3 Independence of Random Variables

隨機變數的獨立性

第三章

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

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

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

第一節 不同的獨立概念

在後面的章節中，我們將討論這學期課程中，最重要的兩個定理：大數定理（定理 4.3.1 ）及中央極限定理（定理 4.5.3 ）。要討論並且敘述這兩個定理前，我們要先定義不同的獨立的概念，其中包含獨立事件（定義 3.1.1 ）、獨立 σ 代數（定義 3.1.5 ）以及獨立隨機變數（定義 3.1.6 ）。

第一小節 獨立事件

在這章節，我們考慮一機率空間 $(\Omega, \mathcal{A}, \mathbb{P})$。

設 $A, B \in \mathcal{A}$，若

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

則我們稱 A 與 B 為獨立事件 (independent events) 獨立概念，換句話說，「知道事件 B 為真，並不影響 A 事件發生的機率」。但要能夠寫出式 (3.1)，我們需要要求 $\mathbb{P}(B) > 0$，且此式子中的條件並不滿足對稱性，因此我們可以改用下列方式來定義兩個獨立事件的概念：

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

從式 (3.2) 我們也可以注意到，若 A 與 B 為獨立事件，則不難驗證

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

所以 A^c 與 B 也是獨立事件。

更一般的情況，當我們有多於兩事件，甚至是無窮多個事件的情況下，我們給出下列定義。

定義 3.1.1 ： 給定任意集合 I，對於所有 $i \in I$，固定 $A_i \in \mathcal{A}$。若對於所有有限的子集合 $J \subseteq I$，我們有

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

則我們說由 I 所標記的事件組 $(A_i)_{i\in I}$ 是獨立的。此外，我們也稱 $(A_i)_{i\in I}$ 為獨立事件 (independent events) 。

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

- $\mathbb{P}(A_1 \cap \cdots \cap A_n) = \mathbb{P}(A_1) \ldots \mathbb{P}(A_n)$;
- 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 $A_1, \ldots, A_n \in \mathcal{A}$, then the following two properties are equivalent.*

(i) The events $A_1, \ldots, A_n$ 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,

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

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 $\{j_1, \ldots, j_p\}$ of $\{1, \ldots, n\}$, when $B_{j_k} = A_{j_k}$ or $A_{j_k}^c$, we have,

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

To show this, it is enough to show that, if $C_1, \ldots, C_p$ are independent events, then $C_1^c, C_2, \ldots, C_p$ 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. □

問題 3.1.2：給定 n 個事件 $A_1, \ldots, A_n \in \mathcal{A}$，試問若只有下列其中一個條件成立，則 $A_1, \ldots, A_n$ 是否會是獨立事件？

- $\mathbb{P}(A_1 \cap \cdots \cap A_n) = \mathbb{P}(A_1) \ldots \mathbb{P}(A_n)$ ；
- 對於任何數對 $1 \leqslant i < j \leqslant n$，我們有 $\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\,\mathbb{P}(A_j)$ 。

在當我們只有有限多個事件的情況下，下列命題給了另一個等價於定義定義 3.1.1 的條件。

命題 3.1.3 ： 給定事件 $A_1, \ldots, A_n \in \mathcal{A}$，則下列兩個敘述等價：

(i) 事件 $A_1, \ldots, A_n$ 為獨立事件；

(ii) 對於每個 $1 \leqslant i \leqslant n$，選擇 B_i 為集合 $\{\varnothing, A_i, A_i^c, \Omega\}$ 中的任意事件，我們有

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

證明：首先我們證明 (ii) $\Longrightarrow$ (i)。我們可以注意到，命題中式 (3.5) 所要求的條件，比定義 3.1.1 中式 (3.4) 所要求的條件還要強；所以當式 (3.5) 成立時，式 (3.4) 也會成立。更確切的說，在式 (3.5) 中，若 $i \in J$，取 $B_i = A_i$；反之則取 $B_i = \Omega$，即可以得到式 (3.4)。

再來我們要證明 (i) $\Longrightarrow$ (ii)。若存在 B_i 使得 $B_i = \varnothing$，則式 (3.5) 左右兩側皆為 0，得證。若存在 B_i 使得 $B_i = \Omega$，則在式 (3.5) 的左側，我們可以移除 B_i 而不影響交集；在其右側，$\mathbb{P}(B_i) = 1$ 並不影響乘積的結果，因此只需要證明：對於 $\{1, \ldots, n\}$ 的任何子集合 $\{j_1, \ldots, j_p\}$，當 $B_{j_k} = A_{j_k}$ 或是 $A_{j_k}^c$ 時，我們有

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

要證明這項結果，只需要證明：若 $C_1, \ldots, C_p$ 為獨立事件，則 $C_1^c, C_2, \ldots, C_p$ 也是獨立事件，這可以透過與式 (3.3) 中相似的計算而得到。□

系理 3.1.4 ： 令 $(A_i)_{i\in I}$ 為獨立事件組。固定子集合 $J \subseteq I$，並定義

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

則 $(B_i)_{i\in I}$ 也是獨立事件組。

證明：這是使用命題 3.1.3 可以得到的直接結果。□

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

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

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

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

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

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

If $X_1, \dots, X_n$ are random variables with values in $(E_1, \mathcal{E}_1), \dots, (E_n, \mathcal{E}_n)$, then the n-tuple $(X_1, \dots, X_n)$ is a random variable with values in $E_1 \times \dots, E_n$ 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 $X_1, \dots, X_n$, the following three properties are equivalent.*

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

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

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

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

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

第二小節 獨立 σ 代數及獨立隨機變數

前面探討的是可測事件的獨立性，接下來我們要討論σ 代數的獨立性以及隨機變數的獨立性。

定義 3.1.5 ： 令 $\mathcal{B}_1, \dots, \mathcal{B}_n$ 為 n 個 $\mathcal{A}$ 的子 σ 代數。若

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

則說 $\mathcal{B}_1, \dots, \mathcal{B}_n$ 為獨立 σ 代數 (independent σ-algebras) 。

定義 3.1.6 ： 令 $X_1, \dots, X_n$ 分別為取值在 $(E_1, \mathcal{E}_1), \dots, (E_n, \mathcal{E}_n)$ 中的隨機變數。若 $\sigma(X_1), \dots, \sigma(X_n)$ 為獨立 σ 代數，我們說 $X_1, \dots, X_n$ 為獨立隨機變數 (independent random variables) 。

註解 3.1.7 ： 透過命題 3.1.3 中的 (2)，我們得知若且唯若事件 $A_1, \dots, A_n$ 獨立，則隨機變數 $\mathbb{1}_{A_1}, \dots, \mathbb{1}_{A_n}$ 也是獨立的。

註解 3.1.8 ： $X_1, \dots, X_n$ 為獨立隨機變數與下列性質等價：

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

若 $X_1, \dots, X_n$ 分別是值域為 $(E_1, \mathcal{E}_1), \dots, (E_n, \mathcal{E}_n)$ 的隨機變數，則 n 元組 $(X_1, \dots, X_n)$ 是個值域為 $E_1 \times \dots, E_n$ 且對 σ 代數 $\mathcal{E}_1 \otimes \cdots \otimes \mathcal{E}_n$ 可測的隨機變數。下面我們有另一種描述獨立隨機變數的方式。

定理 3.1.9 ： 給定隨機變數 $X_1, \dots, X_n$，則下列三個性質等價。

(i) 隨機變數 $X_1, \dots, X_n$ 為獨立隨機變數。

(ii) n 元組隨機變數 $(X_1, \dots, X_n)$ 的分佈為隨機變數 $X_1, \dots, X_n$ 分佈的乘積，也就是說

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

(iii) 對於所有 $i \in \{1, \dots, n\}$ 以及任意在 $(E_i, \mathcal{E}_i)$ 之上的非負可測函數 f_i，我們有

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

Remark 3.1.10 : If the measurable functions f_i are not non-negative, then under the assumption that $\mathbb{E}[|f_i(X_i)|] < \infty$ 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 $i \in \{1, \ldots, n\}$, let $F_i \in \mathcal{E}_i$. We have the following formulas,

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

From the above formulas and Eq. (3.6), $X_1, \ldots, X_n$ are independent random variables if and only if $\mathbb{P}_{(X_1,\ldots,X_n)}$ and $\mathbb{P}_{X_1} \otimes \ldots \otimes \mathbb{P}_{X_n}$ 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 $\mathbb{P}_{(X_1,\ldots,X_n)} = \mathbb{P}_{X_1} \otimes \ldots \otimes \mathbb{P}_{X_n}$.

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

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

and write, using the Fubini's theorem,

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

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 $\mu_1, \ldots, \mu_n$ 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 $Y = (Y_1, \ldots, Y_n)$ 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 $Y_1, \ldots, Y_n$ of the random variable Y are independent random variables and their distributions are respectively $\mu_1, \ldots, \mu_n$.

註解 3.1.10 ： 若可測函數 f_i 未必非負，若對於所有的 $i \in \{1, \cdots, n\}$，我們有 $\mathbb{E}[|f_i(X_i)|] < \infty$，在此假設情況下，式 (3.7) 仍然為真。

註解 3.1.11 ： 我們可以觀察到，若 $X_1, \ldots, X_n$ 為在可積的獨立隨機變數，則他們的乘積 $X_1 \ldots X_n$ 仍然可積；但在一般情況下，可積變數的乘積未必是可積的。

證明： 我們以下列方式來證明 (i) 與 (ii) 是等價的。對於所有 $i \in \{1, \ldots, n\}$，令 $F_i \in \mathcal{E}_i$。我們有下列兩個式子：

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

從上式與式 (3.6) 可以得知，若且唯若 $X_1, \ldots, X_n$ 為獨立隨機變數，則 $\mathbb{P}_{(X_1,\ldots,X_n)}$ 與 $\mathbb{P}_{X_1} \otimes \ldots \otimes \mathbb{P}_{X_n}$ 在所有的 $F_1 \times \cdots \times F_n$ 上有相同的值。單調類引理告訴我們（定理 1.4.1 ），在積可測空間上的測度，是被所有 $F_1 \times \cdots \times F_n$ 所決定，也就是說，獨立性質 $\mathbb{P}_{(X_1,\ldots,X_n)} = \mathbb{P}_{X_1} \otimes \ldots \otimes \mathbb{P}_{X_n}$ 等價。

接著，我們證明 (ii) 與 (iii) 是等價的。對於所有 $1 \leqslant i \leqslant n$，我們固定取值在 $(E_i, \mathcal{E}_i)$ 上的非負可測函數 f_i。我們有

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

並透過富比尼定理，我們有

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

因此我們可以得到等價關係。 □

註解 3.1.12 【有限多個獨立隨機變數的構造】 ： 此定理也間接告訴我們如何構造獨立的隨機變數。考慮實變數的情況，並假設 $\mu_1, \ldots, \mu_n$ 為 $\mathbb{R}$ 上的機率測度，定理 1.4.1 （積測度的構造）以及註解 2.1.11 （隨機變數的正則構造） ，讓我們可以構造值域為 $\mathbb{R}^n$ 的隨機變數 $Y = (Y_1, \ldots, Y_n)$ 使得他的分佈為 $\mu_1 \otimes \cdots \otimes \mu_n$。上述定理告訴我們，這樣構造出來的隨機變數 Y 其中的分量 $Y_1, \ldots, Y_n$ 是分佈分別為 $\mu_1, \ldots, \mu_n$ 的獨立隨機變數。

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 $X_1, \ldots, X_n$ be n random variables.*

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

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

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

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

where functions q_i are non-negative measurable functions on $\mathbb{R}$, then $X_1, \ldots, X_n$ are independent random variables. Moreover, for all $i \in \{1, \ldots, n\}$, 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,

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

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

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

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

系理 3.1.13 ： 若 X_1 及 X_2 為在 L^2 中且互為獨立的實隨機變數，則我們有 $\mathrm{Cov}(X_1, X_2) = 0$。

註解 3.1.14 ： 此系理的逆命題是錯的。當兩個隨機變數的變異數為零，我們稱他們互不相關，但這並不代表他們是獨立的。

問題 3.1.15： 構造兩個隨機變數 X 及 Y 使得 $\mathrm{Cov}(X, Y) = 0$ 但 X 與 Y 並不獨立。在習題 3.20 中，我們會看到在某特定的條件下，$\mathrm{Cov}(X, Y) = 0$ 蘊含 X 與 Y 獨立。

系理 3.1.16 ： 令 $X_1, \ldots, X_n$ 為 n 個隨機變數。

(1) 假設對於所有 $i \in \{1, \ldots, n\}$，X_i 是有密度的分佈，將其密度函數記作 p_i，並假設 $X_1, \ldots, X_n$ 為獨立隨機變數。這樣的情況下，$(X_1, \ldots, X_n)$ 是個有密度的分佈，而且其密度函數可以寫作

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

(2) 反之，假設 $(X_1, \ldots, X_n)$ 是個有密度的分佈，而且其密度函數可以寫作

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

其中函數 q_i 為 $\mathbb{R}$ 上的非負可測函數，則 $X_1, \ldots, X_n$ 為獨立隨機變數，而且對於所有的 $i \in \{1, \ldots, n\}$，X_i 是個有密度的分佈，且存在常數 $C_i > 0$，使得其密度函數 p_i 滿足 $p_i = C_i q_i$。

證明： (1) 為定理 3.1.9 及富比尼定理的應用，因為若 $\mathbb{P}_{X_i}(\mathrm{d}x_i) = p_i(x_i)\,\mathrm{d}x_i$，則我們可以將積測度寫作

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

接著證明 (2)。對於所有 $i \in \{1, \ldots, n\}$，令 $K_i = \int q_i(x)\,\mathrm{d}x \in (0, \infty)$。我們首先注意到，富比尼定理給出

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

接著，根據命題 2.1.18 ，我們可以計算 $X = (X_1, \ldots, X_n)$ 的邊緣分佈，也就是說 X_i 的分佈可以寫作

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

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

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

(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

從上式我們可以得到 $\mathbb{P}_{(X_1,\dots,X_n)} = \mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_n}$，也就是隨機變數的獨立性。□

問題 3.1.17： 令 $X_1, \dots, X_n$ 為實隨機變數，我們有下列等價關係：

(i) $X_1, \dots, X_n$ 為獨立隨機變數。

(ii) 對於任意 $a_1, \dots, a_n \in \mathbb{R}$，我們有 $\mathbb{P}(X_1 \leqslant a_1, \dots, X_n \leqslant a_n) = \prod_{i=1}^n \mathbb{P}(X_i \leqslant a_i)$ 。

(iii) 若 $f_1, \dots, f_n$ 為緊緻支撐 (compactly supported)，由 $\mathbb{R}$ 映射到 $\mathbb{R}_{\geqslant 0}$ 上的連續函數，則

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

(iv) X 的特徵函數可以寫作

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

下列命題是單調類引理的應用，需要點技巧性，但在後續非常實用。

命題 3.1.18 ： 令 $\mathcal{B}_1, \dots, \mathcal{B}_n$ 為 $\mathcal{A}$ 的子 σ 代數。對於所有 $i \in \{1, \dots, n\}$，設 $\mathcal{C}_i \subseteq \mathcal{B}_i$ 為一在有限交集下封閉的子集合，包含元素 Ω 且 $\sigma(\mathcal{C}_i) = \mathcal{B}_i$。若

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

則 $\mathcal{B}_1, \dots, \mathcal{B}_n$ 為獨立 σ 代數。

證明： 首先，讓我們固定 $C_2 \in \mathcal{C}_2, \dots, C_n \in \mathcal{C}_n$ 並設

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

我們有 $\mathcal{C}_1 \subseteq \mathcal{M}_1$ 且我們可以檢查，$\mathcal{M}_1$ 為單調類，因此根據單調類引理，$\mathcal{M}_1$ 包含 $\sigma(\mathcal{C}_1) = \mathcal{B}_1$。因此我們證明了，如果 $\mathcal{C}_1, \dots, \mathcal{C}_n$ 為獨立 σ 代數，則 $\sigma(\mathcal{C}_1), \mathcal{C}_2, \dots, \mathcal{C}_n$ 也是獨立的 σ 代數。接著根據歸納法，將上述證明應用在 $\mathcal{C}_2, \dots, \mathcal{C}_n, \sigma(\mathcal{C}_1)$ 以證明 $\sigma(\mathcal{C}_2), \mathcal{C}_3, \dots, \mathcal{C}_n, \sigma(\mathcal{C}_1)$ 為獨立的 σ 代數，依此類推。□

問題 3.1.19： 令 $\mathcal{B}_1, \dots, \mathcal{B}_n$ 為獨立 σ 代數。對於 $n_0 = 0 < n_1 < \dots < n_p = n$，下列 σ 代數是獨立

following σ-algebras are independent,

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

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 $\{i_1, \ldots, i_p\}$ of I, the σ-algebras $\mathcal{B}_{i_1}, \ldots, \mathcal{B}_{i_p}$ 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 $I_1, \ldots, I_p \subseteq \mathbb{N}$ that are pairwise disjoint, the subsets $(\mathcal{B}_k)_{1\leqslant k\leqslant p}$ defined by,*

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

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

的：

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

第三小節 無窮多個隨機變數的獨立性

接下來，我們要在有無窮多個隨機變數的情況下，定義獨立性的概念。這可以被簡化為在所有有限子集合上的獨立性。

定義 3.1.20 ：

(1) 令 $(B_i)_{i\in I}$ 為由 I 所標記、且由 $\mathcal{A}$ 的子 σ 代數所構成的序列。若對於 I 的任意有限集合 $\{i_1, \ldots, i_p\}$，$\mathcal{B}_{i_1}, \ldots, \mathcal{B}_{i_p}$ 為獨立的 σ 代數，則我們說 $(B_i)_{i\in I}$ 是由獨立 σ 代數構成的集合。

(2) 同樣的，若 $(X_i)_{i\in I}$ 為由任意隨機變數構成的集合，若 $(\sigma(X_i))_{i\in I}$ 由獨立 σ 代數，我們稱 $(X_i)_{i\in I}$ 為由獨立隨機變數構成的集合。

命題 3.1.21 ： 令 $(X_n)_{n\geqslant 1}$ 為獨立隨機變數序列。對於任意正整數 $p \geqslant 1$，以及兩兩互斥的正整數子集合 $I_1, \ldots, I_p \subseteq \mathbb{N}$，由下列式子定義的

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

為獨立 σ 代數。

證明： 對 $1 \leqslant k \leqslant p$，令

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

根據問題 3.1.19 ，對於任意有限子集合 $J_1 \subseteq I_1, \ldots, J_p \subseteq I_p$，$\sigma$ 代數 $\sigma(X_j : j \in J_1), \ldots, \sigma(X_j : j \in J_p)$ 是獨立的。接著，根據命題 3.1.18 ，我們得到 $\mathcal{B}_1 = \sigma(\mathcal{C}_1), \ldots, \mathcal{B}_p = \sigma(\mathcal{C}_p)$ 也是獨立 σ 代數。 □

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 $Y \sim \mathrm{Unif}([0,1])$ 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 $\mathrm{Ber}(\frac{1}{2})$.

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,

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

(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,

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

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

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

系理 3.1.22 ： 令 $(X_n)_{n\geqslant 1}$ 為獨立隨機變數序列，其中對於 $n \geqslant 1$，X_n 取值在可測空間 $(E_n, \mathcal{E}_n)$ 中。令 $(I_n)_{n\geqslant 1}$ 是個由兩兩互斥 $\mathbb{N}$ 的子集合所構成的序列，且 $(f_n)_{n\geqslant 1}$ 為可測函數序列，其中對於 $n \geqslant 1$，f_n 定義在 $F_n := \prod_{i\in I_n} E_i$ 上。那麼下列隨機變數是獨立的：

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

在註解 3.1.12 中，我們解釋了如何構造有限多個獨立隨機變數。再來，我們要解釋當我們有無窮可數多個獨立隨機變數時，如何達成這樣的構造。

引理 3.1.23 ： 令 $Y \sim \mathrm{Unif}([0,1])$ 為 $[0,1]$ 上均勻分佈的隨機變數。那麼 Y 的二元展開式，記作

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

會滿足下列性質。

(1) 式 (3.8) 中的展開式殆必唯一。

(2) 隨機變數 $(Y_n)_{n\geqslant 1}$ 獨立，且每項的分佈都是 $\mathrm{Ber}(\frac{1}{2})$。

證明： 不要忘記，二元展開可以藉由下列方式得到：

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

直接使用歸納法，我們也可以改寫：

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

(1) 對於 $x \in [0,1)$，他會有兩個不同的二元展開式，若且唯若他是個二元有理數 $\frac{m}{2^n}$，其中 $n \geqslant 1$ 以及 $0 \leqslant 1 \leqslant 2^n - 1$ 為整數。這可以藉由下列關係式來得到：

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

最後，我們注意到，由二元有理數構成的子集合，測度為零：

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

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.

(2) 首先，讓我們檢查 Y_1 和 X_1 的分佈。Y_1 跟從 $\mathrm{Ber}(\frac{1}{2})$ 的分佈：

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

對於 $0 \leqslant a < b < 1$，我們有

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

因此，X_1 跟從 $[0,1]$ 上的均勻分佈。使用數學歸納法，我們得知對於所有 $n \geqslant 1$，我們有 $Y_n \sim \mathrm{Ber}(\frac{1}{2})$。

令 $n \geqslant 1$ 為整數。我們想要檢查 $(Y_k)_{1 \leqslant k \leqslant n}$ 是獨立的隨機變數。令 $m_1, \ldots, m_n \in \{0,1\}$ 並計算

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

我們得以總結。□

註解 3.1.24 ： 如同在式 (3.8) 中，讓我們考慮 $Y \sim \mathrm{Unif}([0,1])$ 以及他的二元展開 $Y = 0.Y_1Y_2\ldots$。由於 $\mathbb{N}^2$ 是 $\mathbb{N}$ 等勢的，我們可以找到雙射函數 $f : \mathbb{N}^2 \to \mathbb{N}$。對於每個 $i \in \mathbb{N}$，隨機變數 $(Y_{f(i,j)})_{j \geqslant 1}$ 是獨立的，而且如果我們定義，

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

從定義 3.1.20 以及命題 3.1.21 ，我們得知 $(Z_i)_{i \geqslant 1}$ 是獨立隨機變數；此外，每一項都是 $[0,1]$ 上的均勻分佈。如果我們給定在 $\mathbb{R}$ 上的分佈 $(\mu_i)_{i \geqslant 1}$，我們可以把他們的累積分佈函數記作 $(F_i = F_{\mu_i})_{i \geqslant 1}$，那麼我們可以定義隨機變數 $(X_i)_{i \geqslant 1}$ 如下：

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

他們會是獨立隨機變數，而且對於每個 $i \geqslant 1$，我們有 $X_i \sim \mu_i$，見命題 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) ,

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

若我們需要構造不可數多個獨立隨機變數，則需要用到下列的 Kolmogorov 拓延定理 (Kolmogorov's extension theorem)。

定理 3.1.25 【Kolmogorov 拓延定理】**：** 給定可測空間 $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ 以及任意集合 T。假設下列兩條件成立：

(a) 對於任意 T 的有限子集合 S，我們可以構造在 $((\mathbb{R}^d)^{\otimes S}, \mathcal{B}(\mathbb{R}^d)^{\otimes S})$ 上的機率測度 $\mathbb{P}_S$。

(b) 對於任意有限的子集合 S_1 及 S_2 且 $S_1 \subseteq S_2$，機率測度 $\mathbb{P}_{S_1}$ 以及 $\mathbb{P}_{S_2}$ 是相容 (compatible) 的，也就是說若我們將 S_2 至 S_1 的投影記作 π，則 $\mathbb{P}_{S_1} = \mathbb{P}_{S_2} \circ \pi^{-1}$。

則在 $((\mathbb{R}^d)^{\otimes T}, \mathcal{B}(\mathbb{R}^d)^{\otimes T})$ 上，存在唯一的機率測度 $\mathbb{P}$ 使得對於任意有限子集合 S，將 T 至 S 的投影記作 π，我們有 $\mathbb{P}_S = \mathbb{P} \circ \pi^{-1}$。

證明：我們定義

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

為所有有限維度積 σ 代數中元素構成的集合，且我們知道 $\mathcal{B}(\mathbb{R}^d)^{\otimes T}$ 是由 $\mathcal{C}$ 生成，也就是說

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

存在性的部份，與在 $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ 上構造勒貝格測度相似，可以使用外側度來完成。唯一性的話，可以透過單調類引理而得到，請見系理 1.1.19 。 □

第二節 Borel–Cantelli 引理

前面定義了不同的獨立概念，這個章節中，我們要給出第一個利用獨立概念得到的漸進結果。

第一小節 敘述及證明

定義 3.2.1 ： 給定一個事件序列 $(A_n)_{n \geqslant 1}$，我們定義下列概念

(1) 我們稱下列集合為上極限 (upper limit) ：

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

(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

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

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

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

(2) 我們稱下列集合為下極限 (lower limit)：

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

(3) 若 (A_n) 的上下極限相等，則其上下極限亦可被稱做極限 (limit)，並記作 $\lim_{n\to\infty} A_n := \liminf_{n\to\infty} A_n = \limsup_{n\to\infty} A_n$。

(4) 若 (A_n) 的上下極限不相等，則稱 (A_n) 之極限不存在。

命題 3.2.2 ： 令 (A_n) 為一事件序列，則

(1) $\limsup A_n = \{\omega \mid \omega \in A_n \text{ i.o.}\}$，其中 i.o. 意為 infinitely often，表示存在無限多個 n 使得 $\omega \in A_n$。

(2) $\liminf A_n = \{\omega \mid \omega \in A_n \text{ a.a.}\}$，其中 a.a. 意為 almost all，表示除了有限多個 n 之外，我們皆有 $\omega \in A_n$。

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

證明：參照習題 1.12 。 □

範例 3.2.3 ： 令 $(X_n)_{n\geqslant 1}$ 為隨機變數序列以及 $a \in \mathbb{R}$。

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

引理 3.2.4 ： 令 $(A_n)_{n\geqslant 1}$ 為一事件序列。

(1) 若 $\sum_{n\geqslant 1}\mathbb{P}(A_n) < \infty$，則

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

換句話說，集合 $\{n \in \mathbb{N} \mid \omega \in A_n\}$ 為 a.s. 有限。

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,

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

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,

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

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

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

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

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

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

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

□

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 $\frac{1}{n}$.*

(2) 若 $\sum_{n\geqslant 1} \mathbb{P}(A_n) = \infty$ 且 $(A_n)_{n\geqslant 1}$ 為獨立事件序列，則

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

換句話說，集合 $\{n \in \mathbb{N} \mid \omega \in A_n\}$ 為 a.s. 無限。

問題 3.2.5：解釋為什麼在引理 3.2.4 的 (2) 中，我們必須假設 $(A_n)_{n\geqslant 1}$ 為獨立事件。

證明：

(1) 根據假設，我們有

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

也就是說 $\sum_{n\geqslant 1} \mathbb{1}_{A_n} < \infty$ 殆必為真。

(2) 給定 $n_0 \in \mathbb{N}$，對於所有 $n \geqslant n_0$，我們有

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

由於級數 $\sum_{k\geqslant 1} \mathbb{P}(A_k)$ 發散，我們可以得到

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

由於上述式子對於所有 n_0 皆為真，我們有

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

接著取其補集，我們就得到該證明的結果：

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

□

第二小節 應用

在這小節，我們要利用 Borel–Cantelli 引理證明下列結果，其他更多應用請參見習題。

命題 3.2.6 ：在 $\mathbb{N}$ 上，不存在任何機率測度，使得對於任意整數 $n \geqslant 1$，由其倍數構成的正整數集 $n\mathbb{N}$ 測度為 $\frac{1}{n}$。

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 $p_1, \ldots, p_k \in \mathcal{P}$, 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 $m_1, \ldots, m_p \in \{0,1\}$, almost surely there exist infinitely many $k \in \mathbb{N}$ such that*

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

Proof : For any positive integer $n \geqslant 1$, define the random vector $Z_n = (Y_{np+1}, \ldots, Y_{np+p})$. 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

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

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

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

□

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

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

that is

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

證明：假設這樣的測度存在，並記作 $\mathbb{P}$。設 $\mathcal{P}$ 為質數集，且對於所有 $p \in \mathcal{P}$，令 $A_p = p\mathbb{N}$。我們可以得知 $(A_p)_{p\in\mathcal{P}}$ 為獨立事件，因為對於任意互異質數 $p_1, \ldots, p_k \in \mathcal{P}$，我們有

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

此外，我們也知道 $\sum_{p\in\mathcal{P}} \frac{1}{p} = \infty$，所以根據 Borel–Cantelli 引理中的 (2)，我們有 $\mathbb{P}(\limsup A_p) = 1$；也就是說，在 $\mathbb{P}$ 測度之下，幾乎所有正整數 n 都在無限多個 A_p 集合中，但這是不可能的，因為任意正整數不可能同時為無限多個質數的倍數。 □

命題 3.2.7 ： 令 $Y \sim \mathrm{Unif}([0,1])$ 以及如同在式 (3.8) 中的二元展開式 $Y = 0.Y_1Y_2\ldots$。對於任意整數 $p \geqslant 1$ 以及 $m_1, \ldots, m_p \in \{0,1\}$，則殆必存在無窮多個 $k \in \mathbb{N}$ 使得

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

證明：對於任意正整數 $n \geqslant 1$，定義隨機向量 $Z_n = (Y_{np+1}, \ldots, Y_{np+p})$。根據系理 3.1.22 ，我們知道隨機變數 $(Z_n)_{n\geqslant 1}$ 是獨立的。根據引理 3.1.23 ，我們還知道他們有相同的分佈。對於每個 $n \geqslant 1$，我們有

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

由於事件 $(\{Z_n = (m_1, \ldots, m_p)\})_{n\geqslant 1}$ 是獨立的，且 $\sum_{n\geqslant 1} 2^{-p} = +\infty$，引理 3.2.4 中的 (2) 蘊含

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

□

命題 3.2.8 ： 令 $(X_n)_{n\geqslant 1}$ 為 i.i.d. 隨機變數序列，假設各項皆為 $\mathrm{Ber}(\frac{1}{2})$ 的分佈。設

$$L_n := \max\{k \geqslant 1 : \text{ 存在 } 0 \leqslant i \leqslant n-k \text{ 使得 } X_{i+1} = \cdots = X_{i+k} = 1\}.$$

則我們有

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

也就是

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

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,

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

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,

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

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,

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

證明：我們想要證明，對於所有 $\varepsilon > 0$，

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

首先，讓我們先引入幾個新的記號：

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

也就是從 X_m 開始，最多連續出現 1 的次數。因此我們有

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

我們先證明 (1)。給定 $\varepsilon > 0$，我們有

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

由於 $\sum m^{-(1+\varepsilon)} < \infty$，因此根據引理 3.2.4 中的 (1)，我們得到

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

也就是說

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

因此，我們有

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

接下來，讓我們證明 (2)。給定 $\varepsilon > 0$，我們將 n 個試驗結果分割成長度為 $a_n := \lceil (1-\varepsilon)\log_2(n)\rceil + 1$ 的區間，其中最後一個區間有可能長度長度不滿 a_n。因此，我們區間個數一共有

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

我們將這些區間記作

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

對於所有 $1 \leqslant j \leqslant N_n - 1$，令事件 A_j 為 $\{X_i = 1, \forall i \in I_j\}$，我們有

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

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

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

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

第三章 隨機變數的獨立性

接著，我們來看 L_n 相關的事件機率：

$$\begin{aligned}\mathbb{P}\left(L_n \leqslant (1-\varepsilon)\log_2(n)\right) &\leqslant \mathbb{P}\left(\text{存在至少一個 } i \in I_j \text{ 使得 } X_i \neq 1 : 1 \leqslant j \leqslant N_n - 1\right)\\ &= \prod_{j=1}^{N_n-1}(1-\mathbb{P}(A_j)) \leqslant \left(1-\frac{1}{4}\frac{1}{n^{1-\varepsilon}}\right)^{N_n-1}.\end{aligned}$$

利用式 (3.10)，對於夠大的 n，我們有

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

上述的右項對 n 取和的值為有限，因此根據引理 3.2.4 中的 (1)，我們得到

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

故得證。 □

第三節 獨立隨機變數的和

第一小節 定義及性質

若 μ 及 ν 為兩個在 $\mathbb{R}^d$ 上的機率測度，我們將 $\mu \otimes \nu$ 在映射 $(x, y) \mapsto x + y$ 之下的影像測度記作 $\mu * \nu$，換句話說，$\mu * \nu$ 具有下列性質：對於任意在 $\mathbb{R}^d$ 上的非負可測函數 φ，我們有

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

命題 3.3.1 ： 令 X 及 Y 兩個在 $\mathbb{R}^d$ 上的獨立隨機變數。

(1) 隨機變數 $X+Y$ 的分佈為 $\mathbb{P}_X * \mathbb{P}_Y$。在 X 及 Y 皆有密度函數的情況下，將他們的密度函數記作 p_X 及 p_Y，則 $X+Y$ 也有密度函數，而且可以寫作 $p_X * p_Y$。

(2) 隨機變數 $X+Y$ 的特徵函數為 $\Phi_{X+Y}(\xi) = \Phi_X(\xi)\Phi_Y(\xi)$，換句話說 $\widehat{\mathbb{P}_X * \mathbb{P}_Y} = \widehat{\mathbb{P}}_X\widehat{\mathbb{P}}_Y$。

(3) 若 X 及 Y 為平方可積，則 $K_{X+Y} = K_X + K_Y$。在一維 $d = 1$ 的情況下，我們有 $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$。

證明：

(1) 由於 X 以及 Y 獨立，我們有 $\mathbb{P}_{(X,Y)} = \mathbb{P}_X \otimes \mathbb{P}_Y$，所以對任何在 $\mathbb{R}^d$ 之上的非負可測函數

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 $X = (X_1, \ldots, X_d)$ and $Y = (Y_1, \ldots, Y_d)$. Using their independence, for all $i, j \in \{1, \ldots, d\}$, 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 $\lambda_1, \ldots, \lambda_n$, 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,

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

φ，根據式 (3.11) 中 $*$ 運算子的定義，我們有

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

接著，若 X 及 Y 有密度函數，

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

所以 $p_X * p_Y$ 是 $X+Y$ 的密度函數。我們注意到由於 p_X 及 p_Y 是在 $L^1(\mathbb{R}^d, \lambda)$ 中的函數，所以 $p_X * p_Y$ 幾乎到處定義良好，參見命題 1.4.3 。

(2) 根據定義 2.4.12 以及 X 及 Y 的獨立性，我們有

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

(3) 若將 X 及 Y 寫作 $X = (X_1, \ldots, X_d)$ 及 $Y = (Y_1, \ldots, Y_d)$，根據他們的獨立性，對於所有 $i, j \in \{1, \ldots, d\}$，我們有 $\mathrm{Cov}(X_i, Y_j) = 0$，也會有 $\mathrm{Cov}(X_i + Y_i, X_j + Y_j) = \mathrm{Cov}(X_i, X_j) + \mathrm{Cov}(Y_i, Y_j)$，也就是 $K_{X+Y} = K_X + K_Y$。

□

第二小節 例子

若獨立隨機變數序列 $(X_n)_{n\geqslant 1}$ 中各項皆有相同的分佈，則我們稱之為 i.i.d. 隨機變數序列，意為 independent and identically distributed。對 i.i.d. 隨機變數取和不是太困難的事，通常可以使用特徵函數來計算這種和的分佈，或是直接使用式 (3.11) 中的定義。下面我們給出幾個例子，在習題 3.13 、習題 3.14 及習題 3.15 中有更多其他例子。

命題 3.3.2 ： 若 $(X_k)_{1\leqslant k\leqslant n}$ 為 i.i.d. 隨機變數序列，且各項皆為參數為 λ 的帕松分佈，則 $X_1 + \cdots + X_n$ 為參數為 $n\lambda$ 的帕松分佈。更一般來說，如果 $(X_k)_{1\leqslant k\leqslant n}$ 是個獨立隨機變數序列，且各項分別為參數為 $\lambda_1, \ldots, \lambda_n$ 的帕松分佈，則 $X_1 + \cdots + X_n$ 是個參數為 $\lambda_1 + \cdots + \lambda_n$ 的帕松分佈。

證明：令 $X_i \sim \mathrm{Pois}(\lambda_i)$ 為獨立的帕松分佈隨機變數，我們可以計算他們的特徵函數：

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

因此，$X := X_1 + \cdots + X_n$ 的特徵函數寫作：

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

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

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

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,*

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

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

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

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 (簡單收斂)[1] up to a set of measure zero. [2]

[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.

我們可以認出來這個是 $\mathrm{Pois}(\lambda_1+\cdots+\lambda_n)$ 的特徵函數。因此根據定理 2.4.15 ，我們能夠得知 $X \sim \mathrm{Pois}(\lambda_1+\cdots+\lambda_n)$。 □

命題 3.3.3 ： 若 $(X_k)_{1\leqslant k\leqslant n}$ 為獨立隨機變數序列，且對於所有 $1\leqslant k\leqslant n$，X_k 的分佈為參數為 $(0,\sigma_k^2)$ 的高斯分佈，則 $X_1+\cdots+X_n$ 為參數為 $(0,\sigma_1^2+\cdots+\sigma_n^2)$ 的高斯分佈。

證明：參照習題 3.12 。 □

第三小節 大數定理

下個章節我們才會正式的討論隨機變數的收斂概念，但用我們現在的工具，已經足夠先討論一些非常簡單的收斂例子，順便當作在後續章節討論收斂性質的開端。

定理 3.3.4 【L^2 大數定理】 **：** 設 $(X_n)_{n\geqslant 1}$ 為互無關連的實隨機變數序列，且假設他們有相同的機率分佈。假設 $\mathbb{E}[X_1^2] < \infty$，那麼我們有

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

證明：根據期望值的線性性質，我們有 $\mathbb{E}[\frac{1}{n}(X_1+\ldots X_n)] = \mathbb{E}[X_1]$。接著根據命題 3.3.1 的第三點，我們有

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

所以當 $n\to\infty$，上式趨近於零。 □

註解 3.3.5 ： 在在機率空間中，有不同的收斂概念，在第四章我們會做更仔細的探討。這裡要注意的是，定理 3.3.4 的收斂是在 L^2 空間，而不是在機率空間中的簡單收斂 (simple convergence)[1]。[2]。

下列命題在比較強的條件下，給我們殆必收斂的結果，我們稍後在定理 4.3.1 會證明其加強版。

[1] 也就是殆必收斂，機率為 1 會收斂。
[2] 注意：L^2 收斂和殆必收斂並沒有強弱之分，兩者互不蘊含。

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 $i \in \{1, \ldots, n\}$, 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 on the left-hand side,

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

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,

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

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

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

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),

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

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

Corollary 3.3.7 : *If $(A_n)_{n\geqslant 1}$ 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 $\mathbb{P}(A)$.

命題 3.3.6 ： 令 $(X_n)_{n\geqslant 1}$ 為 i.i.d. 隨機變數序列且滿足 $\mathbb{E}[|X_1|^4] < \infty$。則我們有

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

證明：若對於所有 $i \in \{1, \ldots, n\}$，我們將 X_i 用 $X_i - \mathbb{E}[X_i]$ 取代，我們可以觀察到該證明的收斂式子和原本的是等價的，所以不失一般性，我們可以假設所有的隨機變數 X_i 皆滿足 $\mathbb{E}[X_i] = 0$ 的條件。首先，我們計算左側的四階動差，

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

由於 (X_k) 為獨立且期望值為零的隨機變數，在上式的和中，大部分的項皆為零，除了當四元組 (i_1, i_2, i_3, i_4) 中，四個變數皆相等，或是 $i_1 = i_2, i_3 = i_4$ 及其他排列組合的情況：前者共有 n 種可能，後者有 $3n(n-1)$ 種可能。因此我們得到

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

其中 $C < \infty$ 為一常數。接著我們有

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

上式中我們可以交換期望值與級數和，因為級數中所有項皆非負。這蘊含下列級數和殆必有限 (almost surely finite)，

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

所以級數項殆必收斂至零。 □

系理 3.3.7 ： 若 $(A_n)_{n\geqslant 1}$ 為機率相等的獨立事件序列，則我們有

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

註解 3.3.8 ： 在現代機率發展之前，機率通常會被詮釋為在一連串的隨機試驗中，事件發生的頻率。此系理解釋，用現代機率論的角度出發，可以證明此一事實。

用比較白話的說法來解釋，若我們想要知道某個結果 A 為真的機率，我們可以不斷以獨立的方式重複實驗，並去計算 A 為真的比例，則機率為一（處處收斂，強大數法則）的情況下，此數值會收斂至 $\mathbb{P}(A)$。

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* (捲積半群) 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,

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

第四小節 捲積半群

在第七章探討馬可夫鍊，以及下學期討論連續時間的隨機過程時，我們會用到捲積半群的概念，因此在這裡我們先定義並且探討其基本性質。

假設 $I = \mathbb{Z}_{\geqslant 0}$ 或 $I = \mathbb{R}_{\geqslant 0}$。

定義 3.3.9 ： 令 $(\mu_t)_{t\in I}$ 為在 $\mathbb{R}$ 或 $\mathbb{R}^d$ 上，由 I 標記的機率測度所構成的集合。若

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

則我們稱 $(\mu_t)_{t\in I}$ 為捲積半群 (convolution semigroup) 。

引理 3.3.10 ： 若存在函數 $\varphi : \mathbb{R} \longrightarrow \mathbb{C}$ 使得下列其中一個條件成立：

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

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

則 $(\mu_t)_{t\in I}$ 是個捲積半群。

證明：若 $\widehat{\mu_t}$ 如引理中所敘述，則我們有 $\widehat{\mu_{t+t'}} = \widehat{\mu_t}\widehat{\mu_{t'}} = \widehat{\mu_t * \mu_{t'}}$，接著透過傅立葉變換的單射性質，我們可以得到 $\mu_{t+t'} = \mu_t * \mu_{t'}$ 。 □

範例 3.3.11 ：

(1) 當 $I = \mathbb{Z}_{\geqslant 0}$，給定 $p \in [0, 1]$。對於所有 $n \geqslant 1$，令 μ_n 為二項式分佈 $\mathcal{B}(n, p)$。根據我們對二項式分佈做的詮釋，他會是 i.i.d. 伯努力分佈的和，我們顯然有 $\mu_{n+m} = \mu_n * \mu_m$。不然，我們也可以計算其特徵函數並使用上述引理：$\widehat{\mu_n}(\xi) = (pe^{\mathrm{i}\xi} + 1 - p)^n$。

(2) 當 $I = \mathbb{R}_{\geqslant 0}$，對於所有 $t \geqslant 0$，令 μ_t 為參數為 t 的帕松分佈。我們有

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

(3) 當 $I = \mathbb{R}_{\geqslant 0}$，對於所有 $t \geqslant 0$，令 μ_t 為高斯分佈 $\mathcal{N}(0, t)$。根據引理 2.4.14 ，我們有

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

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 $X \overset{(d)}{=} AZ + B$.

(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*

第四節 複雜一點的隨機變數

這裡我們會使用獨立性來構造一些有趣的機率工具：多元常態分佈與帕松過程。

第一小節 多元常態分佈

這個小節的目的是要把高斯分佈的概念推廣至高維度。下列命題同時定義了多元常態分佈的概念，給出他的重要性質、以及最標準（正則）的構造方式。

命題 3.4.1 ： 令 $X = (X_1, \dots, X_d)$ 為 d 維實隨機變數。我們這裡要證明下列三個條件是等價的，而且當其中一個條件成立時，我們稱 X 為多元常態分佈 (multivariate normal distribution)。

(i) 存在 d 維實隨機變數 $Z = (Z_1, \dots, Z_d)$，其中分量為 i.i.d. 標準常態分佈；存在 $d \times d$ 的方形矩陣 A 以及向量 $B \in \mathbb{R}^d$，使得 $X \overset{(d)}{=} AZ + B$。

(ii) 對於任意 $\alpha \in \mathbb{R}^d$，$\alpha^T X$ 仍然是常態分佈。

(iii) 存在 $d \times d$ 的半正定對稱矩陣 Σ 以及向量 $B \in \mathbb{R}^d$ 使得我們可以將 X 的特徵函數寫作

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

此外，向量 $B = \mathbb{E}[X]$ 為 X 的期望值，矩陣 $\Sigma = AA^T = K_X$ 為向量 X 的共變異數矩陣。

證明： 證明 (i) $\Longrightarrow$ (ii)。我們可以計算 X 的期望值為 B，共變異數矩陣為 AA^T。接著，取 $\alpha \in \mathbb{R}^d$，$\alpha^T X = \sum \alpha_i X_i$ 的分佈會是 $\mathcal{N}(\alpha^T B, \alpha^T AA^T \alpha)$。

證明 (ii) $\Longrightarrow$ (iii)。給定 $\xi \in \mathbb{R}^d$，由於 $\xi^T X$ 仍是常態分佈，將其期望值以及變異數記作 $m = \xi^T\,\mathbb{E}[X]$ 及 $\sigma^2 = \xi^T K_X \xi$，我們知道

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

因此我們可以取 $B = \mathbb{E}[X]$ 以及 $\Sigma = K_X$。

證明 (iii) $\Longrightarrow$ (i)。由於 Σ 為半正定對稱矩陣，存在正交矩陣 P 以及對角矩陣 D 使得 $\Sigma = PDP^T$。給定 i.i.d. 標準常態分佈隨機變數 $Z_1, \dots, Z_d$，我們取 $A = P\sqrt{D}$，我們可以計算 $AZ + B$ 的特徵方程式並證明其為 Φ_X。 □

命題 3.4.2 ： 令 $X = (X_1, \dots, X_d)$ 為 d 維度、期望值為 B，共變異數矩陣為 Σ 的多元常態分

expectation B and covariance matrix Σ. If Σ is invertible, then the density function of X writes,

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

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 $X \overset{(d)}{=} AZ+B$ where $Z = (Z_1,\ldots,Z_d)$ are i.i.d. standard normal random variables. The density function of Z writes

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

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,

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

□

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

佈。若 Σ 為可逆矩陣，則 X 的密度函數寫作

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

證明：由於 Σ 是半正定對稱矩陣，我們取正交矩陣 P 及對角矩陣 D 使得 $\Sigma = PDP^T$ 並定義 $A = P\sqrt{D}$。根據命題 3.4.1 的定義、性質及構造，給定 i.i.d. 標準常態分佈 $Z = (Z_1,\ldots,Z_d)$，我們知道 $X \overset{(d)}{=} AZ+B$。我們將 Z 的密度函數記作

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

給定非負可測函數 $f:\mathbb{R}^d\longrightarrow\mathbb{R}_+$，我們有

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

其中在第二行的等式中，我們利用變數變換 $x = Az+B$，且因為 A 是個可逆矩陣，$\mathbb{R}^d$ 與自己在 $z\mapsto Az+B$ 之下是個微分同胚。由於上式對於所有非負可測函數皆成立，根據註解 2.1.16，我們可以得到 X 的密度函數：

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

□

第二小節 帕松過程

我們想要去描述一些會在時間中發生的隨機事件。我們可能會想知道，例如：什麼時候這些事件會發生，或是在給定固定時間時，總共會有多少個隨機事件是已經發生的。我們可以使用下列的方式來討論這樣的現象。給定隨機變數序列 $(X_i)_{i\geqslant 1}$，其中 X_i 描述的是連續兩個事件 $i-1$ 及 i 之間間隔的等待時間；S_n 描述的是等待的所有時間，才會讓第 n 個事件發生；N_t 描述的則是在時間 t 以前（包含時間 t），總共發生的事件數量。接著，讓我們用數學的方式來討論。

在機率空間 $(\Omega,\mathcal{A},\mathbb{P})$ 上，我們令 $(X_i)_{i\geqslant 1}$ 為 i.i.d. 分佈為 $\mathrm{Exp}(1)$ 的指數隨機變數，並且定義 $S_0 = 0$ 及

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

根據習題 3.14，我們知道 S_n 是個伽瑪分佈 $\Gamma(n,1)$。在這個小節中，我們要討論的是下面這個隨機過

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, $\frac{S_n}{n} \xrightarrow{\text{a.s.}} 1$.

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 $N_t \sim \mathrm{Pois}(t)$ 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.

程 (stochastic process)：

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

稱作帕松過程 (Poisson process)。我們不難看出來，$t \mapsto N_t$ 是個（隨機）非遞減函數。

命題 3.4.3 ： 我們有下列性質：

(1) $(S_n)_{n\geqslant 0}$ 幾乎必然為嚴格遞增數列，並且幾乎必然發散至 ∞。

(2) 我們有大數法則 $\frac{S_n}{n} \xrightarrow{\text{a.s.}} 1$。

證明：

(1) 首先，我們可以注意到，由於

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

加上總共只有可數多個事件，我們可以得到

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

接著，透過 Borel–Cantelli 的第二部份，我們知道給定任何 $\alpha > 0$ 時，我們有

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

使用 $(X_n)_{n\geqslant 1}$ 的獨立性，我們得知存在無窮多個 $n \geqslant 1$ 使得 $X_n \geqslant \alpha$，因此 $S_n \xrightarrow{\text{a.s.}} \infty$。

(2) 我們使用命題 3.3.6 中 L^4 版本的大數法則，得到

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

□

命題 3.4.4 ： 固定 $t > 0$，則 $N_t \sim \mathrm{Pois}(t)$ 是個參數為 t 的帕松分佈。

註解 3.4.5 ： 從範例 3.3.11 (2)，我們直接得知由分佈 $(\mathbb{P}_{N_t})_{t\geqslant 0}$ 構成的序列是個捲積半群。

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,

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

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

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,

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

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

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 $\mathrm{Pois}(T)$, 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.

證明：固定正實數 $t > 0$ 及非負整數 $n \geqslant 0$，我們有

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

其中在第二個等式中，我們使用了 $S_{n+1} = S_n + X_{n+1}$ 為兩個獨立隨機變數的和，且我們有 $S_n \sim \Gamma(n, 1)$；在第三個等式中，我們使用了富比尼定理。 □

命題 3.4.6 ： 隨機過程 $t \mapsto N_t$ 在 $\mathbb{R}_{\geqslant 0}$ 上幾乎必然是個 càdlàg (continue à droite, limite à gauche) 函數，也就是右連續且左極限存在的函數。

註解 3.4.7 ： 我們注意到，下面兩個關於右連續的敘述是不同的：

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

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

一般來講，第一個敘述會比第二個來得強，因為 $\mathbb{R}_{>0}$ 是不可數的。我們必須記住，括號放的位置會影響敘述的意義。

證明：我們需要證明下列兩個性質：

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

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

對於 $T > 0$，讓我們證明 $t \mapsto N_t$ 在 $[0, T]$ 幾乎必然是個 càdlàg 函數。命題 3.4.4 告訴我們，N_T 的分佈是 $\mathrm{Pois}(T)$，因此幾乎必然為有限。令 $\Omega_T \in \mathcal{A}$ 使得 $\mathbb{P}(\Omega_T) = 1$ 以及

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

對於 $\omega \in \Omega_T$，函數 $t \mapsto N_t$ 在 $[0, T]$ 上是非遞減且有界的，因此左極限處處存在。對於 $\omega \in \Omega_T$ 以及 $s \in [0, T)$，令 $n := n(\omega)$ 使得 $N_s = n$，這與 $S_n \leqslant s$ 和 $S_{n+1} > s$ 等價。這代表著，對於 $u \in [s, S_{n+1}) \neq \varnothing$，我們一定也會有 $N_u = n$，這蘊含右連續性。

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

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

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

最後，我們這樣總結：定義 $\Omega' := \bigcap_{T\geqslant 1}\Omega_T$ 並注意到 $\mathbb{P}(\Omega') = 1$ 且在 Ω' 上，對於每個整數 $T \geqslant 1$，函數 $t\mapsto N_t$ 在 $[0,T]$ 是 càdlàg 的。因此，函數 $t\mapsto N_t$ 幾乎必然在 $\mathbb{R}_{\geqslant 0}$ 上會是 càdlàg 的。□

引理 3.4.8 ： 令 $(N_s)_{s\geqslant 0}$ 為帕松過程以及 $t > 0$。定義他的平移過程 $(N_s^{(t)})_{s\geqslant 0}$ 如下：

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

那麼 $(N_s^{(t)})_{s\geqslant 0}$ 是個帕松過程，且會與 N_t 獨立。

證明： 我們固定 $t > 0$ 並且考慮在時間 t 之後的帕松過程。從時間 t 開始計算，需要等待 $S_{N_t+1} - t$ 下一個事件才會發生，接著需要等待 $X_{N_t+2}, X_{N_t+3}, \ldots$。因此我們可以定義

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

對於任意 $s > 0$，我們有

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

我們可以注意到，此定義與式 (3.12) 非常相似。若我們能夠對於所有非負整數 $n \geqslant 0$ 證明：

(a) 事件 $\{N_t = n\}$ 與隨機變數序列 $(X_m^{(t)})_{m\geqslant 1}$ 獨立，而且

(b) 在 $\{N_t = n\}$ 成立之下，$(X_m^{(t)})_{m\geqslant 1}$ 也是 i.i.d. 參數為 1 的指數隨機變數序列，

那麼我們不僅證明了 $N^{(t)} = (N_s^{(t)})_{s\geqslant 0}$ 與 N_t 的獨立性，還證明了 $(N_s^{(t)} = N_{t+s} - N_t)_{s\geqslant 0}$ 與原本的 $(N_t)_{t\geqslant 0}$ 會有相同分佈。

固定非負整數 $n \geqslant 0$。首先，由於 $\{N_t = n\} = \{S_n \leqslant t < S_{n+1}\}$，我們知道 $\{N_t = n\}$ 只取決於 $(X_i)_{1\leqslant i\leqslant n+1}$，因此與 $(X_i)_{i\geqslant n+2}$ 是獨立的。我們需要先討論 $\{N_t = n\}$ 與 $S_{n+1} - t = X_{n+1} + S_n - t$ 的獨立性。固定 $y \geqslant 0$，我們有

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

其中在第一行中，我們使用 $\{S_{n+1} > t + y\} \subseteq \{S_{n+1} > t\}$；在第二行中，我們將 $S_n \sim \Gamma(n,1)$ 的密度函數記作 $\gamma_n(x)$；在第三行中，使用了指數隨機變數的無記憶性質。因此，上面計算告

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.

□

訴我們，$\{N_t = n\}$ 與 $S_{n+1} - t = X_{n+1} + S_n - t$ 是獨立的，且 $S_{n+1} - t \sim \mathrm{Exp}(1)$。由於 $S_{n+1} - t$ 只取決於 $(X_i)_{1\leqslant i\leqslant n+1}$，因此也會與 $(X_i)_{i\geqslant n+2}$ 獨立。

□

命題 3.4.9 ： 帕松過程 $(N_t)_{t\geqslant 0}$ 有下列兩個性質：

(1) 給定時間點 $0 = t_0 < t_1 < \cdots < t_k$，隨機過程的增量 $(N_{t_{i+1}} - N_{t_i})_{0\leqslant i\leqslant k-1}$ 為獨立序列。

(2) 隨機過程的增量為帕松分佈：固定 $0 \leqslant s < t$，我們有

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

證明：

(1) 因此，對任意非負整數序列 $(m_i)_{1\leqslant i\leqslant k}$，根據引理 3.4.8 ，我們有

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

最後透過遞迴，我們可以得到

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

也就是性質 (1)。

(2) 使用性質 (1) 及命題 3.4.4 ，我們可以直接得到此結果。

□