香港中文大學
數學系
統計學練習

1 期望、方差、標準差

1.1 離散隨機變量的定義

隨機變量的期望值（期望）是概率加權的概率實驗所有可能結果的平均值。方差量化隨機變量值圍繞其期望的分散程度。標準差是方差的正平方根，具有與隨機變量相同的單位。

設X 為離散隨機變量，其支撐為$\{x_1, x_2, \ldots, x_n\}$，概率質量函數（PMF）為$P(X = x_i) = p_i$，其中$\sum_{i=1}^{n} p_i = 1,\ p_i \geq 0$。

(1) 期望：

$$\mathbb{E}[X] = \mu_X = \sum_{i=1}^{n} x_i\, p_i$$

(2) 方差（兩個等價公式）：

$$\mathrm{Var}(X) = \sigma_X^2 = \mathbb{E}\big[(X-\mu_X)^2\big] = \sum_{i=1}^{n} (x_i - \mu_X)^2 p_i$$

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - \big(\mathbb{E}[X]\big)^2, \quad \mathbb{E}[X^2] = \sum_{i=1}^{n} x_i^2 p_i$$

(3) 標準差：

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

1.2 示例1：離散隨機變量X

給定離散X 的PMF 表：

x	0	1	2	3	總和
$P(X = x)$	0.010	0.840	0.145	0.005	1

1.2.1 數值計算

1: 計算$\mathbb{E}[X]$

$$\begin{aligned}\mathbb{E}[X] &= (0)(0.010) + (1)(0.840) + (2)(0.145) + (3)(0.005)\\ &= 0 + 0.840 + 0.290 + 0.015\\ &= 1.145\end{aligned}$$

2: 計算$\mathbb{E}[X^2]$（用於方差簡捷公式）

$$\begin{aligned}\mathbb{E}[X^2] &= (0^2)(0.010) + (1^2)(0.840) + (2^2)(0.145) + (3^2)(0.005)\\ &= 0 + (1)(0.840) + (4)(0.145) + (9)(0.005)\\ &= 0 + 0.840 + 0.580 + 0.045\\ &= 1.465\end{aligned}$$

3: 計算方差$\mathrm{Var}(X)$

$$\begin{aligned}\mathrm{Var}(X) &= \mathbb{E}[X^2] - \big(\mathbb{E}[X]\big)^2 \\ &= 1.465 - (1.145)^2 \\ &= 1.465 - 1.311025 \\ &= 0.153975\end{aligned}$$

透過直接平方偏差和驗證：

$$\begin{aligned}\sum (x_i - \mu_X)^2 p_i &= (0-1.145)^2(0.010) + (1-1.145)^2(0.840) \\ &\quad + (2-1.145)^2(0.145) + (3-1.145)^2(0.005) \\ &= (1.311025)(0.010) + (0.021025)(0.840) \\ &\quad + (0.731025)(0.145) + (3.441025)(0.005) \\ &= 0.01311025 + 0.017661 + 0.105998625 + 0.017205125 \\ &= 0.153975 \quad \text{(與簡捷結果匹配)}\end{aligned}$$

4: 計算標準差σ_X

$$\sigma_X = \sqrt{0.153975} \approx 0.39239648$$

1.3 連續隨機變量定義

設X 為連續隨機變量，具有概率密度函數（PDF）$f_X(t)$，支撐於區間$[a,b]$，滿足$\int_a^b f_X(t)dt = 1$，$f_X(t) \geq 0$。

(1) 期望：

$$\mathbb{E}[X] = \mu_X = \int_{-\infty}^{\infty} t\, f_X(t)dt$$

(2) 方差：

$$\mathrm{Var}(X) = \mathbb{E}\big[(X-\mu_X)^2\big] = \int_{-\infty}^{\infty}(t-\mu_X)^2 f_X(t)dt = \mathbb{E}[X^2]-(\mathbb{E}[X])^2, \quad \mathbb{E}[X^2] = \int_{-\infty}^{\infty} t^2 f_X(t)dt$$

(3) 標準差：

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

1.4 期望與方差性質

設$a, b \in \mathbb{R}$ 為常數，X, Y 為隨機變量：

1. $\mathbb{E}[aX+b] = a\mathbb{E}[X] + b$；$\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$
2. $\mathrm{Var}(aX+b) = a^2\,\mathrm{Var}(X)$；$\mathrm{Var}(b) = 0$（常數具有零方差）
3. 標準化隨機變量：定義$Z = \dfrac{X-\mu_X}{\sigma_X}$。由性質1 和2：

$$\mathbb{E}[Z] = \mathbb{E}\left[\frac{X-\mu_X}{\sigma_X}\right] = \frac{1}{\sigma_X}\big(\mathbb{E}[X] - \mu_X\big) = 0$$

$$\mathrm{Var}(Z) = \mathrm{Var}\left(\frac{X-\mu_X}{\sigma_X}\right) = \frac{1}{\sigma_X^2}\,\mathrm{Var}(X) = \frac{\sigma_X^2}{\sigma_X^2} = 1$$

Z 始終具有均值0，方差1。

1.5 示例2：連續隨機變量X，PDF $f(t) = 3t^2,\ 0 \le t \le 1$

定義$Y = 4X + 2$。計算$\mathbb{E}[X], \mathrm{Var}(X), \mathbb{E}[Y], \mathrm{Var}(Y)$。

1.5.1 解析積分

1: 計算$\mathbb{E}[X]$

$$\mathbb{E}[X] = \int_0^1 t \cdot f(t)dt = \int_0^1 t \cdot 3t^2 dt = 3\int_0^1 t^3 dt = 3\left.\frac{t^4}{4}\right|_0^1 = \frac{3}{4} = 0.75$$

2: 計算$\mathbb{E}[X^2]$

$$\mathbb{E}[X^2] = \int_0^1 t^2 \cdot 3t^2 dt = 3\int_0^1 t^4 dt = 3\left.\frac{t^5}{5}\right|_0^1 = \frac{3}{5} = 0.6$$

3: X 的方差

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{3}{5} - \left(\frac{3}{4}\right)^2 = \frac{3}{5} - \frac{9}{16} = \frac{48-45}{80} = \frac{3}{80} = 0.0375$$

4: $Y = 4X + 2$ 的期望（線性性質）

$$\mathbb{E}[Y] = \mathbb{E}[4X+2] = 4\mathbb{E}[X] + 2 = 4 \cdot \frac{3}{4} + 2 = 3 + 2 = 5$$

5: $Y = 4X + 2$ 的方差（方差縮放性質）

$$\mathrm{Var}(Y) = 4^2\,\mathrm{Var}(X) = 16 \cdot \frac{3}{80} = \frac{3}{5} = 0.6$$

積分驗證$\mathbb{E}[Y^2]$（交叉檢查）：

$$\begin{aligned}
\mathbb{E}[Y^2] &= \int_0^1 (4t+2)^2 \cdot 3t^2 dt = 3\int_0^1 (16t^2 + 16t + 4)t^2 dt \\
&= 3\int_0^1 (16t^4 + 16t^3 + 4t^2)dt = 3\left.\left(\frac{16t^5}{5} + \frac{16t^4}{4} + \frac{4t^3}{3}\right)\right|_0^1 \\
&= 3\left(\frac{16}{5} + 4 + \frac{4}{3}\right) = 3\left(\frac{48+60+20}{15}\right) = 3 \cdot \frac{128}{15} = \frac{128}{5} = 25.6
\end{aligned}$$

$$\mathrm{Var}(Y) = \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2 = 25.6 - 25 = 0.6 \quad \text{(與性質結果一致)}$$

2 聯合概率分布

對於兩個離散隨機變量X, Y，其聯合概率質量函數為$p(x,y) = P(X = x, Y = y)$，滿足：

$$\sum_x \sum_y p(x,y) = 1, \quad p(x,y) \ge 0$$

邊際概率質量函數為：

$$p_X(x) = \sum_y p(x,y), \quad p_Y(y) = \sum_x p(x,y)$$

對於連續隨機變量，其聯合概率密度函數$f(x, y)$ 滿足：

$$\iint_{\mathbb{R}^2} f(x, y)\, dx\, dy = 1, \quad f(x, y) \geq 0$$

邊際概率密度函數為：

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy, \quad f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\, dx$$

聯合期望值：$\mathbb{E}[g(X, Y)] = \sum_x \sum_y g(x, y)p(x, y)$（離散情況），$\iint g(x, y)f(x, y)\, dx\, dy$（連續情況）。

3 協方差與相關係數

3.1 協方差的定義

$$\text{Cov}(X, Y) = \mathbb{E}\big[(X - \mu_X)(Y - \mu_Y)\big] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

協方差的性質：

1. $\text{Cov}(X, X) = \text{Var}(X)$
2. $\text{Cov}(aX + b, cY + d) = ac\, \text{Cov}(X, Y)$
3. $\text{Cov}(X_1 + X_2, Y) = \text{Cov}(X_1, Y) + \text{Cov}(X_2, Y)$

3.2 Pearson 相關係數

$$\rho_{XY} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}, \quad -1 \leq \rho_{XY} \leq 1$$

- $\rho = 1$：完全正線性關係
- $\rho = -1$：完全負線性關係
- $\rho = 0$：不相關（無線性關聯；獨立蘊含不相關，反之不一定成立）

4 協方差矩陣

對於隨機向量$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_k \end{bmatrix}$，其均值向量為$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_k \end{bmatrix}$，其中$\mu_i = \mathbb{E}[X_i]$。

協方差矩陣$\mathbf{\Sigma} \in \mathbb{R}^{k \times k}$ 定義為：

$$\mathbf{\Sigma} = \begin{bmatrix} \text{Var}(X_1) & \text{Cov}(X_1, X_2) & \dots & \text{Cov}(X_1, X_k) \\ \text{Cov}(X_2, X_1) & \text{Var}(X_2) & \dots & \text{Cov}(X_2, X_k) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(X_k, X_1) & \text{Cov}(X_k, X_2) & \dots & \text{Var}(X_k) \end{bmatrix}$$

性質：$\mathbf{\Sigma}$ 是對稱矩陣（$\mathbf{\Sigma} = \mathbf{\Sigma}^T$），且為半正定矩陣。

4.1 練習1

從統計教科書中選取一個連續型概率密度函數$f(x)$（例如均勻分布、指數分布、正態分布、Beta 分布），並完成步驟：

1. 驗證$\int f(x)\,dx = 1$（確認為有效PDF）
2. 計算$\mathbb{E}[X] = \int x f(x)\,dx$
3. 計算$\mathbb{E}[X^2] = \int x^2 f(x)\,dx$
4. 計算方差$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$
5. 計算標準差$\sigma_X = \sqrt{\mathrm{Var}(X)}$

5 聯合概率分布

當同時分析兩個或多個隨機變量時，我們研究它們的聯合概率分布，其描述多個事件同時發生的概率。給定聯合分布和一個邊際分布，我們可以求解另一個單一隨機變量的剩餘邊際分布。

5.1 離散隨機變量：聯合概率質量函數（PMF）

聯合PMF 的定義

設X, Y 為離散隨機變量，其支撐值分別為$x_1, x_2, \ldots, x_s$ 和$y_1, y_2, \ldots, y_t$。**聯合概率質量函數**定義為：

$$p_{ij} = p(x_i, y_j) = \mathbb{P}(X = x_i,\ Y = y_j), \quad x_i, y_j \in \mathbb{R},\ i = 1, \ldots, s,\ j = 1, \ldots, t$$

5.2 聯合概率表佈局

X	Y	y_1 p_{i1}	y_2 p_{i2}	$\ldots$ $\ldots$	行和（X 的邊際） $\mathbb{P}(X = x_i)$
x_1		p_{11}	p_{12}	$\ldots$	$\mathbb{P}(X = x_1)$
x_2		p_{21}	p_{22}	$\ldots$	$\mathbb{P}(X = x_2)$
$\vdots$		$\vdots$	$\vdots$	$\ddots$	$\vdots$
x_s		p_{s1}	p_{s2}	$\ldots$	$\mathbb{P}(X = x_s)$
列和（Y 的邊際）		$\mathbb{P}(Y = y_1)$	$\mathbb{P}(Y = y_2)$	$\ldots$	**1**

5.3 邊際PMF 公式

從聯合PMF，透過對另一個變量求和，獲得單一變量的邊際分布：

$$\mathbb{P}(X = x_i) = \sum_{j=1}^{t} \mathbb{P}(X = x_i, Y = y_j) = \sum_{j=1}^{t} p_{ij}$$

$$\mathbb{P}(Y = y_j) = \sum_{i=1}^{s} \mathbb{P}(X = x_i, Y = y_j) = \sum_{i=1}^{s} p_{ij}$$

邊際分布是從聯合表中提取的單一變量單獨的單變量概率分布。

5.4 離散聯合PMF 的性質

1. 非負性：$p(x_i, y_j) \geq 0$ 對所有有效x_i, y_j。
2. 總概率和為1：$\sum_{i=1}^{s} \sum_{j=1}^{t} p(x_i, y_j) = 1$。
3. 矩形概率：$\mathbb{P}(a < X < b,\ c < Y < d) = \displaystyle\sum_{\substack{a<x_i<b \\ c<y_j<d}} p(x_i, y_j)$。

5.5 示例3：彩球超幾何抽取

問題陳述

口袋內容：
- 3 個藍球，2 個紅球，3 個綠球；總計$3 + 2 + 3 = 8$ 個球。無放回抽取2 個球。

定義：
- $X =$ 抽取的藍球數
- $Y =$ 抽取的紅球數

任務：

(1) 推導聯合PMF $p(x, y) = \mathbb{P}(X = x, Y = y)$ 的公式。

(2) 構造完整的聯合概率表。

(3) 計算$\mathbb{P}(X + Y \leq 1)$。

(4) 求X 的邊際分布。

(5) 求Y 的邊際分布。

1: 聯合PMF 公式（超幾何） **總**抽取2 個球的方式：$\binom{8}{2}$。對於x 個藍球，y 個紅球，其餘$2 - x - y$ 個球為綠球。有效範圍：$x = 0, 1, 2;\ y = 0, 1, 2;\ x + y \leq 2$。

$$p(x, y) = \frac{\binom{3}{x}\binom{2}{y}\binom{3}{2-x-y}}{\binom{8}{2}}$$

先計算分母：

$$\binom{8}{2} = \frac{8!}{2!6!} = \frac{8 \cdot 7}{2} = 28$$

個別聯合概率計算

$$p(0,0) = \frac{\binom{3}{0}\binom{2}{0}\binom{3}{2}}{28} = \frac{1 \cdot 1 \cdot 3}{28} = \frac{3}{28}$$
$$p(0,1) = \frac{\binom{3}{0}\binom{2}{1}\binom{3}{1}}{28} = \frac{1 \cdot 2 \cdot 3}{28} = \frac{6}{28} = \frac{3}{14}$$
$$p(0,2) = \frac{\binom{3}{0}\binom{2}{2}\binom{3}{0}}{28} = \frac{1 \cdot 1 \cdot 1}{28} = \frac{1}{28}$$
$$p(1,0) = \frac{\binom{3}{1}\binom{2}{0}\binom{3}{1}}{28} = \frac{3 \cdot 1 \cdot 3}{28} = \frac{9}{28}$$

$$
\begin{aligned}
p(1,1) &= \frac{\binom{3}{1}\binom{2}{1}\binom{3}{0}}{28} = \frac{3\cdot 2\cdot 1}{28} = \frac{6}{28} = \frac{3}{14} \\
p(1,2) &= \binom{3}{1}\binom{2}{2}\binom{3}{-1} \quad \text{(invalid, zero)} = 0 \\
p(2,0) &= \frac{\binom{3}{2}\binom{2}{0}\binom{3}{0}}{28} = \frac{3\cdot 1\cdot 1}{28} = \frac{3}{28} \\
p(2,1) &= p(2,2) = 0 \quad \text{(insufficient balls, negative green count)}
\end{aligned}
$$

2: 聯合概率表

$X \setminus Y$	$y=0$	$y=1$	$y=2$	行和$\mathbb{P}(X=x)$
$x=0$	$\frac{3}{28}$	$\frac{3}{14}$	$\frac{1}{28}$	$\frac{3}{28}+\frac{6}{28}+\frac{1}{28}=\frac{10}{28}=\frac{5}{14}$
$x=1$	$\frac{9}{28}$	$\frac{3}{14}$	0	$\frac{9}{28}+\frac{6}{28}+0=\frac{15}{28}$
$x=2$	$\frac{3}{28}$	0	0	$\frac{3}{28}+0+0=\frac{3}{28}$
列和$\mathbb{P}(Y=y)$	$\frac{3+9+3}{28}=\frac{15}{28}$	$\frac{6+6+0}{28}=\frac{12}{28}=\frac{3}{7}$	$\frac{1+0+0}{28}=\frac{1}{28}$	$\frac{28}{28}=1$

3: 計算$\mathbb{P}(X+Y\le 1)$ 有效(x,y) 對滿足$x+y\le 1$：$(0,0)$, $(0,1)$, $(1,0)$

$$
\begin{aligned}
\mathbb{P}(X+Y\le 1) &= p(0,0)+p(0,1)+p(1,0) \\
&= \frac{3}{28}+\frac{3}{14}+\frac{9}{28} \\
&= \frac{3}{28}+\frac{6}{28}+\frac{9}{28} \\
&= \frac{9}{14}
\end{aligned}
$$

4: X 的邊際分布

$$
\begin{aligned}
\mathbb{P}(X=0) &= p(0,0)+p(0,1)+p(0,2) = \frac{3}{28}+\frac{6}{28}+\frac{1}{28} = \frac{10}{28} = \frac{5}{14} \\
\mathbb{P}(X=1) &= p(1,0)+p(1,1)+p(1,2) = \frac{9}{28}+\frac{6}{28}+0 = \frac{15}{28} \\
\mathbb{P}(X=2) &= p(2,0)+p(2,1)+p(2,2) = \frac{3}{28}+0+0 = \frac{3}{28}
\end{aligned}
$$

x	0	1	2	Sum
$\mathbb{P}(X=x)$	$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

5: Y 的邊際分布

$$
\begin{aligned}
\mathbb{P}(Y=0) &= p(0,0)+p(1,0)+p(2,0) = \frac{3}{28}+\frac{9}{28}+\frac{3}{28} = \frac{15}{28} \\
\mathbb{P}(Y=1) &= p(0,1)+p(1,1)+p(2,1) = \frac{6}{28}+\frac{6}{28}+0 = \frac{12}{28} = \frac{3}{7} \\
\mathbb{P}(Y=2) &= p(0,2)+p(1,2)+p(2,2) = \frac{1}{28}+0+0 = \frac{1}{28}
\end{aligned}
$$

y	0	1	2	Sum
$\mathbb{P}(Y=y)$	$\frac{15}{28}$	$\frac{3}{7}$	$\frac{1}{28}$	1

5.6 連續隨機變量：聯合概率密度函數（PDF）

對於連續隨機變量X, Y，我們使用**聯合密度函數** $f(x,y)$ 並透過雙重積分進行概率計算。

1. 非負性：$f(x,y) \geq 0,\ \forall x, y \in \mathbb{R}$。
2. 總積分等於1：$\displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dxdy = 1$。
3. 矩形概率：$\mathbb{P}(a < X < b,\ c < Y < d) = \displaystyle\int_c^d \int_a^b f(x,y)\,dxdy$。
4. 一般區域概率：對於區域$A \subset \mathbb{R}^2$，
$$\mathbb{P}\big((X,Y) \in A\big) = \iint_A f(x,y)\,dxdy$$
5. 邊際PDF 公式：
$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy, \qquad f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx$$

5.7 示例4：聯合PDF $f(x,y) = x + y,\ 0 < x < 1,\ 0 < y < 1$

問題陳述

連續隨機變量X, Y 的聯合PDF 為

$$f(x,y) = \begin{cases} x + y & 0 < x < 1,\ 0 < y < 1 \\ 0 & \text{其他情況} \end{cases}$$

計算支撐區域$(0,1) \times (0,1)$ 上的邊際密度函數$f_X(x)$ 和$f_Y(y)$。

對$f_X(x)$ 的解析推導

由邊際PDF 定義：

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy$$

在$0 < y < 1$ 之外，$f(x,y) = 0$，因此我們將積分限限制為$y = 0$ 到$y = 1$：

$$f_X(x) = \int_0^1 (x+y)\,dy, \quad 0 < x < 1$$

對y 逐項積分（將x 視為常數）：

$$\begin{aligned}
\int_0^1 (x+y)\,dy &= \int_0^1 x\,dy + \int_0^1 y\,dy \\
&= xy\Big|_0^1 + \frac{1}{2}y^2\Big|_0^1 \\
&= x(1-0) + \left(\frac{1}{2}(1)^2 - \frac{1}{2}(0)^2\right) \\
&= x + \frac{1}{2}
\end{aligned}$$

X 的最終邊際：

$$f_X(x) = \begin{cases} x + \dfrac{1}{2} & 0 < x < 1 \\ 0 & \text{其他地方} \end{cases}$$

對 $f_Y(y)$ 的解析推導

對稱計算，積分消去 x：

$$f_Y(y) = \int_0^1 (x+y)\,dx, \quad 0 < y < 1$$

對 x 逐項積分（將 y 視為常數）：

$$\begin{aligned} \int_0^1 (x+y)\,dx &= \int_0^1 x\,dx + \int_0^1 y\,dx \\ &= \frac{1}{2}x^2\Big|_0^1 + y\,(x)\,\big|_0^1 \\ &= \left(\frac{1}{2}(1)^2 - \frac{1}{2}(0)^2\right) + y(1-0) \\ &= \frac{1}{2} + y \end{aligned}$$

Y 的最終邊際：

$$f_Y(y) = \begin{cases} y + \dfrac{1}{2} & 0 < y < 1 \\ 0 & \text{其他地方} \end{cases}$$

一致性檢查（驗證邊際積分等於1）

檢查 $\int_0^1 f_X(x)dx = 1$：

$$\int_0^1 \left(x + \frac{1}{2}\right) dx = 1$$

檢查 $\int_0^1 f_Y(y)dy = 1$：

$$\int_0^1 \left(y + \frac{1}{2}\right) dy = 1$$

兩個邊際PDF 均滿足有效密度函數的歸一化規則。

6 協方差矩陣

矩陣記號提供了一種緊湊的方式來量化一組隨機變量之間的成對關係。**協方差矩陣**彙集了集合中每對隨機變量的方差（自分散）和協方差（交叉線性關聯）。

總體協方差矩陣的定義

設$\{X_1, X_2, \ldots, X_p\}$ 為p 個隨機變量的集合。協方差矩陣$\mathbf{\Sigma} \in \mathbb{R}^{p\times p}$ 按元素定義：

$$\mathbf{\Sigma}_{ij} = \begin{cases} \mathrm{Var}(X_i) & i = j \quad \text{(主對角線：}X_i\text{ 的方差)} \\ \mathrm{Cov}(X_i, X_j) & i \neq j \quad \text{(非對角線：}X_i\text{ 與}X_j\text{ 的協方差)} \end{cases}$$

矩陣展開：

$$\mathbf{\Sigma} = \begin{bmatrix} \mathrm{Var}(X_1) & \mathrm{Cov}(X_1, X_2) & \ldots & \mathrm{Cov}(X_1, X_p) \\ \mathrm{Cov}(X_2, X_1) & \mathrm{Var}(X_2) & \ldots & \mathrm{Cov}(X_2, X_p) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(X_p, X_1) & \mathrm{Cov}(X_p, X_2) & \ldots & \mathrm{Var}(X_p) \end{bmatrix}$$

性質：協方差對稱$\mathrm{Cov}(X_i, X_j) = \mathrm{Cov}(X_j, X_i)$，因此$\mathbf{\Sigma} = \mathbf{\Sigma}^T$（對稱方陣）。簡單解釋：對角線元素量度個別變量分散；非對角線元素量度兩個不同變量之間的線性共變動。

6.1 示例5：樣本協方差矩陣計算

給定樣本數據集

5 個變量（A, B, C, D, E），每個變量$n = 6$ 個樣本觀測值：

變量	觀測1	觀測2	觀測3	觀測4	觀測5	觀測6
A	1	2	3	4	5	6
B	2	3	5	6	1	9
C	3	5	5	5	10	8
D	10	20	30	40	50	55
E	7	8	9	4	6	10

樣本協方差公式（無偏，分母$n - 1$）：

$$\widehat{\mathrm{Cov}}(X_i, X_j) = \frac{1}{n-1}\sum_{k=1}^{n}\big(X_{i,k} - \bar{X}_i\big)\big(X_{j,k} - \bar{X}_j\big), \quad \widehat{\mathrm{Var}}(X_i) = \widehat{\mathrm{Cov}}(X_i, X_i)$$

$n = 6$，因此$n - 1 = 5$。

1: 計算每個變量的樣本均值

$$\bar{A} = \frac{1+2+3+4+5+6}{6} = \frac{21}{6} = 3.5$$
$$\bar{B} = \frac{2+3+5+6+1+9}{6} = \frac{26}{6} \approx 4.3333333$$
$$\bar{C} = \frac{3+5+5+5+10+8}{6} = \frac{36}{6} = 6$$
$$\bar{D} = \frac{10+20+30+40+50+55}{6} = \frac{205}{6} \approx 34.1666667$$
$$\bar{E} = \frac{7+8+9+4+6+10}{6} = \frac{44}{6} \approx 7.3333333$$

2: 中心化數據矩陣（從每個條目中減去行均值） 中心化行= 原始行− 行均值。

$$\tilde{A} = [1-3.5,\ 2-3.5,\ 3-3.5,\ 4-3.5,\ 5-3.5,\ 6-3.5] = [-2.5,\ -1.5,\ -0.5,\ 0.5,\ 1.5,\ 2.5]$$
$$\tilde{B} = \left[2-\tfrac{26}{6},\ 3-\tfrac{26}{6},\ 5-\tfrac{26}{6},\ 6-\tfrac{26}{6},\ 1-\tfrac{26}{6},\ 9-\tfrac{26}{6}\right] = \left[-\tfrac{14}{6},\ -\tfrac{8}{6},\ \tfrac{4}{6},\ \tfrac{10}{6},\ -\tfrac{20}{6},\ \tfrac{28}{6}\right]$$
$$\tilde{C} = [3-6,\ 5-6,\ 5-6,\ 5-6,\ 10-6,\ 8-6] = [-3,\ -1,\ -1,\ -1,\ 4,\ 2]$$
$$\tilde{D} = \left[10-\tfrac{205}{6},\ 20-\tfrac{205}{6},\ 30-\tfrac{205}{6},\ 40-\tfrac{205}{6},\ 50-\tfrac{205}{6},\ 55-\tfrac{205}{6}\right] = \left[-\tfrac{145}{6},\ -\tfrac{85}{6},\ -\tfrac{25}{6},\ \tfrac{35}{6},\ \tfrac{95}{6},\ \tfrac{125}{6}\right]$$
$$\tilde{E} = \left[7-\tfrac{44}{6},\ 8-\tfrac{44}{6},\ 9-\tfrac{44}{6},\ 4-\tfrac{44}{6},\ 6-\tfrac{44}{6},\ 10-\tfrac{44}{6}\right] = \left[\tfrac{2}{6},\ \tfrac{8}{6},\ \tfrac{14}{6},\ -\tfrac{20}{6},\ -\tfrac{8}{6},\ \tfrac{16}{6}\right]$$

3: 透過中心化矩陣的樣本協方差矩陣公式 設$\widetilde{\mathbf{X}}$ 為5×6 中心化數據矩陣。無偏樣本協方差矩陣：

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{n-1}\widetilde{\mathbf{X}}\,\widetilde{\mathbf{X}}^T$$

$n-1=5$，因此將外積縮放$\frac{1}{5}$。

經驗證的數值樣本協方差矩陣（匹配Sage 輸出）

$$\widehat{\boldsymbol{\Sigma}} = \begin{bmatrix} 3.5000000 & 3.0000000 & 0.4000000 & 32.5000000 & 0.4000000 \\ 3.0000000 & 8.6666667 & 0.4000000 & 25.3333333 & 2.4666667 \\ 0.4000000 & 0.4000000 & 38.0000000 & 0.4000000 & 0.4000000 \\ 32.5000000 & 25.3333333 & 0.4000000 & 304.1666667 & 1.3333333 \\ 0.4000000 & 2.4666667 & 0.4000000 & 1.3333333 & 4.6666667 \end{bmatrix}$$