機率論基礎

第二章

第一節 基礎定義

在這個章節，我們將定義並且討論在機率論中最基本的概念，包含機率空間、隨機變數、期望值、特徵函數等等。

機率論可以被視為測度論的特例，也就是說當測度總質量為有限或者為 1 的情況，所以在第一章中複習過的很多概念以及定理，在機率論中還是有效的。此外，從機率論的角度來看，機率空間本身並不是這麼重要，我們感興趣的是隨機變數可以取的值，以及這些值出現的頻率。

在機率論以及測度論中的不同概念有下列對應：

機率空間	$\longleftrightarrow$	測度空間
事件	$\longleftrightarrow$	σ 代數中的元素
隨機變數	$\longleftrightarrow$	可測函數
期望值	$\longleftrightarrow$	積分

再來我們會將上述不同名詞給出明確的定義，以及提供不同例子。

第一小節 機率空間

我們在第 1.1 節中回顧了測度空間的定義，在機率論中，我們需要定義的機率空間其實只是它的一個特例。

定義 2.1.1 ： 令 $(\Omega, \mathcal{A})$ 為一個可測空間，且令 $\mathbb{P}$ 為一個在 $(\Omega, \mathcal{A})$ 之上總質量為 1 的測度，稱作機率測度 (probability measure) ，那麼我們將 $(\Omega, \mathcal{A}, \mathbb{P})$ 稱為一個機率空間 (probability space) 。

- 集合 Ω 稱作樣本空間 (sample space) ，可以視為隨機試驗中，描述「隨機單位」的空間。
- 集合 $\mathcal{A}$ 包含了所有的可測事件，也可以簡稱做事件，也就是所有 Ω 可以被「機率」所測量的子集合。換句話說，元素 $A \in \mathcal{A}$ 是個 Ω 的子集合，其中包含了所有滿足特定條件的「隨機單位」。
- 對於任何 $A \in \mathcal{A}$，$\mathbb{P}(A)$ 描述可測事件 A 發生的機率。
- 若樣本空間 Ω 為離散（有限或可數無限），則我們可以考慮 $\mathcal{A} = \mathcal{P}(\Omega)$，並將任何定義在可測空間 $(\Omega, \mathcal{A})$ 上的機率測度 $\mathbb{P}$ 稱作離散機率 (discrete probability)，機率空間 $(\Omega, \mathcal{A}, \mathbb{P})$ 則稱作離散機率空間 (discrete probability space)。

範例 2.1.2 【離散機率空間】： 給定一顆六面數字分別為 1 至 6 的公正骰子。

(1) 考慮骰子丟兩次的機率實驗，那麼我們可以令機率空間為 $\Omega = \{1, \ldots, 6\}^2$，可以測量的事件集合為 $\mathcal{A} = \mathcal{P}(\Omega)$，機率測度 $\mathbb{P}$ 則有下列性質：對於任何 $A \in \mathcal{A}$，我們有 $\mathbb{P}(A) = \frac{|A|}{36}$。

(2) 考慮一顆不公正的銅板，其出現正面的機率是反面的兩倍，考慮投擲此銅板三次的實驗，那麼我們可以令機率空間為

$$\begin{aligned}\Omega = \{&(\text{正},\text{正},\text{正}),(\text{正},\text{正},\text{反}),(\text{正},\text{反},\text{正}),(\text{正},\text{反},\text{反}),\\ &(\text{反},\text{正},\text{正}),(\text{反},\text{正},\text{反}),(\text{反},\text{反},\text{正}),(\text{反},\text{反},\text{反})\},\end{aligned}$$

可以測量的事件集合為 $\mathcal{A} = \mathcal{P}(\Omega)$，機率測度 $\mathbb{P}$ 則有下列性質：若對於任何 $\omega \in \Omega$，我們記正面及反面出現次數為 $H(\omega)$ 及 $T(\omega)$，那麼我們有 $\mathbb{P}(\omega) = (\frac{2}{3})^{H(\omega)}(\frac{1}{3})^{T(\omega)}$。

範例 2.1.3 【一般機率空間】： 平面上的單位圓可以記作

$$\mathbb{S}^1 := \{z \in \mathbb{C} : |z| = 1\} \simeq \mathbb{R}/(2\pi\mathbb{Z}),$$

且可以詮釋為二維空間中的單位方向。若我們考慮可測空間 $(\mathbb{S}^1, \mathcal{B}(\mathbb{S}^1))$ 及機率測度

$$\mathbb{P}([a,b]) = \frac{b-a}{2\pi}, \quad a \leqslant b, |b-a| \leqslant 2\pi$$

則 $\mathbb{P}$ 會是個在 $\mathbb{S}^1 = \mathbb{R}/(2\pi\mathbb{Z})$ 上的均勻機率。這也可以視為由在 $\mathbb{R}$ 上的勒貝格測度「引導」出來在商空間 $\mathbb{S}^1$ 上的（機率）測度。

第二小節 隨機變數

在機率論中，其實機率空間本身顯得並沒有這麼重要，因為我們感興趣的是在隨機試驗中，可以觀察且測量的量，以及這些不同結果發生的頻率。因此，我們在下面定義隨機變數的概念，並將它視為「在一個隨機試驗中，隨機單位事件給出的結果」。

定義 2.1.4 ： 令 $(\Omega, \mathcal{A}, \mathbb{P})$ 為機率空間及 $(E, \mathcal{E})$ 為可測空間，則我們稱可測函數 $X : \Omega \longrightarrow E$ 為取值在 E 中的隨機變數 (random variable)。

範例 2.1.5 ： 我們取上章節範例 2.1.2 中的機率空間，並考慮在上面定義的隨機變數。

(1) 定義 $X((i,j)) = i + j$，那麼 X 是個值域為 $\{2, \ldots, 12\}$ 的隨機變數。

(2) 定義 $H(\omega)$ 為「在 ω 之中，正面出現的次數」，那麼 H 是個值域為 $\{0, 1, 2, 3\}$ 的隨機變

數。

範例 2.1.6 ： 考慮範例 2.1.3 定義的機率空間，並考慮隨機變數

$$X(\omega) = \cos(\omega), \qquad Y(\omega) = \sin(\omega), \qquad \forall \omega \in \mathbb{S}^1.$$

則隨機變數 X 及 Y 取值在 $[-1, 1]$ 之中並可被視為均勻隨機方向在 x 軸及 y 軸上的投影。

隨機變數 $X : \Omega \longrightarrow E$ 的定義域（也就是機率空間）本身並沒這麼重要，我們比較在意的是它的值域，也就是說，他取不同值的機率。換句話說，我們想要理解機率測度 $\mathbb{P}$ 在 X 函數作用下的影像測度。

定義 2.1.7 ： 定義 $\mathbb{P}_X$ 為測度 $\mathbb{P}$ 在隨機變數 X 之下的影像測度 (image measure) ，我們稱之為隨機變數 X 的分佈 (distribution) 或是律 (law)。換句話說，$\mathbb{P}_X$ 是個在可測空間 $(E, \mathcal{E})$ 之上的機率測度，可以寫作

$$\mathbb{P}_X(B) := \mathbb{P}(X^{-1}(B)), \qquad \forall B \in \mathcal{E}.$$

在機率論的語言中，通常我們也會將上述縮寫為

$$\mathbb{P}_X(B) = \mathbb{P}(X \in B) = \mathbb{P}(\omega \in \Omega : X(\omega) \in B), \qquad \forall B \in \mathcal{E}.$$

$$\begin{tikzcd}
(\Omega, \mathcal{A}, \mathbb{P}) \arrow[dd, "X"'] \arrow[dr, "\mathbb{P}"] & \\
& {[0,1]} \\
(E, \mathcal{E}) \arrow[ur, "\mathbb{P}_X"] &
\end{tikzcd}$$

圖 2.1: 描述影像測度 $\mathbb{P}_X$ 的示意圖，此影像測度也可以看作為機率測度 $\mathbb{P}$ 透過隨機變數 X 向前推進 (pushforward) 所得到的測度，也記作 $X_*\mathbb{P}$。我們真正在意的會是機率測度 $\mathbb{P}_X$ 而不是 $\mathbb{P}$。

我們應該以下面方式來理解影像測度：對於任意機率空間中的一點 $\omega \in \Omega$，$X(\omega)$ 是個在 E 中的點，那麼 $\mathbb{P}_X(B)$ 則表示這個隨機點在 B 中的機率。

範例 2.1.8 ： 令 $(\Omega, \mathcal{A}, \mathbb{P})$ 為機率空間，其中 $\Omega = [0, \pi]$、$\mathcal{A}$ 是伯雷爾 σ 代數，以及 $\mathbb{P} = \frac{1}{\pi}\lambda$。考慮隨機變數 $X : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}$ 定義做：

$$X(\omega) = \sin(\omega), \quad \omega \in \Omega = [0, \pi].$$

影像測度 $\mathbb{P}_X$ 是定義在可測空間 $([0,1], \mathcal{B}([0,1]))$ 上。我們可以透過 $\mathbb{P}_X([a,b])$ 對於所有 $0 \leqslant a \leqslant b \leqslant 1$ 的值，來刻劃這個影像測度。對於 $0 \leqslant a \leqslant b \leqslant 1$，我們有

$$\begin{aligned}
\mathbb{P}_X([a,b]) &= \mathbb{P}(\omega \in \Omega : \sin(\omega) \in [a,b]) \\
&= \mathbb{P}(\omega \in \Omega : \omega \in [\sin^{-1} a, \sin^{-1} b]) + \mathbb{P}(\omega \in \Omega : \omega \in [\pi - \sin^{-1} b, \pi - \sin^{-1} a]) \\
&= \frac{2}{\pi}(\sin^{-1} b - \sin^{-1} a).
\end{aligned}$$

註解 2.1.9 ： 若隨機變數 $X : (\Omega_1, \mathcal{A}_1, \mathbb{P}_1) \longrightarrow (E, \mathcal{E})$ 及 $Y : (\Omega_2, \mathcal{A}_2, \mathbb{P}_2) \longrightarrow (E, \mathcal{E})$ 有相同的分佈，也就是說 $\mathbb{P}_{1,X} := X_*\mathbb{P}_1 = Y_*\mathbb{P}_2 =: \mathbb{P}_{2,Y}$，則我們可以記作

$$X \overset{(\mathrm{d})}{=} Y \qquad \text{或} \qquad X \sim Y \qquad \text{或} \qquad X \sim \mathbb{P}_{2,Y}.$$

若我們要說隨機變數 X 的分佈是 μ，也就是說 $\mathbb{P}_{1,X} = \mu$，則我們也可以記作

$$X \sim \mu,$$

並說 X 遵從 μ 的分佈。

註解 2.1.10 ： 在 $(E, \mathcal{E}) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ 的情況下，$\mathbb{P}_X$ 是個在 $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ 上的機率測度，若他對 $\mathbb{R}^d$ 上的勒貝格測度 λ 絕對連續，則根據 Radon–Nikodym 定理（定理 1.3.16 ），我們知道存在密度函數 $g : \mathbb{R}^d \longrightarrow \mathbb{R}_+$ 使得我們可以將 $\mathbb{P}_X$ 寫作 $\mathbb{P}_X = g \cdot \lambda$。我們稱 g 為隨機變數 X 的機率密度函數 (probability density function)。

註解 2.1.11 【隨機變數的正則構造】 ： 若 μ 是個在 $\mathbb{R}^d$ 上的機率測度，我們可以用下列方式來構造一個分佈為 μ 的隨機變數：取 $\Omega = \mathbb{R}^d$，$\mathcal{A} = \mathcal{B}(\mathbb{R}^d)$，$\mathbb{P} = \mu$ 並且令 $X(\omega) = \omega$，我們可以檢查 X 的分佈是 μ。稍後在命題 2.1.23 中，當我們給定 $\mathbb{R}$ 上的機率測度 μ 時，我們會看到該如何從均勻測度出發，來構造分佈為 μ 的實數隨機變數。

第三小節 期望值

有了隨機變數之後，第一個重要的量是其期望值。

定義 2.1.12 ： 令 X 為一值域為 $\mathbb{R}$ 的隨機變數（也稱為實隨機變數），在下列情況下：

- X 非負；
- X 可取正負值，且 $\int |X| \, \mathrm{d}\mathbb{P} < \infty$，

我們可以定義其期望值 (expectation)，並記作

$$\mathbb{E}[X] = \int_\Omega X(\omega) \, \mathbb{P}(\mathrm{d}\omega). \tag{2.1}$$

若 $X = (X_1, \ldots, X_d)$ 是個 d 維的實隨機變數（也就是其值域為 $\mathbb{R}^d$），若所有的期望值 $\mathbb{E}[X_i]$ 都是有定義的，我們可以定義 X 的期望值為 $\mathbb{E}[X] = (\mathbb{E}[X_1], \ldots, \mathbb{E}[X_d])$。

註解 2.1.13 ： 令 B 為一可測子集以及 $X = \mathbb{1}_B$，則 $\mathbb{E}[X] = \mathbb{P}(B)$。一般來講，我們可以將 $\mathbb{E}[X]$ 視為隨機變數 X 的「平均值」，例如在 Ω 為有限集合，且 $\mathbb{P}$ 為均勻分佈的情況下，$\mathbb{E}[X]$ 代表的是 X 所有可能取的值的（加權）平均值。

命題 2.1.14 ： 令 X 為一值域為 $(E, \mathcal{E})$ 的隨機變數，則對於任何可測函數 $f : E \longrightarrow [0, \infty]$，我們有

$$\mathbb{E}[f(X)] = \int_E f(x) \mathbb{P}_X(\mathrm{d}x).$$

證明： 我們首先證明對於任意可測集合 $B \in \mathcal{E}$，指標函數 $f = \mathbb{1}_B$ 滿足此式，接著透過線性組合，證明對於所有的非負簡單函數皆為真。再來，由於非負可測函數可以寫作非負簡單函數序列的非遞減極限（命題 1.2.14），根據單調收斂定理，得證。 □

註解 2.1.15 ： 若 f 不恆正，只要期望值 $\mathbb{E}[|f(X)|]$ 是有限的，上述命題依然為真。不要忘記可取正負值的一般函數，他的積分也是以相同方式定義的，請見定義 1.2.18。

註解 2.1.16 ： 機率分佈 $\mathbb{P}_X$ 可以讓我們輕易計算 $f(X)$ 類型的隨機變數的期望值。反過來看，對於任意可測集合 $B \in \mathcal{E}$，若我們考慮 $f = \mathbb{1}_B$，則 $\mathbb{E}[f(X)] = \mathbb{P}(X \in B) = \mathbb{P}_X(B)$。不要忘記，機率測度 $\mathbb{P}_X$ 是定義在 $\mathcal{E}$ 上的，也就是說，如果我們知道 $(\mathbb{P}_X(B))_{B \in \mathcal{E}}$，則此機率測度是被唯一決定的。換句話說，若是我們有辦法對「夠多的」可測函數 f 寫出下列式子

$$\mathbb{E}[f(X)] = \int f \, \mathrm{d}\nu, \tag{2.2}$$

則透過上面的討論，我們知道隨機變數 X 的分佈會是被機率測度 ν 所描述的。在實際操作上，我們可以對所有非負可測函數，或是在 $\mathcal{C}_c(\mathbb{R}^d)$ 中所有的函數來計算式 (2.2)。

範例 2.1.17 ： 我們考慮與範例 2.1.8 中相同的範例，並且比較兩邊的計算。回顧 $(\Omega, \mathcal{A}, \mathbb{P})$ 是個機率空間，其中 $\Omega = [0, \pi]$、$\mathcal{A}$ 是伯雷爾 σ 代數，以及 $\mathbb{P} = \frac{1}{\pi}\lambda$。隨機變數 $X : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}$ 定義做：

$$X(\omega) = \sin(\omega), \quad \omega \in \Omega = [0, \pi].$$

我們可以透過對於所有非負可測函數 $f : \mathbb{R} \to \mathbb{R}$，計算 $\mathbb{E}[f(X)]$ 的值，來刻劃 X 的分佈。對於非負可測函數 $f : \mathbb{R} \to \mathbb{R}$，我們記

$$\begin{aligned}
\mathbb{E}[f(X)] = \mathbb{E}[f(X(\omega))] &= \frac{1}{\pi}\int_0^{\pi} f(\sin\omega)\,\mathrm{d}\omega \\
&= \frac{2}{\pi}\int_0^{\pi/2} f(\sin\omega)\,\mathrm{d}\omega \\
&= \frac{2}{\pi}\int_0^{1} f(x)\frac{\mathrm{d}x}{\sqrt{1-x^2}},
\end{aligned}$$

其中在第二行，我們使用積分項在 $\omega \mapsto \pi - \omega$ 下的對稱性；在第三行，我們使用變數變換 $x = \sin\omega$。這代表著，影像測度 $\mathbb{P}_X$ 寫做：

$$\mathbb{P}_X(\mathrm{d}x) = \frac{2}{\pi}\frac{\mathrm{d}x}{\sqrt{1-x^2}}.$$

這是個對於勒貝格測度來說，是絕對連續的測度，他的密度函數為 $x \mapsto \frac{2}{\pi}\frac{1}{\sqrt{1-x^2}}$。

下列命題給我們一個重要的例子，來幫助理解這樣的概念。

命題 2.1.18 ： 令 $(X_1, \ldots, X_d)$ 為一值域為 $\mathbb{R}^d$ 的隨機變數。假設其機率密度函數為 $p(x_1, \ldots, x_d)$，則對於任何 $1 \leqslant j \leqslant d$，隨機變數 X_j 的機率密度函數也存在，且可以寫作

$$p_j(x) = \int_{\mathbb{R}^{d-1}} p(x_1, \ldots, x_{j-1}, x, x_{j+1}, \ldots, x_d)\,\mathrm{d}x_1 \ldots \mathrm{d}x_{j-1}\,\mathrm{d}x_{j+1} \ldots \mathrm{d}x_d.$$

註解 2.1.19 ： 考慮一個二維隨機變數的情況下（$d = 2$），我們有

$$p_1(x) = \int_{\mathbb{R}} p(x, y)\,\mathrm{d}y, \qquad p_2(y) = \int_{\mathbb{R}} p(x, y)\,\mathrm{d}x.$$

證明：令 π_j 為對於 j 座標的投影：$\pi_j(x_1,\dots,x_d)=x_j$，對於所有非負可測函數 $f:\mathbb{R}\longrightarrow\mathbb{R}_+$，我們可以使用富比尼定理：

$$\begin{aligned}\mathbb{E}[f(X_j)]=\mathbb{E}[f(\pi_j(X))]&=\int_{\mathbb{R}^d} f(x_j)p(x_1,\dots,x_d)\,\mathrm{d}x_1\dots\mathrm{d}x_d\\&=\int_{\mathbb{R}} f(x_j)\Big(\int_{\mathbb{R}^{d-1}} p(x_1,\dots,x_d)\,\mathrm{d}x_1\dots\mathrm{d}x_{j-1}\,\mathrm{d}x_{j+1}\dots\mathrm{d}x_d\Big)\,\mathrm{d}x_j\\&=\int_{\mathbb{R}} f(x_j)p_j(x_j)\,\mathrm{d}x_j.\end{aligned}$$

□

註解 2.1.20： 若 $X=(X_1,\dots,X_d)$ 是個值域 d 維的實隨機變數，我們稱 X_j 的分佈為邊緣分佈 (marginal distribution) ，並記為 $\mathbb{P}_{X_j}$。根據上述命題，我們可以輕易得到 $\mathbb{P}_{X_j}=(\pi_j)_*\mathbb{P}_X$，也就是說，從 X 的分佈我們可以輕易求得任意 X_j 的分佈。需要知道的是，若我們知道所有的邊緣分佈 $\mathbb{P}_{X_j}$，我們是無法找回 X 的分佈的。

問題 2.1.21：試著構造兩個二維的實隨機變數 $X=(X_1,X_2)$ 及 $X'=(X'_1,X'_2)$ 使得對於 $j=1,2$，X_j 及 X'_j 有著相同的邊緣分佈，但 X 及 X' 的分佈卻不相同。

第四小節 累積分佈函數

定義 2.1.22 ： 令 X 為一實隨機變數，我們定義函數 $F_X:\mathbb{R}\longrightarrow[0,1]$ 為

$$F_X(t)=\mathbb{P}(X\leqslant t)=\mathbb{P}_X((-\infty,t]),\qquad t\in\mathbb{R}.$$

我們稱他為隨機變數 X 的累積分佈函數 (cumulative distribution function) ，記作 c.d.f.，也可以稱作分佈函數 (distribution function) 。

我們提醒在定理 1.2.29 中有看到，F_X 是個非遞減、右連續，且在 $-\infty$ 及 $+\infty$ 的極限分別為 0 及 1 的函數。此外我們知道，逆命題也成立，也就是說如果 F 是個滿足上列條件的函數，則我們可以找到唯一的機率測度 μ 使得對於所有 $t\in\mathbb{R}$，我們有 $F(t)=\mu((-\infty,t])$。也就是說 F 是個實隨機變數的累積分佈函數。

此外，累積分佈函數 F_X 也可以完整描述隨機變數 X 的分佈 $\mathbb{P}_X$，其中我們有：

$$\begin{aligned}\mathbb{P}(a\leqslant X\leqslant b)&=F_X(b)-F_X(a-), &\quad a\leqslant b,\\\mathbb{P}(a<X<b)&=F_X(b-)-F_X(a), &\quad a<b,\end{aligned}$$

且 F_X 中的不連續點為 $\mathbb{P}_X$ 的原子。

若 $X := (X_1, \dots, X_d)$ 是個取值在 $\mathbb{R}^d$ 中的隨機變數，我們也能夠以類似的方式來討論上面的概念。對於所有 $(t_1, \dots, t_d) \in \mathbb{R}^d$，我們可以定義

$$F_X(t_1, \dots, t_d) := \mathbb{P}(X_1 \leqslant t_1, \dots, X_d \leqslant t_d) = \mathbb{P}_X\big((-\infty, t_1] \times \dots \times (-\infty, t_d]\big).$$

命題 2.1.23 ： 令 $F : \mathbb{R} \longrightarrow [0,1]$ 是個非遞減、右連續，且在 $-\infty$ 及 $+\infty$ 的極限分別為 0 及 1 的函數。定義函數 $h : (0,1) \longrightarrow \mathbb{R}$ 為

$$h(y) := \inf\{z \in \mathbb{R} : F(z) \geqslant y\}, \qquad \forall y \in (0,1). \tag{2.3}$$

若 Y 是個隨機變數使得 $\mathbb{P}_Y$ 的分佈與 $[0,1]$ 上勒貝格測度相同，則隨機變數 $h(Y)$ 的分配函數會是 F，也就是說 $F_{h(Y)} = F$。

註解 2.1.24 ： 在此命題的敘述中，我們注意到，若 F 是個雙射函數（也就是嚴格遞增），則我們有 $h = F^{-1}$。

證明：固定 $y \in (0,1)$ 及 $x \in \mathbb{R}$。若 $F(x) \geqslant y$，根據 h 的定義，我們可以得到 $x \geqslant h(y)$。反過來，若 $F(x) < y$，使用 F 的右連續性，我們可以找到 $\varepsilon > 0$ 使得 $F(x+\varepsilon) < y$，再根據單調性，對於所有 $z \leqslant x + \varepsilon$，我們會有 $F(z) < y$，因此 $h(y) \geqslant x + \varepsilon > x$。因此，我們有下列的等價關係：

$$x \geqslant h(y) \;\Leftrightarrow\; F(x) \;\geqslant y.$$

令 Y 為隨機變數使得 $\mathbb{P}_Y$ 是在 $[0,1]$ 上的勒貝格測度，以及 $Z := h(Y)$。我們注意到，根據命題 1.2.2 ，Z 也會是一個隨機變數。由於 $\mathbb{P}(Y \in (0,1)) = 1$，我們有

$$F_Z(x) = \mathbb{P}(h(Y) \leqslant x) = \mathbb{P}(Y \leqslant F(x)) = F_Y(F(x)) = F(x).$$

□

註解 2.1.25 ： 從命題 2.1.23 ，我們得知若均勻隨機變數存在（這是我們的假設，也是測度論會去構造的東西，這裡不深入討論），則對於任何在 $\mathbb{R}$ 上給定的分佈，我們可以透過他的分佈函數來構造有此分佈的實隨機變數。

第二節　隨機變數的生成 σ 代數

令 $(E, \mathcal{E})$ 為一可測空間。

定義 2.2.1 ： 若 $X:(\Omega,\mathcal{A})\longrightarrow(E,\mathcal{E})$ 為隨機變數，我們可以定義 $\sigma(X)$ 為 $\mathcal{A}$ 最小的子 σ 代數，使得 X 是可測的，我們稱之為由 X 生成的 σ 代數：

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

註解 2.2.2 ： 我們可以將 $\sigma(X)$ 視為需要能夠正確描述隨機變數 X 的最小的 σ 代數。

範例 2.2.3 ： 固定機率空間 $(\Omega,\mathcal{A},\mathbb{P})$ 並考慮隨機變數 $X:(\Omega,\mathcal{A},\mathbb{P})\longrightarrow(\mathbb{R},\mathcal{B}(\mathbb{R}))$。

(1) 若隨機變數 X 取值恆為常數，則 $\sigma(X)=\{\varnothing,\Omega\}$ 為平凡 σ 代數。

(2) 給定可測集合 $A\in\mathcal{A}$，若隨機變數 X 寫作 $X=\mathbb{1}_A$，則 $\sigma(X)=\{\varnothing,A,A^c,\Omega\}$。

(3) 若隨機變數 X 只會取 0 及 1 兩個值，則存在 $A\in\mathcal{A}$ 使得 $X=\mathbb{1}_A$，因此 $\sigma(X)$ 的描述與上面情況像同。

(4) 我們不難驗證，對於任意 $A,B\in\mathcal{A}$ 滿足 $A\notin\{\varnothing,B,B^c,\Omega\}$，隨機變數 $\mathbb{1}_A$ 並不會是 $\sigma(\mathbb{1}_B)$ 可測的，我們看到在樣本空間上選擇不同的 σ 代數，也會造成隨機變數的概念改變。

範例 2.2.4 ： 考慮機率空間 $(\Omega,\mathcal{A},\mathbb{P}):=([0,2),\mathcal{B}([0,2)),\frac{1}{2}\lambda)$，其中 λ 為勒貝格測度。我們定義隨機變數 $X,Y:(\Omega,\mathcal{A},\mathbb{P})\longrightarrow([0,2),\mathcal{B}([0,2)))$ 做 $X(\omega)=\omega$ 及 $Y(\omega)=\lfloor\omega\rfloor$，則

$$\sigma(X)=\mathcal{B}([0,2)),\qquad \text{且}\qquad \sigma(Y)=\{\varnothing,[0,1),[1,2),[0,2)\}.$$

我們可以注意到 $([0,2),\sigma(Y),\mathbb{P})$ 可以被視為一個離散機率空間：定義

$$\Omega':=\{0,1\},\quad \mathcal{A}':=\{\varnothing,\{0\},\{1\},\{0,1\}\},\quad \text{且}\quad \mathbb{P}':=\tfrac{1}{2}\delta_0+\tfrac{1}{2}\delta_1,$$

則 $(\Omega,\mathcal{A},\mathbb{P})$ 與 $(\Omega',\mathcal{A}',\mathbb{P}')$ 有相同的結構。

註解 2.2.5 ： 我們可以將這個定義拓展到任意集合的隨機變數 $X=(X_i)_{i\in I}$。假設對任意 $i\in I$，X_i 的值域為 $(E_i,\mathcal{E}_i)$，我們可以定義

$$\sigma(X)=\sigma(X_i^{-1}(B_i):B_i\in\mathcal{E}_i,i\in I).$$

命題 2.2.6 ： 令 X 為值域在 $(E,\mathcal{E})$ 中的隨機變數，Y 為實隨機變數，則下列性質等價：

(1) Y 對 $\sigma(X)$ 可測。

(2) 存在一可測函數 $f : (E, \mathcal{E}) \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ 使得 $Y = f(X)$。

證明：根據命題 1.2.2 ，若 (2) 成立，則 (1) 成立。

假設 Y 對 $\sigma(X)$ 是可測的，我們想要證明 (2). 首先處理當 Y 為簡單函數的狀況：對於 $i \in \{1, \ldots, n\}$，令 $\lambda_i \in \mathbb{R}$ 以及 $A_i \in \sigma(X)$，且 Y 寫作

$$Y = \sum_{i=1}^{n} \lambda_i \mathbb{1}_{A_i}.$$

對於所有 $i \in \{1, \ldots, n\}$，我們可以找到 $B_i \in \mathcal{E}$ 使得 $A_i = X^{-1}(B_i)$，所以 Y 可以重新寫作

$$Y = \sum_{i=1}^{n} \lambda_i \mathbb{1}_{A_i} = \sum_{i=1}^{n} \lambda_i \mathbb{1}_{B_i} \circ X = f \circ X,$$

其中 $f = \sum_{i=1}^{n} \lambda_i \mathbb{1}_{B_i}$ 是個對於 $\mathcal{E}$ 可測的函數。

在一般情況下，存在對 $\sigma(X)$ 可測的簡單函數序列 (Y_n) 使得 Y_n 簡單收斂至 Y（命題 1.2.14），因此根據此證明前半部所述，對於所有 n，我們可以找到可測函數 $f_n : E \longrightarrow \mathbb{R}$ 使得 $Y_n = f_n(X)$。對於所有 $x \in E$，我們可以設

$$f(x) = \begin{cases} \lim_{n\to\infty} f_n(x) & \text{如果極限存在}, \\ 0 & \text{其他情況}. \end{cases}$$

根據命題 1.2.6 ，函數 f 仍然為可測。此外，我們注意到，對於所有 $\omega \in \Omega$，我們有

$$\lim_{n\to\infty} f_n(X(\omega)) = \lim_{n\to\infty} Y_n(\omega) = Y(\omega).$$

若 $x = X(\omega)$，則在上面的定義中，$\lim f_n(x)$ 的極限是存在的，而且我們有

$$f(X(\omega)) = \lim_{n\to\infty} f_n(X(\omega)) = Y(\omega),$$

也就是說 $Y = f(X)$。 □

第三節 常用機率分佈

在這章節中，我們介紹一些常用的隨機變數分佈。

在第 2.3.1 小節中，我們會刻劃兩種類別的機率分佈：離散分佈以及絕對連續分佈。在第 2.3.2 小節中，我們給出常用的離散隨機變數分佈的例子，在第 2.3.3 小節中我們給出常用的絕對連續隨機變數的例子。

第一小節 機率分佈的分類

我們這裡探討兩個不同類型的隨機變數：離散型隨機變數以及絕對連續型隨機變數。

令 $(\Omega, \mathcal{A}, \mathbb{P})$ 為一機率空間且 $(E, \mathcal{E})$ 為一可測空間，考慮一個值域為 E 的隨機變數 $X : \Omega \longrightarrow E$。

定義 2.3.1 【離散型隨機變數】：當 E 為一可數集合，且 $\mathcal{E} = \mathcal{P}(E)$ 時，X 的分佈可以寫作

$$\mathbb{P}_X = \sum_{x \in E} p_x \delta_x,$$

其中 $p_x = \mathbb{P}(X = x)$，δ_x 代表的是在 x 的狄拉克測度。上述式子可以透過下列式子來理解：

$$\mathbb{P}_X(B) = \mathbb{P}(X \in B) = \mathbb{P}(\bigcup_{x \in B} \{X = x\}) = \sum_{x \in B} \mathbb{P}(X = x) = \sum_{x \in E} p_x \delta_x(B).$$

從上述性質可以得知，如果想要知道一離散隨機變數 X 的分佈，我們需要知道的是對於所有 $x \in E$，機率 $\mathbb{P}(X = x)$ 的值。

當機率空間不為離散時，我們這裡只討論下面的情況，也就是可以定義密度函數的情況。

定義 2.3.2 【絕對連續型隨機變數】：假設 $(E, \mathcal{E}) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$，若 $\mathbb{P}_X$ 對於勒貝格測度 λ 是絕對連續的話，我們說 $\mathbb{P}_X$ 是個絕對連續機率分佈 (absolutely continuous probability distribution)，而 X 是個絕對連續隨機變數 (absolutely continuous random variable)。在這樣的情況下，我們可以使用定理 1.3.16 中的 Radon–Nikodym 定理，以建構一個非負可測函數 $p : \mathbb{R}^d \to \mathbb{R}_+$ 使得

$$\mathbb{P}_X(B) = \int_B p(x)\,\mathrm{d}x.$$

除了在一測度為零的集合之上，此函數 p 的定義是唯一的，我們稱之為 X 的密度函數 (density function) 或機率密度函數 (probability density function)。此外，在一維的情況下 $d = 1$，對於所有 $\alpha \leqslant \beta$，我們有

$$\mathbb{P}(\alpha \leqslant X \leqslant \beta) = \int_\alpha^\beta p(x)\,\mathrm{d}x.$$

註解 2.3.3：若 $\mathbb{P}_X$ 對於任意 $\mathbb{R}^d$ 中的單點測度皆為零，則我們說 $\mathbb{P}_X$ 是個連續測度或連續分佈，且 X 是個連續隨機變數。我們不難驗證，絕對連續測度（或隨機變數）也會是連續的；但在習題 1.33 中我們看到，Cantor 分佈是個連續但不是絕對連續的（機率）測度。

第二小節 離散型隨機變數

下面我們介紹在未來會常常使用的離散機率分佈。令 $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ 為離散機率空間，考慮值域空間 E 以及隨機變數 $X : \Omega \longrightarrow E$，將對應到機率分佈 $\mathbb{P}_X$ 的質量函數記作 f_X，根據不同的下列狀況我們會有不同的分佈。

均勻分佈 (Uniform distribution)　若 E 為有限集合，$n = |E|$ 為其元素個數，且質量函數 f_X 滿足以下條件：

$$f_X(x) = \mathbb{P}(X = x) = \frac{1}{n}, \qquad \forall x \in E,$$

我們稱 X 為在 E 之上的均勻分佈，記作 $X \sim \mathrm{Unif}(E)$。

> 例：投擲一顆六面的公正骰子，點數分別為 $\{1, \ldots, 6\}$。

伯努力分佈 (Bernoulli distribution)　給定參數 $p \in [0, 1]$，若 $E = \{0, 1\}$ 且 f_X 滿足下列條件：

$$f_X(1) = \mathbb{P}(X = 1) = p, \qquad f_X(0) = \mathbb{P}(X = 0) = 1 - p,$$

我們稱 X 為參數為 p 的伯努力分佈，記作 $X \sim \mathrm{Ber}(p)$。

> 例：投擲一顆不公正硬幣，得到正面（記作 1）的機率為 p，反面（記作 0）的機率為 $1 - p$。

二項式分佈 (Binomial distribution)　給定參數 $n \in \mathbb{N}_0$ 及 $p \in [0, 1]$，若 $E = \{0, \ldots, n\}$ 且 f_X 滿足下列條件：

$$f_X(k) = \mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \qquad \forall k \in E,$$

我們稱 X 為參數為 (n, p) 的二項式分佈，記作 $X \sim \mathrm{Bin}(n, p)$。

> 例：考慮上述不公正硬幣，投擲 n 次，則正面出現的次數滿足二項式分佈。

圖 2.2: 二項式分佈的質量函數，左圖的參數為 $n = 30, p = 0.2$，右圖的參數為 $n = 30, p = 0.5$。

幾何分佈 (Geometric distribution)　給定參數 $p \in (0,1)$，若 $E = \mathbb{N}_0$ 且 f_X 滿足下列條件：

$$f_X(k) = \mathbb{P}(X = k) = (1-p)^k p, \qquad \forall k \in \mathbb{N}_0,$$

我們稱 X 為參數為 p 的幾何分佈，記作 $X \sim \mathrm{Geo}(p)$。

> 例：投擲上述不公正硬幣，在得到第一次正面之前反面出現的次數，滿足幾何分佈。若我們把正面視為「成功」，反面視為「失敗」，則此分佈描述成功前必須要經過的失敗次數。

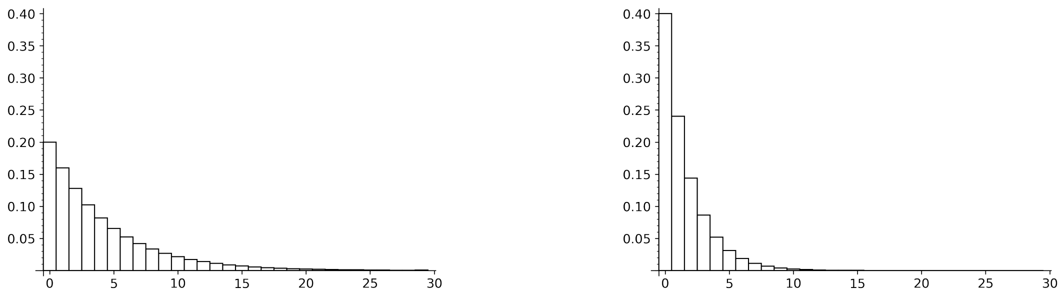

圖 2.3: 幾何分佈的質量函數，左圖的參數為 $n = 30, p = 0.2$，右圖的參數為 $n = 30, p = 0.4$。

超幾何分佈 (Hypergeometric distribution)　給定非負整數 N, K, n 滿足 $0 \leqslant K \leqslant N$，取 $E = \{(n-N+K) \vee 0, \ldots, K \wedge n\}$ 且 f_X 滿足下列條件：

$$f_X(k) = \mathbb{P}(X = k) = \binom{K}{k}\binom{N-K}{n-k}\Bigg/\binom{N}{n}, \qquad \forall k \in E,$$

我們稱 X 為參數為 (N, K, n) 的超幾何分佈，記作 $X \sim \mathrm{Hypergeo}(N, K, n)$。

> 例：在一個球箱中，有 N 顆球，其中 K 顆是白色的，當我們進行不重複抽樣取 n 顆球時，抽到的白球數量會滿足超幾何分佈 $\mathrm{Hypergeo}(N, K, n)$。

帕松分佈 (Poisson distribution)　令 $\lambda > 0$ 為一參數，$E = \mathbb{N}_0$ 且 $\mathbb{P}$ 滿足下列條件：

$$\mathbb{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \qquad \forall k \in \mathbb{N}_0.$$

我們稱 X 為參數為 λ 的帕松分佈，記作 $X \sim \mathrm{Pois}(\lambda)$。

圖 2.4: 帕松分佈的質量函數，左圖的參數為 $\lambda = 5$，右圖的參數為 $\lambda = 10$。

註解 2.3.4 ： 不管從理論或是應用的角度來看，帕松分佈都具有重要意義。在稍後的章節，當我們探討機率分佈的收斂時，我們可以證明帕松分佈可以被理解為二項式分佈的極限。換句話說，令 X_n 為滿足二項式分佈 $\mathrm{Bin}(n, p_n)$ 的隨機變數，且當 n 趨近無窮大時，我們有 $np_n \to \lambda$，則

$$\lim_{n\to\infty} \mathbb{P}(X_n = k), \quad \forall k \in \mathbb{Z}_{\geqslant 0}.$$

第三小節 有密度的隨機變數

在這個章節中，我們定義常見的連續隨機變數分佈，並且討論他們的性質。我們將所要討論的連續隨機變數記作 X，並將他的機率密度函數記作 f_X。

均勻分佈 (Uniform distribution)　令 $a < b$，若 p 可寫為

$$p(x) = \frac{1}{b-a}\mathbb{1}_{[a,b]}(x).$$

則稱 X 為在區間 $[a, b]$ 上的均勻分佈，記作 $X \sim \mathrm{Unif}([a, b])$。

指數分佈 (Exponential distribution)　令 $\lambda > 0$ 為一額外參數，若 f_X 滿足：

$$f_X(x) = \lambda e^{-\lambda x}\mathbb{1}_{\mathbb{R}_{\geqslant 0}}(x), \qquad \forall x \in \mathbb{R},$$

則稱 X 為參數為 λ 的指數分佈，記作 $X \sim \mathcal{E}(\lambda)$ 或 $X \sim \mathrm{Exp}(\lambda)$。我們不難驗證，若 $X \sim \mathrm{Exp}(\lambda)$，則

$$\mathbb{E}[X] = \frac{1}{\lambda} \qquad \text{以及} \qquad \mathrm{Var}(X) = \frac{1}{\lambda^2}.$$

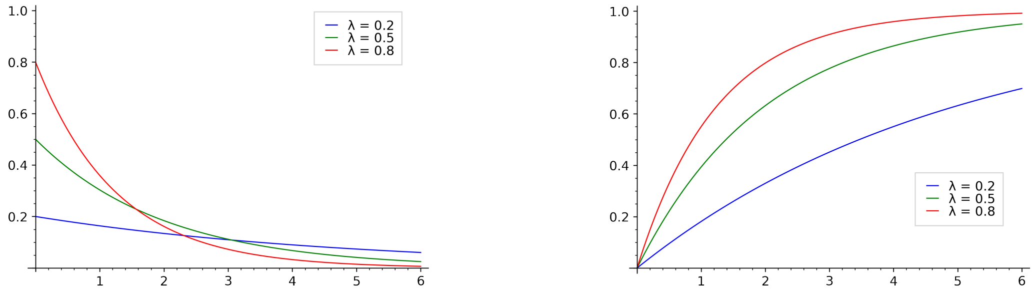

圖 2.5: 左邊是不同參數指數分佈的機率密度函數，右邊是他們的分佈函數。

常態分佈 (Normal distribution) 或高斯分佈 (Gaussian distribution)　令 $m \in \mathbb{R}$ 及 $\sigma > 0$ 為兩額外參數，若 f_X 滿足：

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{(x-m)^2}{2\sigma^2}\Big),$$

則稱 X 為均值為 m 變異數為 σ^2 的高斯分佈，記作 $X \sim \mathcal{N}(m,\sigma^2)$。我們不難驗證，若 $X \sim \mathcal{N}(m,\sigma^2)$，則

$$\mathbb{E}[X] = m \qquad \text{以及} \qquad \mathrm{Var}(X) = \sigma^2.$$

當 $m = 0$，我們稱 X 置中 (centered)；當 $\sigma = 1$，我們稱 X 約化 (reduced)；當兩者皆成立，我們稱之為標準常態分佈 (standard normal distribution)。

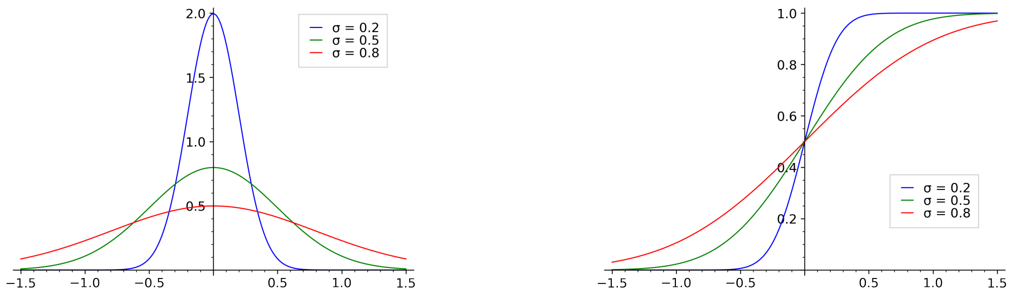

圖 2.6: 左邊是不同高斯分佈的機率密度函數，右邊是他們的分佈函數，這裡我們取 $\mu = 0$。

柯西分佈 (Cauchy distribution)　給定參數 $\gamma > 0$ 及 $x_0 \in \mathbb{R}$，若 f_X 滿足：

$$f_X(x) = \frac{1}{\pi\gamma(1+(\frac{x-x_0}{\gamma})^2)} = \frac{1}{\pi}\frac{\gamma}{(x-x_0)^2+\gamma^2}, \qquad \forall x \in \mathbb{R},$$

則稱 X 為參數為 (x_0,γ) 的柯西分佈，記作 $X \sim \mathrm{Cauchy}(x_0,\gamma)$。我們可以檢查，柯西分佈的期望值是沒有定義的。（參照習題 2.15）

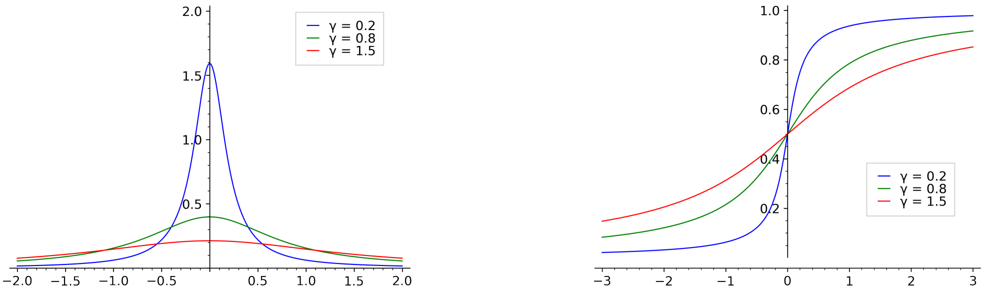

圖 2.7: 左邊是不同柯西分佈的機率密度函數，右邊是他們的分佈函數，這裡我們取 $x_0 = 0$。

珈瑪分佈 (Gamma distribution)　給定參數 $\alpha, \beta > 0$，若 f_X 滿足：

$$f_X(x) = \frac{1}{\Gamma(\alpha)}\beta^\alpha x^{\alpha-1}e^{-\beta x}\mathbb{1}_{x>0}, \qquad \forall x \in \mathbb{R},$$

其中 $\Gamma(\alpha)$ 為珈瑪函數，定義做

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}\,\mathrm{d}x,$$

則我們稱 X 為參數為 (α, β) 的珈瑪分佈，記作 $X \sim \Gamma(\alpha, \beta)$ 或 $X \sim \mathrm{Gamma}(\alpha, \beta)$。我們回顧珈瑪函數的重要性質：對於 $\alpha > 0$，我們有 $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$（變數變換）、$\Gamma(1) = 1$，以及 $\Gamma(\frac{1}{2}) = \sqrt{\pi}$。

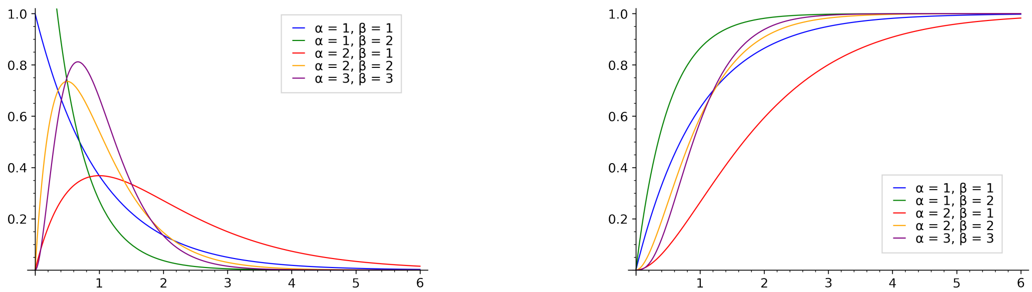

圖 2.8: 左邊是不同珈瑪分佈的機率密度函數，右邊是他們的分佈函數。

貝塔分佈 (Beta distribution)　給定參數 $\alpha, \beta > 0$，若 f_X 滿足：

$$f_X(x) = \frac{1}{B(\alpha, \beta)}x^{\alpha-1}(1-x)^{\beta-1}\mathbb{1}_{x\in(0,1)}, \qquad \forall x \in \mathbb{R},$$

其中 $B(\alpha, \beta)$ 為貝塔函數，定義做

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\,\mathrm{d}x = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)},$$

則我們稱 X 為參數為 (α, β) 的貝塔分佈，記作 $X \sim \mathrm{Beta}(\alpha, \beta)$。

註解 2.3.5 ： 機率分佈 $\mathrm{Beta}(1,1)$ 為在 $[0,1]$ 上的均勻分佈，且我們有下列對稱性：若 $X \sim \mathrm{Beta}(\alpha,\beta)$，則 $1-X \sim \mathrm{Beta}(\beta,\alpha)$。

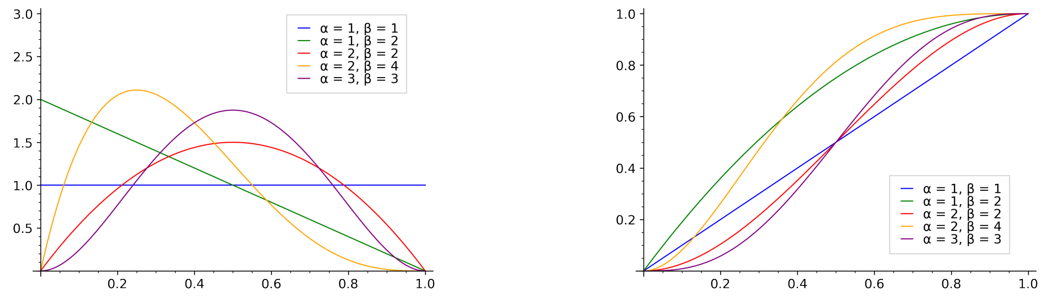

圖 2.9: 左邊是不同貝塔分佈的機率密度函數，右邊是他們的分佈函數。

第四節 隨機變數的動差

動差是隨機變數冪次的期望值，可以讓我們得到一些非線性的特性，例如在統計上有重要意義的變異數。在機率論中，動差方法 (method of moments) 有很多重要的應用，甚至可以拿來證明圖論、數論中各種存在性的問題，可以參見例如習題 2.24 和習題 2.25 。

第一小節 動差及變異數

定義 2.4.1 ： 固定正整數 $k \geqslant 1$，若實隨機變數 X 滿足 $\mathbb{E}[|X|^k] < \infty$，也就是說 $X \in L^k$，則我們說 X 的 k 階動差 (moment) 為有限。我們稱 $\mathbb{E}[X^k]$ 為 X 的 k 階動差。

定義 2.4.2 ： 令 $X \in L^2(\Omega, \mathcal{P}(\Omega), \mathbb{P})$，則 X 的變異數 (variance) 定義為

$$\mathrm{Var}(X) := \mathbb{E}\left[(X-\mathbb{E}[X])^2\right] \geqslant 0. \tag{2.4}$$

其標準差 (standard deviation) 則記作

$$\sigma_X = \sqrt{\mathrm{Var}(X)}.$$

註解 2.4.3 ： 我們可以這樣理解上述定義中的變異數 $\mathrm{Var}(X)$：它是一種可以量測隨機變數 X 在其期望值 $\mathbb{E}[X]$ 附近分散程度的數學量。

在式 (2.4) 中，取平方是重要的，不然我們會有

$$\mathbb{E}[X-\mathbb{E}[X]] = \mathbb{E}[X]-\mathbb{E}[X] = 0,$$

而無法正確描述我們的隨機變數 X 到底離他的期望值多遠。

若要避免相消為零的現象發生，我們可以考慮 $\mathbb{E}[|X-\mathbb{E}[X]|]$，他的好處是，我們不需要要求隨機變數 $X \in L^2$，我們只需要 $X \in L^1$ 的假設即足夠，但缺點是，絕對值的規律性比平方來得差，且在計算上比較不容易，因此才會考慮平方。在後面的章節中，我們也會看到取平方在機率理論中的重要性，這與中央極限定理有關係。

我們可以這樣理解上述定義中的變異數 $\mathrm{Var}(X)$：它量測隨機變數 X 在其平均值 $\mathbb{E}[X]$ 附近分散的程度，另外我們可以注意到，若且為若 $\mathrm{Var}(X)=0$ 則 X 幾乎必然為常數。

命題 2.4.4 ： 變異數 $\mathrm{Var}(X)$ 會滿足下列最佳化問題：

$$\mathrm{Var}(X)=\inf_{a\in\mathbb{R}} \mathbb{E}\left[(X-a)^2\right].$$

證明：我們只需要證明，對於所有實數 a，我們有

$$\mathbb{E}[(X-a)^2]=\mathrm{Var}(X)+(\mathbb{E}[X]-a)^2.$$

透過直接的計算，我們有

$$\begin{aligned}\mathbb{E}\left[(X-a)^2\right] &= \mathbb{E}\left[\left((X-\mathbb{E}[X])+(\mathbb{E}[X]-a)\right)^2\right] \\ &= \mathrm{Var}(X)+2(\mathbb{E}[X]-a)\,\mathbb{E}[X-\mathbb{E}[X]]+(\mathbb{E}[X]-a)^2 \\ &= \mathrm{Var}(X)+(\mathbb{E}[X]-a)^2.\end{aligned}$$

□

我們能利用期望值及變異數估計隨機變數取值的機率：

馬可夫不等式 (Markov's inequality)　　若 $X \in L^1_+(\Omega,\mathcal{A},\mathbb{P})$ 且 $a>0$，則

$$\mathbb{P}(X \geqslant a) \leqslant \frac{1}{a}\,\mathbb{E}[X].$$

柴比雪夫不等式 (Bienaymé-Chebyshev inequality)　　若 $X \in L^2(\Omega,\mathcal{A},\mathbb{P})$ 且 $a>0$，則

$$\mathbb{P}(|X-\mathbb{E}[X]| \geqslant a) \leqslant \frac{1}{a^2}\,\mathrm{Var}(X).$$

定義 2.4.5 ： 令 $X,Y \in L^2(\Omega,\mathcal{P}(\Omega),\mathbb{P})$，則它們的共變異數 (covariance) 定義為

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

我們注意到，共變異數的定義是變異數的推廣：$\mathrm{Var}(X)=\mathrm{Cov}(X,X)$。

註解 2.4.6 ： 兩隨機變數 X 及 Y 之間的共變異數測量它們之間的關聯性 (correlation)，這和隨機變數的獨立性是不一樣的。參閱第 3.1.2 小節。

註解 2.4.7 ： 給定離散機率空間 $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$，我們不難檢查共變異數函數

$$\mathrm{Cov}(\cdot,\cdot) : L^2(\Omega, \mathcal{A}, \mathbb{P}) \times L^2(\Omega, \mathcal{A}, \mathbb{P}) \longrightarrow \mathbb{R}$$

是個對稱、雙線性的運算子。換句話說，對於所有 $X, Y, Z \in L^2(\Omega, \mathcal{A}, \mathbb{P})$ 及 $a, b \in \mathbb{R}$，我們有

$$\begin{aligned} \mathrm{Cov}(X, Y) &= \mathrm{Cov}(Y, X), \\ \mathrm{Cov}(aX + bY, Z) &= a\,\mathrm{Cov}(X, Z) + b\,\mathrm{Cov}(Y, Z). \end{aligned}$$

這讓我們可以使用柯西不等式，進而得到

$$|\mathrm{Cov}(X, Y)| \leqslant \sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}.$$

此外，若 X 及 Y 其中之一幾乎必然為常數，則 $\mathrm{Cov}(X, Y) = 0$。

定義 2.4.8 ： 若 $X = (X_1, \ldots, X_d)$ 為一值域為 $\mathbb{R}^d$ 的隨機變數，且假設其所有的分量都在 $L^2(\Omega, \mathcal{A}, \mathbb{P})$ 之中，則 X 的共變異數矩陣 (covariance matrix) 定義為

$$K_X = \big(\mathrm{Cov}(X_i, X_j)\big)_{1 \leqslant i,j \leqslant d}.$$

問題 2.4.9： 試證明當 X 是個 d 維的實隨機變數時，其共變異數矩陣 K_X 是個半正定 (positive semi-definite) 對稱矩陣，換句話說，證明對於所有的 $\lambda = (\lambda_1, \ldots, \lambda_d) \in \mathbb{R}^d$，我們有 $\lambda K_X \lambda^T \geqslant 0$。

問題 2.4.10： 若 A 是個 $n \times d$ 維的矩陣，且 X 為 d 維實隨機變數，定義 $Y = AX$，試證明 $K_Y = AK_XA^T$。

第二小節 線性回歸

令 $X, Y_1, \ldots, Y_n$ 為在 $L^2(\Omega, \mathcal{A}, \mathbb{P})$ 中的隨機變數，我們想要找出近似 X 而且能被寫為 $1, Y_1, \ldots, Y_n$ 線性組合的隨機變數。換句話說，我們想要找出實數 $\beta_0, \ldots, \beta_n$ 將下列數值最小化：

$$\mathbb{E}[(X - (\beta_0 + \beta_1 Y_1 + \ldots \beta_n Y_n))^2].$$

命題 2.4.11 ： 當 (α_j) 為下列線性系統

$$\sum_{j=1}^{n} \alpha_j \,\mathrm{Cov}(Y_j, Y_k) = \mathrm{Cov}(X, Y_k), \qquad 1 \leqslant k \leqslant n, \tag{2.5}$$

的任意一個解時，設

$$Z = \mathbb{E}[X] + \sum_{j=1}^{n} \alpha_j (Y_j - \mathbb{E}[Y_j]), \tag{2.6}$$

則我們有

$$\inf_{\beta_0,\ldots,\beta_n \in \mathbb{R}} \mathbb{E}\left[(X - (\beta_0 + \beta_1 Y_1 + \ldots \beta_n Y_n))^2 \right] = \mathbb{E}[(X - Z)^2].$$

證明： 令 H 為由 $1, Y_1, \ldots, Y_n$ 生成在 $L^2(\Omega, \mathcal{A}, \mathbb{P})$ 中的子向量空間，由於 Z 會讓泛函函數 $U \in H \mapsto \|X - U\|_2$ 達到最小值，所以 Z 是 X 在 H 之上的正交投影。若我們把 Z 寫作

$$Z = \alpha_0 + \sum_{j=1}^{n} \alpha_j (Y_j - \mathbb{E}[Y_j]), \tag{2.7}$$

透過正交投影的性質，我們有 $\mathbb{E}[(X - Z) \cdot 1] = 0$，也就是說 $\alpha_0 = \mathbb{E}[X]$。同樣的，對於所有 $1 \leqslant k \leqslant n$，我們有

$$\mathbb{E}\left[(X - Z) \cdot (Y_k - \mathbb{E}[Y_k]) \right] = 0,$$

也就是說 $\mathrm{Cov}(X, Y_k) = \mathrm{Cov}(Z, Y_k)$；接著利用式 (2.7) 中 Z 的線性組合，我們可以得到式 (2.5)。

反過來說，若 (α_i) 滿足式 (2.5)，則式 (2.6) 中定義的 Z 是個在 H 中的元素，而且 $X - Z$ 和 H 是正交的，也就是說 Z 是 X 在 H 上的正交投影。 □

第三小節 特徵函數

定義 2.4.12 ： 若 X 是個值域為 $\mathbb{R}^d$ 的隨機變數，則其特徵函數 (characteristic function) $\Phi_X : \mathbb{R}^d \longrightarrow \mathbb{C}$ 定義為

$$\Phi_X(\xi) = \mathbb{E}\left[\exp(\mathrm{i}\, \xi \cdot X) \right], \qquad \xi \in \mathbb{R}^d.$$

註解 2.4.13 ： 上述定義也可以重新寫作

$$\Phi_X(\xi) = \int_{\mathbb{R}^d} e^{\mathrm{i}\, \xi \cdot x} \mathbb{P}_X(\mathrm{d}x).$$

換句話說，我們可以把 Φ_X 視作 $\mathbb{P}_X$ 的傅立葉變換 (Fourier transform) ：$\Phi_X(\xi) = \hat{\mathbb{P}}_X(\xi)$。另外，從勒貝格控制收斂定理可以得知 Φ_X 為一在 $\mathbb{R}^d$ 上的連續有界函數。

我們這裡列出以第 2.3 節中提到的機率分佈為例子，所對應到的特徵函數。

隨機變數 X	特徵函數 $\Phi_X(\xi)$
$\text{Ber}(p)$	$1-p+pe^{\mathrm{i}\,\xi}$
$\text{Bin}(n,p)$	$(1-p+pe^{\mathrm{i}\,\xi})^n$
$\text{Geo}(p)$	$\dfrac{p}{1-(1-p)e^t}$
$\text{Pois}(\lambda)$	$\exp(\lambda(e^{\mathrm{i}\,\xi}-1))$
$\text{Unif}([a,b])$	$\dfrac{e^{\mathrm{i}\,\xi b}-e^{\mathrm{i}\,\xi a}}{\mathrm{i}\,\xi(b-a)}$
$\mathcal{E}(\lambda)$	$\dfrac{\lambda}{\lambda-\mathrm{i}\,\xi}$
$\mathcal{N}(m,\sigma^2)$	$\exp(\mathrm{i}\,m\xi-\frac{1}{2}\sigma^2\xi^2)$

在這小節，我們將會證明特徵函數可以完全決定隨機變數的機率分佈。

首先，我們證明常態分佈在傅立葉變換 $\mathcal{F}:\mathbb{P}_X\mapsto\hat{\mathbb{P}}_X$ 之下的不變性 (invariance)。

引理 2.4.14 ： 若 X 為一分佈為 $\mathcal{N}(0,\sigma^2)$ 的隨機變數，則

$$\Phi_X(\xi)=\exp\Big(-\frac{\sigma^2\xi^2}{2}\Big),\qquad \xi\in\mathbb{R}.$$

證明：透過被積分式的奇偶性，我們可以得到 $\Phi_X(\xi)$ 的虛數部份為零，所以我們要計算

$$f(\xi)=\int_{\mathbb{R}}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}\cos(\xi x)\,\mathrm{d}x.$$

我們對上述式子微分，由於可積函數 $x\mapsto|x|e^{-x^2/2}$ 是 $x\mapsto x\sin(\xi x)e^{-x^2/2}$ 的上界，所以微分算子可以移至積分之內，也就是說

$$f'(\xi)=-\int_{\mathbb{R}}\frac{1}{\sqrt{2\pi}}xe^{-x^2/2}\sin(\xi x)\,\mathrm{d}x.$$

我們對上述式子進行部份積分，可以得到

$$f'(\xi)=-\xi\int_{\mathbb{R}}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}\cos(\xi x)\,\mathrm{d}x=-\xi f(\xi).$$

因此我們證明出來 f 是微分方程 $f'(\xi)=-\xi f(\xi)$ 且初始值為 $f(0)=1$ 的解，由於唯一性，我們得到 $f(\xi)=\exp(-\xi^2/2)$。 □

定理 2.4.15 ： 任一隨機變數 X 的特徵函數可以完全定義其分佈，換句話說，傅立葉變換 $\mathcal{F}:\mathbb{P}_X\mapsto\hat{\mathbb{P}}_X$ 在 $\mathbb{R}^d$ 之上的分佈所構成的空間是單射 (injective) 的。

證明：首先，假設一維的情況 $d=1$。對於所有的 $\sigma>0$，我們可以定義常態分佈 $\mathcal{N}(0,\sigma^2)$ 的密度函數

$$g_\sigma(x)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\Big(-\frac{x^2}{2\sigma^2}\Big).$$

若 μ 為一在 $\mathbb{R}$ 之上的機率測度，我們定義：

$$\begin{aligned}f_\sigma(x)&=\int_{\mathbb{R}}g_\sigma(x-y)\mu(\mathrm{d}y)\overset{\text{(def)}}{=}g_\sigma*\mu(x),\\ \mu_\sigma(\mathrm{d}x)&=f_\sigma(x)\,\mathrm{d}x.\end{aligned}$$

要證明此定理，我們將分兩步驟：

(1) μ_σ 可以完全被 $\widehat{\mu}$ 所定義；

(2) 對於任何的函數 $\varphi\in\mathcal{C}_b(\mathbb{R})$，當 $\sigma\to0$ 我們有下列收斂 $\int\varphi(x)\mu_\sigma(\mathrm{d}x)\longrightarrow\int\varphi(x)\mu(\mathrm{d}x)$。

首先，我們來證明第一點。透過引理 2.4.14 ，我們可以得到，對於所有的 $x\in\mathbb{R}$，有

$$\sqrt{2\pi\sigma^2}g_\sigma(x)=\exp\Big(-\frac{x^2}{2\sigma^2}\Big)=\int_{\mathbb{R}}e^{\mathrm{i}\,\xi\cdot x}g_{1/\sigma}(\xi)\,\mathrm{d}\xi.$$

接著，我們可以把 $f_\sigma(x)$ 重新寫作

$$\begin{aligned}f_\sigma(x)=\int_{\mathbb{R}}g_\sigma(x-y)\mu(\mathrm{d}y)&=\frac{1}{\sqrt{2\pi\sigma^2}}\int_{\mathbb{R}}\Big(\int_{\mathbb{R}}e^{\mathrm{i}\,\xi(x-y)}g_{1/\sigma}(\xi)\,\mathrm{d}\xi\Big)\mu(\mathrm{d}y)\\&=\frac{1}{\sqrt{2\pi\sigma^2}}\int_{\mathbb{R}}e^{\mathrm{i}\,\xi x}g_{1/\sigma}(\xi)\Big(\int_{\mathbb{R}}e^{-\mathrm{i}\,\xi y}\mu(\mathrm{d}y)\Big)\,\mathrm{d}\xi\\&=\frac{1}{\sqrt{2\pi\sigma^2}}\int_{\mathbb{R}}e^{\mathrm{i}\,\xi x}g_{1/\sigma}(\xi)\widehat{\mu}(-\xi)\,\mathrm{d}\xi.\end{aligned}\tag{2.8}$$

在上式的倒數第二個等式當中，我們用富比尼定理（定理 1.4.3 ），因為 μ 是個機率測度且 $g_{1/\sigma}$ 對勒貝格測度來說，是個可積函數。

接著證明第二點，首先注意到，對於所有的 $\varphi\in C_b(\mathbb{R})$，我們有

$$\int\varphi(x)\mu_\sigma(\mathrm{d}x)=\int\varphi(x)\Big(\int g_\sigma(y-x)\mu(\mathrm{d}y)\Big)\,\mathrm{d}x=\int g_\sigma*\varphi(y)\mu(\mathrm{d}y),\tag{2.9}$$

其中我們用到富比尼定理以及 g_σ 是個偶函數的特性。接著，我們使用 g_σ 的其他特性：

$$\int g_\sigma(x)\,\mathrm{d}x=1,\quad \text{以及}\quad \lim_{\sigma\to0}\int_{\{|x|>\varepsilon\}}g_\sigma(x)\,\mathrm{d}x=0,\quad\forall\varepsilon>0,$$

這讓我們簡單得到

$$\forall y\in\mathbb{R},\qquad \lim_{\sigma\to0}g_\sigma*\varphi(y)=\varphi(y).\tag{2.10}$$

由於對於所有的 $\sigma > 0$，$|g_\sigma * \varphi| \leqslant \sup|\varphi|$，透過勒貝格收斂定理，我們有

$$\lim_{\sigma\to 0}\int \varphi(x)\mu_\sigma(\mathrm{d}x) = \lim_{\sigma\to 0}\int g_\sigma * \varphi(y)\mu(\mathrm{d}y) = \int \varphi(x)\mu(\mathrm{d}x).$$

因此此定理在一維的情況下為真。

在 d 維的情況下，證明相似。我們定義下列函數

$$g_\sigma^{(d)}(x_1,\ldots,x_d) = \prod_{j=1}^d g_\sigma(x_j),$$

並且利用對於所有 $\xi \in \mathbb{R}^d$，我們有

$$\int_{\mathbb{R}^d} g_\sigma^{(d)}(x)e^{\mathrm{i}\,\xi\cdot x}\,\mathrm{d}x = \prod_{j=1}^d \int_{\mathbb{R}} g_\sigma(x_j)e^{\mathrm{i}\,\xi_j x_j}\,\mathrm{d}x_j = (2\pi\sigma^2)^{d/2} g_{1/\sigma}^{(d)}(\xi).$$

□

問題 2.4.16： 將 $C_b(\mathbb{R})$ 中緊緻支撐 (compactly supported) 函數構成的集合記作 $C_c(\mathbb{R})$。證明若 $\varphi \in C_c(\mathbb{R})$，則可以強化式 (2.10) 中的收斂為在 $\mathbb{R}$ 上的均勻收斂；換句話說，證明

$$\forall \varphi \in C_c(\mathbb{R}), \qquad \|g_\sigma * \varphi - \varphi\|_\infty \xrightarrow[\sigma\to 0]{} 0.$$

命題 2.4.17 ： 令 $X = (X_1,\ldots,X_n)$ 為一 n 維的實隨機變數，假設 $\|X\|_2^2$ 是個可積的隨機變數，則 Φ_X 是個 $\mathcal{C}^2$ 函數，而且當 $\xi = (\xi_1,\ldots,\xi_n)$ 趨近於 0 時，我們有

$$\Phi_X(\xi) = 1 + \mathrm{i}\sum_{j=1}^d \xi_j\,\mathbb{E}[X_j] - \frac{1}{2}\sum_{j,k=1}^d \xi_j\xi_k\,\mathbb{E}[X_jX_k] + o(\|\xi\|^2).$$

證明： 下面三個性質並不難檢查：

(a) 對於所有 $\xi \in \mathbb{R}^n$，函數 $x \mapsto e^{\mathrm{i}\,\xi\cdot x}$ 對於 $\mathbb{P}_X$ 是可積的。

(b) 對於所有 $x \in \mathbb{R}^n$，函數 $\xi \mapsto e^{\mathrm{i}\,\xi\cdot x}$ 對於 ξ 是可微的，而且他在 ξ 的微分寫做：

$$\eta = (\eta_1,\ldots,\eta_n) \mapsto \sum_{i=j}^n \mathrm{i}\,\eta_j x_j e^{\mathrm{i}\,\xi\cdot x} \quad \text{或} \quad \sum_{j=1}^n \mathrm{i}\,x_j e^{\mathrm{i}\,\xi\cdot x}\,\mathrm{d}\xi_j.$$

(c) 對於所有 $j = 1,\ldots,n$，偏微分 $x \mapsto \mathrm{i}\,x_j e^{\mathrm{i}\,\xi\cdot x}$ 是連續的，而且可以被 x_j 所控制住，由於 $X_j \in L^2 \subseteq L^1$，這個控制函數對 $\mathbb{P}_X$ 是可積的。

因此，定理 1.2.24 讓我們可以交換微分算子和積分算子的順序，讓我們改寫 Φ_X 的微分，也就是說：

$$\forall j = 1,\ldots,n, \qquad \frac{\partial \Phi_X}{\partial \xi_j}(\xi) = \mathrm{i}\,\mathbb{E}[X_j e^{\mathrm{i}\,\xi\cdot X}].$$

此外，定理 1.2.22 告訴我們這些偏微分都是連續的。

對於二階微分，我們可以使用相似的方法。更確切來說，我們會需要用到 $\mathbb{E}[|X_j X_k|] \leqslant \mathbb{E}[X_j]^2\,\mathbb{E}[X_k]^2 < \infty$ 來解釋二階微分和積分可以交換。 □

要小心的是，不同於特徵函數可以完全描述機率分佈，在一般情況下，只知道動差 $(\mathbb{E}[X^k])_{k\geqslant 0}$ 並不足以描述 X 的機率分佈。在習題 2.22 中，我們會看到反例，也就是說，我們會構造出（至少）兩個不同的機率分佈，使得他們的所有動差皆是相等的。接著在習題 2.23 中，我們會討論在什麼樣的情況之下，動差可以唯一決定機率分佈。

第四小節 生成函數

當離散隨機變數的值域為非負整數 $\mathbb{N}_0$ 時，我們可以定義他的生成函數 (generating function) 。

定義 2.4.18 ： 令 X 為一值域為 $\mathbb{N}_0$ 的隨機變數，其生成函數 G_X 定義在下列級數會收斂的 $s \in \mathbb{C}$ 上：

$$G_X(s) = \mathbb{E}[s^X] = \sum_{n=0}^{\infty} \mathbb{P}(X = n)s^n. \tag{2.11}$$

反過來說，若我們知道離散隨機變數 X 的生成函數，透過讀取他的係數，我們也可以描述 X 的分佈。

由式 (2.11) 定義出來的級數會滿足一些基本性質，這部份屬於微積分的先備知識，這裡只做回顧。

收斂性　存在收斂半徑 (radius of convergence) $0 \leqslant R \leqslant \infty$ 使得當 $|s| < R$ 時，級數 $G_X(s)$ 會收斂；當 $|s| > R$ 時，級數 $G_X(s)$ 會發散。此外，對於所有 $R' < R$，級數 $G_X(s)$ 在 $\{s : |s| \leqslant R'\}$ 上會均勻收斂。這裡我們注意到，由於 $G_X(1) = 1$，所以顯然有 $R \geqslant 1$。

微分性　在定義域 $\{s : |s| < R\}$ 上，我們可以一項一項對級數 $G_X(s)$ 微分或積分無限多次。

唯一性　若存在 $0 < R' \leqslant R$ 使得當 $|s| < R'$ 時，等式 $G_X(s) = G_Y(s)$ 成立，則對於所有 $n \in \mathbb{N}_0$，我們有 $\mathbb{P}(X = n) = \mathbb{P}(Y = n)$。此外，我們有

$$\mathbb{P}(X = n) = \frac{1}{n!} G_X^{(n)}(0), \qquad \forall n \geqslant 0.$$

連續性　【Abel 定理】由於各項 $\mathbb{P}(X = n)$ 皆為非負，若 $R < \infty$，我們有

$$\lim_{s\uparrow R} G_X(s) = \sum_{n=0}^{\infty} \mathbb{P}(X = n)R^n.$$

命題 2.4.19 ： 我們可以透過 G_X 的微分來計算 X 的期望值：

$$G_X'(1) := \lim_{s\uparrow 1} G_X(s) = \mathbb{E}[X] \in [0,\infty].$$

更一般來說，對於所有正整數 $p \geqslant 1$，我們有

$$G_X^{(p)} := \lim_{s\uparrow 1} G_X^{(p)}(s) = \mathbb{E}[X(X-1)\dots(X-p+1)],$$

也就是說透過 X 的生成函數，我們可以簡單計算 X 的所有動差。此外，對於 $X \in L^2(\Omega,\mathcal{P}(\Omega),\mathbb{P})$，我們還有

$$\mathrm{Var}(X) = G_X''(1) + G_X'(1) - G_X'(1)^2.$$

證明： 對於所有正整數 $p \geqslant 1$，根據 $G_X^{(p)}(s)$ 在 $\{s : |s| < R\}$ 上的連續性（已知 $R \geqslant 1$），我們有

$$\lim_{s\uparrow 1} G_X^{(p)}(s) = \sum_{n=0}^{\infty} \mathbb{P}(X=n)\cdot n(n-1)\dots(n-p+1) = \sum_{n=0}^{\infty} \mathbb{P}(X=n)\cdot \varphi(n),$$

其中 $\varphi(n) = n(n-1)\dots(n-p+1)$，得證。

令 $X \in L^2(\Omega,\mathcal{P}(\Omega),\mathbb{P})$，我們根據前面性質，可以得到

$$\mathbb{E}[X^2] = \mathbb{E}[X(X-1)] + \mathbb{E}[X] = G_X''(1) + G_X'(1),$$

因此 $\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = G_X''(1) + G_X'(1) - G_X'(1)^2$，得證。 □

範例 2.4.20 ： 若 X 是滿足二項式分佈 $\mathrm{Bin}(n,p)$ 的隨機變數，則他的生成函數寫作：

$$G_X(s) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} s^k = \left((1-p)+ps\right)^n.$$

我們可以計算他的期望值

$$\mathbb{E}[X] = G_X'(s)_{s=1} = \left[n((1-p)+ps)^{n-1}p\right]_{s=1} = np.$$

若要計算他的變異數，我們先計算

$$G_X''(1) = \left[n(n-1)((1-p)+ps)^{n-2}p^2\right]_{s=1} = n(n-1)p^2,$$

接著帶入

$$\mathrm{Var}(X) = n(n-1)p^2 + np - (np)^2 = np - np^2 = np(1-p).$$

當我們的目的是要計算 X 的動差時，我們可以直接定義他的動差生成函數 (moment generating function) ，這可以簡化我們的計算。

定義 2.4.21 ： 給定取值在 $\mathbb{N}_0$ 中的離散隨機變數 X，我們定義他的動差生成函數 (moment generating function) ，也稱作指數生成函數 (exponential generating function) 為

$$M_X(t) := G_X(e^t) = \mathbb{E}[e^{tX}],$$

其中我們將 G_X 的收斂半徑記作 R，並要求 $e^t < R$。

註解 2.4.22 ： 若我們以 $M_X(t) := \mathbb{E}[e^{tX}]$ 來定義動差生成函數，則我們只需要要求 X 為實數離散隨機變數即可。

命題 2.4.23 ： 離散隨機變數 X 的動差可以透過 M_X 的微分來計算。更確切的說，對於所有非負整數 $k \geqslant 0$，我們有

$$\mathbb{E}[X^k] = M_X^{(k)}(0).$$

證明：我們有

$$\begin{aligned} M_X(t) &= \sum_{k=0}^{\infty} e^{tk}\,\mathbb{P}(X=k) = \sum_{k=0}^{\infty}\sum_{n=0}^{\infty} \frac{(tk)^n}{n!}\,\mathbb{P}(X=k) \\ &= \sum_{n=0}^{\infty} \frac{t^n}{n!}\Big(\sum_{k=0}^{\infty} k^n\,\mathbb{P}(X=k)\Big) = \sum_{n=0}^{\infty} \frac{t^n}{n!}\,\mathbb{E}[X^n]. \end{aligned}$$

□

我們也能夠對 X 做拉普拉斯轉換 (Laplace transform)：

$$L_X(\lambda) := \mathbb{E}[e^{-\lambda X}].$$

要注意的是，與指數生成函數相同，這並不是對於所有 $\lambda \in \mathbb{R}$ 都是有定義的。

第五小節 尾端機率

給定隨機變數 X，$\mathbb{P}(|X| > \lambda)$ 或 $\mathbb{P}(|X| > \lambda)$ 稱作尾端機率 (tail probability) 。一般來講，若我們能夠有效控制 X 越高階的動差，則代表尾端機率會越小，且反之亦然。

定義 2.4.24 ： 根據尾端機率，我們可以將隨機變數分佈歸類為：

(1) 次高斯分佈 (sub-Gaussian distribution)：存在 $C, c > 0$ 使得

$$\mathbb{P}(|X| > x) \leqslant C\exp(-cx^2), \qquad x \to \infty.$$

(2) 重尾分佈 (heavy-tailed distribution)：存在 $C, \alpha > 0$ 使得

$$\mathbb{P}(|X| > x) \sim Cx^{-\alpha}, \qquad x \to \infty.$$

註解 2.4.25： 我們可以注意到，柯西分佈會是個重尾分佈，因為如果 $X \sim \mathrm{Cauchy}(0,1)$，則

$$\mathbb{P}(|X| > x) \sim \frac{2}{\pi x}, \qquad x \to \infty.$$

此外，高斯分佈也是個次高斯分佈。

問題 2.4.26：給定隨機變數 X，則下列三敘述等價：

(1) X 是個次高斯分佈；

(2) 存在 $C, c > 0$ 使得對於所有 $t \in \mathbb{R}$，我們有 $\mathbb{E}[e^{tX}] \leqslant C\exp(ct^2)$；

(3) 存在 $C > 0$ 使得對於所有 $k \geqslant 1$，我們有 $\mathbb{E}[|X|^k] \leqslant (Ck)^{k/2}$。

問題 2.4.27：令 X 為實隨機變數且 $k > 0$。若 X 在 L^k 中，證明當 $\lambda \longrightarrow \infty$ 時，我們有

$$\lambda^k\, \mathbb{P}(|X| > \lambda) \longrightarrow 0.$$