a set and $\mathcal{C} \subseteq \mathcal{P}(\Omega)$ be a subset. The *σ-algebra generated* by $\mathcal{C}$, $\sigma(\mathcal{C})$, is the smallest σ-algebra over Ω containing all the elements of $\mathcal{C}$. Moreover, if $\{\mathcal{A}_n : \mathcal{C} \subseteq \mathcal{A}_n, 1 \leq n \leq N\}$, $N \in \mathbb{N} \cup \{\infty\}$, are all the σ-algebras over Ω containing $\mathcal{C}$, then:
$$\sigma(\mathcal{C}) = \bigcap_{n=1}^{N} \mathcal{A}_n$$

Theorem 10. Let Ω be a set and $\mathcal{C}, \mathcal{B} \subseteq \mathcal{P}(\Omega)$ be subsets. Suppose:

1. $\mathcal{B}$ is a σ-algebra.
2. $\mathcal{C} \subseteq \mathcal{B}$.

Then, $\sigma(\mathcal{C}) \subseteq \mathcal{B}$.

Definition 11. Let $(\Omega, \mathcal{T})$ be a topological space. The *Borel σ-algebra* over $(\Omega, \mathcal{T})$, $\mathcal{B}((\Omega, \mathcal{T}))$, is the σ-algebra generated by the open sets of $(\Omega, \mathcal{T})$:
$$\mathcal{B}((\Omega, \mathcal{T})) := \sigma(\mathcal{T})$$
In particular, the Borel σ-algebra over $\mathbb{R}$ (together with the usual topology) is:
$$\mathcal{B}(\mathbb{R}) := \sigma(\{U \subseteq \mathbb{R} : U \text{ is open}\})$$

Proposition 12. Let $(\Omega, \mathcal{T})$ be a topological space. Then:
$$\mathcal{B}((\Omega, \mathcal{T})) = \sigma(\{C \subseteq \Omega : C \text{ is closed}\})$$

Proposition 13. Consider the Borel σ-algebra over $\mathbb{R}$, $\mathcal{B}(\mathbb{R})$. Then:

1. $\mathcal{B}(\mathbb{R}) = \sigma(\{(a, b) \subset \mathbb{R} : a, b \in \mathbb{R}, a < b\})$
2. $\mathcal{B}(\mathbb{R}) = \sigma(\{[a, b] \subset \mathbb{R} : a, b \in \mathbb{R}, a < b\})$
3. $\mathcal{B}(\mathbb{R}) = \sigma(\{[a, b) \subset \mathbb{R} : a, b \in \mathbb{R}, a < b\})$
4. $\mathcal{B}(\mathbb{R}) = \sigma(\{(a, b] \subset \mathbb{R} : a, b \in \mathbb{R}, a < b\})$
5. $\mathcal{B}(\mathbb{R}) = \sigma(\{(a, \infty) \subset \mathbb{R} : a \in \mathbb{R}\})$
6. $\mathcal{B}(\mathbb{R}) = \sigma(\{(-\infty, a) \subset \mathbb{R} : a \in \mathbb{R}\})$
7. $\mathcal{B}(\mathbb{R}) = \sigma(\{[a, \infty) \subset \mathbb{R} : a \in \mathbb{R}\})$
8. $\mathcal{B}(\mathbb{R}) = \sigma(\{(-\infty, a] \subset \mathbb{R} : a \in \mathbb{R}\})$

Probability

Definition 14 (Sample space). The *sample space* Ω of an experiment is the set of all possible outcomes of that experiment.

Definition 15 (Kolmogorov axioms). Let Ω be a set and $\mathcal{A}$ be a σ-algebra over Ω. A *probability* is any function

$$\mathbb{P} : \mathcal{A} \longrightarrow [0, \infty)$$

satisfying the following properties:

- $\mathbb{P}(\Omega) = 1$.
- *σ-additivity*: If $\{A_n : n \geq 1\} \subset \mathcal{A}$ are pairwise disjoint, then:

$$\mathbb{P}\left(\bigsqcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n)$$

Definition 16. Let Ω be a set and $\mathcal{A}$ be a σ-algebra over Ω. An *event* $A \in \mathcal{A}$ is a subset of Ω for which we want to calculate the probability.

Definition 17. A *probability space* is a triplet $(\Omega, \mathcal{A}, \mathbb{P})$ where Ω is any set, $\mathcal{A}$ is a σ-algebra over Ω and $\mathbb{P}$ is a probability over $\mathcal{A}$.

Proposition 18. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $A, B \in \mathcal{A}$. Then, we have the following properties:

1. $\mathbb{P}(\varnothing) = 0$.
2. If $A_i \in \mathcal{A}$, $i = 1, \ldots, n$, is a finite set of pairwise disjoint events, then:

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

3. $\mathbb{P}(A \setminus B) = \mathbb{P}(A) - \mathbb{P}(A \cap B)$.
4. If $B \subset A$, then $\mathbb{P}(A \setminus B) = \mathbb{P}(A) - \mathbb{P}(B)$.
5. If $B \subset A$, then $\mathbb{P}(B) \leq \mathbb{P}(A)$.
6. $\mathbb{P}(A) \leq 1$.
7. $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.
8. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.
9. If $A_1, \ldots, A_n \in \mathcal{A}$, then:

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathbb{P}(A_i) - \\ - \sum_{\substack{i,j=1 \\ i<j}}^{n} \mathbb{P}(A_i \cap A_j) + \sum_{\substack{i,j,k=1 \\ i<j<k}}^{n} \mathbb{P}(A_i \cap A_j \cap A_k) - \cdots + \\ + (-1)^{n+1} \mathbb{P}(A_1 \cap \cdots \cap A_n)$$

10. If $A_1, \ldots, A_n \in \mathcal{A}$, then:

$$\mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathbb{P}(A_i) - \\ - \sum_{\substack{i,j=1 \\ i<j}}^{n} \mathbb{P}(A_i \cup A_j) + \sum_{\substack{i,j,k=1 \\ i<j<k}}^{n} \mathbb{P}(A_i \cup A_j \cap A_k) - \cdots + \\ + (-1)^{n+1} \mathbb{P}(A_1 \cup \cdots \cup A_n)$$

11. *Finite subadditivity*: If $A_1, \ldots, A_n \in \mathcal{A}$, then:

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

Proposition 19. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space such that Ω is finite and all its elements are equiprobable. Let $A \in \mathcal{A}$ be an event. Then:

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

Theorem 20 (Continuity from below). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $(A_n) \subset \mathcal{A}$ be an increasing sequence of events, that is:

$$A_1 \subseteq A_2 \subseteq \cdots \subseteq A_n \subseteq \cdots$$

Let $A := \bigcup_{n=1}^{\infty} A_n$. Then:

$$\mathbb{P}(A) = \lim_{n \to \infty} \mathbb{P}(A_n)$$

Corollary 21 (Continuity from above). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $(A_n) \subset \mathcal{A}$ be a decreasing sequence of events, that is:

$$A_1 \supseteq A_2 \supseteq \cdots \supseteq A_n \supseteq \cdots$$

Let $A := \bigcap_{n=1}^{\infty} A_n$. Then:

$$\mathbb{P}(A) = \lim_{n \to \infty} \mathbb{P}(A_n)$$

Proposition 22 (Countable subadditivity). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $(A_n) \subset \mathcal{A}$ be a sequence of events. Then:

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mathbb{P}(A_n)$$

Corollary 23. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $(A_n) \subset \mathcal{A}$ be a sequence of events with probability 0. Then:

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = 0$$

Corollary 24. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $(A_n) \subset \mathcal{A}$ be a sequence of events with probability 1. Then:

$$\mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n\right) = 1$$

Conditional probability

Definition 25. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $A \in \mathcal{A}$ be an event such that $\mathbb{P}(A) > 0$. The *conditional probability* that $B \in \mathcal{A}$ occurs given that A occurs is defined as:

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

Proposition 26. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $A \in \mathcal{A}$ be an event such that $\mathbb{P}(A) > 0$. Then, the function

$$\begin{aligned} \mathbb{P}(\cdot \mid A) : \mathcal{A} &\longrightarrow [0, \infty] \\ B &\longmapsto \mathbb{P}(B \mid A) \end{aligned}$$

is a probability.

Proposition 27 (Compound probability formula). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $A \in \mathcal{A}$ be an event such that $\mathbb{P}(A) > 0$. Then, $\forall B \in \mathcal{A}$:

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

Proposition 28 (Generalized compound probability formula). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $A_1, \ldots, A_n \in \mathcal{A}$, $n \geq 2$, be events such that $\mathbb{P}(A_1 \cap \cdots \cap A_{n-1}) > 0$. Then:

$$\begin{aligned} \mathbb{P}(A_1 \cap \cdots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2 \mid A_1)\mathbb{P}(A_3 \mid A_2 \cap A_1) \cdots \\ \cdots \mathbb{P}(A_n \mid A_1 \cap \cdots \cap A_{n-1}) \end{aligned}$$

Definition 29. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $A = \{A_n : 1 \leq n \leq N\} \subset \mathcal{A}$, $N \in \mathbb{N} \cup \{\infty\}$, be a collection of events. We say that A is a *partition* of Ω if:

$$\Omega = \bigsqcup_{n=1}^{N} A_n$$

Proposition 30 (Law of total probability). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $\{A_n : 1 \leq n \leq N\} \subset \mathcal{A}$, $N \in \mathbb{N} \cup \{\infty\}$, be a partition of Ω such that $\mathbb{P}(A_n) > 0$ for all $1 \leq n \leq N$. Then, $\forall A \in \mathcal{A}$:

$$\mathbb{P}(A) = \sum_{n=1}^{N} \mathbb{P}(A_n)\mathbb{P}(A \mid A_n)$$

Proof.

$$\begin{aligned} \mathbb{P}(A) = \mathbb{P}(A \cap \Omega) = \mathbb{P}\left(\bigsqcup_{n=1}^{N} (A \cap A_n)\right) = \\ = \sum_{n=1}^{N} \mathbb{P}(A \cap A_n) = \sum_{n=1}^{N} \mathbb{P}(A_n)\mathbb{P}(A \mid A_n) \end{aligned}$$

□

Proposition 31 (Bayes' formula). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $\{A_n : 1 \leq n \leq N\} \subset \mathcal{A}$, $N \in \mathbb{N} \cup \{\infty\}$, be a partition of Ω such that $\mathbb{P}(A_n) > 0$ for all $1 \leq n \leq N$. Let $A \in \mathcal{A}$ with $\mathbb{P}(A) > 0$. Then, $\forall k \leq N$:

$$\mathbb{P}(A_k \mid A) = \frac{\mathbb{P}(A_k)\mathbb{P}(A \mid A_k)}{\sum_{n=1}^{N} \mathbb{P}(A_n)\mathbb{P}(A \mid A_n)}$$

Independence of events

Definition 32. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. We say that $A, B \in \mathcal{A}$ are *independent events* if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$$

Proposition 33. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. Then:

1. $\varnothing$ and Ω are independent of any event.
2. If $A \in \mathcal{A}$ satisfies either $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$, then A is independent of any other event $B \in \mathcal{A}$.
3. If an event $A \in \mathcal{A}$ is independent of itself, then either $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$.

Proposition 34. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $A, B \in \mathcal{A}$ be two events. The following statements are equivalent:

- A and B are independent.
- A^c and B are independent.
- A and B^c are independent.
- A^c and B^c are independent.

Definition 35. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $n \in \mathbb{N}$. We say that $A_1, \ldots, A_n \in \mathcal{A}$ are *independent events* if for any $i_1, \ldots, i_k \in \{1, \ldots, n\}$, we have:

$$\mathbb{P}\left(\bigcap_{r=1}^{k} A_{i_r}\right) = \prod_{r=1}^{k} \mathbb{P}(A_{i_r})$$

Definition 36. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and I be an arbitrary index set. We say that $\{A_i : i \in I\} \subset \mathcal{A}$ are *independent events* if for any finite subset $\{A_{i_1}, \ldots, A_{i_k} : i_r \in I \text{ for } r = 1, \ldots, k\}$ of different events, we have:

$$\mathbb{P}\left(\bigcap_{r=1}^{k} A_{i_r}\right) = \prod_{r=1}^{k} \mathbb{P}(A_{i_r})$$

2. | Lebesgue integration

Definition 37. Let $A \subset \mathbb{R}^n$ be a subset. Then, A is a *null set* (or a *set of zero-content*) if $\forall \varepsilon > 0$ there exists a collection $\{R_k \subset \mathbb{R}^n : R_k \text{ is a rectangle } \forall k \in \mathbb{N}\}$ of rectangles such that:

$$A \subset \bigcup_{k=1}^{\infty} R_k \quad \text{and} \quad \sum_{k=1}^{\infty} \text{vol}(R_k) < \varepsilon$$

Definition 38. Let E be a set and $\mathcal{E}$ be a σ-algebra over E. We say that the function:

$$\begin{aligned} \mu : \mathcal{E} &\longrightarrow [0, \infty] \\ A &\longmapsto \mu(A) \end{aligned}$$

is a *measure* if:

1. There exists $A \in \mathcal{E}$ such that $\mu(A) < \infty$.

2. If $\{A_n \in \mathcal{E} : n \in \mathbb{N}\}$ is a collection of pairwise disjoint sets, then:

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n)$$

The triplet $(E, \mathcal{E}, \mu)$ is called a *measure space*.

Definition 39. The *σ-algebra of all Lebesgue measurable sets in $\mathbb{R}^n$*, $\mathcal{L}_n \subset \mathcal{P}(\mathbb{R}^n)$, is defined as:

$$\mathcal{L}_n := \{A \subseteq \mathbb{R}^n : A = B \cup N\}$$

where $B \in \mathcal{B}(\mathbb{R}^n)$ and N is a null set.

Theorem 40. We can extend the concept of volume on rectangles in $\mathbb{R}^n$ to all the elements in $\mathcal{L}_n$. This extension is called *Lebesgue measure* (or simply *volume*) in $\mathbb{R}^n$.

Definition 41. Let $(E, \mathcal{E}, \mu)$ be a measure space and $f : E \to \mathbb{R}$ be a function. We say that f is *measurable* if $\forall B \in \mathcal{B}(\mathbb{R})$ we have $f^{-1}(B) \in \mathcal{E}$. The *Lebesgue integral* of f over E with respect to μ is denoted by:

$$\int_E f \mathrm{d}\mu$$

Proposition 42. Let $(E, \mathcal{E}, \mu)$ be a measure space and $f : E \to \mathbb{R}$ be a measurable function such that $f(x) \geq 0$ $\forall x \in E$. Then, we can always define the integral

$$\int_E f \mathrm{d}\mu$$

taking into account that may be $+\infty$.

Definition 43. Let $(E, \mathcal{E}, \mu)$ be a measurable space and $f : E \to \mathbb{R}$ be a measurable function. We say that f is *Lebesgue integrable* with respect to μ if:

$$\int_E |f| \mathrm{d}\mu < \infty$$

Moreover if $G \in \mathcal{E}$, we define:

$$\int_G f \mathrm{d}\mu := \int_E f \mathbf{1}_G \mathrm{d}\mu$$

Proposition 44. Consider the measurable space $(\mathbb{R}^n, \mathcal{L}_n, m_n)$, where m_n is the volume in $\mathbb{R}^n$. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a Riemann integrable function satisfying:

$$\int_{\mathbb{R}^n} |f(x_1, \ldots, x_n)| \, \mathrm{d}x_1 \cdots \mathrm{d}x_n < \infty$$

Then, f is Lebesgue integrable and:

$$\int_{\mathbb{R}^n} |f(x_1, \ldots, x_n)| \, \mathrm{d}x_1 \cdots \mathrm{d}x_n = \int_{\mathbb{R}^n} f \, \mathrm{d}m_n$$

Theorem 45 (Tonelli's theorem). Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a non-negative Lebesgue measurable function. Then:

$$\begin{aligned}\int_{\mathbb{R}^2} f(x,y) \, \mathrm{d}x \, \mathrm{d}y &= \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} f(x,y) \, \mathrm{d}x \right) \mathrm{d}y \\ &= \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} f(x,y) \, \mathrm{d}y \right) \mathrm{d}x\end{aligned}$$

Theorem 46 (Fubini's theorem). Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a Lebesgue measurable function such that at least one of the following integrals is finite.

$$\int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} |f(x,y)| \, \mathrm{d}x \right) \mathrm{d}y$$

$$\int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} |f(x,y)| \, \mathrm{d}y \right) \mathrm{d}x$$

Then, f is Lebesgue integrable and:

$$\begin{aligned}\int_{\mathbb{R}^2} f(x,y) \, \mathrm{d}x \, \mathrm{d}y &= \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} f(x,y) \, \mathrm{d}x \right) \mathrm{d}y \\ &= \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} f(x,y) \, \mathrm{d}y \right) \mathrm{d}x\end{aligned}$$

3. | Random variables and random vectors

Random variables

Definition 47. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. A *real random variable* (or simply *random variable*) X is a function $X : \Omega \to \mathbb{R}$ satisfying for all $B \in \mathcal{B}(\mathbb{R})$:

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

Proposition 48. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $\mathcal{C}$ be a collection of subsets of $\mathbb{R}$ such that $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{C})$ and let $X : \Omega \to \mathbb{R}$ be a function. Then, X is a random variable if and only if $X^{-1}(B) \in \mathcal{A}$, $\forall B \in \mathcal{C}$.

Definition 49. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $a, b \in \mathbb{R}$ and $B \in \mathcal{B}(\mathbb{R})$. We define the following set:

$$\{X \in B\} := \{\omega \in \Omega : X(\omega) \in B\} = X^{-1}(B)$$

In particular:

$$\begin{aligned}\{X \leq a\} &:= \{\omega \in \Omega : X(\omega) \leq a\} = X^{-1}((-\infty, a]) \\ \{X > b\} &:= \{\omega \in \Omega : X(\omega) > b\} = X^{-1}((b, \infty)) \\ \{a < X \leq b\} &:= \{\omega \in \Omega : a < X(\omega) \leq b\} = X^{-1}([b, a)) \\ \{X = a\} &:= \{\omega \in \Omega : X(\omega) = a\} = X^{-1}(\{a\})\end{aligned}$$

Proposition 50. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, X, Y be random variables and $a \in \mathbb{R}$. Then:

1. $X + Y$ is also a random variable.

2. aX is also a random variable.

3. XY is also a random variable.

4. $\frac{1}{X}$ is also a random variable if $X(\omega) \neq 0\ \forall \omega \in \Omega$.

Proposition 51. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and (X_n) be a sequence of random variables. Then, the following quantities are also random variables provided that they are finite for all $\omega \in \Omega$:

1. $\sup X_n$

2. $\inf X_n$

3. $\limsup_{n\to\infty} X_n$

4. $\liminf_{n\to\infty} X_n$

Corollary 52. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and (X_n) be a sequence of random variables such that $\forall \omega \in \Omega$ the following limit exists and it is finite:

$$X(\omega) := \lim_{n\to\infty} X_n(\omega)$$

Then, X is a random variable.

Distribution of a random variable

Definition 53. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. The *distribution* of a random variable X is the function:

$$\begin{aligned} \mathbb{P}_X : \mathcal{B}(\mathbb{R}) &\longrightarrow [0,1] \\ B &\longmapsto \mathbb{P}(\{X \in B\}) \end{aligned}$$ [1]

Proposition 54. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. Then, for any random variable X, the function $\mathbb{P}_X$ is a probability over $\mathcal{B}(\mathbb{R})$. Hence, $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_X)$ is a probability space.

Definition 55. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. We say that two random variables X, Y are *equal in distribution* (denoted by $X \overset{\mathrm{d}}{=} Y$) if they satisfy:

$$\mathbb{P}_X(B) = \mathbb{P}_Y(B) \qquad \forall B \in \mathcal{B}(\mathbb{R})$$

That is, $X \overset{\mathrm{d}}{=} Y$ if they have the same distribution functions.

Definition 56. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. We say that two random variables X, Y are *equal almost surely* (denoted by $X \overset{\text{a.s.}}{=} Y$) if $\mathbb{P}(X = Y) = 1$, or equivalently, if $\mathbb{P}(X \neq Y) = 0$.

Proposition 57. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and X, Y be two random variables such that $X \overset{\text{a.s.}}{=} Y$. Then, $X \overset{\mathrm{d}}{=} Y$.

Definition 58 (Cumulative distribution function). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and X be a random variable. We define the *cumulative distribution function* (*cdf*) as:

$$\begin{aligned} F_X : \mathbb{R} &\longrightarrow [0,1] \\ x &\longmapsto \mathbb{P}(X \leq x) = \mathbb{P}_X((-\infty, x]) \end{aligned}$$

Theorem 59. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, X be a random variable and F_X be its cdf. Then:

1. If $x < y$, then $F_X(x) \leq F_X(y)$.
2. F_X is *càdlàg*[2].
3. $\lim_{x\to-\infty} F_X(x) = 0$ and $\lim_{x\to\infty} F_X(x) = 1$.

Reciprocally, if there is a function F satisfying these properties[3], then there exists a random variable X on $(\Omega, \mathcal{A}, \mathbb{P})$ such that F is its cdf.

Proposition 60. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, X be a random variable and F_X be its cdf. Then:

1. F_X has at most a countable number of discontinuities.
2. $\forall x, y \in \mathbb{R}$ such that $s < t$, we have:
$$\begin{aligned} \mathbb{P}(x < X \leq y) &= F(y) - F(x) \\ \mathbb{P}(x < X < t) &= \lim_{t\to y^-} F_X(y) - F(x) \\ \mathbb{P}(x \leq X \leq y) &= F(y) - \lim_{t\to x^-} F_X(x) \\ \mathbb{P}(x \leq X < y) &= \lim_{t\to y^-} F_X(y) - \lim_{t\to x^-} F(x) \end{aligned}$$
3. $\forall x \in \mathbb{R}$, $\mathbb{P}(X < x) = \lim_{t\to x^-} F_X(t)$.
4. For all $x \in \mathbb{R}$:
$$\mathbb{P}(X = x) = F_X(x) - \lim_{t\to x^-} F_X(t)$$
Hence, F_X is discontinuous at $x \iff \mathbb{P}(X = x) > 0$.

Theorem 61. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. Then, the cdf completely determine a distribution of a random variable X. That is, if X and Y are random variables such that $F_X(t) = F_Y(t)\ \forall t \in \mathbb{R}$, then $X \overset{\mathrm{d}}{=} Y$.

Discrete random variables

Definition 62. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. We say that a random variable X is *discrete* if there exists a finite or countable set $S \subset \mathbb{R}$ such that $\mathbb{P}(X \in S) = 1$[4]. In that case, S is called the *support* of X[5].

Definition 63 (Probability mass function). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and X be a discrete random variable with support points $S_X = \{x_i : i \in I\}$, where $I \subseteq \mathbb{N}$. The *probability mass function* (*pmf*) of the random variable X is:

$$\begin{aligned} p_X : S_X &\longrightarrow [0,1] \\ x_i &\longmapsto \mathbb{P}(X = x_i) \end{aligned}$$

[1] From now on, in order to simplify the notation, we will write $\mathbb{P}(X \in B) := \mathbb{P}(\{X \in B\})$.
[2] From French "continue à droite, limite à gauche" (right continuous with left limits).
[3] Such kind of functions are called *distribution functions*.
[4] By agreement, we will suppose that S only contains points x such that $\mathbb{P}(X = x) > 0$.
[5] In general we will denote S by S_X.

Proposition 64. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, X be a discrete random variable with support points $S_X = \{x_i : i \in I\}$ and p_X be its pmf. Then:

1. $p_X(x_i) > 0\ \forall i \in I$.
2. $\displaystyle\sum_{i \in I} p_X(x_i) = 1$.
3. $\forall B \in \mathcal{B}(\mathbb{R})$, we have:
$$\mathbb{P}(X \in B) = \sum_{i \in I : x_i \in B} p_X(x_i)$$

Corollary 65. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, X be a discrete random variable with support points $S_X = \{x_i : i \in I\}$, F_X be its cdf and p_X be its pmf. Then $\forall x \in \mathbb{R}$ we have:
$$F_X(x) = \mathbb{P}(X \leq x) = \sum_{i \in I : x_i \leq x} p_X(x_i)$$

Definition 66 (Degenerated distribution). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. The *degenerated distribution* consists in taking a constant random variable X so that
$$\mathbb{P}(X = a) = 1$$
for some $a \in \mathbb{R}$. Here we have $S_X = \{a\}$.

Definition 67 (Bernoulli distribution). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. The *Bernoulli distribution* is the one whose random variable X can only take two values (1 and 0)[6] with probabilities p and $q := 1 - p$:
$$\mathbb{P}(X = 0) = p \qquad \mathbb{P}(X = 1) = q$$
Here we have $S_X = \{0, 1\}$. If X follows a Bernoulli distribution of parameter p, we will write $X \sim \mathrm{Ber}(p)$.

Definition 68 (Discrete uniform distribution). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. The *discrete uniform distribution* is the one whose random variable X takes values on $S_X = \{x_1, \ldots, x_n\}$ each of these with probability $1/n$:
$$\mathbb{P}(X = x_i) = \frac{1}{n} \qquad \forall i = 1, \ldots, n$$
If X follows a discrete uniform distribution, we will write $X \sim U(\{x_1, \ldots, x_n\})$. The probability space $(S, \mathcal{P}(S), \mathbb{P}_X)$ is an *equiprobable space*.

Definition 69 (Binomial distribution). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $A \in \mathcal{A}$. Suppose $\mathbb{P}(A) = p$. The *binomial distribution* is the one whose random variable X is the number of successes of A in a sequence of n repetitions. Thus, $S_X = \{0, 1, \ldots, n\}$ and:
$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k} \qquad \forall k = 0, 1, \ldots, n$$
If X follows a binomial distribution of parameters n and p, we will write $X \sim \mathrm{B}(n, p)$[7].

Definition 70 (Poisson distribution). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $\lambda \in \mathbb{R}_{>0}$. The *Poisson distribution* of parameter λ is the one whose random variable X has support $S_X = \mathbb{N} \cup \{0\}$ and:
$$\mathbb{P}(X = k) = \mathrm{e}^{-\lambda} \frac{\lambda^k}{k!} \qquad \forall k \in \mathbb{N} \cup \{0\}$$
If X follows a Poisson distribution of parameter λ, we will write $X \sim \mathrm{Pois}(\lambda)$.

Theorem 71. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. Let $(p_n) \subset (0, 1)$ be a sequence such that:
$$\lim_{n \to \infty} n p_n = \lambda > 0$$
For each $n \geq 1$, consider $X_n \sim \mathrm{B}(n, p_n)$. Then, $\forall k \in \mathbb{N} \cup \{0\}$ we have:
$$\lim_{n \to \infty} \mathbb{P}(X_n = k) = \lim_{n \to \infty} \binom{n}{k} {p_n}^k (1 - p_n)^{n-k} = \mathrm{e}^{-\lambda} \frac{\lambda^k}{k!}$$

Corollary 72. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and suppose $n \in \mathbb{N}$ and $p \in (0, 1)$ are such that $n \gg 1$ and $p \ll 1$. Then, $\mathrm{B}(n, p) \simeq \mathrm{Pois}(np)$[8].

Definition 73 (Geometric distribution). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $A \in \mathcal{A}$. Suppose $\mathbb{P}(A) = p$. The *geometric distribution* is the one whose random variable X is the number of trials needed to get one success. Thus, $S_X = \mathbb{N}$ and:
$$\mathbb{P}(X = k) = (1 - p)^{k-1} p \qquad \forall k \in \mathbb{N}$$
If X follows a geometric distribution of parameter p, we will write $X \sim \mathrm{Geo}(p)$.

Definition 74 (Discrete memorylessness property). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and X be a discrete random variable whose support is $\mathbb{N}$ and such that $\mathbb{P}(X > m) > 0\ \forall m \in \mathbb{N}$. The distribution of X is *memoryless* if $\forall m, n \in \mathbb{N}$, we have:
$$\mathbb{P}(X > m + n \mid X > m) = \mathbb{P}(X > n)$$

Proposition 75. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, X be a discrete random variable and $p \in (0, 1)$. Then, $X \sim \mathrm{Geo}(p)$ if and only if the distribution of X is memoryless.

Definition 76 (Hypergeometric distribution). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. Suppose we have a population of size N of whom K have a special feature (success). Let X be the random variable that counts the number of successes that we have obtained in n draws (without replacement). Thus, the support of X is:
$$S_X = \{\max\{n + K - N, 0\}, \ldots, \min\{n, K\}\}$$
And the pmf is given by:
$$\mathbb{P}(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$
This type of distribution is called *hypergeometric distribution* and it is denoted by $X \sim \mathrm{HG}(N, p, n)$, where $p = \frac{K}{N}$ is the proportion of successes in the population.

[6] Also called *success/true* or *failure/false*, respectively.

[7] Note that, a Bernoulli distribution of parameter p may be considered as a Binomial distribution of parameters $n = 1$ and p. Hence, $\mathrm{Ber}(p) = \mathrm{B}(1, p)$.

[8] In practice, the approximation is good enough for $n \geq 10$ and $p \leq 0.05$.

Theorem 77. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $X \sim \mathrm{HG}(N, p, n)$ such that when $N \to \infty$, p remains constant. Then:

$$\lim_{N\to\infty} \mathbb{P}(X = k) = \lim_{N\to\infty} \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}} = \binom{n}{k} p^k (1-p)^{n-k}$$

which is the pmf of a binomial distribution $\mathrm{B}(n, p)$.

Definition 78 (Negative binomial distribution). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $A \in \mathcal{A}$. Suppose $\mathbb{P}(A) = p$. The *negative binomial distribution* is the one whose random variable X is the number of trials needed to get $r \geq 1$ successes. Thus, $S_X = \{r, r+1, \ldots\}$ and:

$$\mathbb{P}(X = k) = \binom{k-1}{r-1} p^r (1-p)^{k-r} \qquad \forall k \geq r$$

If X follows a negative binomial distribution of parameters r and p, we will write $X \sim \mathrm{NB}(r, p)$.

Absolutely continuous random variables

Definition 79. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. We say that a random variable X is *absolutely continuous* if there exists a function $f : \mathbb{R} \to \mathbb{R}$ satisfying:

1. $f(x) \geq 0$, $\forall x \in \mathbb{R}$.
2. f is integrable over $\mathbb{R}$ and:
$$\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x = 1$$
3. For all $a, b \in \mathbb{R} \cup \{\pm\infty\}$ with $a \leq b$, we have:
$$\mathbb{P}(a \leq X \leq b) = \int_a^b f(x)\,\mathrm{d}x$$

The function f, denoted by f_X, is called *probability density function* (*pdf*) of X. In general, a function satisfying the first two properties is called a *density function*.

Proposition 80. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, X be an absolutely continuous random variable and F_X be its cdf. Then:

1. $\mathbb{P}(X = a) = 0, \quad \forall a \in \mathbb{R}$.
2. $\mathbb{P}(X \in B) = \displaystyle\int_B f_X(x)\,\mathrm{d}x\,, \quad \forall B \in \mathcal{B}(\mathbb{R})$.
3. $F_X(b) = \mathbb{P}(X \leq b) = \displaystyle\int_{-\infty}^b f_X(x)\,\mathrm{d}x\,, \quad \forall b \in \mathbb{R}$.
4. F_X is continuous on $\mathbb{R}$.
5. If $a, b \in \mathbb{R}$ are such that $a < b$, then:
$$\begin{aligned}\mathbb{P}(a < X < b) &= \mathbb{P}(a \leq X < b) = \\ &= \mathbb{P}(a < X \leq b) = \mathbb{P}(a \leq X \leq b)\end{aligned}$$

Definition 81 (Continuous uniform distribution). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. We say that an absolutely continuous random variable X follows a *continuous uniform distribution* on (a, b) (also $[a, b]$), and we denoted it by $X \sim U(a, b)$, if X has the pdf

$$f_X(x) = \frac{1}{b-a}\mathbf{1}_{(a,b)}(x)$$

where $\mathbf{1}_{(a,b)}$ is the indicator function. Therefore, its cdf is:

$$F_X(x) = \frac{x-a}{b-a}\mathbf{1}_{(a,b)}(x) + \mathbf{1}_{[b,\infty)}(x)$$

Definition 82 (Exponential distribution). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. We say that an absolutely continuous random variable X follows an *exponential distribution* of parameter $\lambda > 0$, and we denoted it by $X \sim \mathrm{Exp}(\lambda)$, if X has the pdf:

$$f_X(x) = \lambda \mathrm{e}^{-\lambda x}\mathbf{1}_{(0,\infty)}$$

Furthermore, its cdf is:

$$F_X(x) = (1 - \mathrm{e}^{-\lambda x})\mathbf{1}_{(0,\infty)}(x)$$

Definition 83 (Continuous memorylessness property). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and X be an absolutely continuous random variable such that $\mathbb{P}(X > s) > 0$ $\forall s \in \mathbb{R}_{\geq 0}$. The distribution of X is *memoryless* if $\forall s, t \in \mathbb{R}_{\geq 0}$, we have:

$$\mathbb{P}(X > s + t \mid X > s) = \mathbb{P}(X > t)$$

Proposition 84. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, X be an absolutely continuous random variable and $\lambda \in \mathbb{R}_{>0}$. Then, $X \sim \mathrm{Exp}(\lambda)$ if and only if the distribution of X is memoryless.

Definition 85 (Standard normal distribution). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. We say that an absolutely continuous random variable Z follows a *standard normal distribution*, and we denoted it by $Z \sim N(0, 1)$, if Z has the pdf:

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

Definition 86 (Normal distribution). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $\mu \in \mathbb{R}$ and $\sigma \in \mathbb{R}_{>0}$. We say that an absolutely continuous random variable X follows a *normal distribution*, and we denoted it by $X \sim N(\mu, \sigma^2)$, if X has the pdf:

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

μ is called the *mean* or *expectation* of X; σ^2, its *variance*, and σ, its *standard deviation*.

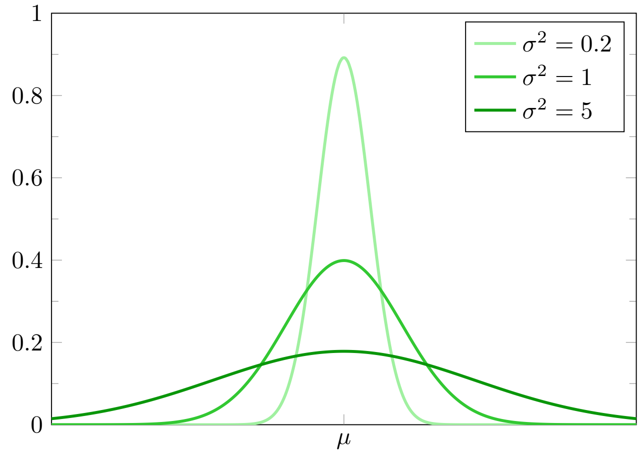

Figure 1: Probability density function of a normal distribution

Proposition 87. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $X \sim N(\mu, \sigma^2)$ and $Z \sim N(0,1)$ be absolutely continuous random variables. Then:

$$\mu + \sigma Z \stackrel{\mathrm{d}}{=} X$$

Therefore, $\forall x \in \mathbb{R}$ we have:

$$\mathbb{P}(X \leq x) = \mathbb{P}\left(Z \leq \frac{x-\mu}{\sigma}\right)$$

In this case, Z is called the *standardized form* of X.

Definition 88 (Gamma distribution). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. We say that an absolutely continuous random variable X follows a *gamma distribution* of parameters $\alpha, \beta \in \mathbb{R}_{>0}$, and we denoted it by $X \sim \mathrm{Gamma}(\alpha, \beta)$, if X has the pdf:

$$f_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \mathrm{e}^{-\beta x} \mathbf{1}_{(0,\infty)}(x)$$

The parameter α is called *shape*; β, *rate*, and $1/\beta$, *scale*.

Definition 89. Let $a, b \in \mathbb{R}_{>0}$. The *beta function* is defined as[9]:

$$\mathrm{B}(a,b) := \int_0^1 x^{a-1}(1-x)^{b-1} \, \mathrm{d}x$$

Proposition 90. For all $a, b \in \mathbb{R}_{>0}$, we have:

$$\mathrm{B}(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

Definition 91 (Beta distribution). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. We say that an absolutely continuous random variable X follows a *beta distribution* of parameters $a, b \in \mathbb{R}_{>0}$, and we denoted it by $X \sim \mathrm{Beta}(a,b)$, if X has the pdf:

$$f_X(x) = \frac{1}{\mathrm{B}(a,b)} x^{a-1}(1-x)^{b-1} \mathbf{1}_{(0,1)}(x)$$

Definition 92 (Cauchy distribution). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $\mu \in \mathbb{R}$ and $\sigma \in \mathbb{R}_{>0}$. We say that an absolutely continuous random variable X follows a *Cauchy distribution* of parameters $x_0 \in \mathbb{R}$ and $\gamma \in \mathbb{R}_{>0}$, and we denoted it by $X \sim \mathrm{C}(x_0, \gamma)$, if X has the pdf:

$$f_X(x) = \frac{1}{\pi\gamma \left[1 + \left(\frac{x-x_0}{\gamma}\right)^2\right]}$$

Definition 93. A *mixed random variable* is a random variable whose cdf is neither piecewise-constant (a discrete random variable) nor absolutely continuous.

Theorem 94. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and X be a random variable with cdf F_X. Suppose that:

1. F_X is continuous.
2. F_X is differentiable at any point except for, maybe, a finite number of points.
3. F_X is continuously differentiable at any point except for, maybe, a finite number of points.

Then:

$$F_X(x) = \int_{-\infty}^{x} F'(t) \, \mathrm{d}t \qquad \forall x \in \mathbb{R}$$

That is, ${F_X}'$ is the pdf of X.

Transformations of random variables

Proposition 95. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, X be an absolutely continuous random variable with pdf f_X and U, V be open sets such that $\mathbb{P}(X \in U) = 1$. Let $h : U \to V$ be a diffeomorphism of class $\mathcal{C}^1$. Then, $Y := h(X)$ is also an absolutely continuous random variable and:

$$f_Y(y) = f_X(h^{-1}(y))|(h^{-1})'(y)|\mathbf{1}_V(y)$$

Proposition 96. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, X be an absolutely continuous random variable with pdf f_X, $U_1, \ldots, U_k$ be pairwise disjoint open intervals such that $\mathbb{P}(X \in U_1 \sqcup \cdots \sqcup U_k) = 1$. Let $h : U_1 \sqcup \cdots \sqcup U_k \to \mathbb{R}$ and denote $h_i = h|_{U_i}$. Then, if $h_i : U_i \to V_i$ are diffeomorphisms of class $\mathcal{C}^1$ for $i = 1, \ldots, k$, then $Y := h(X)$ is also an absolutely continuous random variable and:

$$f_Y(y) = \sum_{i=1}^{k} f_X({h_i}^{-1}(y))|({h_i}^{-1})'(y)|\mathbf{1}_{V_i}(y)$$

Random vectors

Definition 97. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. A *random vector* $\mathbf{X}$ is a function $\mathbf{X} = (X_1, \ldots, X_n) : \Omega \to \mathbb{R}^n$ satisfying for all $B \in \mathcal{B}(\mathbb{R}^n)$:

$$\{\mathbf{X} \in B\} = \{\omega \in \Omega : \mathbf{X}(\omega) \in B\} \in \mathcal{A}$$

Proposition 98. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. $\mathbf{X} = (X_1, \ldots, X_n) : \Omega \to \mathbb{R}^n$ is a random vector if and only if $X_i : \Omega \to \mathbb{R}$ is a random variable for $i = 1, \ldots, n$.

[9] Beta function should not be confused with binomial distribution, although we have used the same notation.

Definition 99. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $\mathbf{X} = (X_1, \ldots, X_n)$ be a random vector. For all $B_1 \times \cdots \times B_n \in \mathcal{B}(\mathbb{R}^n)$, we have that:

$$\{\mathbf{X} \in B_1 \times \cdots \times B_n\} = \{X_1 \in B_1\} \cap \cdots \cap \{X_n \in B_n\}$$

We will denote:

$$\{X_1 \in B_1, \ldots, X_n \in B_n\} := \{X_1 \in B_1\} \cap \cdots \cap \{X_n \in B_n\}$$

Definition 100. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $\mathbf{X} = (X_1, \ldots, X_n)$ be a random vector. Then, the *distribution* of $\mathbf{X}$ is the function:

$$\begin{aligned} \mathbb{P}_{\mathbf{X}} : \mathcal{B}(\mathbb{R}^n) &\longrightarrow [0,1] \\ B &\longmapsto \mathbb{P}(\mathbf{X} \in B) \end{aligned}$$

Definition 101. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $\mathbf{X} = (X_1, \ldots, X_n)$ be a random vector. We say that $\mathbf{X}$ is *discrete* if there exists a finite or countable subset $S \subset \mathbb{R}^n$ such that $\mathbb{P}(\mathbf{X} \in S) = 1$[10]. In that case, S is called the *support* of $\mathbf{X}$[11].

Proposition 102. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $\mathbf{X} = (X_1, \ldots, X_n)$ be a random vector. Then, $\mathbf{X}$ is discrete if and only if X_i is a discrete random variable for $i = 1, \ldots, n$.

Definition 103 (Joint probability mass function). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $\mathbf{X} = (X_1, \ldots, X_n)$ be a discrete random vector. Then, the *joint probability mass function* (*joint pmf*) of $\mathbf{X}$ is:

$$\begin{aligned} p_{\mathbf{X}} : S_{X_1} \times \cdots \times S_{X_n} &\longrightarrow [0,1] \\ (x_1, \ldots, x_n) &\longmapsto \mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) \end{aligned}$$

Proposition 104. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $\mathbf{X}$ be a discrete random vector. Then, the joint pmf of $\mathbf{X}$ determines the distribution of $\mathbf{X}$.

Definition 105 (Marginal probability mass functions). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $\mathbf{X} = (X_1, \ldots, X_n)$ be a discrete random vector with support $S_{\mathbf{X}} = S_{X_1} \times \cdots \times S_{X_n}$. Then, the *marginal probability mass functions* (*marginal pmf*) of $\mathbf{X}$ are:

$$\begin{aligned} p_{X_i}(x_i) &= \mathbb{P}(X_i = x_i) \\ &= \sum_{\substack{y_j \in S_{X_j} \\ j \neq i}} p_{\mathbf{X}}(y_1, \ldots, y_{i-1}, x_i, y_{i+1}, \ldots, y_n) \end{aligned}$$

for $i = 1, \ldots, n$.

Definition 106 (Multinomial distrbution). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $A_1, \ldots, A_r \in \mathcal{A}$. Suppose $\mathbb{P}(A_i) = p_i$ for $i = 1, \ldots, r$ such that $p_1 + \cdots + p_r = 1$. The *multinomial distribution* is the one whose i-th random variable X_i is the number of successes of A_i in a sequence of n repetitions, for $i = 1, \ldots, r$, that is, $X_i \sim \mathrm{B}(n, p_i)$. Thus, for all $n_1, \ldots, n_r \in \{0, 1, \ldots, n\}$ such that $n_1 + \cdots + n_r = n$ we have:

$$\mathbb{P}(X_1 = n_1, \ldots, X_r = n_r) = \frac{n!}{n_1! \cdots n_r!} {p_1}^{n_1} \cdots {p_r}^{n_r}$$

If $\mathbf{X} = (X_1, \ldots, X_r)$ follows a multinomial distribution of parameters n and $p_1, \ldots, p_r$, we will write $\mathbf{X} \sim \mathrm{Mult}(n; p_1, \ldots, p_r)$.

Definition 107. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $\mathbf{X} = (X_1, \ldots, X_n)$ be a random vector. We say that $\mathbf{X}$ is *absolutely continuous* if exists a function $f : \mathbb{R}^n \to \mathbb{R}$ such that:

1. $f(x) \geq 0$, $\forall x \in \mathbb{R}^n$.
2. $$\int_{-\infty}^{+\infty} \overset{(n)}{\cdots} \int_{-\infty}^{+\infty} f(x_1, \ldots, x_n) \, \mathrm{d}x_1 \cdots \mathrm{d}x_n = 1$$
3. For all $B \in \mathcal{B}(\mathbb{R}^n)$ we have:
$$\mathbb{P}(\mathbf{X} \in B) = \int_B f(x) \, \mathrm{d}x$$

The function f, denoted by $f_{\mathbf{X}}$, is called *joint probability density function* (*joint pdf*) of $\mathbf{X}$.

Proposition 108 (Marginal probability density functions). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $\mathbf{X} = (X_1, \ldots, X_n)$ be an absolutely continuous random vector with density $f_{\mathbf{X}}$. Then, X_i is an absolutely random variable with pdf:

$$\begin{aligned} f_{X_i}(x_i) = \int_{-\infty}^{+\infty} \overset{(n-1)}{\cdots} \int_{-\infty}^{+\infty} f(x_1, \ldots, x_n) \, \mathrm{d}x_1 \cdots \mathrm{d}x_{i-1} \cdot \\ \cdot \mathrm{d}x_{i+1} \cdots \mathrm{d}x_n \end{aligned}$$

for $i = 1, \ldots, n$. These functions f_{X_i} are called *marginal probability density functions* (*marginal pdf*) of $\mathbf{X}$.

Definition 109 (Multivariate standard normal distribution). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $\mathbf{X} = (X_1, \ldots, X_n)$ be an absolutely continuous random vector. We say that $\mathbf{X}$ follows a *multivariate normal distribution*, denoted by $\mathbf{X} \sim N(0, 1)$, if $\mathbf{X}$ has a joint pdf:

$$f_{\mathbf{X}}(x_1, \ldots, x_n) = \frac{1}{(2\pi)^{\frac{n}{2}}} \mathrm{e}^{-\frac{{x_1}^2 + \cdots + {x_n}^2}{2}}$$

Moreover, $X_i \sim N(0, 1)$ for $i = 1, \ldots, n$.

Definition 110 (Multivariate uniform distribution). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $\mathbf{X} = (X_1, \ldots, X_n)$ be an absolutely continuous random vector. We say that $\mathbf{X}$ has a *multivariate uniform distribution* over $B \in \mathcal{B}(\mathbb{R}^n)$, with $\mathrm{vol}(B) < \infty$, if it has joint pdf:

$$f_{\mathbf{X}}(x) = \frac{1}{\mathrm{vol}(B)} \mathbf{1}_B(x)$$

If $\mathbf{X} = (X_1, \ldots, X_r)$ follows a multivariate uniform, we will write $\mathbf{X} \sim U(B)$.

[10] By agreement, we will suppose that S only contains points x such that $\mathbb{P}(\mathbf{X} = x) > 0$.

[11] In general we will denote S by $S_{\mathbf{X}}$. Moreover, note that $S_{\mathbf{X}} = S_{X_1} \times \cdots \times S_{X_n}$, where S_{X_i} is the support of the random variable X_i.

Definition 111. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $\mathbf{X} = (X_1, \ldots, X_n)$ be a random vector. The *multivariate cumulative distribution function* (*multivariate cdf*) of $\mathbf{X}$ is defined as:

$$F_{\mathbf{X}}(x_1, \ldots, x_n) := \mathbb{P}(X_1 \leq x_1, \ldots, X_n \leq x_n)$$

for $(x_1, \ldots, x_n) \in \mathbb{R}^n$.

Theorem 112. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $\mathbf{X}$ be a random vector. Then, $F_{\mathbf{X}}$ determines the distribution of $\mathbf{X}$.

Proposition 113. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $\mathbf{X} = (X_1, \ldots, X_n)$ be a random vector. Then, the multivariate cdf $F_{\mathbf{X}}$ of $\mathbf{X}$ has the following properties:

1. It is monotonically increasing in each of its variables[12].
2. It is right-continuous in each of its variables.
3. For all $i = 1, \ldots, n$ we have:
$$\lim_{x_i \to -\infty} F_{\mathbf{X}}(x_1, \ldots, x_n) = 0$$
$$\lim_{x_1, \ldots, x_n \to +\infty} F_{\mathbf{X}}(x_1, \ldots, x_n) = 1$$
4. For all $i = 1, \ldots, n$ we have:
$$\lim_{x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n \to +\infty} F_{\mathbf{X}}(x_1, \ldots, x_n) = F_{X_i}(x_i)$$
5. If $\mathbf{X}$ is absolutely continuous, then:
$$F_{\mathbf{X}}(x_1, \ldots, x_n) = \\ = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_{\mathbf{X}}(s_1, \ldots, s_n) \, \mathrm{d}s_1 \cdots \mathrm{d}s_n$$
6. If $\mathbf{X}$ is absolutely continuous, then:
$$\mathbb{P}(a_1 < X_1 < b_1, \ldots, a_n < X_n < b_n) = \\ = \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} f_{\mathbf{X}}(s_1, \ldots, s_n) \, \mathrm{d}s_1 \cdots \mathrm{d}s_n$$

Proposition 114. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $\mathbf{X} = (X_1, \ldots, X_n)$ be an absolutely continuous random vector. If $\mathbf{X}$ has a continuous joint pdf, then:

$$f_{\mathbf{X}}(x_1, \ldots, x_n) = \frac{\partial^n F_{\mathbf{X}}}{\partial x_1 \cdots \partial x_n}(x_1, \ldots, x_n)$$

Transformations of random vectors

Definition 115. We say that a function $\mathbf{h} : \mathbb{R}^n \to \mathbb{R}^m$ is *Borel measurable* if $\forall B \in \mathcal{B}(\mathbb{R}^m)$ we have:

$$\mathbf{h}^{-1}(B) \in \mathcal{B}(\mathbb{R}^n)$$

Proposition 116. Let $\mathbf{h} : \mathbb{R}^n \to \mathbb{R}^m$ be a continuous function. Then, $\mathbf{h}$ is Borel measurable.

Proposition 117. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $\mathbf{X} = (X_1, \ldots, X_n)$ be a random vector and $\mathbf{h} : \mathbb{R}^n \to \mathbb{R}^m$ be a Borel measurable function. Then, $\mathbf{Y} := \mathbf{h}(\mathbf{X})$ is also a random vector.

Proposition 118. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $U, V \subseteq \mathbb{R}^n$ be open sets and $\mathbf{X} = (X_1, \ldots, X_n)$ be an absolutely continuous random vector with joint pdf $f_{\mathbf{X}}$ such that $\mathbb{P}(\mathbf{X} \in U) = 1$. Let $\mathbf{h} : U \to V$ be a diffeomorphism of class $\mathcal{C}^1$. Then, $\mathbf{Y} := \mathbf{h}(\mathbf{X})$ is absolutely continuous and

$$f_{\mathbf{Y}}(y) = f_{\mathbf{X}}(\mathbf{h}^{-1}(y)) |J\mathbf{h}^{-1}(y)| \mathbf{1}_V(y)$$

where $J\mathbf{h}^{-1}(y) = \det \mathbf{D}\mathbf{h}^{-1}(y)$ is the Jacobian of $\mathbf{h}^{-1}$ evaluated at y.

Definition 119 (Multivariate normal distribution). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $\mathbf{X} \sim N(0, 1)$ be an absolutely continuous random vector and $\mathbf{h} : \mathbb{R}^n \to \mathbb{R}^n$ be a function defined as

$$\mathbf{h}(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{b} \qquad \mathbf{x} \in \mathbb{R}^n$$ [13]

where $\mathbf{A} \in \mathrm{GL}_n(\mathbb{R})$ and $\mathbf{b} \in \mathbb{R}^n$. Then, $\mathbf{Y} = \mathbf{h}(\mathbf{X})$ is an absolutely continuous random vector and it has joint pdf:

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{(2\pi)^{\frac{n}{2}}} \frac{1}{\det \mathbf{A}} \mathrm{e}^{-\frac{\|\mathbf{A}^{-1}(\mathbf{y}-\mathbf{b})\|^2}{2}}$$

In this case, we write $\mathbf{Y} \sim N(\mathbf{b}, \mathbf{A}\mathbf{A}^{\mathrm{T}})$. The vector $\mathbf{b}$ is called *mean vector* and the matrix $\mathbf{A}\mathbf{A}^{\mathrm{T}}$, *covariance matrix*.

Independent random variables

Definition 120. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $X_1, \ldots, X_n$ be random variables. We say that they are *independent* if $\forall B_1, \ldots, B_n \in \mathcal{B}(\mathbb{R})$, we have:

$$\mathbb{P}(X_1 \in B_1, \ldots, X_n \in B_n) = \prod_{i=1}^{n} \mathbb{P}(X_i \in B_i)$$

That is, the events $\{X_1 \in B_1\}, \ldots, \{X_n \in B_n\}$ are independent. Moreover if they have the same distribution, we will say that $X_1, \ldots, X_n$ are *independent and identically distributed* (abbreviated as *i.i.d.*).

Proposition 121. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $X_1, \ldots, X_n$ be independent random variables and $g_1, \ldots, g_n : \mathbb{R} \to \mathbb{R}$ be Borel measurable functions. Then, $Y_1 = g_1(X_1), \ldots, Y_n = g_n(X_n)$ are also independent random variables.

---

[12] For the 2-dimensional case, we have that for all $x < x'$ and $y < y'$:

$$F_{\mathbf{X}}(x', y') - F_{\mathbf{X}}(x, y') - F_{\mathbf{X}}(x', y) + F_{\mathbf{X}}(x, y) \geq 0$$

This positive quantity is called *increment* of $F_{\mathbf{X}}$ in the rectangle $(x, x'] \times (y, y']$. In general a function $f : \mathbb{R}^2 \to \mathbb{R}$ satisfying

$$f(x', y') - f(x, y') - f(x', y) + f(x, y) \geq 0 \qquad \forall x < x' \text{ and } \forall y < y'$$

is said to be *increasing*.

[13] Here, we have thought $(x_1, \ldots, x_n)$ as the vector $\mathbf{x}$ in $\mathbb{R}^n$.

Proposition 122. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $X_1, \ldots, X_n$ be random variables. Then, $X_1, \ldots, X_n$ are independent if and only if

$$F_{(X_1,\ldots,X_n)}(x_1, \ldots, x_n) = F_{X_1}(x_1) \cdots F_{X_n}(x_n)$$

for all $(x_1, \ldots, x_n) \in \mathbb{R}^n$.

Proposition 123. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $\mathbf{X} = (X_1, \ldots, X_n)$ be a discrete random vector with support $S_{\mathbf{X}}$. Then, $X_1, \ldots, X_n$ are independent if and only if

$$p_{\mathbf{X}}(x_1, \ldots, x_n) = p_{X_1}(x_1) \cdots p_{X_n}(x_n)$$

for all $(x_1, \ldots, x_n) \in S_{\mathbf{X}}$.

Proposition 124. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $\mathbf{X} = (X_1, \ldots, X_n)$ be an absolutely continuous random vector. Then, $X_1, \ldots, X_n$ are independent if and only if

$$f_{\mathbf{X}}(x_1, \ldots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n)$$

for all $(x_1, \ldots, x_n) \in \mathbb{R}^n$, except for, maybe, a null set.

Proposition 125. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $X_1, \ldots, X_n$ be absolutely continuous and independent random variables. Then, $\mathbf{X} := (X_1, \ldots, X_n)$ is an absolutely continuous random vector and

$$f_{\mathbf{X}}(x_1, \ldots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n)$$

for all $(x_1, \ldots, x_n) \in \mathbb{R}^n$, except for, maybe, a null set.

Conditional distributions

Definition 126. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, (X, Y) be a discrete random vector with support $S_X \times S_Y$ and $y \in S_Y$. The *conditional probability mass function* of X given $Y = y$ is defined as:

$$p_{X|Y}(x \mid y) := \mathbb{P}(X = x \mid Y = y) = \frac{p_{(X,Y)}(x, y)}{p_Y(y)}$$

for all $x \in S_X$.

Proposition 127. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and (X, Y) be a discrete random vector with support $S_X \times S_Y$. Then, the pmf of Y together with the pmf of X conditioned to $Y = y$ determine the pmf of X in the following way:

$$\mathbb{P}(X = x) = \sum_{y \in S_Y} p_{X|Y}(x \mid y) p_Y(y) \quad \forall x \in S_X$$

Definition 128. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, (X, Y) be an absolutely continuous random vector and $y \in S_Y$. The *conditional probability density function* of X given $Y = y$ is defined as:

$$f_{X|Y}(x \mid y) := \begin{cases} \frac{f_{(X,Y)}(x,y)}{f_Y(y)} & \text{if } f_Y(y) > 0 \\ a & \text{if } f_Y(y) = 0 \end{cases}$$

where $x \in \mathbb{R}$ and $a \in \mathbb{R}$ is an arbitrary value[14].

Proposition 129. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, (X, Y) be an absolutely continuous random vector, $y \in S_Y$ and $a, b \in \mathbb{R} \cup \{\pm\infty\}$ such that $a < b$. Then:

$$\mathbb{P}(X \in (a, b) \mid Y = y) = \int_a^b f_{X|Y}(x \mid y) \, \mathrm{d}x$$

Proposition 130. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and (X, Y) be an absolutely continuous random vector. Then, the pdf of Y together with the pdf of X conditioned to $Y = y$ determine the pdf of X in the following way:

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X|Y}(x \mid y) f_Y(y) \, \mathrm{d}y \quad \forall x \in \mathbb{R}$$

4. | Expectation

Expectation of simple random variables

Definition 131. A *simple random variable* is a random variable that takes a finite or countable number of values[15].

Definition 132. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and X be a simple random variable whose outcomes are $\{x_i : i \in I\}$[16], where I is a finite or countable index set. We say that X has *finite expectation* or that it is *integrable* if:

$$\sum_{i \in I} |x_i| \mathbb{P}(X = x_i) < \infty$$

If so, we define the *expectation* of X as:

$$\mathbb{E}_{\mathrm{s}}(X) := \sum_{i \in I} x_i \mathbb{P}(X = x_i)$$

If the series of above is not absolutely convergent, we will say that X is *not integrable*.

Lemma 133. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and suppose that a random variable X can be expressed as

$$X = \sum_{n=1}^{N} a_n \mathbf{1}_{A_n}$$

where $N \in \mathbb{N} \cup \{\infty\}$, $\{A_n : n = 1, \ldots, N\} \subseteq \mathcal{A}$ is a partition of Ω and $\{a_n : n = 1, \ldots, N\}$ are not necessarily distinct values. Then, X has finite expectation if and only if

$$\sum_{n=1}^{N} |a_n| \mathbb{P}(A_n) < \infty$$

and in that case, $\mathbb{E}_{\mathrm{s}}(X) = \sum_{n=1}^{N} a_n \mathbb{P}(A_n)$.

Proposition 134. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and X, Y be simple and integrable random variables. Then, $X + Y$ is also a simple and integrable random variable and:

$$\mathbb{E}_{\mathrm{s}}(X + Y) = \mathbb{E}_{\mathrm{s}}(X) + \mathbb{E}_{\mathrm{s}}(Y)$$

[14] Usually chosen equal to 0.

[15] Note that a simple random variable is a particular case of a discrete random variable.

[16] Note that we can write X as $X = \sum_{i \in I} x_i \mathbf{1}_{\{X = x_i\}}$, where the events $\{X = x_i\}$, $i \in I$, form a partition of Ω.

Proposition 135. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, X be a simple and integrable random variable and $c \in \mathbb{R}$. Then, cX is also a simple and integrable random variable and:

$$\mathbb{E}_s(cX) = c\mathbb{E}_s(X)$$

Proposition 136. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, X be a simple random variable such that $|X| \le C$ for some $C \in \mathbb{R}$. Then, X is integrable and:

$$|\mathbb{E}_s(X)| \le C$$

Proposition 137. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and X be a random variable. For all $n \in \mathbb{N}$ and all $\omega \in \Omega$ $\exists! k \in \mathbb{Z}$ such that $X(\omega) \in \left[\frac{k}{2^n}, \frac{k+1}{2^n}\right)$. We define $X_n(\omega) := \frac{k}{2^n}$. Then,

$$X_n = \sum_{k \in \mathbb{Z}} \frac{k}{2^n} \mathbf{1}_{\left\{\frac{k}{2^n} \le X < \frac{k+1}{2^n}\right\}}$$

is a simple random variable such that $X_n(\omega) \le X(\omega)$ $\forall n \in \mathbb{N}$ and $\omega \in \Omega$[17].

Extension of the expectation

Proposition 138. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and X be a random variable. Then, there exists a sequence (X_n) of simple random variables that converges uniformly to X, that is:

$$\lim_{n \to \infty} \sup\{|X_n(\omega) - X(\omega)| : \omega \in \Omega\} = 0$$

Furthermore, $\forall \omega \in \Omega$ and $\forall n \in \mathbb{N}$ we have that $X_n(\omega) \le X_{n+1}(\omega)$.

Theorem 139. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and X be a random variable such that there exists a sequence (X_n) of simple and integrable random variables that converges uniformly to X. Then, the following statements are satisfied:

1. The limit $\lim_{n \to \infty} \mathbb{E}_s(X_n)$ exists.
2. The limit $\lim_{n \to \infty} \mathbb{E}_s(X_n)$ does not depend on the sequence (X_n).
3. If X is simple, then $\mathbb{E}_s(X) = \lim_{n \to \infty} \mathbb{E}_s(X_n)$.

Definition 140 (Expectation). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and X be a random variable. We say that X has *finite expectation* or that it is *integrable* if there exists a sequence (X_n) of simple and integrable random variables that converges uniformly to X. In that case, we define the *expectation* of X as:

$$\mathbb{E}(X) := \lim_{n \to \infty} \mathbb{E}_s(X_n)$$ [18]

Proposition 141. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and X, Y be random variables and $c \in \mathbb{R}$. Then:

1. If X and Y are integrable, then $X + Y$ is also integrable and:
$$\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$$
2. If X is integrable, then cX is also integrable and:
$$\mathbb{E}(cX) = c\mathbb{E}(X)$$
3. If $|X| \le C$ for some $C \in \mathbb{R}$, then X is integrable.
4. If X is integrable and $X \ge 0$, then $\mathbb{E}(X) \ge 0$.
5. If X and Y are integrable and $Y \ge X$, then $\mathbb{E}(Y) \ge \mathbb{E}(X)$.
6. If $m \le X \le M$ for some $m, M \in \mathbb{R}$, then X is integrable and
$$m \le \mathbb{E}(X) \le M$$
7. *Comparison test*: If X is integrable and $|Y| \le X$, then Y is integrable.
8. X is integrable $\iff$ $|X|$ is integrable.
9. If $\mathbb{P}(A) = 0$ for some $A \in \mathcal{A}$, then for any random variable X, we have that $X\mathbf{1}_A$ is integrable and $\mathbb{E}(X\mathbf{1}_A) = 0$.
10. If $X \overset{\text{a.s.}}{=} Y$ and one of them is integrable, then so will be the other one and, furthermore, $\mathbb{E}(X) = \mathbb{E}(Y)$.

Corollary 142. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, X be a random variable and $A \in \mathcal{A}$. Then, $\mathbb{E}(\mathbf{1}_{\{X \in A\}}) = \mathbb{P}(X \in A)$.

Proposition 143. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and X be a random variable with support $\mathbb{N} \cup \{0\}$. Then:

$$\mathbb{E}(X) = \sum_{k=0}^{\infty} \mathbb{P}(X > k)$$

Proof.

$$\sum_{k=0}^{\infty} \mathbb{P}(X > k) = \sum_{k=0}^{\infty} \sum_{n=k+1}^{\infty} \mathbb{P}(X = n) =$$
$$= \sum_{n=1}^{\infty} \sum_{k=0}^{n-1} \mathbb{P}(X = n) = \sum_{n=1}^{\infty} n\mathbb{P}(X = n) = \mathbb{E}(X)$$

□

Theorem 144 (Monotone convergence theorem). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and (X_n) be an increasing sequence of non-negative random variables such that $\lim_{n \to \infty} X_n \overset{\text{a.s.}}{=} X$[19], for some random variable X. Then:

$$\lim_{n \to \infty} \mathbb{E}(X_n) = \mathbb{E}(X)$$

Sketch of the proof. Check the proof of ?? ??. □

[17] The partition $\mathbb{R} = \bigsqcup_{k \in \mathbb{Z}} \left[\frac{k}{2^n}, \frac{k+1}{2^n}\right)$ is called *dyadic partition of order n*.

[18] Theorem 139 proves that this definition is well-defined.

[19] See Theorem 196.

Theorem 145 (**Dominated convergence theorem**). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and (X_n) be sequence of random variables such that $\lim_{n\to\infty} X_n \overset{\text{a.s.}}{=} X$, for some random variable X. Suppose that there exists an integrable random variable Y such that

$$|X| \leq Y \quad \forall n \geq 1$$

Then:

$$\lim_{n\to\infty} \mathbb{E}(X_n) = \mathbb{E}(X)$$

Sketch of the proof. Check the proof of **?? ??**. □

Theorem 146 (**Fatou's lemma**). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and (X_n) be a sequence of non-negative random variables. Then:

$$\mathbb{E}(\liminf_{n\to\infty} X_n) \leq \liminf_{n\to\infty} \mathbb{E}(X_n)$$

Sketch of the proof. Check the proof of **?? ??**. □

Expectation of absolutely continuous random variables

Theorem 147. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and X be an absolutely continuous random variable with density f_X. Then, X is integrable if and only if

$$\int_{-\infty}^{+\infty} |x| f_X(x) \, \mathrm{d}x < \infty$$

In that case, we have:

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} x f_X(x) \, \mathrm{d}x < \infty$$

Expectation of transformations of random vectors

Proposition 148. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $\mathbf{X} = (X_1, \ldots, X_n)$ be a discrete random vector with support $S_{\mathbf{X}}$ and $h : \mathbb{R}^n \to \mathbb{R}$ be a function. Then, $Y := h(\mathbf{X})$ is an integrable random variable if and only if

$$\sum_{(x_1,\ldots,x_n)\in S_{\mathbf{X}}} |h(x_1, \ldots, x_n)| \mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) < \infty$$

In that case, we have:

$$\mathbb{E}(Y) = \\ = \sum_{(x_1,\ldots,x_n)\in S_{\mathbf{X}}} h(x_1, \ldots, x_n) \mathbb{P}(X_1 = x_1, \ldots, X_n = x_n)$$

Proposition 149. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $\mathbf{X} = (X_1, \ldots, X_n)$ be an absolutely continuous random vector with density $f_{\mathbf{X}}$ and $h : \mathbb{R}^n \to \mathbb{R}$ be a Borel measurable function. Then, $Y := h(\mathbf{X})$ is an integrable random variable if and only if

$$\int_{\mathbb{R}^n} |h(x_1, \ldots, x_n)| f_{\mathbf{X}}(x_1, \ldots, x_n) \, \mathrm{d}x_1 \cdots \mathrm{d}x_n < \infty$$

In that case, we have:

$$\mathbb{E}(Y) = \int_{\mathbb{R}^n} h(x_1, \ldots, x_n) f_{\mathbf{X}}(x_1, \ldots, x_n) \, \mathrm{d}x_1 \cdots \mathrm{d}x_n$$

Expectation of non-negative and mixed random variables

Proposition 150. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $X \geq 0$ be a non-negative random variable and $(X_n) \geq 0$ be a sequence of simple random variables that converges uniformly to X. Then:

- If X_n is integrable $\forall n \in \mathbb{N}$, then:

$$\mathbb{E}(X) = \lim_{n\to\infty} \mathbb{E}(X_n)$$

- If $\exists m \in \mathbb{N}$ such that X_m isn't integrable, then:

$$\mathbb{E}(X) = \infty$$

Definition 151. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and X be a mixed random variable. Suppose that the discrete part of X has support S_X and "pmf" $p_X(x)$[20]. Moreover, suppose that the absolutely continuous part of X has "pdf" $f_X(x)$[21] We say that X has *finite expectation* if:

$$\sum_{x\in S} |x| p_X(x) + \int_{-\infty}^{+\infty} |x| f_X(x) \, \mathrm{d}x < \infty$$

If so, we define the *expectation* of X as:

$$\mathbb{E}(X) := \sum_{x\in S_X} x p_X(x) + \int_{-\infty}^{+\infty} x f_X(x) \, \mathrm{d}x$$

Moments

Definition 152 (**Moment**). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, X be a random variable and $k \in \mathbb{N}$. We say that X has *finite moment of order k* (or *finite k-th moment*) if X^k has finite expectation. We denote by μ_k the k-th moment of X:

$$\mu_k = \mathbb{E}(X^k)$$

Proposition 153. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, X, Y be random variables such that they have finite k-th moment. Then:

1. X and Y have finite r-th moment $\forall r \in \{1, \ldots, k\}$.
2. $X + Y$ has finite k-th moment.

Proposition 154. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, X, Y be random variables such that they have finite $2k$-th moment. Then, XY has finite k-th moment.

[20] We write pmf in quotation marks because $p_X(x)$ is not exactly a probability mass function since $\sum_{x\in S_X} p_X(x) < 1$.

[21] Again, we write pdf in quotation marks because $f_X(x)$ is not exactly a probability density function since $\int_{-\infty}^{+\infty} f_X(x) \, \mathrm{d}x < 1$.

Theorem 155 (Cauchy-Schwarz inequality). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and X, Y be two random variables such that $\mathbb{E}(X^2) < \infty$. Then:

$$\mathbb{E}(|XY|) \leq \left(\mathbb{E}(X^2)\mathbb{E}(Y^2)\right)^{1/2}$$

Definition 156 (Variance). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and X be a random variable such that $\mathbb{E}(X^2) < \infty$. We define the *variance* of X as:

$$\mathrm{Var}(X) := \mathbb{E}\left((X - \mathbb{E}(X))^2\right) \geq 0$$

Proposition 157. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and X be a random variable such that $\mathbb{E}(X^2) < \infty$. Then:

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

Definition 158 (Standard deviation). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and X be a random variable such that $\mathbb{E}(X^2) < \infty$. We define the *standard deviation* (or *standard error*) of X as:

$$\sigma(X) := \sqrt{\mathrm{Var}(X)}$$ [22]

Proposition 159. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and X be a random variable such that $\mathbb{E}(X^2) < \infty$. If $\mathrm{Var}(X) = 0$, then $X \overset{\text{a.s.}}{=} \mathbb{E}(X)$.

Proposition 160. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and X, Y be independent random variables with finite expectation. Then, XY has finite expectation and:

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$$

Definition 161 (Covariance). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and X, Y be two random variables with finite 2nd moments. We define the *covariance* between X and Y as:

$$\mathrm{Cov}(X, Y) := \mathbb{E}\left([X - \mathbb{E}(X)][Y - \mathbb{E}(Y)]\right)$$

Proposition 162. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and X, Y be two random variables with finite 2nd moments. Then:

$$\mathrm{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

Proposition 163. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and X, Y be independent random variables with finite expectation. Then, $\mathrm{Cov}(X, Y)$ is well-defined and $\mathrm{Cov}(X, Y) = 0$.

Definition 164. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. We say that two random variables X, Y are *uncorrelated* if $\mathrm{Cov}(X, Y) = 0$.

Proposition 165. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, X be a random variable such that $\mathbb{E}(X^2) < \infty$ and $a, b \in \mathbb{R}$. Then:

$$\mathrm{Var}(aX + b) = a^2 \mathrm{Var}(X)$$

Proposition 166. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $X_1, \ldots, X_n$ be random variables with finite 2nd moments. Then:

$$\mathrm{Var}(X_1 + \cdots + X_n) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2 \sum_{1 \leq i < j \leq n} \mathrm{Cov}(X_i, X_j)$$

Corollary 167. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $X_1, \ldots, X_n$ be random variables with finite 2nd moments such that they are pairwise uncorrelated. Then:

$$\mathrm{Var}(X_1 + \cdots + X_n) = \sum_{i=1}^{n} \mathrm{Var}(X_i)$$

Definition 168 (Pearson correlation coefficient). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and X, Y be non-constant random variables with finite 2nd moments. We define the *Pearson correlation coefficient* (or simply *correlation coefficient*)[23] as:

$$\rho(X, Y) := \frac{\mathrm{Cov}(X, Y)}{\sigma(X)\sigma(Y)} = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}}$$

Proposition 169. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, X, Y be non-constant random variables with finite 2nd moments and $\rho := \rho(X, Y)$. Then:

1. $|\rho| \leq 1$
2. If $\rho = 1$, then $\exists a, b \in \mathbb{R}$ with $a > 0$ such that $Y = aX + b$.
3. If $\rho = -1$, then $\exists a, b \in \mathbb{R}$ with $a < 0$ such that $Y = aX + b$.

Theorem 170 (Markov's inequality). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $X \geq 0$ be a non-negative random variable with finite expectation and $\lambda \in \mathbb{R}_{>0}$. Then:

$$\mathbb{P}(X > \lambda) \leq \mathbb{P}(X \geq \lambda) \leq \frac{\mathbb{E}(X)}{\lambda}$$

Corollary 171. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $X \geq 0$ be a non-negative random variable such that $\mathbb{E}(X) = 0$. Then, $X \overset{\text{a.s.}}{=} 0$.

Corollary 172. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $p \in \mathbb{R}_{>0}$ and X be a random variable such that $\mathbb{E}(|X|^p) < \infty$. Then, for all $a \in \mathbb{R}_{>0}$:

$$\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}(|X|^p)}{a^p}$$

Corollary 173 (Chebyshev's inequality). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, X be a random variable such that $\mathbb{E}(X^2) < \infty$ and $\delta \in \mathbb{R}_{>0}$. Then:

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq \delta) \leq \frac{\mathrm{Var}(X)}{\delta^2}$$

Furthermore, if $\sigma := \sigma(X)$ and we take $\delta = k\sigma$, $k \in \mathbb{R}_{>0}$, then:

$$\mathbb{P}(|X - \mathbb{E}(X)| < k\sigma) \geq 1 - \frac{1}{k^2}$$

[22] Therefore, $\mathrm{Var}(X)$ is sometimes expressed as $\sigma^2(X)$.

[23] The correlation coefficient measures the linear correlation between two random variables.

Definition 174 (Moment-generating function). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and X be a random variable. The *moment-generating function* of X is the function ψ_X defined as:

$$\begin{aligned} \psi_X : \mathbb{R} &\longrightarrow \mathbb{R}_{>0} \cup \{+\infty\} \\ t &\longmapsto \mathbb{E}(\mathrm{e}^{tX}) \end{aligned}$$

Theorem 175. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, X, Y be random variables such that $\psi_X(t), \psi_Y(t) < +\infty$ in a neighbourhood of 0 and such that ψ_X, ψ_Y are equal in another neighbourhood of 0. Then, $X \stackrel{\mathrm{d}}{=} Y$.

Theorem 176. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and X be a random variable such that $\psi_X(t) < +\infty$ in a neighbourhood of 0. Then, X has moment of order k $\forall k \in \mathbb{N}$ and:

$$\mathbb{E}(X^k) = {\psi_X}^{(k)}(0)$$

Theorem 177. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, X, Y be independent random variables. Then, for all t such that $\psi_X(t), \psi_Y(t) < +\infty$, the function $\psi_{X+Y}(t)$ is finite and:

$$\psi_{X+Y}(t) = \psi_X(t)\psi_Y(t)$$

Proof.

$$\psi_{X+Y}(t) = \mathbb{E}(\mathrm{e}^{t(X+Y)}) = \mathbb{E}(\mathrm{e}^{tX})\mathbb{E}(\mathrm{e}^{tY}) = \psi_X(t)\psi_Y(t)$$

□

where the second equality is because of the independence of X and Y.

Proposition 178. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $p \in (0, 1)$, $\lambda \in \mathbb{R}_{>0}$ and $X_1, \ldots, X_n$ be i.i.d. random variables.

- If $X_1 \sim \mathrm{Ber}(p)$, then $X_1 + \cdots + X_n \sim \mathrm{B}(n, p)$
- If $X_1 \sim \mathrm{Geo}(p)$, then $X_1 + \cdots + X_n \sim \mathrm{NB}(n, p)$
- If $X_1 \sim \mathrm{Exp}(\lambda)$, then $X_1 + \cdots + X_n \sim \mathrm{Gamma}(n, \lambda)$

Proposition 179. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $X_i \sim \mathrm{Pois}(\lambda_i)$ be random variables for some $\lambda_i > 0$, $i = 1, \ldots, n$. Suppose that $X_1, \ldots, X_n$ are independent. Then:

$$X_1 + \cdots + X_n \sim \mathrm{Pois}(\lambda_1 + \cdots + \lambda_n)$$

Proposition 180. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $\mu_i \in \mathbb{R}$ and $\sigma_i \in \mathbb{R}$ for $i = 1, \ldots, n$. Let $X_i \sim N(\mu_i, {\sigma_i}^2)$ be independent random variables for $i = 1, \ldots, n$. Then:

$$X_1 + \cdots + X_n \sim N(\mu_1 + \cdots + \mu_n, {\sigma_1}^2 + \cdots + {\sigma_n}^2)$$

Conditional expectation

Definition 181. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, (X, Y) be a discrete random vector with support $S_X \times S_Y$ and $y \in S_Y$. The *conditional expectation* of X given $Y = y$ is defined as:

$$\mathbb{E}(X \mid Y = y) := \sum_{x \in S_X} x\mathbb{P}(X = x \mid Y = y)$$

provided that the series is absolutely convergent. More generally, if $\mathbf{X}$ is a discrete random vector with support $S_{\mathbf{X}}$ and $h : S_{\mathbf{X}} \to \mathbb{R}$ is a function, then the conditional expectation of $h(\mathbf{X})$ given $Y = y$ is defined as:

$$\mathbb{E}(h(\mathbf{X}) \mid Y = y) := \sum_{\mathbf{x} \in S_{\mathbf{X}}} h(\mathbf{x})\mathbb{P}(\mathbf{X} = \mathbf{x} \mid Y = y)$$

provided that the series is absolutely convergent.

Proposition 182 (Law of total expectation). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $\mathbf{X}$ be a discrete random vector with support $S_{\mathbf{X}}$, Y be a random variable with support S_Y and $h : S_{\mathbf{X}} \to \mathbb{R}$ be a function. If $h(\mathbf{X})$ has finite expectation, then:

$$\mathbb{E}(h(\mathbf{X})) = \sum_{y \in S_Y} \mathbb{E}(h(\mathbf{X}) \mid Y = y)\mathbb{P}(Y = y)$$

Proof. We have that:

$$\begin{aligned} \mathbb{E}(h(\mathbf{X})) &= \sum_{\mathbf{x} \in S_{\mathbf{X}}} h(\mathbf{x})\mathbb{P}(\mathbf{X} = \mathbf{x}) \\ &= \sum_{\mathbf{x} \in S_{\mathbf{X}}} \sum_{y \in S_Y} h(\mathbf{x})\mathbb{P}(\mathbf{X} = \mathbf{x} \mid Y = y)\mathbb{P}(Y = y) \\ &= \sum_{y \in S_Y} \sum_{\mathbf{x} \in S_{\mathbf{X}}} h(\mathbf{x})\mathbb{P}(\mathbf{X} = \mathbf{x} \mid Y = y)\mathbb{P}(Y = y) \\ &= \sum_{y \in S_Y} \mathbb{E}(h(\mathbf{X}) \mid Y = y)\mathbb{P}(Y = y) \end{aligned}$$

where in the second equality we have used the 30 Law of total probability and in the third step we can rearrange the terms due to the finite expectation of $h(\mathbf{X})$. □

Definition 183. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, (X, Y) be an absolutely continuous random vector with support $S_X \times S_Y$ and $y \in S_Y$. The *conditional expectation* of X given $Y = y$ is defined as:

$$\mathbb{E}(X \mid Y = y) := \int_{S_X} x f_{X|Y}(x \mid y)\,\mathrm{d}x$$

provided that the integral is absolutely convergent. More generally, if $\mathbf{X}$ is an absolutely continuous random vector with support $S_{\mathbf{X}}$ and $h : S_{\mathbf{X}} \to \mathbb{R}$ is a function, then the conditional expectation of $h(\mathbf{X})$ given $Y = y$ is defined as:

$$\mathbb{E}(h(\mathbf{X}) \mid Y = y) := \int_{S_{\mathbf{X}}} h(\mathbf{x}) f_{\mathbf{X}|Y}(\mathbf{x} \mid y)\,\mathrm{d}\mathbf{x}$$

provided that the integral is absolutely convergent.

Proposition 184 (Law of total expectation). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $\mathbf{X}$ be an absolutely continuous random vector with support $S_{\mathbf{X}}$, Y be a random variable with support S_Y and $h : S_{\mathbf{X}} \to \mathbb{R}$ be a function. If $h(\mathbf{X})$ has finite expectation, then:

$$\mathbb{E}(h(\mathbf{X})) = \int_{-\infty}^{+\infty} \mathbb{E}(h(\mathbf{X}) \mid Y = y) f_Y(y)\,\mathrm{d}y$$

Sketch of the proof. Adapt the proof of 182 Law of total expectation. □

X	$\mathbb{E}(X)$	$\text{Var}(X)$
$c \in \mathbb{R}$	c	0
$U(\{x_1,\ldots,x_n\})$	$\frac{1}{n}\sum_{i=1}^n x_i$	$\frac{1}{n}\sum_{i=1}^n ({x_i}^2 - x_i)$
$\text{B}(n,p)$	np	$np(1-p)$
$\text{Pois}(\lambda)$	λ	λ
$\text{Geo}(p)$	$1/p$	$\dfrac{1-p}{p^2}$
$\text{HG}(N,p,n)$	np	$np(1-p)\dfrac{N-n}{N-1}$
$\text{NB}(r,p)$	$\dfrac{r}{p}$	$n\dfrac{1-p}{p^2}$
$U(a,b)$	$(a+b)/2$	$(b-a)^2/12$
$\text{Exp}(\lambda)$	$1/\lambda$	$1/\lambda^2$
$N(\mu,\sigma^2)$	μ	σ^2
$\text{Gamma}(\alpha,\beta)$	α/β	α/β^2
$\text{Beta}(a,b)$	$\dfrac{a}{a+b}$	$\dfrac{ab}{(a+b)^2(a+b+1)}$
$\text{C}(x_0,\gamma)$	$+\infty$	$+\infty$

Table 1: Expectations and variances of common distributions.

5. | Convergence of random variables

Convergence in probability

Definition 185. Let $(\Omega,\mathcal{A},\mathbb{P})$ be a probability space, (X_n) be a sequence of random variables and X be a random variable. We say that (X_n) *converges in probability* to X, and we denote it by $X_n \xrightarrow{\mathbb{P}} X$, if $\forall\varepsilon>0$ we have:

$$\lim_{n\to\infty}\mathbb{P}(|X_n - X| \geq \varepsilon) = 0$$

Or equivalently:

$$\lim_{n\to\infty}\mathbb{P}(|X_n - X| < \varepsilon) = 1$$

Proposition 186. Let $(\Omega,\mathcal{A},\mathbb{P})$ be a probability space. Then, the limit in probability is unique almost surely. That is, if (X_n) is a sequence of random variables and X, Y are a random variables such that $X_n \xrightarrow{\mathbb{P}} X$ and $X_n \xrightarrow{\mathbb{P}} Y$, then $\mathbb{P}(X\neq Y)=0$.

Proposition 187. Let $(\Omega,\mathcal{A},\mathbb{P})$ be a probability space, (X_n) be a sequence of random variables, X be a random variable such that $X_n \xrightarrow{\mathbb{P}} X$ and $f:\mathbb{R}\to\mathbb{R}$ be continuous function. Then, $f(X_n)\xrightarrow{\mathbb{P}} f(X)$.

Proposition 188. Let $(\Omega,\mathcal{A},\mathbb{P})$ be a probability space, $(X_{1n}),\ldots,(X_{mn})$ be m sequences of random variables, $X_1,\ldots,X_m$ be a random variable such that $X_{in}\xrightarrow{\mathbb{P}} X_i$ $\forall i=1,\ldots,m$ and $f:\mathbb{R}^m\to\mathbb{R}$ be continuous function. Then:

$$f(X_{1n},\ldots,X_{mn})\xrightarrow{\mathbb{P}} f(X_1,\ldots,X_m)$$

Corollary 189. Let $(\Omega,\mathcal{A},\mathbb{P})$ be a probability space, $(X_{1n}),\ldots,(X_{mn})$ be m sequences of random variables, $X_1,\ldots,X_m$ be a random variable such that $X_{in}\xrightarrow{\mathbb{P}} X_i$ $\forall i=1,\ldots,m$ and $f:\mathbb{R}^m\to\mathbb{R}$ be continuous function. Then:

- $X_{1n}+\cdots+X_{mn}\xrightarrow{\mathbb{P}} X_1+\cdots+X_m$.
- $X_{1n}\cdots X_{mn}\xrightarrow{\mathbb{P}} X_1\cdots X_m$.

Lemma 190. Let $(\Omega,\mathcal{A},\mathbb{P})$ be a probability space. Then, the set $\mathcal{L}^0$ of all random variables of $(\Omega,\mathcal{A},\mathbb{P})$ is a vector space. Moreover, the relation $\sim$ defined in $\mathcal{L}^0$ as

$$X\sim Y \iff X \overset{\text{a.s.}}{=} Y \quad \forall X,Y\in\mathcal{L}^0$$

is an equivalence relation. The quotient set $\mathcal{L}^0/\sim$ is denoted by L^0[24].

Proposition 191. Let $(\Omega,\mathcal{A},\mathbb{P})$ be a probability space. We define the function $d_{\mathbb{P}}$ in L^0 as:

$$\begin{aligned} d_{\mathbb{P}}: L^0\times L^0 &\longrightarrow \mathbb{R} \\ (X,Y) &\longmapsto \mathbb{E}\left(\tfrac{|X-Y|}{1+|X-Y|}\right)\end{aligned}$$

Then, $(L^0,d_{\mathbb{P}})$ is a metric space.

Proposition 192. Let $(\Omega,\mathcal{A},\mathbb{P})$ be a probability space, (X_n) be a sequence of random variables and X be a random variable. Then:

$$X_n\xrightarrow{\mathbb{P}} X \iff \lim_{n\to\infty} d_{\mathbb{P}}(X_n,X)=0$$

Because of that, the convergence in probability is said to be *metrizable*.

Definition 193. Let $(\Omega,\mathcal{A},\mathbb{P})$ be a probability space, (X_n) be a sequence of random variables and X be a random variable. We say that (X_n) satisfies the *Cauchy condition in probability* (or is *Cauchy in probability*) if $\forall\varepsilon>0$ we have:

$$\lim_{n,m\to\infty}\mathbb{P}(|X_n-X_m|\geq\varepsilon)=0$$

Proposition 194. Let $(\Omega,\mathcal{A},\mathbb{P})$ be a probability space, (X_n) be a sequence of random variables and X be a random variable. Then:

$$X_n\xrightarrow{\mathbb{P}} X \iff (X_n) \text{ is Cauchy in probability}$$

Thus, $(L^0,d_{\mathbb{P}})$ is a complete metric space.

Proposition 195. Let $(\Omega,\mathcal{A},\mathbb{P})$ be a probability space, (X_n) be a sequence of random variables and X be a random variable such that $X_n\xrightarrow{\mathbb{P}} X$. Then, all subsequence (X_{n_k}) of (X_n) converges in probability to X.

Almost surely convergence

Definition 196. Let $(\Omega,\mathcal{A},\mathbb{P})$ be a probability space, (X_n) be a sequence of random variables and X be a random variable. We say that (X_n) *converges almost surely* to X, and we denote it by $X_n\xrightarrow{\text{a.s.}} X$, if

$$\mathbb{P}\left(\lim_{n\to\infty} X_n = X\right)=1$$

That is,

$$\lim_{n\to\infty} X_n(\omega)=X(\omega)$$

[24]If $X\in\mathcal{L}^0$, we will use the same notation for its equivalence class.

for all $\omega \in \Omega$ except for maybe a set of probability zero. Another equivalent expression is the following one: $X_n \xrightarrow{\text{a.s.}} X$ if and only if

$$\mathbb{P}\left(\bigcap_{\varepsilon \in \mathbb{Q}_{>0}} \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} \{\omega \in \Omega : |X_k(\omega) - X(\omega)| < \varepsilon\}\right) = 1$$

Proposition 197. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, (X_n) be a sequence of random variables and X be a random variable. Then, $X_n \xrightarrow{\text{a.s.}} X$ if and only if $\forall \varepsilon > 0$ we have:

$$\lim_{n\to\infty} \mathbb{P}\left(\bigcup_{k=n}^{\infty} \{|X_k - X| \geq \varepsilon\}\right) = 0$$

Proposition 198. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, (X_n) be a sequence of random variables and X be a random variable. Then:

$$X_n \xrightarrow{\text{a.s.}} X \implies X_n \xrightarrow{\mathbb{P}} X$$

Proposition 199. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, (X_n) be a sequence of random variables, X be a random variable such that $X_n \xrightarrow{\text{a.s.}} X$ and $f : \mathbb{R} \to \mathbb{R}$ be continuous function. Then, $f(X_n) \xrightarrow{\text{a.s.}} f(X)$.

Proposition 200. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, (X_n) be a sequence of random variables and X be a random variable. Suppose that $X_n \xrightarrow{\mathbb{P}} X$. Then, there exists a subsequence (X_{n_k}) of (X_n) such that $X_{n_k} \xrightarrow{\text{a.s.}} X$.

Proposition 201. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, (X_n) be a sequence of random variables and X be a random variable. Suppose that $\forall \varepsilon > 0$ we have:

$$\sum_{n=1}^{\infty} \mathbb{P}(|X_n - X| \geq \varepsilon) < \infty$$

Then, $X_n \xrightarrow{\text{a.s.}} X$.

Definition 202. Let Ω be a set and $(A_n) \subset \Omega$ be a sequence of subsets. We define the *limit superior* of (A_n) as:

$$\limsup_{n\to\infty} A_n := \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$$

That is:

$$\omega \in \limsup_{n\to\infty} A_n \iff \forall n \geq 1 \ \exists k \geq n \text{ such that } \omega \in A_k$$

We can express that as:

$$\limsup_{n\to\infty} A_n = \{\omega \in \Omega : \omega \in A_n \text{ infinitely often}\}$$

Definition 203. Let Ω be a set and $(A_n) \subset \Omega$ be a sequence of subsets. We define the *limit inferior* of (A_n) as:

$$\liminf_{n\to\infty} A_n := \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k$$

That is:

$$\omega \in \liminf_{n\to\infty} A_n \iff \exists n \geq 1 \text{ such that } \forall k \geq n,\ \omega \in A_k$$

We can express that as:

$$\liminf_{n\to\infty} A_n = \{\omega \in \Omega : \omega \in A_n \text{ eventually}\}$$

Proposition 204. Let Ω be a set and $(A_n) \subset \Omega$ be a sequence of subsets. Then:

1. $\liminf_{n\to\infty} A_n \subseteq \limsup_{n\to\infty} A_n$
2. $\left(\limsup_{n\to\infty} A_n\right)^c = \liminf_{n\to\infty} {A_n}^c$

Definition 205. Let Ω be a set and $(A_n) \subset \Omega$ be a sequence of subsets. We say that (A_n) has *limit* if:

$$\liminf_{n\to\infty} A_n = \limsup_{n\to\infty} A_n$$

In that case, $A := \limsup_{n\to\infty} A_n$ is called the limit of the sequence.

Lemma 206 (First Borel-Cantelli lemma). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $(A_n) \subset \mathcal{A}$ be a sequence of events such that:

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$$

Then, $\mathbb{P}\left(\limsup_{n\to\infty} A_n\right) = 0$.

Proof. Let $B_n := \bigcup_{k\geq n} A_k$ and note that $B_{n+1} \subseteq B_n$. Thus, using the definition of $\limsup$ and 21 Continuity from above we have that

$$\mathbb{P}\left(\limsup_{n\to\infty} A_n\right) = \lim_{n\to\infty} \mathbb{P}(B_n) \leq \lim_{n\to\infty} \sum_{k\geq n} \mathbb{P}(A_n) = 0$$

because it is the tail of a convergent sequence. □

Lemma 207 (Second Borel-Cantelli lemma). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $(A_n) \subset \mathcal{A}$ be a sequence of independent events such that:

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$$

Then, $\mathbb{P}\left(\limsup_{n\to\infty} A_n\right) = 1$.

Proof. We will prove that $\mathbb{P}\left(\left[\limsup_{n\to\infty} A_n\right]^c\right) = 0$. From 20 Continuity from below, if $B_n := \bigcap_{k\geq n} {A_k}^c$ we have:

$$\mathbb{P}\left(\left[\limsup_{n\to\infty} A_n\right]^c\right) = \lim_{n\to\infty} \mathbb{P}(B_n)$$

Now, $\forall N \geq n$ we have $\mathbb{P}(B_n) \leq \mathbb{P}\left(\bigcap_{k=n}^{N} {A_n}^c\right)$. Using the independence and the inequality $1 + x \leq \mathrm{e}^x$, we get:

$$\mathbb{P}(B_n) \leq \prod_{k=n}^{N} (1 - \mathbb{P}(A_n)) \leq \mathrm{e}^{-\sum_{k=n}^{N}} \xrightarrow{N\to\infty} 0$$

□

Convergence in mean

Definition 208. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $p \geq 1$, (X_n) be a sequence of random variables such that $\mathbb{E}(|X_n|^p) < \infty$ and X be a random variable such that $\mathbb{E}(|X|^p) < \infty$. We say that (X_n) *converges in the p-th mean* to X, and we denote it by $X_n \xrightarrow{L^p} X$, if

$$\lim_{n\to\infty} \mathbb{E}(|X_n - X|^p) = 0$$

Proposition 209. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $p \geq 1$, (X_n) be a sequence of random variables such that $\mathbb{E}(|X_n|^p) < \infty$ and X be a random variable such that $\mathbb{E}(|X|^p) < \infty$. Then:

$$X_n \xrightarrow{L^p} X \implies X_n \xrightarrow{\mathbb{P}} X$$

Theorem 210 (Dominated convergence theorem). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $p \geq 1$, X be a random variable and (X_n) be a sequence of random variables such that $X_n \xrightarrow{\mathbb{P}} X$ or $X_n \xrightarrow{\text{a.s.}} X$. Suppose that there exists a random variable Y such that $|X_n| \leq Y$ $\forall n \geq 1$ and $\mathbb{E}(|Y|^p) < \infty$. Then, $X_n \xrightarrow{L^p} X$.

Lemma 211. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $p \geq 1$. Define the set $\mathcal{L}^p$ of all random variables of $(\Omega, \mathcal{A}, \mathbb{P})$ such that $\mathbb{E}(|X|^p) < \infty$. Then, $\mathcal{L}^p$ is a vector space. Moreover, the relation $\sim$ defined in $\mathcal{L}^p$ as

$$X \sim Y \iff X \overset{\text{a.s.}}{=} Y \quad \forall X, Y \in \mathcal{L}^p$$

is an equivalence relation. The quotient set $\mathcal{L}^p/\sim$ is denoted by L^p[25].

Proposition 212. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $p \geq 1$. We define the function:

$$\begin{aligned} \|\cdot\|_p : L^p &\longrightarrow \mathbb{R} \\ X &\longmapsto \mathbb{E}(|X|^p) \end{aligned}$$

Then, $(L^p, \|\cdot\|_p)$ is a normed vector space. Moreover, the norm $\|\cdot\|_p$ induces a distance d_p defined as:

$$d_p(X,Y) := \|X - Y\|_p \quad \forall X, Y \in L^p$$

Proposition 213. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $p \geq 1$, $(X_n) \subset L^p$ be a sequence of random variables and $X \in L^p$. Then:

$$X_n \xrightarrow{L^p} X \iff \lim_{n\to\infty} d_p(X_n, X) = 0$$

Therefore, the convergence in p-th mean is metrizable.

Definition 214. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $p \geq 1$, $(X_n) \subset L^p$ be a sequence of random variables and $X \in L^p$. We say that (X_n) satisfies the *Cauchy condition in p-th mean* (or is *Cauchy in p-th mean*) if:

$$\lim_{n,m\to\infty} \mathbb{E}(|X_n - X_m|^p) = 0$$

Proposition 215. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $p \geq 1$, $(X_n) \subset L^p$ be a sequence of random variables and $X \in L^p$. Then:

$$X_n \xrightarrow{L^p} X \iff (X_n) \text{ is Cauchy in } p\text{-th mean}$$

Thus, L^p is a Banach space and L^2 is a Hilbert space.

Proposition 216. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $p \geq 1$, $X, Y \in L^p$ and $(X_n), (Y_n) \subset L^p$ be sequences of random variables such that $X_n \xrightarrow{L^p} X$ and $Y_n \xrightarrow{L^p} Y$. Then:

$$X_n + Y_n \xrightarrow{L^p} X + Y$$

Proposition 217. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $X, Y \in L^2$ and $(X_n), (Y_n) \subset L^2$ be sequences of random variables such that $X_n \xrightarrow{L^2} X$ and $Y_n \xrightarrow{L^2} Y$. Then:

$$X_n Y_n \xrightarrow{L^1} XY$$

Proposition 218. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $p \geq 1$, $X \in L^p$ and $(X_n) \subset L^p$ be a sequence of random variables such that $X_n \xrightarrow{L^p} X$. Then:

1. $\lim_{n\to\infty} \|X_n\|_p = \|X\|_p$
2. If $1 \leq r < p$, then:
$$X_n \xrightarrow{L^p} X \implies X_n \xrightarrow{L^r} X$$
3. If $p = 1$, then:
$$\lim_{n\to\infty} \mathbb{E}(X_n) = \mathbb{E}(X)$$

Convergence in distribution

Definition 219. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, (X_n) be a sequence of random variables and X be a random variable. We say that (X_n) *converges in distribution* to X, and we denote it by $X_n \xrightarrow{\text{d}} X$, if $\forall B \in \mathcal{B}(\mathbb{R})$ such that $\mathbb{P}(X \in \partial B) = 0$ we have:

$$\lim_{n\to\infty} \mathbb{P}(X_n \in B) = \mathbb{P}(X \in B)$$

Definition 220. Let $A \subseteq \mathbb{R}$ be a set and $f : A \to \mathbb{R}$ be a function. We denote by $C(f)$ the set of points where f is continuous.

Proposition 221. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, (X_n) be a sequence of random variables with cdfs F_{X_n} $\forall n \in \mathbb{N}$ and X be a random variable be a random variable with cdf F_X. Then:

$$X_n \xrightarrow{\text{d}} X \iff \lim_{n\to\infty} F_{X_n}(t) = F_X(t) \quad \forall t \in C(F_X)$$[26]

Proposition 222. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, (X_n) be a sequence of random variables and X, Y be random variables such that $X_n \xrightarrow{\text{d}} X$ and $X_n \xrightarrow{\text{d}} Y$. Then, $X \overset{\text{d}}{=} Y$.

[25] If $X \in \mathcal{L}^p$, we will use the same notation for its equivalence class.

[26] That is, $X_n \xrightarrow{\text{d}} X \iff \forall t \in \mathbb{R}$ such that $\mathbb{P}(X = t) = 0$ we have $\lim_{n\to\infty} \mathbb{P}(X_n \leq t) = \mathbb{P}(X \leq t)$.

Theorem 223 (Skorokhod's representation theorem). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, (X_n) be a sequence of random variables and X be a random variable be such that $X_n \xrightarrow{\text{d}} X$. Then, there exists a probability space $(\Omega', \mathcal{A}', \mathbb{P}')$, and random variables (X'_n) and X defined on Ω' such that:

1. $X_n \overset{\text{d}}{=} X'_n \ \forall n \geq 1$
2. $X \overset{\text{d}}{=} X'$
3. $X'_n \xrightarrow{\text{a.s.}} X'$

Theorem 224. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, (X_n) be a sequence of random variables and X be a random variable. Then, $X_n \xrightarrow{\text{d}} X$ if and only if for any continuous and bounded function $f : \mathbb{R} \to \mathbb{R}$ we have:

$$\lim_{n\to\infty} \mathbb{E}(f(X_n)) = \mathbb{E}(f(X))$$

Lemma 225. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, (X_n) be a sequence of random variables and X be a random variable be such that both (X_n) and X take values in $\mathbb{N}$. Then, $X_n \xrightarrow{\text{d}} X$ if and only if $\forall k \in \mathbb{N}$, we have:

$$\lim_{n\to\infty} \mathbb{P}(X_n = k) = \mathbb{P}(X = k)$$

Proposition 226. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, (X_n) be a sequence of random variables, X be a random variable and $a \in \mathbb{R}$. Then:

1. $X_n \xrightarrow{\mathbb{P}} X \implies X_n \xrightarrow{\text{d}} X$
2. $X_n \xrightarrow{\text{d}} a \implies X_n \xrightarrow{\mathbb{P}} a$

Proposition 227. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, (X_n) be a sequence of random variables, X be a random variable such that $X_n \xrightarrow{\text{d}} X$ and $f : \mathbb{R} \to \mathbb{R}$ be a continuous function. Then, $f(X_n) \xrightarrow{\text{d}} f(X)$.

Corollary 228. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, (X_n) be a sequence of random variables and X be a random variable. Then:

1. $X_n + a \xrightarrow{\text{d}} X + a$
2. $aX_n \xrightarrow{\text{d}} aX$

Theorem 229 (Slutsky's theorem). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, (X_n), (Y_n) be sequences of random variables and X be a random variable and $a \in \mathbb{R}$ such that $X_n \xrightarrow{\text{d}} X$ and $Y_n \xrightarrow{\text{d}} a$. Then:

1. $X_n + Y_n \xrightarrow{\text{d}} X + a$
2. $X_n Y_n \xrightarrow{\text{d}} aX$
3. $\frac{X_n}{Y_n} \xrightarrow{\text{d}} \frac{X}{a}$ provided that $a \neq 0$.

6. | Laws of large numbers

Definition 230. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and (X_n) be a sequence of random variables. We define the sequence of partial sums (S_n) as:

$$S_n := \sum_{i=1}^n X_i$$

Weak laws

Theorem 231 (Weak law of large numbers). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and (X_n) be a sequence of i.i.d. random variables with finite 2nd moment. Then:

$$\frac{S_n}{n} \xrightarrow{\mathbb{P}} \mathbb{E}(X_1) \quad \text{and} \quad \frac{S_n}{n} \xrightarrow{L^2} \mathbb{E}(X_1)$$

Theorem 232 (Weak law of large numbers). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and (X_n) be a sequence of pairwise uncorrelated random variables with finite 2nd moment. Suppose that:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n \mathbb{E}(X_i) = \mu < \infty \text{ and } \lim_{n\to\infty} \frac{1}{n^2} \sum_{i=1}^n \mathrm{Var}(X_i) = 0$$

Then:

$$\frac{S_n}{n} \xrightarrow{\mathbb{P}} \mu \quad \text{and} \quad \frac{S_n}{n} \xrightarrow{L^2} \mu$$

Strong laws

Theorem 233 (Kolmogorov's strong law of large numbers). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and (X_n) be a sequence of i.i.d. random variables.

1. If $\mathbb{E}(X_1) < \infty$, then:
$$\frac{S_n}{n} \xrightarrow{\text{a.s.}} \mathbb{E}(X_1)$$
2. If $\mathbb{E}(X_1) = \infty$, then:
$$\limsup_{n\to\infty} \frac{|S_n|}{n} \overset{\text{a.s.}}{=} +\infty$$

Theorem 234 (Strong law of large numbers). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and (X_n) be a sequence of i.i.d. random variables such that $\mathbb{E}({X_1}^4) < \infty$. Then:

$$\frac{S_n}{n} \xrightarrow{\text{a.s.}} \mathbb{E}(X_1)$$

Corollary 235. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $A \in \mathcal{A}$. Let (X_n) be a sequence of i.i.d. random variables such that $X_n \sim \mathrm{Ber}(\mathbb{P}(A))\ \forall n \in \mathbb{N}$. Then:

$$\frac{S_n}{n} \xrightarrow{\text{a.s.}} \mathbb{P}(A)$$

Definition 236. Let $x \in [0, 1)$ and $b \in \mathbb{N}_{\geq 2}$. Suppose the expression of x in base b is $x_b = 0.a_1a_2a_3\cdots$. Let $N_{x,b}(k, n)$ denote the number of times the digit $k \in \{0, 1, \ldots, b-1\}$ appears in the decimal expansion of x_b in the first n digits. We say that x is *simply normal* if there exists $b \in \mathbb{N}_{\geq 2}$ such that

$$\lim_{n\to\infty} \frac{N_{x,b}(k, n)}{n} = \frac{1}{b} \quad \forall k \in \{0, 1, \ldots, b-1\}$$

We say that x is *normal* if

$$\lim_{n\to\infty} \frac{N_{x,b}(k, n)}{n} = \frac{1}{b} \quad \forall k \in \{0, 1, \ldots, b-1\}, \forall b \in \mathbb{N}_{\geq 2}$$

Theorem 237 (Borel's theorem). All the numbers in $[0, 1)$, except for a null set, are normal.

7. | Central limit theorem

Characteristic function

Definition 238. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. A *complex random variable* is a function $Z : \Omega \to \mathbb{C}$ such that $\mathrm{Re}(Z)$ and $\mathrm{Im}(Z)$ are real random variables. Therefore, Z may be written as $Z = X + \mathrm{i}Y$, where X and Y are real random variables.

Proposition 239. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $Z = X + \mathrm{i}Y$ be a complex random variable. Then[27]:

1. $\mathbb{E}(Z) = \mathbb{E}(X) + \mathrm{i}\mathbb{E}(Y)$
2. $\overline{\mathbb{E}(Z)} = \mathbb{E}(\overline{Z})$
3. $|\mathbb{E}(Z)| \leq \mathbb{E}(|Z|)$

Definition 240 (Characteristic function). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and X be a real random variable. The *characteristic function* of X is the function φ_X defined as:

$$\begin{aligned} \varphi_X : \mathbb{R} &\longrightarrow \mathbb{C} \\ t &\longmapsto \mathbb{E}(\mathrm{e}^{\mathrm{i}tX}) \end{aligned}$$

Proposition 241. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and X be a discrete random variable with support S_X. Then:

$$\varphi_X(t) = \sum_{x \in S_X} \mathrm{e}^{\mathrm{i}tx} \mathbb{P}(X = x)$$

Proposition 242. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and X be an absolutely continuous random variable with density f_X. Then:

$$\varphi_X(t) = \int_{-\infty}^{+\infty} \mathrm{e}^{\mathrm{i}tx} f_X(x) \, \mathrm{d}x$$

Proposition 243. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and X be a random variable. Then:

1. $\varphi_X(0) = 1$
2. $|\varphi_X(t)| \leq 1 \ \forall t \in \mathbb{R}$
3. $\overline{\varphi_X(t)} = \varphi_X(-t) \ \forall t \in \mathbb{R}$
4. If $Y = aX + b$ for some $a, b \in \mathbb{R}$, then:
$$\varphi_Y(t) = \mathrm{e}^{\mathrm{i}tb} \varphi_X(at) \quad \forall t \in \mathbb{R}$$
5. φ_X is uniformly continuous.

Theorem 244. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $X_1, \ldots, X_n$ be independent random variables. Let $Y := \sum_{i=1}^n X_i$. Then:

$$\varphi_Y(t) = \prod_{i=1}^n \varphi_{X_i}(t) \quad \forall t \in \mathbb{R}$$

Theorem 245. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and X, Y be random variables. Then:

$$X \overset{\mathrm{d}}{=} Y \iff \varphi_X(t) = \varphi_Y(t)$$

Theorem 246. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, (X_n) be a sequence of random variables and X be a random variable. Then:

$$X_n \xrightarrow{\mathrm{d}} X \iff \lim_{n \to \infty} \varphi_{X_n}(t) = \varphi_X(t) \quad \forall t \in \mathbb{R}$$

Proposition 247. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and X be random variables with finite n-th moment for some $n \in \mathbb{N}$. Then, there exists the derivative of order n of φ_X and it satisfies:

$${\varphi_X}^{(n)}(t) = \mathrm{i}^n \mathbb{E}\left(X^n \mathrm{e}^{\mathrm{i}tX}\right) \quad \forall t \in \mathbb{R}$$

In particular, ${\varphi_X}^{(n)}(0) = \mathrm{i}^n \mathbb{E}(X^n)$.

Central limit theorem

Theorem 248 (Lévy-Lindeberg central limit theorem). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and (X_n) be a sequence of i.i.d. random variables with finite 2nd moments. Let $\mu := \mathbb{E}(X_1)$ and $\sigma^2 := \mathrm{Var}(X_1)$. Then:

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{\mathrm{d}} Z$$

where $Z \sim N(0, 1)$.

Theorem 249 (Lyapunov central limit theorem). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and (X_n) be a sequence of independent random variables each with finite expectation $\mu_i := \mathbb{E}(X_i)$ and variance ${\sigma_i}^2 := \mathrm{Var}(X_i)$ $\forall i = 1, \ldots, n$. Then:

$$\frac{\sum_{i=1}^n (X_i - \mu_i)}{\sqrt{\sum_{i=1}^n {\sigma_i}^2}} \xrightarrow{\mathrm{d}} Z$$

where $Z \sim N(0, 1)$.

Corollary 250. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and (X_n) be a sequence of i.i.d. random variables with finite 2nd moments. Let $\mu := \mathbb{E}(X_1)$ and $\sigma^2 := \mathrm{Var}(X_1)$. Then, $\forall s, t \in \mathbb{R}$ such that $s < t$ we have

$$\lim_{n \to \infty} \mathbb{P}\left(s < \frac{S_n - n\mu}{\sigma\sqrt{n}} \leq t\right) = F_Z(t) - F_Z(s)$$

where $Z \sim N(0, 1)$.

Definition 251. Let $(\Omega, \mathcal{A}, \mathbb{P})$ and $X_1, \ldots, X_n$ be random variables. We define the *sample mean* of $X_1, \ldots, X_n$ as:

$$\overline{X}_n := \frac{1}{n} S_n = \frac{1}{n} \sum_{i=1}^n X_i$$

If the value of n is fixed, we denoted $\overline{X}_n$ by $\overline{X}$.

Proposition 252. Let $(\Omega, \mathcal{A}, \mathbb{P})$ and $X_1, \ldots, X_n$ be i.i.d. random variables. Then:

$$\mathbb{E}(\overline{X}_n) = \mathbb{E}(X_1) \quad \text{and} \quad \mathrm{Var}(\overline{X}_n) = \frac{1}{n} \mathrm{Var}(X_1)$$

[27]Here we have only exposed two properties of the expectation of a complex random variable but in general all the properties of the expectation that we've already seen in Section 4 can be extended conveniently to complex random variables.

Corollary 253. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $X_1, \ldots, X_n$ be i.i.d. random variables. Let $\mu := \mathbb{E}(X_1)$ and $\sigma^2 := \mathrm{Var}(X_1)$. Then:

$$\overline{X}_n \overset{\mathrm{d}}{\simeq} N\left(\mu, \frac{\sigma^2}{n}\right) \qquad \text{for } n \text{ large enough}$$

Corollary 254 (De Moivre-Laplace theorem). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $X_1, \ldots, X_n$ be i.i.d. random variables such that $X_n \sim \mathrm{Ber}(p)$ $\forall n \in \mathbb{N}$. Then:

$$\mathrm{B}(n,p) \overset{\mathrm{d}}{\simeq} N\left(np, np(1-p)\right) \qquad \text{for } n \text{ large enough}^{28}$$

Definition 255 (Continuity correction). The *continuity correction* is an adjustment that is made when a discrete distribution is approximated by a continuous distribution. For example if $X \sim \mathrm{B}(n,p)$ is a random variable and $np(1-p)$ is large enough, then $\mathbb{P}(X \leq k)$ is well approximated by $\mathbb{P}(Z \leq k + \frac{1}{2})$, where $Z \sim N(0,1)$ which is even better than the approximation given by $\mathbb{P}(Z \leq k)$.

X	Moment-generating function	Characteristic function
$c \in \mathbb{R}$	e^{tc}	$\mathrm{e}^{\mathrm{i}tc}$
$U(\{x_1, \ldots, x_n\})$	$\frac{1}{n}\sum_{i=1}^n \mathrm{e}^{tx_i}$	$\frac{1}{n}\sum_{i=1}^n \mathrm{e}^{\mathrm{i}tx_i}$
$\mathrm{B}(n,p)$	$(p\mathrm{e}^t + 1 - p)^n$	$(p\mathrm{e}^{\mathrm{i}t} + 1 - p)^n$
$\mathrm{Pois}(\lambda)$	$\mathrm{e}^{\lambda(\mathrm{e}^t - 1)}$	$\mathrm{e}^{\lambda(\mathrm{e}^{\mathrm{i}t} - 1)}$
$\mathrm{Geo}(p)$	$\dfrac{p\mathrm{e}^t}{1-(1-p)\mathrm{e}^t}$	$\dfrac{p\mathrm{e}^{\mathrm{i}t}}{1-(1-p)\mathrm{e}^{\mathrm{i}t}}$
$\mathrm{NB}(r,p)$	$\left(\dfrac{1-p}{1-p\mathrm{e}^t}\right)^r$ for $t < -\ln p$	$\left(\dfrac{1-p}{1-p\mathrm{e}^{\mathrm{i}t}}\right)^r$
$U(a,b)$	$\begin{cases} \frac{\mathrm{e}^{tb}-\mathrm{e}^{ta}}{t(b-a)} & \text{if } t \neq 0 \\ 1 & \text{if } t = 0 \end{cases}$	$\begin{cases} \frac{\mathrm{e}^{\mathrm{i}tb}-\mathrm{e}^{\mathrm{i}ta}}{\mathrm{i}t(b-a)} & \text{if } t \neq 0 \\ 1 & \text{if } t = 0 \end{cases}.$
$\mathrm{Exp}(\lambda)$	$\dfrac{\lambda}{\lambda - t}$ for $t < \lambda$	$\dfrac{\lambda}{\lambda - \mathrm{i}t}$
$N(\mu, \sigma^2)$	$\mathrm{e}^{\mu t + \frac{\sigma^2 t^2}{2}}$	$\mathrm{e}^{\mathrm{i}\mu t - \frac{\sigma^2 t^2}{2}}$
$\mathrm{Gamma}(\alpha, \beta)$	$\left(\dfrac{\beta}{\beta - t}\right)^\alpha$ for $t < \beta$	$\left(\dfrac{\beta}{\beta - \mathrm{i}t}\right)^\alpha$
$\mathrm{C}(x_0, \gamma)$	Does not exist	$\mathrm{e}^{\mathrm{i}tx_0 - \gamma\lvert t\rvert}$

Table 2: Moment-generating functions and characteristic functions of common distributions.

[28] In practice, the approximation is good enough for $np(1-p) \geq 18$.