3 Independence of Random Variables

3.1 Different Notions of Independence

In the following chapters, we will discuss two important theorems in the lecture of this semester: the law of large numbers (Theorem 4.3.1) and the central limit theorem (Theorem 4.5.3). Before stating these two theorems, we will need to define the notion of independence for different objects: events (Definition 3.1.1), σ-algebras (Definition 3.1.5) and random variables (Definition 3.1.6).

3.1.1 Independent Events

In this chapter, we consider a probability space $(\Omega, \mathcal{A}, \mathbb{P})$.

Let $A, B \in \mathcal{A}$. We say that A and B are *independent events* (獨立事件) if

$$\mathbb{P}(A \mid B) \overset{\text{(def)}}{=} \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \mathbb{P}(A). \tag{3.1}$$

In other words, "knowing that the event B holds does not change the probability that the event A occurs". However, in order to write Eq. (3.1), we need $\mathbb{P}(B) > 0$; additionally, this formula is not symmetric. This is the reason why we would rather define the notion of *two independent events* using the following condition,

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B). \tag{3.2}$$

From Eq. (3.2) we may also notice that, if A and B are independent events, then the following computation

$$\begin{aligned} \mathbb{P}(A^c \cap B) &= \mathbb{P}(B) - \mathbb{P}(A \cap B) = \mathbb{P}(B) - \mathbb{P}(A)\,\mathbb{P}(B) \\ &= (1 - \mathbb{P}(A))\,\mathbb{P}(B) = \mathbb{P}(A^c)\,\mathbb{P}(B), \end{aligned} \tag{3.3}$$

leads to the property that A^c and B are also independent events.

In general, when we deal with more than two events, even infinitely many events, we give the following definition.

Definition 3.1.1 : Given any set I and fix $A_i \in \mathcal{A}$ for any $i \in I$. We say that the family of events $(A_i)_{i \in I}$ indexed by I are independent if

$$\mathbb{P}\left(\bigcap_{j \in J} A_j \right) = \prod_{j \in J} \mathbb{P}(A_j), \tag{3.4}$$

for all finite subset $J \subseteq I$. And we call $(A_i)_{i \in I}$ *independent events* (獨立事件) .

Question 3.1.2: Given n events $A_1, \dots, A_n \in \mathcal{A}$. Are $A_1, \dots, A_n$ independent events if only one of the following conditions holds?

- $\mathbb{P}(A_1 \cap \cdots \cap A_n) = \mathbb{P}(A_1) \dots \mathbb{P}(A_n)$;
- for any pair $1 \leqslant i < j \leqslant n$, we have, $\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\,\mathbb{P}(A_j)$.

The following proposition gives an equivalent definition to Definition 3.1.1 when we only have finitely many events to consider.

Proposition 3.1.3 : *Given events $A_1, \dots, A_n \in \mathcal{A}$, then the following two properties are equivalent.*

(i) The events $A_1, \dots, A_n$ are independent.

(ii) For each $1 \leqslant i \leqslant n$, choose B_i among events in the set $\{\varnothing, A_i, A_i^c, \Omega\}$, then we have,

$$\mathbb{P}(B_1 \cap \dots \cap B_n) = \mathbb{P}(B_1) \dots \mathbb{P}(B_n). \tag{3.5}$$

Proof : First, let us prove (ii) $\Longrightarrow$ (i). We notice that the condition imposed by Eq. (3.5) is stronger than the condition imposed by Eq. (3.4) in Definition 3.1.1. Hence, when Eq. (3.5) holds, Eq. (3.4) also holds. More precisely, we may take $B_i = A_i$ if $i \in J$ and $B_i = \Omega$ otherwise in Eq. (3.5) so as to obtain Eq. (3.4).

Now, we want to prove (i) $\Longrightarrow$ (ii). If there exists B_i such that $B_i = \varnothing$, then we get 0 on both sides of Eq. (3.5). If there exists B_i such that $B_i = \Omega$, then on the left side of Eq. (3.5), we can remove B_i without changing the intersection; on the right side, $\mathbb{P}(B_i) = 1$ does not change the product either. Therefore, it is enough to prove that, for any subset $\{j_1, \dots, j_p\}$ of $\{1, \dots, n\}$, when $B_{j_k} = A_{j_k}$ or $A_{j_k}^c$, we have,

$$\mathbb{P}(B_{j_1} \cap \cdots \cap B_{j_p}) = \mathbb{P}(B_{j_1}) \dots \mathbb{P}(B_{j_p}).$$

To show this, it is enough to show that, if $C_1, \dots, C_p$ are independent events, then $C_1^c, C_2, \dots, C_p$ are also independent events. This can be obtained in the same way as the computation in Eq. (3.3). □

Corollary 3.1.4 : *Let $(A_i)_{i\in I}$ be a family of independent events. Fix a subset $J \subseteq I$, and define*

$$\forall i \in I, \qquad B_i = \begin{cases} A_i^c & \text{if } i \in J, \\ A_i & \text{if } i \notin J, \end{cases}$$

then $(B_i)_{i\in I}$ is also a family of independent events.

Proof : It is a direct consequence of Proposition 3.1.3. □

3.1.2 Independent σ-algebras and Independent Random Variables

We discussed the independence of measurable events above, in what follows, we are going to discuss the *independence of σ-algebras* and *independence of random variables*.

Definition 3.1.5 : Let $\mathcal{B}_1, \ldots, \mathcal{B}_n$ be n sub-σ algebras of $\mathcal{A}$. We say that $\mathcal{B}_1, \ldots, \mathcal{B}_n$ are *independent σ-algebras* (獨立 σ 代數) if

$$\mathbb{P}(A_1 \cap \cdots \cap A_n) = \mathbb{P}(A_1) \ldots \mathbb{P}(A_n), \qquad \forall A_1 \in \mathcal{B}_1, \ldots, \forall A_n \in \mathcal{B}_n.$$

Definition 3.1.6 : Let $X_1, \ldots, X_n$ be random variables with values in $(E_1, \mathcal{E}_1), \ldots, (E_n, \mathcal{E}_n)$ respectively. We say that $X_1, \ldots, X_n$ are *independent random variables* (獨立隨機變數) if the σ-algebras $\sigma(X_1), \ldots, \sigma(X_n)$ are independent.

Remark 3.1.7 : By the second point from Proposition 3.1.3, we know that the random variables $\mathbb{1}_{A_1}, \ldots, \mathbb{1}_{A_n}$ are independent if and only if the events $A_1, \ldots, A_n$ are independent.

Remark 3.1.8 : The fact that $X_1, \ldots, X_n$ are independent random variables is equivalent to the following property,

$$\forall F_1 \in \mathcal{E}_1, \ldots, \forall F_n \in \mathcal{E}_n, \qquad \mathbb{P}(\{X_1 \in F_1\} \cap \ldots \cap \{X_n \in F_n\}) = \mathbb{P}(X_1 \in F_1) \ldots \mathbb{P}(X_n \in F_n) \tag{3.6}$$

If $X_1, \ldots, X_n$ are random variables with values in $(E_1, \mathcal{E}_1), \ldots, (E_n, \mathcal{E}_n)$, then the n-tuple $(X_1, \ldots, X_n)$ is a random variable with values in $E_1 \times \ldots, E_n$ and is measurable with respect to the σ-algebra $\mathcal{E}_1 \otimes \cdots \otimes \mathcal{E}_n$. Below we provide another criterion to check the independence between random variables.

Theorem 3.1.9 : *Given random variables $X_1, \ldots, X_n$, the following three properties are equivalent.*

(i) The random variables $X_1, \ldots, X_n$ are independent.

(ii) The distribution of the n-tuple random variable $(X_1, \ldots, X_n)$ is the product distribution of $X_1, \ldots, X_n$, i.e.,

$$\mathbb{P}_{(X_1, \ldots, X_n)} = \mathbb{P}_{X_1} \otimes \ldots \otimes \mathbb{P}_{X_n}.$$

(iii) For all $i \in \{1, \ldots, n\}$ and any non-negative measurable function f_i defined on $(E_i, \mathcal{E}_i)$, we have,

$$\mathbb{E}\left[\prod_{i=1}^{n} f_i(X_i)\right] = \prod_{i=1}^{n} \mathbb{E}\left[f_i(X_i)\right]. \tag{3.7}$$

Remark 3.1.10 : If the measurable functions f_i are not non-negative, then under the assumption that $\mathbb{E}\left[|f_i(X_i)|\right] < \infty$ for all $i \in \{1, \cdots, n\}$, Eq. (3.7) still holds.

Remark 3.1.11 : We can observe that, if $X_1, \ldots, X_n$ are integrable and independent random variables, then their product $X_1 \ldots X_n$ is still integrable. However, in general, *a product of integrable random variables is not necessarily integrable.*

Proof : To show the equivalence between (i) and (ii), we proceed as below. For all $i \in \{1, \dots, n\}$, let $F_i \in \mathcal{E}_i$. We have the following formulas,

$$\mathbb{P}_{(X_1,\dots,X_n)}(F_1 \times \cdots \times F_n) = \mathbb{P}(\{X_1 \in F_1\} \cap \dots \cap \{X_n \in F_n\}),$$
$$\mathbb{P}_{\mathrm{X}_1} \otimes \dots \otimes \mathbb{P}_{\mathrm{X}_n}(F_1 \times \cdots \times F_n) = \prod_{i=1}^{n} \mathbb{P}_{X_i}(F_i) = \prod_{i=1}^{n} \mathbb{P}(X_i \in F_i).$$

From the above formulas and Eq. (3.6), $X_1, \dots, X_n$ are independent random variables if and only if $\mathbb{P}_{(X_1,\dots,X_n)}$ and $\mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_n}$ take the same values on all measurable sets of the form $F_1 \times \cdots \times F_n$. Thanks to the monotone class theorem (Theorem 1.4.1), we know that a measure on a product measurable space is characterized by its values on all $F_1 \times \cdots \times F_n$. This concludes the proof since we deduce that independence is equivalent to $\mathbb{P}_{(X_1,\dots,X_n)} = \mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_n}$.

Then, we show the equivalence between (ii) and (iii). Fix, for each $1 \leqslant i \leqslant n$, a non-negative measurable function f_i defined on $(E_i, \mathcal{E}_i)$. We write

$$\mathbb{E}\left[\prod_{i=1}^{n} f_i(X_i)\right] = \int_{E_1 \times \dots \times E_n} \prod_{i=1}^{n} f_i(x_i) \mathbb{P}_{(X_1,\dots,X_n)}(\mathrm{d}x_1 \dots \mathrm{d}x_n),$$

and write, using the Fubini's theorem,

$$\begin{aligned}
\prod_{i=1}^{n} \mathbb{E}\left[f_i(X_i)\right] &= \prod_{i=1}^{n} \int_{E_i} f_i(x_i)\, \mathbb{P}_{X_i}(\mathrm{d}x_i) \\
&= \int_{E_1 \times \dots \times E_n} \prod_{i=1}^{n} f_i(x_i)\, \mathbb{P}_{X_1}(\mathrm{d}x_1) \dots \mathbb{P}_{X_n}(\mathrm{d}x_n) \\
&= \int_{E_1 \times \dots \times E_n} \prod_{i=1}^{n} f_i(x_i)\, \mathbb{P}_{\mathrm{X}_1} \otimes \dots \otimes \mathbb{P}_{\mathrm{X}_n}(\mathrm{d}x_1 \dots \mathrm{d}x_n),
\end{aligned}$$

giving us the equivalence. □

Remark 3.1.12 (Construction of finitely many independent random variables) **:** As a consequence of the above theorem, we can construct independent random variables. Consider the case of real-valued random variables and assume that $\mu_1, \dots, \mu_n$ are probability measures on $\mathbb{R}$. Using Theorem 1.4.1 (construction of product measures) and Remark 2.1.11 (canonical construction of random variables), we can construct a random variable $Y = (Y_1, \dots, Y_n)$ with values in $\mathbb{R}^n$ whose distribution is given by $\mu_1 \otimes \cdots \otimes \mu_n$. The above theorem tells use that the components $Y_1, \dots, Y_n$ of the random variable Y are independent random variables and their distributions are respectively $\mu_1, \dots, \mu_n$.

Corollary 3.1.13 : *If X_1 and X_2 are independent real random variables in L^2, then we have* $\mathrm{Cov}(X_1, X_2) = 0$.

Remark 3.1.14 : The converse of this corollary is false. When the variance between two random variables is zero, we say that they are *uncorrelated*, but it does not imply their independence.

Question 3.1.15: Construct two random variables X and Y such that $\mathrm{Cov}(X,Y) = 0$ without X and Y being independent. In Exercise 3.20, we will see a specific condition under which $\mathrm{Cov}(X,Y) = 0$ implies the independence between X and Y.

Corollary 3.1.16 : *Let $X_1, \dots, X_n$ be n random variables.*

(1) Assume that, for all $i \in \{1, \dots, n\}$, X_i has a density, denoted p_i. If $X_1, \dots, X_n$ are independent random variables, then $(X_1, \dots, X_n)$ also has a density, given by,

$$p(x_1, \dots, x_n) = \prod_{i=1}^{n} p_i(x_i).$$

(2) Conversely, if $(X_1, \dots, X_n)$ has a probability density which writes,

$$p(x_1, \dots, x_n) = \prod_{i=1}^{n} q_i(x_i),$$

where functions q_i are non-negative measurable functions on $\mathbb{R}$, then $X_1, \dots, X_n$ are independent random variables. Moreover, for all $i \in \{1, \dots, n\}$, the random variable X_i also has a probability density and there exists a constant $C_i > 0$ such that its density function p_i writes $p_i = C_i q_i$.

Proof : (1) is an application of Theorem 3.1.9 and Fubini's theorem. Since $\mathbb{P}_{X_i}(\mathrm{d}x_i) = p_i(x_i)\,\mathrm{d}x_i$, we can write the product measure as,

$$\mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_n}(\mathrm{d}x_1 \dots \mathrm{d}x_n) = \Big(\prod_{i=1}^{n} p_i(x_i) \Big)\, \mathrm{d}x_1 \dots \mathrm{d}x_n.$$

Next, we prove (2). Let $K_i = \int q_i(x)\,\mathrm{d}x \in (0, \infty)$ for all $i \in \{1, \dots, n\}$. We first note that, Fubini's theorem gives,

$$\prod_{i=1}^{n} K_i = \prod_{i=1}^{n} \Big(\int q_i(x)\,\mathrm{d}x \Big) = \int_{\mathbb{R}^n} p(x_1, \dots, x_n)\,\mathrm{d}x_1 \dots \mathrm{d}x_n = 1.$$

Then, from Proposition 2.1.18, we can compute the marginal distributions of $X = (X_1, \dots, X_n)$, i.e., the marginal distribution of X_i writes,

$$p_i(x_i) = \int_{\mathbb{R}^{n-1}} p(x_1, \dots, x_n)\,\mathrm{d}x_1 \dots \mathrm{d}x_{i-1}\,\mathrm{d}x_{i+1} \dots \mathrm{d}x_n = \Big(\prod_{j \neq i} K_j \Big) q_i(x_i) = \frac{1}{K_i} q_i(x_i).$$

From above we know that $\mathbb{P}_{(X_1,\dots,X_n)} = \mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_n}$, which means that the random variables are independent. □

Question 3.1.17: Let $X_1, \dots, X_n$ be real-valued random variables. The following properties are equivalent.

(i) $X_1, \dots, X_n$ are independent random variables.

(ii) For any $a_1, \dots, a_n \in \mathbb{R}$, we have $\mathbb{P}(X_1 \leqslant a_1, \dots, X_n \leqslant a_n) = \prod_{i=1}^{n} \mathbb{P}(X_i \leqslant a_i)$.

(iii) If $f_1, \dots, f_n$ are continuous and compactly supported (**緊緻支撐**) functions from $\mathbb{R}$ to $\mathbb{R}_{\geqslant 0}$, then

$$\mathbb{E}\left[\prod_{i=1}^{n} f_i(X_i)\right] = \prod_{i=1}^{n} \mathbb{E}\left[f_i(X_i)\right].$$

(iv) The characteristic function of X writes,

$$\Phi_X(\xi_1, \dots, \xi_n) = \prod_{i=1}^{n} \Phi_{X_i}(\xi_i).$$

The following proposition is an application of the monotone class theorem. It is slightly technical but will be very useful later.

Proposition 3.1.18 : *Let $\mathcal{B}_1, \dots, \mathcal{B}_n$ be sub-σ-algebras of $\mathcal{A}$. For all $i \in \{1, \dots, n\}$, let $\mathcal{C}_i \subseteq \mathcal{B}_i$ be a subset that is closed under finite intersections, containing Ω and such that $\sigma(\mathcal{C}_i) = \mathcal{B}_i$. If*

$$\forall C_1 \in \mathcal{C}_1, \dots, \forall C_n \in \mathcal{C}_n, \quad \mathbb{P}(C_1 \cap \dots \cap C_n) = \mathbb{P}(C_1) \dots \mathbb{P}(C_n),$$

then $\mathcal{B}_1, \dots, \mathcal{B}_n$ are independent σ-algebras.

Proof : First, we fix $C_2 \in \mathcal{C}_2, \dots, C_n \in \mathcal{C}_n$ and let

$$\mathcal{M}_1 = \{B_1 \in \mathcal{B}_1 : \mathbb{P}(B_1 \cap C_2 \cap \dots \cap C_n) = \mathbb{P}(B_1)\,\mathbb{P}(C_2) \dots \mathbb{P}(C_n)\}.$$

We can easily check that $\mathcal{C}_1 \subseteq \mathcal{M}_1$ and that $\mathcal{M}_1$ is a monotone class. Thus, from the monotone class lemma, $\mathcal{M}_1$ contains $\sigma(\mathcal{C}_1) = \mathcal{B}_1$. Now, we have shown that if $\mathcal{C}_1, \dots, \mathcal{C}_n$ are independent σ-algebras, then $\sigma(\mathcal{C}_1), \mathcal{C}_2, \dots, \mathcal{C}_n$ are also independent σ-algebras. By induction, we apply the same proof to $\mathcal{C}_2, \dots, \mathcal{C}_n, \sigma(\mathcal{C}_1)$ to show that $\sigma(\mathcal{C}_2), \mathcal{C}_3, \dots, \mathcal{C}_n, \sigma(\mathcal{C}_1)$ are independent σ-algebras. This copmletes the proof. □

Question 3.1.19: Let $\mathcal{B}_1, \dots, \mathcal{B}_n$ be independent σ-algebras and $n_0 = 0 < n_1 < \dots < n_p = n$. Then, the following σ-algebras are independent,

$$\begin{aligned}
\mathcal{D}_1 &= \mathcal{B}_1 \vee \dots \vee \mathcal{B}_{n_1} \overset{\text{(def)}}{=} \sigma(\mathcal{B}_1, \dots, \mathcal{B}_{n_1}),\\
\mathcal{D}_2 &= \mathcal{B}_{n_1+1} \vee \dots \vee \mathcal{B}_{n_2},\\
&\vdots\\
\mathcal{D}_p &= \mathcal{B}_{n_{p-1}+1} \vee \dots \vee \mathcal{B}_{n_p}.
\end{aligned}$$

3.1.3 Independence for Infinitely Many Random Variables

We are going to define the notion of independence when we have an infinite family of random variables. This can be reduced to the independence condition on all the finite subsets.

Definition 3.1.20 :

(1) Let $(B_i)_{i \in I}$ be a collection of sub-σ-algebras of $\mathcal{A}$ indexed by I. If for any *finite* subset $\{i_1, \dots, i_p\}$ of I, the σ-algebras $\mathcal{B}_{i_1}, \dots, \mathcal{B}_{i_p}$ are independent, then we say that $(B_i)_{i \in I}$ is a

collection of independent σ-algebras.

(2) Similarly, for any collection of random variables $(X_i)_{i\in I}$, if $(\sigma(X_i))_{i\in I}$ is a collection of independent σ-algebras, then we say that $(X_i)_{i\in I}$ is a *collection of independent random variables.*

Proposition 3.1.21 : *Let $(X_n)_{n\geqslant 1}$ be a sequence of independent random variables. Then for any positive integer $p \geqslant 1$ and subsets of integers $I_1, \dots, I_p \subseteq \mathbb{N}$ that are pairwise disjoint, the subsets $(\mathcal{B}_k)_{1\leqslant k\leqslant p}$ defined by,*

$$\forall k = 1, \dots, p, \quad \mathcal{B}_k = \sigma(X_i : i \in I_k)$$

are independent σ-algebras.

Proof : For $1 \leqslant k \leqslant p$, let

$$\mathcal{C}_k = \bigcup_{\substack{J\subseteq I_k \\ J \text{ is finite}}} \sigma(X_j : j \in J) \subseteq \mathcal{B}_k.$$

From Question 3.1.19, for any finite subsets $J_1 \subseteq I_1, \dots, J_p \subseteq I_p$, the σ-algebras $\sigma(X_j : j \in J_1), \dots, \sigma(X_j : j \in J_p)$ are independent. Then from Proposition 3.1.18, we know that $\mathcal{B}_1 = \sigma(\mathcal{C}_1), \dots, \mathcal{B}_p = \sigma(\mathcal{C}_p)$ are also independent σ-algebras. □

Corollary 3.1.22 : *Let $(X_n)_{n\geqslant 1}$ be a sequence of independent random variables, where X_n takes values in a measurable space $(E_n, \mathcal{E}_n)$ for $n \geqslant 1$. Let $(I_n)_{n\geqslant 1}$ be a sequence of pairwise disjoint subsets of $\mathbb{N}$ and $(f_n)_{n\geqslant 1}$ be a sequence of measurable functions, where f_n is defined on $F_n := \prod_{i\in I_n} E_i$ for $n \geqslant 1$. Then, the following random variables are independent*

$$\forall n \geqslant 1, \quad Z_n = f_n(X_i : i \in I_n).$$

In Remark 3.1.12, we explained how to construct finitely many independent random variables. Below, we explain how to achieve this for *countably infinitely many* independent random variables.

Lemma 3.1.23 : *Let $Y \sim \text{Unif}([0,1])$ be a random variable with uniform distribution on $[0,1]$. Then, the dyadical expansion of Y, denoted*

$$Y = 0.Y_1Y_2\cdots = \sum_{n\geqslant 1} 2^{-n}Y_n, \quad Y_n \in \{0,1\},\ \forall n \geqslant 1 \tag{3.8}$$

satisfies the following properties.

(1) The expansion in Eq. (3.8) is almost surely unique.

(2) The random variables $(Y_n)_{n\geqslant 1}$ are independent and each of them follows the distribution $\text{Ber}(\frac{1}{2})$.

Proof : We recall that a dyadical expansion can be obtained by

$$\begin{aligned} Y_1 &= \lfloor 2Y \rfloor, & X_1 &= 2Y - \lfloor 2Y \rfloor = 2Y - Y_1, \\ \forall n \geqslant 1, \quad Y_{n+1} &= \lfloor 2X_n \rfloor, & X_{n+1} &= 2X_n - Y_{n+1}. \end{aligned}$$

By a direct induction, we may also rewrite,

$$\forall n \geqslant 1, \quad Y_n = \left\lfloor 2^n Y - \sum_{k=1}^{n-1} 2^{n-k} Y_k \right\rfloor. \tag{3.9}$$

(1) For $x \in [0,1)$, it has two distinct dyadical expansion if and only if it is a dyadical rational $\frac{m}{2^n}$ for some integers $n \geqslant 1$ and $0 \leqslant 1 \leqslant 2^n - 1$. This can be checked by the following fact,

$$\forall n \in \mathbb{N}, \qquad \frac{1}{2^n} = \sum_{k \geqslant n+1} \frac{1}{2^k}.$$

To conclude, we note that the subset consisting of all the dyadical rationals has measure zero,

$$\mathbb{P}\left(\bigcup_{n \geqslant 1} \bigcup_{m=0}^{2^n - 1} \left\{ \frac{m}{2^n} \right\} \right) = 0.$$

(2) First, let us check the distribution of Y_1 and X_1. Y_1 follows $\mathrm{Ber}(\frac{1}{2})$,

$$\mathbb{P}(Y_1 = 0) = \mathbb{P}(Y \in [0, \tfrac{1}{2})) = \tfrac{1}{2}.$$

For $0 \leqslant a < b < 1$, we have

$$\begin{aligned} \mathbb{P}(X_1 \in [a,b]) &= \mathbb{P}(X_1 \in [a,b], Y_1 = 0) + \mathbb{P}(X_1 \in [a,b], Y_1 = 1) \\ &= \mathbb{P}(Y \in [\tfrac{a}{2}, \tfrac{b}{2}]) + \mathbb{P}(Y \in [\tfrac{a+1}{2}, \tfrac{b+1}{2}]) \\ &= (\tfrac{b}{2} - \tfrac{a}{2}) + (\tfrac{b+1}{2} - \tfrac{a+1}{2}) = b - a. \end{aligned}$$

Therefore, X_1 follows the uniform distribution on $[0,1]$. We conclude by induction that $Y_n \sim \mathrm{Ber}(\frac{1}{2})$ for all $n \geqslant 1$.

Let $n \geqslant 1$ be an integer. We want to check that $(Y_k)_{1 \leqslant k \leqslant n}$ are independent random variables. Let $m_1, \ldots, m_n \in \{0,1\}$ and compute

$$\begin{aligned} &\mathbb{P}(Y_\ell = m_\ell, \forall \ell = 1, \ldots n) = \mathbb{P}\left(\left\lfloor 2^\ell Y - \sum_{k=1}^{\ell-1} 2^{\ell-k} m_k \right\rfloor = m_\ell, \forall \ell = 1, \ldots n \right) \\ &= \mathbb{P}\left(m_\ell \leqslant 2^\ell Y - \sum_{k=1}^{\ell-1} 2^{\ell-k} m_k < m_\ell + 1, \forall \ell = 1, \ldots n \right) \\ &= \mathbb{P}\left(\sum_{k=1}^{\ell} 2^{-k} m_k \leqslant Y < \sum_{k=1}^{\ell} 2^{-k} m_k + 2^{-\ell}, \forall \ell = 1, \ldots n \right) \\ &= \mathbb{P}\left(\sum_{k=1}^{n} 2^{-k} m_k \leqslant Y < \sum_{k=1}^{n} 2^{-k} m_k + 2^{-n} \right) = 2^{-n}. \end{aligned}$$

This allows us to conclude the independence. □

Remark 3.1.24 : Let us consider $Y \sim \mathrm{Unif}([0,1])$ and its dyadical expansion $Y = 0.Y_1Y_2\ldots$ as in Eq. (3.8). Since $\mathbb{N}^2$ and $\mathbb{N}$ are equipotent, we may find a bijective function $f : \mathbb{N}^2 \to \mathbb{N}$. For every $i \in \mathbb{N}$, the random variables $(Y_{f(i,j)})_{j\geqslant 1}$ are independent, and by defining

$$\forall i \in \mathbb{N}, \quad Z_i = 0.Y_{f(i,1)}Y_{f(i,2)}\cdots = \sum_{n\geqslant 1} 2^{-n} Y_{f(i,n)},$$

it follows from Definition 3.1.20 and Proposition 3.1.21 that $(Z_i)_{i\geqslant 1}$ are independent random variables. Moreover, every Z_i follows the uniform distribution on $[0,1]$. If we are given distributions $(\mu_i)_{i\geqslant 1}$ on $\mathbb{R}$, we may denote their cumulative distribution function by $(F_i = F_{\mu_i})_{i\geqslant 1}$, then the random variables $(X_i)_{i\geqslant 1}$ defined by

$$\forall i \geqslant 1, \quad X_i = \inf\{y \in \mathbb{R} : F_i(y) \geqslant Z_i\}$$

are independent random variables with $X_i \sim \mu_i$ for $i \geqslant 1$, see Proposition 2.1.23.

If we want to construct uncountably many independent random variables, we need to use the following Kolmogorov's extension theorem (Kolmogorov 拓延定理).

Theorem 3.1.25 (Kolmogorov's extension theorem) **:** *Given the measurable space $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ and any set T. Assume that the two following conditions hold.*

(a) For any finite *subset S of T, we can construct a probability measure $\mathbb{P}_S$ on $((\mathbb{R}^d)^{\otimes S}, \mathcal{B}(\mathbb{R}^d)^{\otimes S})$.*

(b) For any finite *subsets S_1 and S_2 such that $S_1 \subseteq S_2$, the probability measures $\mathbb{P}_{S_1}$ and $\mathbb{P}_{S_2}$ are compatible (相容), meaning that $\mathbb{P}_{S_1} = \mathbb{P}_{S_2} \circ \pi^{-1}$, where π denotes the projection from S_2 to S_1.*

Then there exists a unique measure $\mathbb{P}$ on $((\mathbb{R}^d)^{\otimes T}, \mathcal{B}(\mathbb{R}^d)^{\otimes T})$ such that for any finite subset S, we have $\mathbb{P}_S = \mathbb{P} \circ \pi^{-1}$, where π is the projection from T to S.

Proof : We define

$$\mathcal{C} := \bigcup_{\substack{S\subseteq T \\ S \text{ is finite}}} \mathcal{B}(\mathbb{R}^d)^{\otimes S},$$

which is the set consisting of the elements in finite-dimensional product σ-algebras. Moreover, we know that $\mathcal{B}(\mathbb{R}^d)^{\otimes T}$ can be generated by $\mathcal{C}$, that is,

$$\mathcal{B}(\mathbb{R}^d)^{\otimes T} = \sigma(\mathcal{C}).$$

The proof of the existence can be achieved using the outer measure, as for the construction of the Lebesgue measure. For the uniqueness, it is a direct consequence of the monotone class lemma, see Corollary 1.1.19. □

3.2 Borel–Cantelli Lemma

Previously, we defined different notions of independence. In this section, we will use the independence to deduce some asymptotic results.

3.2.1 Statement and Proof

Definition 3.2.1 : Given a sequence $(A_n)_{n\geqslant 1}$ of events, we define the following notions.

(1) The following set is called the *upper limit* (上極限) ,

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

(2) The following set is called the *lower limit* (下極限) ,

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

(3) If the upper limit and the lower limit of (A_n) coincide, then we call this common value its *limit* (極限) , denoted $\lim_{n\to\infty} A_n := \liminf_{n\to\infty} A_n = \limsup_{n\to\infty} A_n$.

(4) If the upper limit and the lower limit of (A_n) differ, then we say that the limit of (A_n) does not exist.

Proposition 3.2.2 : *Let (A_n) be a sequence of events. Then,*

(1) $\limsup A_n = \{\omega \mid \omega \in A_n \text{ i.o.}\}$*, where i.o. stands for "infinitely often", meaning that there exists an infinity of n such that $\omega \in A_n$.*

(2) $\liminf A_n = \{\omega \mid \omega \in A_n \text{ a.a.}\}$*, where a.a. stands for "almost all", meaning that except for a finite number of n, we have $\omega \in A_n$.*

(3) $\liminf A_n \subseteq \limsup A_n$.

Proof : See Exercise 1.12. □

Example 3.2.3 : Let $(X_n)_{n\geqslant 1}$ be a sequence of random variables and $a \in \mathbb{R}$.

(1) $\omega \in \liminf\{X_n \leqslant a\} \Rightarrow \limsup X_n(\omega) \leqslant a.$

(2) $\omega \in \liminf\{X_n \geqslant a\} \Rightarrow \liminf X_n(\omega) \geqslant a.$

(3) $\omega \in \limsup\{X_n \leqslant a\} \Rightarrow \liminf X_n(\omega) \leqslant a.$

(4) $\omega \in \limsup\{X_n \geqslant a\} \Rightarrow \limsup X_n(\omega) \geqslant a.$

Lemma 3.2.4 : *Let $(A_n)_{n\geqslant 1}$ be a sequence of events.*

(1) If $\sum_{n\geqslant 1} \mathbb{P}(A_n)_{n\geqslant 1} < \infty$, then

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

In other words, the set $\{n \in \mathbb{N} \mid \omega \in A_n\}$ is a.s. finite.

(2) If $\sum_{n\geqslant 1} \mathbb{P}(A_n) = \infty$ and $(A_n)_{n\geqslant 1}$ is a sequence of independent events, then

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

In other words, the set $\{n \in \mathbb{N} \mid \omega \in A_n\}$ is a.s. infinite.

Question 3.2.5: Explain why it is necessary to assume that $(A_n)_{n\geqslant 1}$ is an independent sequence of events in (2) of Lemma 3.2.4.

Proof :

(1) From the assumption, we have,

$$\mathbb{E}\left[\sum_{n\geqslant 1} \mathbb{1}_{A_n}\right] = \sum_{n\geqslant 1} \mathbb{P}(A_n) < \infty,$$

meaning that $\sum_{n\geqslant 1} \mathbb{1}_{A_n} < \infty$ almost surely.

(2) Given $n_0 \in \mathbb{N}$, for all $n \geqslant n_0$, we have,

$$\mathbb{P}\left(\bigcap_{k=n_0}^{n} A_k^c\right) = \prod_{k=n_0}^{n} \mathbb{P}(A_k^c) = \prod_{k=n_0}^{n} (1 - \mathbb{P}(A_k)).$$

Since the series $\sum_{k\geqslant 1} \mathbb{P}(A_k)$ diverges, we get,

$$\mathbb{P}\left(\bigcap_{k=n_0}^{\infty} A_k^c\right) = \lim_{n\to\infty} \downarrow \mathbb{P}\left(\bigcap_{k=n_0}^{n} A_k^c\right) = 0.$$

Since the above formula holds for all n_0, we have,

$$\mathbb{P}\left(\bigcup_{n_0=1}^{\infty} \left(\bigcap_{k=n_0}^{\infty} A_k^c\right)\right) = 0.$$

We take its complement and we obtain what needs to be proved,

$$\mathbb{P}\left(\bigcap_{n_0=1}^{\infty} \left(\bigcup_{k=n_0}^{\infty} A_k\right)\right) = 1.$$

□

3.2.2 Applications

In this subsection, we will apply Borel–Cantelli lemma to prove the following result. Other applications are treated in exercises.

Proposition 3.2.6 : *There does not exist any probability measure on $\mathbb{N}$, such that for any positive integer $n \geqslant 1$, the set of its multiples $n\mathbb{N}$ has measure $\frac{1}{n}$.*

Proof : Assume that such a measure exists and is denoted $\mathbb{P}$. Let $\mathcal{P}$ be the set of prime numbers and let $A_p = p\mathbb{N}$ for all $p \in \mathcal{P}$. We know that $(A_p)_{p\in\mathcal{P}}$ are independent events. Indeed, for any distinct prime numbers $p_1, \dots, p_k \in \mathcal{P}$, we have,

$$\mathbb{P}\left(A_{p_1} \cap \ldots \cap A_{p_n}\right) = \mathbb{P}\left(p_1\mathbb{N} \cap \ldots \cap p_n\mathbb{N}\right) = \mathbb{P}\left((p_1 \ldots p_n)\mathbb{N}\right) = \frac{1}{p_1 \ldots p_k} = \prod_{j=1}^{k} \mathbb{P}(A_{p_j}).$$

Moreover, due to the fact that $\sum_{p\in\mathcal{P}} \frac{1}{p} = \infty$, we get, from (2) of Borel–Cantelli lemma, that $\mathbb{P}(\limsup A_p) = 1$, meaning that under the measure $\mathbb{P}$, almost every integer n appears in an infinite number of A_p This is impossible, since a positive integer cannot be a multiple of infinitely many prime numbers. □

Proposition 3.2.7 : *Let $Y \sim \mathrm{Unif}([0,1])$ with dyadical expansion $Y = 0.Y_1Y_2\ldots$ as in Eq. (3.8). For any integer $p \geqslant 1$ and $m_1, \dots, m_p \in \{0,1\}$, almost surely there exist infinitely many $k \in \mathbb{N}$ such that*

$$X_{k+1} = m_1, \dots, X_{k+p} = m_p.$$

Proof : For any positive integer $n \geqslant 1$, define the random vector $Z_n = (Y_{np+1}, \dots, Y_{np+p})$. From Corollary 3.1.22, the random variables $(Z_n)_{n\geqslant 1}$ are independent. They are also identically distributed thanks to Lemma 3.1.23. For every $n \geqslant 1$, we have

$$\mathbb{P}(Z_n = (m_1, \dots, m_p)) = 2^{-p}.$$

Since the events $(\{Z_n = (m_1, \dots, m_p)\})_{n\geqslant 1}$ are independent and $\sum_{n\geqslant 1} 2^{-p} = +\infty$, Lemma 3.2.4 (2) implies that

$$\mathbb{P}\left(\limsup_{n\to\infty} \{Z_n = (m_1, \dots, m_p)\}\right) = 1.$$

□

Proposition 3.2.8 : *Let $(X_n)_{n\geqslant 1}$ be a sequence of i.i.d. random variables with distribution $\mathrm{Ber}(\frac{1}{2})$. Let*

$$L_n := \max\{k \geqslant 1 : \text{ there exists } 0 \leqslant i \leqslant n-k \text{ such that } X_{i+1} = \cdots = X_{i+k} = 1\}.$$

Then we have

$$\limsup_{n\to\infty} \frac{L_n}{\ln_2(n)} \leqslant 1 \leqslant \liminf_{n\to\infty} \frac{L_n}{\ln_2(n)}, \qquad \text{a.s.}$$

that is

$$\frac{L_n}{\ln_2(n)} \xrightarrow[n\to\infty]{} 1, \qquad \text{a.s.}$$

Proof : We want to show that the following hold for all $\varepsilon > 0$,

$$(1)\ \limsup_{n\to\infty} \frac{L_n}{\ln_2(n)} \leqslant 1+\varepsilon,\ \text{a.s.}, \qquad (2)\ \liminf_{n\to\infty} \frac{L_n}{\ln_2(n)} \geqslant 1-\varepsilon,\ \text{a.s.}$$

First, let us introduce a few more notations,

$$\forall m \geqslant 1, \quad \ell_m = \sup\{k \geqslant 0 : X_m = \cdots = X_{m+k-1} = 1\},$$

which is the maximal number of occurrences of consecutive 1's. Therefore, we have

$$L_n = \sup_{1\leqslant m\leqslant n} \{\ell_m \wedge (n-m+1)\} \leqslant \sup_{1\leqslant m\leqslant n} \{\ell_m\} =: \widetilde{L}_n.$$

Let us start with proving (1). Given $\varepsilon > 0$, we have,

$$\mathbb{P}\left(\ell_m \geqslant (1+\varepsilon)\log_2(m)\right) = \mathbb{P}\left(\ell_m \geqslant \lceil (1+\varepsilon)\log_2(m)\rceil\right) = \left(\frac{1}{2}\right)^{\lceil (1+\varepsilon)\log_2(m)\rceil} \leqslant m^{-(1+\varepsilon)}.$$

Since $\sum m^{-(1+\varepsilon)} < \infty$, so from Lemma 3.2.4 (1), we find,

$$\begin{aligned} &\mathbb{P}\left(\limsup\{\omega : \ell_m(\omega) \geqslant (1+\varepsilon)\log_2(m)\}\right) = 0, \\ \implies \quad &\mathbb{P}\left(\liminf\{\omega : \ell_m(\omega) < (1+\varepsilon)\log_2(m)\}\right) = 1. \end{aligned}$$

This implies

$$\limsup_{m\to\infty} \frac{\ell_m}{\log_2(m)} \leqslant 1+\varepsilon, \quad \text{a.s.}$$

As a consequence,

$$\limsup_{n\to\infty} \frac{L_n}{\log_2(n)} \leqslant \limsup_{n\to\infty} \frac{\widetilde{L}_n}{\log_2(n)} \leqslant 1+\varepsilon, \quad \text{a.s.}$$

Next, let us show (2). Given $\varepsilon > 0$, we divide the n experiments into intervals of length $a_n := \lceil (1-\varepsilon)\log_2(n)\rceil + 1$, where the last one can be shorter than a_n. So the total number of intervals is

$$N_n := \left\lceil \frac{n}{a_n} \right\rceil \xrightarrow[n\to\infty]{} \frac{n}{(1-\varepsilon)\log_2(n)} > \frac{n}{\log_2(n)}. \tag{3.10}$$

We denote these intervals as follow,

$$\begin{aligned} \forall k = 1, \ldots, N_n - 1, \quad I_k &= \{(k-1)a_n + 1, \ldots, ka_n\}, \\ I_{N_n} &= \{(N_n - 1)a_n + 1, \ldots, n\}. \end{aligned}$$

For all $1 \leqslant j \leqslant N_n - 1$, let A_j be the event that $\{X_i = 1, \forall i \in I_j\}$. We have,

$$\mathbb{P}(A_j) = \left(\frac{1}{2}\right)^{a_n} \geqslant \left(\frac{1}{2}\right)^{(1-\varepsilon)\log_2(n)+2} = \frac{1}{4}\frac{1}{n^{1-\varepsilon}}.$$

Next, we look at the probability of an event related to L_n,

$$\mathbb{P}\left(L_n \leqslant (1-\varepsilon)\log_2(n)\right) \leqslant \mathbb{P}\left(\text{there exists at least an } i \in I_j \text{ such that } X_i \neq 1 : 1 \leqslant j \leqslant N_n - 1\right)$$
$$= \prod_{j=1}^{N_n-1} (1 - \mathbb{P}(A_j)) \leqslant \left(1 - \frac{1}{4}\frac{1}{n^{1-\varepsilon}}\right)^{N_n-1}.$$

Using Eq. (3.10), for large enough n, we have,

$$\mathbb{P}\left(L_n \leqslant (1-\varepsilon)\log_2(n)\right) \leqslant \left(1 - \frac{1}{4}\frac{1}{n^{1-\varepsilon}}\right)^{n/\log_2(n)} \sim \exp\left(-\frac{1}{4}\frac{n^{\varepsilon}}{\log_2(n)}\right).$$

If we sum up the right side of the above formula in n, we obtain a finite sum. Using Lemma 3.2.4 (1), we find,

$$\liminf_{n\to\infty} \frac{L_n}{\log_2(n)} \geqslant 1-\varepsilon, \quad \text{a.s.}$$

□

3.3 Sum of Independent Random Variables

3.3.1 Definition and Properties

If μ and ν are both probability measures on $\mathbb{R}^d$, we write $\mu * \nu$ for the image measure of $\mu \otimes \nu$ under the function $(x, y) \mapsto x + y$. This means that the measure $\mu * \nu$ has the following property: for any non-negative measurable function φ on $\mathbb{R}^d$, we have,

$$\int_{\mathbb{R}^d} \varphi(z)\mu * \nu(\mathrm{d}z) = \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \varphi(x+y)\mu(\mathrm{d}x)\nu(\mathrm{d}y). \tag{3.11}$$

Proposition 3.3.1 : *Let X and Y be two independent random variables on $\mathbb{R}^d$.*

(1) The distribution of the random variable $X + Y$ is $\mathbb{P}_X * \mathbb{P}_Y$. In the case that both X and Y have a density, denoted respectively p_X and p_Y, then $X + Y$ also has a density which writes $p_X * p_Y$.

(2) The characteristic function of the random variable $X+Y$ writes $\Phi_{X+Y}(\xi) = \Phi_X(\xi)\Phi_Y(\xi)$. Equivalently, $\widehat{\mathbb{P}_X * \mathbb{P}_Y} = \widehat{\mathbb{P}}_X\widehat{\mathbb{P}}_Y$.

(3) If X and Y are both square integrable, then $K_{X+Y} = K_X + K_Y$. In the one-dimensional case $d = 1$, we have $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$.

Proof :

(1) Since X and Y are independent, we have $\mathbb{P}_{(X,Y)} = \mathbb{P}_X \otimes \mathbb{P}_Y$. So for any non-negative measurable function φ on $\mathbb{R}^d$, from the definition of the operator $*$ in Eq. (3.11), we have,

$$\mathbb{E}[\varphi(X+Y)] = \int \varphi(x+y)\mathbb{P}_{(X,Y)}(\mathrm{d}x\,\mathrm{d}y) = \int\int \varphi(x+y)\mathbb{P}_X(\mathrm{d}x)\mathbb{P}_Y(\mathrm{d}y) = \int \varphi(z)\mathbb{P}_X * \mathbb{P}_Y(\mathrm{d}z)$$

Next, if X and Y have a density,

$$\mathbb{E}[\varphi(X+Y)] = \int\int \varphi(x+y)p_X(x)p_Y(y)\,\mathrm{d}x\,\mathrm{d}y = \int \varphi(z)\int \underbrace{\big(p_X(x)p_Y(z-x)\,\mathrm{d}x\big)}_{p_X * p_Y(z)}\,\mathrm{d}z,$$

so $p_X * p_Y$ is the density function of $X+Y$. We notice that since p_X and p_Y are functions in $L^1(\mathbb{R}^d,\lambda)$, so $p_X * p_Y$ is well defined almost everywhere, see Proposition 1.4.3.

(2) From Definition 2.4.12 and the independence between X and Y, we have,

$$\Phi_{X+Y}(\xi) = \mathbb{E}[\exp(\mathrm{i}\,\xi\cdot(X+Y))] = \mathbb{E}[\exp(\mathrm{i}\,\xi\cdot X)]\,\mathbb{E}[\exp(\mathrm{i}\,\xi\cdot Y)] = \Phi_X(\xi)\Phi_Y(\xi)$$

(3) We write X and Y as $X=(X_1,\dots,X_d)$ and $Y=(Y_1,\dots,Y_d)$. Using their independence, for all $i,j\in\{1,\dots,d\}$, we have $\mathrm{Cov}(X_i,Y_j)=0$ and $\mathrm{Cov}(X_i+Y_i,X_j+Y_j)=\mathrm{Cov}(X_i,X_j)+\mathrm{Cov}(Y_i,Y_j)$, meaning that $K_{X+Y}=K_X+K_Y$.

□

3.3.2 Examples

If all the terms in a sequence of independent random variables $(X_n)_{n\geqslant 1}$ have the same distribution, we call it an i.i.d. sequence of random variables, standing for *independent and identically distributed*. It is not hard to find the distribution of a sum of i.i.d. random variables, one may proceed using the characteristic function or directly the definition in Eq. (3.11). Below we give a few examples, more examples being available in Exercise 3.13, Exercise 3.14, and Exercise 3.15.

Proposition 3.3.2 : *If $(X_k)_{1\leqslant k\leqslant n}$ is a sequence of i.i.d. random variables where each term follows the Poisson distribution of parameter λ, then $X_1+\cdots+X_n$ is a Poisson distribution of parameter $n\lambda$. More generally, if $(X_k)_{1\leqslant k\leqslant n}$ is a sequence of independent Poisson random variables with parameters $\lambda_1,\dots,\lambda_n$, then $X_1+\cdots+X_n$ is a Poisson distribution of parameter $\lambda_1+\cdots+\lambda_n$.*

Proof : Let $X_i\sim\mathrm{Pois}(\lambda_i)$ be independent random variables with Poisson distribution. We may compute their characteristic functions,

$$\Phi_{X_i}(\xi) = \sum_{k\geqslant 0} e^{-\lambda_i}\frac{\lambda^k}{k!}e^{\mathrm{i}\,\xi k} = \sum_{k\geqslant 0} e^{-\lambda_i}\frac{(\lambda e^{\mathrm{i}\,\xi})^k}{k!} = e^{\lambda_i(e^{\mathrm{i}\,\xi}-1)}.$$

Therefore, the characteristic function of $X:=X_1+\cdots+X_n$ writes,

$$\Phi_X(\xi) = \prod_{i=1}^n \Phi_{X_i}(\xi) = e^{(\lambda_1+\cdots+\lambda_n)(e^{\mathrm{i}\,\xi}-1)},$$

which we recognize as the characteristic function of $\mathrm{Pois}(\lambda_1+\cdots+\lambda_n)$. By Theorem 2.4.15, we deduce that $X\sim\mathrm{Pois}(\lambda_1+\cdots+\lambda_n)$. □

Proposition 3.3.3 : *If $(X_k)_{1\leqslant k\leqslant n}$ is a sequence of independent random variables such that for all $1 \leqslant k \leqslant n$, X_k has the Gaussian distribution of parameter $(0, \sigma_k^2)$, then $X_1 + \cdots + X_n$ has the Gaussian distribution of parameter $(0, \sigma_1^2 + \cdots + \sigma_n^2)$.*

Proof : See Exercise 3.12. □

3.3.3 Law of Large Numbers

We will only discuss formally different notions of convergence of random variables in the next chapter. However, with what we have learned so far, we can motivate and state some eaiser versions of this law.

Theorem 3.3.4 (Law of large numbers in L^2) **:** *Let $(X_n)_{n\geqslant 1}$ be a sequence of uncorrelated real-valued random variables with the same distribution. Suppose that $\mathbb{E}[X_1^2] < \infty$. Then we have,*

$$\frac{1}{n}(X_1 + \dots X_n) \xrightarrow[n\to\infty]{L^2} \mathbb{E}[X_1].$$

Proof : Using the linearity of the expectation, we have $\mathbb{E}[\frac{1}{n}(X_1 + \dots X_n)] = \mathbb{E}[X_1]$. Then from (3) of Proposition 3.3.1, we have,

$$\mathbb{E}\left[\Big(\frac{1}{n}(X_1 + \cdots + X_n) - \mathbb{E}[X_1]\Big)^2\right] = \frac{1}{n^2}\operatorname{Var}(X_1 + \cdots + X_n) = \frac{1}{n}\operatorname{Var}(X_1),$$

So when $n \to \infty$, the above formula goes to 0. □

Remark 3.3.5 : There are several different notions of convergence in a probability space that we will discuss further in detail in Chapter 4. What we need to notice here is that, Theorem 3.3.4 is a *weak* version of the law of large numbers. We note that the above convergence takes place only in L^2 space, but it is not a simple convergence (簡単収斂)[1] up to a set of measure zero. [2]

Proposition 3.3.6 : *Let $(X_n)_{n\geqslant 1}$ be an i.i.d. sequence of random variables with $\mathbb{E}[|X_1|^4] < \infty$. Then,*

$$\frac{1}{n}(X_1 + \cdots + X_n) \xrightarrow[n\to\infty]{a.s.} \mathbb{E}[X_1].$$

Proof : If we replace X_i with $X_i - \mathbb{E}[X_i]$ for all $i \in \{1, \dots, n\}$, we notice that the new convergence result that needs to be shown is equivalent to the original one. Thus, without loss of generality, we can assume that all the random variables X_i satisfy $\mathbb{E}[X_i] = 0$. First, we compute the fourth-order moment

[1] Or almost sure convergence, meaning that the convergence takes place with probability 1.

[2] We also note that there is not any implication between L^2 convergence and a.s. convergence, so technically speaking one is not stronger or weaker than the other.

on the left-hand side,

$$\mathbb{E}\left[\Big(\frac{1}{n}(X_1+\cdots+X_n)\Big)^4\right]=\frac{1}{n^4}\sum_{1\leqslant i_1,\ldots,i_4\leqslant n}\mathbb{E}[X_{i_1}X_{i_2}X_{i_3}X_{i_4}].$$

Since (X_k) is a sequence of independent random variables with zero expectation, in the above summation, most of the terms are zero. Indeed, when the quadruplet (i_1,i_2,i_3,i_4) is such that all the indices are equal or $i_1=i_2, i_3=i_4$ along with its permuations, the corresponding summation is not zero. The former case appears n times and the latter case $3n(n-1)$ times. Hence, we obtain,

$$\mathbb{E}\left[\Big(\frac{1}{n}(X_1+\cdots+X_n)\Big)^4\right]=\frac{1}{n^4}\Big(n\,\mathbb{E}\left[X_1^4\right]+3n(n-1)\,\mathbb{E}\left[X_1^2X_2^2\right]\Big)\leqslant\frac{C}{n^2},$$

where $C<\infty$ is a constant. Then, we have,

$$\mathbb{E}\left[\sum_{n=1}^{\infty}\Big(\frac{1}{n}(X_1+\cdots+X_n)\Big)^4\right]=\sum_{n=1}^{\infty}\mathbb{E}\left[\Big(\frac{1}{n}(X_1+\cdots+X_n)\Big)^4\right]<\infty,$$

where in the above formula, we inverted the expectation and series since all the terms in the series are non-negative. This implies that the following series converges and is almost surely finite (殆必有限),

$$\sum_{n=1}^{\infty}\Big(\frac{1}{n}(X_1+\cdots+X_n)\Big)^4<\infty,\qquad\text{a.s.},$$

so the terms in the series converges almost surely to zero. □

Corollary 3.3.7 : *If $(A_n)_{n\geqslant1}$ are independent events of the same probability, then we have,*

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_{A_i}\xrightarrow[n\to\infty]{a.s.}\mathbb{P}(A_i).$$

Remark 3.3.8 : Before the development of the modern probability, the *probability* that an event occurs used to be interpreted as the *frequency* of its occurence in a series of independent random experiments. This corollary shows that this interpretation does make sense using the modern approach.

Alternatively, if we want to determine the probability that a result A holds, we can conduct this experiment repeatedly in an independent manner and compute the proportion of times where A holds. Then, with probability one (almost sure convergence, strong law of large numbers), this quantity tends to $\mathbb{P}(A)$.

3.3.4 Convolution Semigroups

When we discuss Markov chains in Chapter 7 or continuous-time stochastic processes in the next semester, the notion of *convolution semigroup* will be important. We introduce this notion and elementary properties in this section.

Assume that $I = \mathbb{Z}_{\geqslant 0}$ or $I = \mathbb{R}_{\geqslant 0}$.

Definition 3.3.9 : Let $(\mu_t)_{t\in I}$ be a family of probability measures on $\mathbb{R}$ or $\mathbb{R}^d$ and indexed by I. We call $(\mu_t)_{t\in I}$ a *convolution semigroup* (畳積半群) if

$$\forall t, t' \in I, \qquad t + t' \in I \quad \text{and} \quad \mu_t * \mu_{t'} = \mu_{t+t'}.$$

Lemma 3.3.10 : *If there exists a function $\varphi : \mathbb{R} \longrightarrow \mathbb{C}$ such that one of the following conditions holds,*

(i) If $I = \mathbb{Z}_{\geqslant 0}$, $\widehat{\mu_t}(\xi) = \varphi(\xi)^t, \quad \forall t \in I,$

(ii) If $I = \mathbb{R}_{\geqslant 0}$, $\widehat{\mu_t}(\xi) = \exp(-t\varphi(\xi)), \quad \forall t \in I,$

then, $(\mu_t)_{t\in I}$ is a convolution semigroup.

Proof : If $\widehat{\mu_t}$ is as described in the above lemma, then we have $\widehat{\mu_{t+t'}} = \widehat{\mu_t}\widehat{\mu_{t'}} = \widehat{\mu_t * \mu_{t'}}$. We can use the injectivity of the Fourier transform to deduce $\mu_{t+t'} = \mu_t * \mu_{t'}$. □

Example 3.3.11 :

(1) Suppose $I = \mathbb{Z}_{\geqslant 0}$. Given $p \in [0, 1]$. For all $n \geqslant 1$, let μ_n be the Binomial distribution $\mathcal{B}(n, p)$. From the interpretation of the Binomial distribution as a sum of i.i.d. Bernoulli random variables, we clearly have $\mu_{n+m} = \mu_n * \mu_m$. Otherwise, we can also compute its characteristic function and apply the above lemma, $\widehat{\mu_n}(\xi) = (pe^{\mathrm{i}\,\xi} + 1 - p)^n$.

(2) Suppose $I = \mathbb{R}_{\geqslant 0}$. For all $t \geqslant 0$, let μ_t be the Poisson distribution of parameter t. We have,

$$\forall t \geqslant 0, \quad \forall \xi \in \mathbb{R}, \quad \widehat{\mu_t}(\xi) = \sum_{k=0}^{\infty} \frac{t^k}{k!} e^{\mathrm{i}\,k\xi} e^{-t} = \exp(-t(1 - e^{\mathrm{i}\,\xi})).$$

(3) Suppose $I = \mathbb{R}_{\geqslant 0}$. For all $t \geqslant 0$, let μ_t be the Gaussian distribution $\mathcal{N}(0, t)$. From Lemma 2.4.14, we have,

$$\forall t \geqslant 0, \quad \forall \xi \in \mathbb{R}, \quad \widehat{\mu_t}(\xi) = \exp\big(-\frac{t\xi^2}{2}\big).$$

3.4 Some More Complicated Random Variables

Here we will use independence to construct some interesting tools in probability: multivariate normal distribution and Poisson process.

3.4.1 Multivariate Normal Distribution

The goal of this subsection is to extend the notion of Gaussian distribution to higher dimensions. The following proposition defines the notion of *multivariate normal distribution*, gives its important properties, and the canonical way to construct it.

Proposition 3.4.1 : *Let $X = (X_1, \dots, X_d)$ be a d-dimensional real-valuded random variable. We want to show that the following three conditions are equivalent. Moreover, when one of the conditions is satisfied, we say that X has a multivariate normal distribution* (多元常態分佈).

(i) There exist a d-dimensional real-valued random variable $Z = (Z_1, \dots, Z_d)$ such that the comonents are i.i.d. standard normal distributions, a square matrix A of size $d \times d$ and a vector $B \in \mathbb{R}^d$ such that $X \overset{(\mathrm{d})}{=} AZ + B$.

(ii) For any $\alpha \in \mathbb{R}^d$, the random variable $\alpha^T X$ also has a normal distribution.

(iii) There exist a semi-definite symmetric matrix Σ of size $d \times d$ and a vector $B \in \mathbb{R}^d$ such that the characteristic function of X writes,

$$\Phi_X(\xi) = \mathbb{E}\left[e^{\mathrm{i}\,\xi \cdot X}\right] = \exp\left(\mathrm{i}\,\xi^T B - \tfrac{1}{2}\xi^T \Sigma \xi\right).$$

Moreover, the vector $B = \mathbb{E}[X]$ is the expectation of X, the matrix $\Sigma = AA^T = K_X$ is the covariance matrix of X.

Proof : Show that (i) $\Longrightarrow$ (ii). We can show that the expecatation of X is given by B and the covariance matrix by AA^T. Then, take $\alpha \in \mathbb{R}^d$, the distribution of $\alpha^T X = \sum \alpha_i X_i$ will be $\mathcal{N}(\alpha^T B, \alpha^T AA^T \alpha)$.

Show that (ii) $\Longrightarrow$ (iii). Given $\xi \in \mathbb{R}^d$. Since $\xi^T X$ is still a normal distribution, write $m = \xi^T \mathbb{E}[X]$ and $\sigma^2 = \xi^T K_X \xi$ for its expecatation and variance. We know that

$$\Phi_X(\xi) = \exp(\mathrm{i}\,\xi^T \mathbb{E}[X] - \tfrac{1}{2}\xi^T K_X \xi).$$

Hence, we can take $B = \mathbb{E}[X]$ and $\Sigma = K_X$.

Show that (iii) $\Longrightarrow$ (i). Since Σ is a semi-definite symmetric matrix, there exists an orthogonal matrix P and a diagonal matrix D such that $\Sigma = PDP^T$. Given i.i.d. random variables $Z_1, \dots, Z_d$ with the standard normal distribution. Consider $A = P\sqrt{D}$, then we can compute the characteristic function of $AZ + B$ and show that it is equal to Φ_X. □

Proposition 3.4.2 : *Let $X = (X_1, \dots, X_d)$ be a d-dimensional multivariate normal distribution with expectation B and covariance matrix Σ. If Σ is invertible, then the density function of X writes,*

$$\mathbb{P}_X(\mathrm{d}x) = \frac{\exp\left(-\frac{1}{2}\langle x - B, \Sigma^{-1}(x - B)\rangle\right)}{(2\pi)^{d/2}|\det(A)|}\,\mathrm{d}x_1 \dots \mathrm{d}x_n.$$

Proof : Since Σ is a semi-definite matrix, we may find an orthogonal matrix P and a diagonal matrix D such that $\Sigma = PDP^T$, then we define $A = P\sqrt{D}$. By the definition, the properties, and the construction in Proposition 3.4.1, we know that $X \overset{(\mathrm{d})}{=} AZ + B$ where $Z = (Z_1, \dots, Z_d)$ are i.i.d.

standard normal random variables. The density function of Z writes

$$\forall z \in \mathbb{R}^d, \qquad p_Z(z) = \frac{\exp\left(-\frac{1}{2}\|z\|_2^2\right)}{(2\pi)^{d/2}} = \frac{\exp\left(-\frac{1}{2}(z_1^2+\cdots+z_n^2)\right)}{(2\pi)^{d/2}}.$$

Given a non-negative measurable function $f : \mathbb{R}^d \longrightarrow \mathbb{R}_+$, we have

$$\begin{aligned}\mathbb{E}[f(X)] = \mathbb{E}[f(AZ+B)] &= \int_{\mathbb{R}^d} f(Az+B)p_Z(z)\,\mathrm{d}z_1\ldots\mathrm{d}z_n \\ &= \int_{\mathbb{R}^d} f(x)p_Z(A^{-1}(x-B))|\det(A)|^{-1}\,\mathrm{d}x_1\ldots\mathrm{d}x_n,\end{aligned}$$

where in the second equality, we apply the change of variables $x = Az+B$, and the fact that under the map $z \mapsto Az+B$, the set $\mathbb{R}^d$ is diffeomorphic with itself, since A is invertible. Since the above identity holds for all non-negative measurable functions, we deduce the density function of X by Remark 2.1.16,

$$\begin{aligned}\forall x \in \mathbb{R}^d, \qquad p_X(x) &= \frac{p_Z(A^{-1}(x-B))}{|\det(A)|} \\ &= \frac{\exp\left(-\frac{1}{2}\langle x-B, (A^{-1})^T A^{-1}(x-B)\rangle\right)}{(2\pi)^{d/2}|\det(A)|} \\ &= \frac{\exp\left(-\frac{1}{2}\langle x-B, \Sigma^{-1}(x-B)\rangle\right)}{(2\pi)^{d/2}|\det(A)|}.\end{aligned}$$

□

3.4.2 Poisson Process

We want to describe the behavior of random events occurring in time. We may want to know, for instance, when these events occur; or at a given fixed time, how many random events have already occured. This can be formulated as below. Given a sequence of random variable $(X_i)_{i\geqslant 1}$ where each X_i describes the waiting time between two successive events $i-1$ and i; S_n gives the total waiting time for the n-th event to happen; N_t gives the total number of events that have occured by time t (including time t). We give a mathematical formulation below.

Let $(X_i)_{i\geqslant 1}$ be i.i.d. random variables with exponential distribution $\mathrm{Exp}(1)$ defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Define $S_0 = 0$ and

$$\forall n \in \mathbb{N}, \qquad S_n := X_1 + \cdots + X_n.$$

We know from Exercise 3.14 that S_n follows the Gamma distribution $\Gamma(n,1)$. In this subsection, we will discuss the following stochastic process (隨機過程),

$$\forall t \geqslant 0, \qquad N_t := \max\{n \geqslant 0 : S_n \leqslant t\} = \max\{n \geqslant 0 : X_1 + \cdots + X_n \leqslant t\} \in \mathbb{N}\cup\{0\}, \tag{3.12}$$

called Poisson process (帕松過程). It is not hard to see that $t \mapsto N_t$ is a (random) non-decreasing function.

Proposition 3.4.3 : *We have the following properties.*

(1) $(S_n)_{n\geqslant 0}$ is almost surely a strictly increasing sequence, and diverges to ∞ almost surely.

(2) The law of large numbers holds, $\frac{S_n}{n} \xrightarrow{a.s.} 1$.

Proof :

(1) First, due to

$$\forall i \geqslant 1, \qquad \mathbb{P}(X_i > 0) = 1,$$

and the fact that there are countably many events, we find

$$\mathbb{P}(\forall i \geqslant 0, S_i < S_{i+1}) = 1 \qquad \Longleftrightarrow \qquad 0 = S_0 < S_1 < S_2 < \ldots, \text{a.s.}$$

Then, by the second part of the Borel–Cantelli lemma, we know that for any fixed $\alpha > 0$, we have

$$\sum_{i \geqslant 1} \mathbb{P}(X_i \geqslant \alpha) = \infty,$$

using the independence of $(X_n)_{n \geqslant 1}$, we deduce that there exist infinitely many $n \geqslant 1$ such that $X_n \geqslant \alpha$, implying $S_n \xrightarrow{\text{a.s.}} \infty$.

(2) We use the law of large numbers in L^4 as stated in Proposition 3.3.6 to conclude

$$\frac{S_n}{n} \xrightarrow{\text{a.s.}} \mathbb{E}[X_1] = 1.$$

□

Proposition 3.4.4 : *Fix $t > 0$, then $N_t \sim \text{Pois}(t)$ follows the Poisson distribution of parameter t.*

Remark 3.4.5 : As a direct consequence of Example 3.3.11 (2), the family of distributions $(\mathbb{P}_{N_t})_{t \geqslant 0}$ forms a convolution semigroup.

Proof : Fix a positive real number $t > 0$ and a non-negative integer $n \geqslant 0$, we have

$$\begin{aligned}
\mathbb{P}(N_t = n) &= \mathbb{P}(S_n \leqslant t < S_{n+1}) \\
&= \int_{\mathbb{R}_{\geqslant 0}} \int_{\mathbb{R}_{\geqslant 0}} \frac{x^{n-1} e^{-x}}{\Gamma(n)} e^{-y} \mathbb{1}_{x \leqslant t < x+y} \, \mathrm{d}x \, \mathrm{d}y \\
&= \int_{\mathbb{R}_{\geqslant 0}} \frac{x^{n-1} e^{-x}}{\Gamma(n)} \mathbb{1}_{x \leqslant t} \int_{\mathbb{R}_{\geqslant 0}} e^{-y} \mathbb{1}_{y > t-x} \, \mathrm{d}y \, \mathrm{d}x \\
&= \int_{\mathbb{R}_{\geqslant 0}} \frac{x^{n-1} e^{-x}}{\Gamma(n)} \mathbb{1}_{x \leqslant t} e^{-(t-x)} \, \mathrm{d}x = e^{-t} \frac{t^n}{n!},
\end{aligned}$$

where in the second equality, we use the property that $S_{n+1} = S_n + X_{n+1}$ is the sum of two independent random variables with $S_n \sim \Gamma(n, 1)$; in the third equality, we use the Fubini's theorem. □

Proposition 3.4.6 : *The stochastic process $t \mapsto N_t$ is almost surely a càdlàg (continue à droite, limite à gauche) function on $\mathbb{R}_{\geqslant 0}$, which means* right-continuous with left limits.

Remark 3.4.7 : We note that the two following properties on the right continuity are different,

$$\left(\forall s \geqslant 0, \quad \lim_{\substack{t \to s \\ t > s}} N_t = N_s\right) \text{ a.s.} \quad \Leftrightarrow \quad \mathbb{P}\left(\forall s \geqslant 0, \quad \lim_{\substack{t \to s \\ t > s}} N_t = N_s\right) = 1,$$

$$\forall s > 0, \quad \left(\lim_{\substack{t \to s \\ t < s}} N_t \quad \text{exists} \quad \text{a.s.}\right) \quad \Leftrightarrow \quad \forall s \geqslant 0, \quad \mathbb{P}\left(\lim_{\substack{t \to s \\ t > s}} N_t = N_s\right) = 1.$$

In general, the first one is stronger than the second one due to the fact that $\mathbb{R}_{>0}$ is uncountable. We need to bear in mind the way we put the parentheses can change the meaning of a statement.

Proof : We need to prove the following two properties,

$$\left(\forall s \geqslant 0, \quad \lim_{\substack{t \to s \\ t > s}} N_t = N_s\right) \text{ a.s.},$$

$$\left(\forall s > 0, \quad \lim_{\substack{t \to s \\ t < s}} N_t \quad \text{exists}\right) \text{ a.s.}$$

For $T > 0$, let us show that $t \mapsto N_t$ is almost surely a càdlàg function on $[0, T]$. Proposition 3.4.4 tells us that N_T follows $\text{Pois}(T)$, which is finite almost surely. Let $\Omega_T \in \mathcal{A}$ with $\mathbb{P}(\Omega_T) = 1$ such that

$$\forall \omega \in \Omega_T, \quad N_T < \infty.$$

For $\omega \in \Omega_T$, the function $t \mapsto N_t$ is non-decreasing and bounded on $[0, T]$, so the left limit exists everywhere. For $\omega \in \Omega_T$ and $s \in [0, T)$, let $n := n(\omega)$ such that $N_s = n$, which is equivalent to $S_n \leqslant s$ and $S_{n+1} > s$. This means that for $u \in [s, S_{n+1}) \neq \varnothing$, we need to have $N_u = n$ as well, which implies the right-continuity.

We conclude by noting that $\mathbb{P}(\Omega') = 1$ with $\Omega' := \bigcap_{T \geqslant 1} \Omega_T$, and on Ω', the function $t \mapsto N_t$ is càdlàg on $[0, T]$ for every integer $T \geqslant 1$. In consequence, the function $t \mapsto N_t$ is càdlàg on $\mathbb{R}_{\geqslant 0}$ almost surely. □

Lemma 3.4.8 : *Let $(N_s)_{s \geqslant 0}$ be a Poisson process and $t > 0$. Define the shifted process $(N_s^{(t)})_{s \geqslant 0}$ as below,*

$$\forall s \geqslant 0, \qquad N_s^{(t)} := N_{t+s} - N_t.$$

Then, $(N_s^{(t)})_{s \geqslant 0}$ is a Poisson process that is independent from N_t.

Proof : We fix $t > 0$ and consider the Poisson process from time t. Starting from time t, we need to wait $S_{N_t+1} - t$ for the following event to occur, then we need to wait $X_{N_t+2}, X_{N_t+3}, \ldots$. Thus, let us define

$$X_1^{(t)} = S_{N_t+1} - t,$$
$$\forall n \geqslant 2, \qquad X_n^{(t)} = X_{N_t+n}.$$

For any $s > 0$, we have

$$N_s^{(t)} := N_{t+s} - N_t = \max\{m \geqslant 0 : X_1^{(t)} + \cdots + X_m^{(t)} \leqslant s\},$$

We may note that this definition is very similar to that of Eq. (3.12). If we can prove that for all non-negative integer $n \geqslant 0$, the following hold:

(a) the event $\{N_t = n\}$ and the sequence $(X_m^{(t)})_{m\geqslant 1}$ of random variables are independent, and

(b) under the event $\{N_t = n\}$, the sequence $(X_m^{(t)})_{m\geqslant 1}$ is also an i.i.d. sequence of random variables with exponential distribution of parameter 1,

then not only we will have proven the independence between $N^{(t)} = (N_s^{(t)})_{s\geqslant 0}$ and N_t, but also the equality in distribution between $(N_s^{(t)} = N_{t+s} - N_t)_{s\geqslant 0}$ and the original process $(N_t)_{t\geqslant 0}$.

Fix a non-negative integer $n \geqslant 0$. First, due to the equality $\{N_t = n\} = \{S_n \leqslant t < S_{n+1}\}$, we know that $\{N_t = n\}$ only depends on $(X_i)_{1\leqslant i\leqslant n+1}$, so is independent from $(X_i)_{i\geqslant n+2}$. Then, we discuss the independence between $\{N_t = n\}$ and $S_{n+1} - t = X_{n+1} + S_n - t$. Fix $y \geqslant 0$, we have

$$\begin{aligned}
\mathbb{P}(N_t = n, S_{n+1} - t > y) &= \mathbb{P}(S_n \leqslant t, X_{n+1} > t + y - S_n) \\
&= \int_{\mathbb{R}_{\geqslant 0}} \gamma_n(x) \mathbb{1}_{x\leqslant t}\, \mathbb{P}(X_{n+1} > t + y - x)\, \mathrm{d}x \\
&= \int_{\mathbb{R}_{\geqslant 0}} \gamma_n(x) \mathbb{1}_{x\leqslant t}\, \mathbb{P}(X_{n+1} > t - x)\, \mathbb{P}(X_{n+1} > y)\, \mathrm{d}x \\
&= e^{-y}\, \mathbb{P}(S_n \leqslant t, X_{n+1} > t - S_n) \\
&= e^{-y}\, \mathbb{P}(S_n \leqslant t < S_{n+1}),
\end{aligned}$$

where in the first line, we use the inclusion $\{S_{n+1} > t + y\} \subseteq \{S_{n+1} > t\}$; in the second line, we write $\gamma_n(x)$ for the density function of $S_n \sim \Gamma(n, 1)$; in the third line, we use the memoryless property of exponential random variables. Therefore, the above computation implies that $\{N_t = n\}$ and $S_{n+1} - t = X_{n+1} + S_n - t$ are independent, and that $S_{n+1} - t \sim \mathrm{Exp}(1)$. Since $S_{n+1} - t$ only depends on $(X_i)_{1\leqslant i\leqslant n+1}$, so is also independent from $(X_i)_{i\geqslant n+2}$. □

Proposition 3.4.9 : *The Poisson process $(N_t)_{t\geqslant 0}$ has the two following properties.*

(1) Fix time points $0 = t_0 < t_1 < \cdots < t_k$, the increments $(N_{t_{i+1}} - N_{t_i})_{0\leqslant i\leqslant k-1}$ of the stochastic process is an independent sequence.

(2) The increments of the Poisson process follow a Poisson distribution. More precisely, fix $0 \leqslant s < t$, we have

$$\forall n \geqslant 0, \qquad \mathbb{P}(N_t - N_s = n) = e^{-(t-s)} \frac{(t-s)^n}{n!}.$$

Proof :

(1) In consequence, for any sequence $(m_i)_{1\leqslant i\leqslant k}$ of non-negative integers, by Lemma 3.4.8, we find

$$\begin{aligned}
&\mathbb{P}\left(N_{t_{i+1}} - N_{t_i} = m_{i+1}, 0 \leqslant i \leqslant k-1\right) \\
&= \mathbb{P}\left(N_{t_1} = m_1, N^{(t_1)}_{t_{i+1}} - N^{(t_1)}_{t_i} = m_{i+1}, 1 \leqslant i \leqslant k-1\right) \\
&= \mathbb{P}\left(N_{t_1} = m_1\right)\mathbb{P}\left(N^{(t_1)}_{t_{i+1}} - N^{(t_1)}_{t_i} = m_{i+1}, 1 \leqslant i \leqslant k-1\right),
\end{aligned}$$

Finally, by induction, we find

$$\mathbb{P}\left(N_{t_{i+1}} - N_{t_i} = m_{i+1}, 0 \leqslant i \leqslant k-1\right) = \prod_{i=0}^{k-1}\mathbb{P}\left(N_{t_{i+1}} - N_{t_i} = m_{i+1}\right).$$

which is exactly what we need to show for the property (1).

(2) We use the property (1) and Proposition 3.4.4 to deduce the result.

□