單元 40: 隨機變數線性函數的期望值與變異數

(課本 §5.8)

隨機變數 $Y_1, Y_2, \ldots, Y_n$ 的線性函數爲

$$a_1Y_1 + a_2Y_2 + \cdots + a_nY_n$$

其中 $a_1, a_2, \ldots, a_n$ 爲常數, 於第九章與第十一章中會常出現, 故有必要探討其期望值, 變異數, 與二線性組合間的共變異數如下述.

定理 5.12. 令 $Y_1, Y_2, \ldots, Y_n$ 的期望值

$$E(Y_i) = \mu_i,\ i = 1, 2, \ldots, n$$

且 $X_1, X_2, \ldots, X_m$ 的期望值

$$E(X_j) = \xi_j,\ j = 1, 2, \ldots, m$$

定義

$$U_1 = \sum_{i=1}^{n} a_iY_i$$

且

$$U_2 = \sum_{j=1}^{m} b_jX_j$$

其中 $a_1, a_2, \ldots, a_n$ 與 $b_1, b_2, \ldots, b_m$ 爲常數. 則下列各式成立:

(a) 期望值的線性性質,

$$E(U_1) = \sum_{i=1}^{n} a_i \mu_i \overset{\text{或}}{=} \sum_{i=1}^{n} a_i E(Y_i)$$

亦即, 線性組合的期望值等於期望值的線性組合.

(b) 線性組合的變異數,

$$\begin{aligned} \mathsf{Var}(U_1) &= \sum_{i=1}^{n} a_i^2 \mathsf{Var}(Y_i) + \\ & \qquad 2 \sum_{1 \le i < j \le n} a_i a_j \mathsf{Cov}(Y_i, Y_j) \end{aligned}$$

(c) 二線性組合間的共變異數,

$$\mathsf{Cov}(U_1, U_2) = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j \mathsf{Cov}(Y_i, X_j)$$

<證> (a) 多次提過的結論, 可透過多變量隨機變數函數的期望值公式以及積分 (連續情況) 和累加 (離散情況) 的線性性質證之, 略.

(b) 根據變異數的定義, U_1 的定義 以及 (a),

$$\begin{aligned}\text{Var}(U_1) &= E[(U_1 - E(U_1))^2] \\ &= E\left[\left(\sum_{i=1}^{n} a_iY_i - \sum_{i=1}^{n} a_i\mu_i\right)^2\right]\end{aligned}$$

接著, 合併上式中小括號內的兩個累加和並乘開, 得

$$\begin{aligned}\text{Var}(U_1) &= E\left[\left(\sum_{i=1}^{n} a_i(Y_i - \mu_i)\right)^2\right] \\ &= E\left[\sum_{i=1}^{n} a_i^2(Y_i - \mu_i)^2 + \right. \\ &\qquad \left.\sum_{1\le i\ne j\le n} a_ia_j(Y_i - \mu_i)(Y_j - \mu_j)\right]\end{aligned}$$

再根據 (a) 的期望值線性性質, 上式相當於

$$\begin{aligned}\text{Var}(U_1) &= \sum_{i=1}^{n} a_i^2E[(Y_i - \mu_i)^2] + \\ &\qquad \sum_{1\le i\ne j\le n} a_ia_jE[(Y_i - \mu_i)(Y_j - \mu_j)] \\ &= \sum_{i=1}^{n} a_i^2\text{Var}(Y_i) + \\ &\qquad \sum_{1\le i\ne j\le n} a_ia_j\text{Cov}(Y_i, Y_j) \quad (1)\end{aligned}$$

其中最後一個等號成立乃根據變異數與共變異數的定義.
又因爲

$$\begin{aligned}\mathrm{Cov}(Y_i, Y_j) &= E[(Y_i-\mu_i)(Y_j-\mu_j)] \\ &= E[(Y_j-\mu_j)(Y_i-\mu_i)] \\ &= \mathrm{Cov}(Y_j, Y_i) \qquad (2)\end{aligned}$$

故將 (1) 式中等號右邊第二個累加和中相同的項合併, 並把小的足標放在前面, 得

$$\begin{aligned}\mathrm{Var}(U_1) &= \sum_{i=1}^{n} a_i^2 \mathrm{Var}(Y_i) + \\ &\qquad 2\sum_{1\le i<j\le n} a_i a_j \mathrm{Cov}(Y_i, Y_j)\end{aligned}$$

(c) 根據共變異數的定義, 期望値的線性性質, U_1 與 U_2 的定義, 並經整理合併後, 得

$$\begin{aligned}&\mathrm{Cov}(U_1, U_2) \\ &= E\{[U_1-E(U_1)][U_2-E(U_2)]\} \\ &= E\left\{\left[\sum_{i=1}^{n} a_i(Y_i-\mu_i)\right]\left[\sum_{j=1}^{m} b_j(X_j-\xi_j)\right]\right\} \\ &= E\left[\sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j (Y_i-\mu_i)(X_j-\xi_j)\right] \\ &= \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j E[(Y_i-\mu_i)(X_j-\xi_j)]\end{aligned}$$

其中第三個等號成立乃是將第二個等號中大括號內的式子乘開後的結果. 最後, 再根據共變異數的定義, 由上式得

$$\text{Cov}(U_1, U_2) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j \text{Cov}(Y_i, X_j)$$

亦即, 兩個線性組合的共變異數等於此兩個線性組合相乘配對後的共變異數的線性組合.

註. 對任一隨機變數 Z, 由共變異數與變異數的定義,

$$\begin{aligned}\text{Cov}(Z, Z) &= E\{[Z - E(Z)][Z - E(Z)]\} \\ &= \text{Var}(Z) \qquad (3)\end{aligned}$$

故由 (3) 式, (c) 的公式, 以及 (2) 式, 經整理後, 得

$$\begin{aligned}\text{Var}(U_1) &= \text{Cov}(U_1, U_1) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \text{Cov}(Y_i, Y_j) \\ &= \sum_{i=1}^{n} a_i^2 \text{Cov}(Y_i, Y_i) + \\ &\qquad 2 \sum_{1\leq i<j\leq n} a_i a_j \text{Cov}(Y_i, Y_j) \\ &= \sum_{i=1}^{n} a_i^2 \text{Var}(Y_i) + \\ &\qquad 2 \sum_{1\leq i<j\leq n} a_i a_j \text{Cov}(Y_i, Y_j)\end{aligned}$$

此乃 (b) 的結果. 所以, (b) 爲 (c) 的一特例.

例 1. 承接石油庫存與銷售的例子, 其中 (Y_1, Y_2) 的 joint pdf

$$f(y_1, y_2) = \begin{cases} 3y_1, & 0 \le y_2 \le y_1 \le 1 \\ 0, & \text{其它} \end{cases}$$

試求剩餘石油比率 $Y_1 - Y_2$ 的變異數.

<解> 首先, 由定理 5.12 (b), 得

$$\begin{aligned} &\text{Var}(Y_1 - Y_2) \\ &= \text{Var}(Y_1) + \text{Var}(Y_2) - 2\text{Cov}(Y_1, Y_2) \end{aligned}$$

又由前例, 知 Y_1 的 marginal pdf

$$f_1(y_1) = \begin{cases} 3y_1^2, & 0 \le y_1 \le 1 \\ 0, & \text{其它} \end{cases}$$

因此, Y_1 的二階動差

$$\begin{aligned} E(Y_1^2) &= \int_{-\infty}^{\infty} y_1^2 f_1(y_1) dy_1 \\ &= \int_0^1 y_1^2 (3y_1^2) dy_1 \\ &= \frac{3}{5} y_1^5 \Big|_0^1 = \frac{3}{5} \end{aligned}$$

由前例, 又知

$$E(Y_1) = \frac{3}{4}$$

故

$$\mathrm{Var}(Y_1) = \frac{3}{5} - \left(\frac{3}{4}\right)^2 = \frac{48-45}{80} = \frac{3}{80}$$

同理, 由前例, 知 Y_2 的 marginal pdf

$$f_2(y_2) = \begin{cases} \frac{3}{2}(1-y_2^2), & 0 \le y_2 \le 1 \\ 0, & \text{其它} \end{cases}$$

所以, Y_2 的二階動差

$$\begin{aligned} E(Y_2^2) &= \int_{-\infty}^{\infty} y_2^2 f_2(y_2) dy_2 \\ &= \int_0^1 y_2^2 \cdot \frac{3}{2}(1-y_2^2) dy_2 \\ &= \frac{3}{2}\left(\frac{1}{3}y_2^3 - \frac{1}{5}y_2^5\right)\Big|_0^1 \\ &= \frac{3}{2}\left(\frac{1}{3} - \frac{1}{5}\right) = \frac{6}{30} = \frac{1}{5} \end{aligned}$$

由前例, 亦知

$$E(Y_2) = \frac{3}{8}$$

故

$$\mathrm{Var}(Y_2) = \left(\frac{1}{5}\right) - \left(\frac{3}{8}\right)^2 = \frac{64-45}{320} = \frac{19}{320}$$

最後, 由前例得

$$\mathrm{Cov}(Y_1, Y_2) = \frac{3}{160}$$

因此,

$$\begin{aligned}
&\mathrm{Var}(Y_1 - Y_2) \\
&= \mathrm{Var}(Y_1) + \mathrm{Var}(Y_2) - 2\mathrm{Cov}(Y_1, Y_2) \\
&= \frac{3}{80} + \frac{19}{320} - 2 \cdot \frac{3}{160} \\
&= \frac{12 + 19 - 12}{320} = \frac{19}{320}
\end{aligned}$$

例 2. 令 $Y_1, Y_2, \ldots, Y_n$ 相互獨立且

$$E(Y_i) = \mu,\ i = 1, 2, \ldots, n$$

以及

$$\mathrm{Var}(Y_i) = \sigma^2,\ i = 1, 2, \ldots, n$$

定義

$$\overline{Y} \stackrel{\mathrm{def}}{=} \frac{1}{n} \sum_{i=1}^{n} Y_i$$

此乃一 Y_i 的線性組合, 係數均爲 $\frac{1}{n}$. 試證

$$E(\overline{Y}) = \mu$$

且

$$\text{Var}(\overline{Y}) = \frac{\sigma^2}{n}$$

<證> 根據定理 5.12 (a) 的期望值線性性質,

$$\begin{aligned} E(\overline{Y}) &= \sum_{i=1}^{n} \frac{1}{n} E(Y_i) \\ &= \frac{1}{n} \sum_{i=1}^{n} E(Y_i) \\ &= \frac{1}{n} \sum_{i=1}^{n} \mu = \frac{1}{n} \cdot n\mu = \mu \end{aligned}$$

其中第三個等號成立乃根據假設有共同的期望值 μ 所致.

註. 在推導期望值的過程中沒有用到獨立性的假設, 亦即, 對任意的隨機變數, 只要它們有共同的期望值 μ, 則這些隨機變數平均值的期望值 $E(\overline{Y}) = \mu$.

接著, 根據定理 5.12 (b) 的線性組合變異數的公式,

$$\begin{aligned} \text{Var}(\overline{Y}) &= \sum_{i=1}^{n} \left(\frac{1}{n}\right)^2 \text{Var}(Y_i) + \\ & \qquad 2 \sum_{1 \le i < j \le n} \left(\frac{1}{n}\right)^2 \text{Cov}(Y_i, Y_j) \end{aligned}$$

又因爲當 $i \neq j$ 時, Y_i 與 Y_j 相互獨立, 而導致

$$\mathrm{Cov}(Y_i, Y_j) = 0$$

以及根據假設它們有共同的變異數 σ^2, 故由上式可得

$$\begin{aligned} \mathrm{Var}(\overline{Y}) &= \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 + 2\sum_{1\le i<j\le n}\left(\frac{1}{n^2}\right)(0) \\ &= \frac{1}{n^2}\cdot n\sigma^2 = \frac{\sigma^2}{n} \end{aligned}$$

註. 事實上, 在推導變異數的過程中, 關鍵乃在於對於任意的 $i \neq j$, $\mathrm{Cov}(Y_i, Y_j) = 0$ 這一較弱的條件, 即可推導出結論, 而不需要 "相互獨立" 這一較強的條件.

例 3. 設離散隨機變數

$$Y \sim \text{hypergeometric}(N, r, n)$$

亦即, 以不放回方式, 由 r 個紅球和 $(N-r)$ 個黑球中, 任選 n 個, 則 Y 爲選出的紅球數. 試求 $E(Y)$ 與 $\mathrm{Var}(Y)$.

<解> "以不放回方式選出 n 個, 觀察紅球數" 相當於 "按順序, 以不放回方式一次選一個, 觀察紅球數, 共 n 次", 亦即,

$$Y = X_1 + X_2 + \cdots + X_n$$

其中對於 $1 \le i \le n$,

$$X_i = \begin{cases} 1, & \text{若第 } i \text{ 次選出紅球} \\ 0, & \text{其它} \end{cases}$$

亦相當於 "由事先排列好的 n 個球中, 一個一個地觀察紅球數". 所以, 對於 $1 \le i \le n$, 事件

$$(X_i = 1) = \text{事先排列好的 } n \text{ 個球中, 第 } i \text{ 個爲紅球}$$

亦即,

$$\begin{array}{cccccccc} & \underline{\text{一球}} & \cdots & \underline{\text{一球}} & \underline{\text{紅球}} & \underline{\text{一球}} & \cdots & \underline{\text{一球}} \\ \text{位置:} & 1 & \cdots & i-1 & i & i+1 & \cdots & n \end{array}$$

其中 "一球" 的顏色不限, 可爲紅色或黑色. 又此種排列的方式爲, 先由 r 個紅球中選一個放於第 i 個位置, 再將剩下的 $(N-1)$ 個球排列到其他的 $(n-1)$ 個位置上, 故對於 $1 \le i \le n$,

$$\begin{aligned} & P(X_i = 1) \\ = & \ \frac{r \cdot (N-1)(N-2) \cdot (N-(n-1))}{N(N-1)\cdots(N-(n-1))} \\ = & \ \frac{r}{N} \end{aligned} \tag{4}$$

同理, 對於 $1 \le i \ne j \le n$, 事件

$$(X_i = 1, X_j = 1) = \text{事先排列好的 } n \text{ 個球中, 第 } i \text{ 個與第 } j \text{ 個必爲紅球}$$

亦即,

$$\begin{array}{cccccccc} & \underline{\text{一球}} & \cdots & \underline{\text{紅球}} & \cdots & \underline{\text{紅球}} & \cdots & \underline{\text{一球}} \\ \text{位置:} & 1 & \cdots & i & \cdots & j & \cdots & n \end{array}$$

其中 "一球" 的顏色不限, 可為紅色或黑色. 又此種排列的方式為, 先由 r 個紅球中選出二個排列於第 i 個與第 j 個位置, 再將剩下的 $(N-2)$ 個球排列到其他的 $(n-2)$ 個位置上, 故對於 $1 \leq i \neq j \leq n$,

$$\begin{aligned} & P(X_i = 1, X_j = 1) \\ & \quad = \frac{r(r-1) \cdot (N-2)(N-3) \cdots (N-(n-1))}{N(N-1)(N-2) \cdots (N-(n-1))} \\ & \quad = \frac{r(r-1)}{N(N-1)} \end{aligned} \tag{5}$$

註. 對於 $i \neq j$, 由 (4) 式與 (5) 式, 得

$$\begin{aligned} P(X_i = 1, X_j = 1) &= \frac{r(r-1)}{N(N-1)} \\ &\neq \left(\frac{r}{N}\right)\left(\frac{r}{N}\right) \\ &= P(X_i = 1)P(X_j = 1) \end{aligned}$$

故 X_i 與 X_j 不相互獨立.

接著, 由 (4) 式知, 對於 $1 \leq i \leq n$,

$$X_i \sim b\left(1, \frac{r}{N}\right)$$

因此, X_i 的期望值

$$E(X_i) = (1)\left(\frac{r}{N}\right) = \frac{r}{N} \tag{6}$$

且變異數

$$\begin{aligned}\text{Var}(X_i) &= (1)\left(\frac{r}{N}\right)\left(1-\frac{r}{N}\right) \\ &= \frac{r}{N}\left(1-\frac{r}{N}\right)\end{aligned} \tag{7}$$

由 (5) 式, 對於 $1 \leq i \neq j \leq n$,

$$\begin{aligned}&E(X_iX_j) \\ &= 1 \cdot P(X_iX_j = 1) + 0 \cdot P(X_iXj = 0) \\ &= 1 \cdot P(X_i = 1, X_j = 1) \\ &= \frac{r(r-1)}{N(N-1)}\end{aligned} \tag{8}$$

所以, 由 (6) 式及 (8) 式, 對於 $1 \leq i \neq j \leq n$,

$$\begin{aligned}\text{Cov}(X_i, X_j) &= E(X_iX_j) - E(X_i)E(X_j) \\ &= \frac{r(r-1)}{N(N-1)} - \left(\frac{r}{N}\right)\left(\frac{r}{N}\right) \\ &= \frac{r}{N}\left(\frac{r-1}{N-1} - \frac{r}{N}\right) \\ &= -\frac{r}{N}\left(1-\frac{r}{N}\right)\left(\frac{1}{N-1}\right)\end{aligned} \tag{9}$$

最後, 由 (6) 式, 得

$$E(Y) = \sum_{i=1}^{n} E(X_i) = n\left(\frac{r}{N}\right)$$

以及由 (7) 式和 (9) 式, 得

$$\begin{aligned}
&\text{Var}(Y) \\
= \; & \sum_{i=1}^{n} \text{Var}(X_i) + 2 \sum_{1 \le i < j \le n} \text{Cov}(X_i, X_j) \\
= \; & n \cdot \frac{r}{N}\left(1 - \frac{r}{N}\right) + 2 \cdot \frac{n(n-1)}{2} \cdot \\
& \qquad \left(-\frac{r}{N}\right)\left(1 - \frac{r}{N}\right)\left(\frac{1}{N-1}\right) \\
= \; & n\left(\frac{r}{N}\right)\left(1 - \frac{r}{N}\right)\left(1 - \frac{n-1}{N-1}\right) \\
= \; & n\left(\frac{r}{N}\right)\left(1 - \frac{r}{N}\right)\left(\frac{N-n}{N-1}\right)
\end{aligned}$$

其中第二個等號成立乃因爲前一式的第一個累加符號中,

$$1 \le i \le n$$

共有 n 項, 以及第二個累加符號中,

$$1 \le i < j \le n$$

中共有

$$\binom{n}{2} = \frac{n(n-1)}{2}$$

項所致.