單元 41: 多項機率分布
(課本 §5.9)

多項實驗是二項實驗的推廣, 其定義及相關的分布如下述.

定義. 一多項實驗的性質如下:

(1) 由 n 個相同的試驗 (trials) 所組成.

(2) 每一試驗會產生 k 種結果中的一種 (或落入 k 個盒中 (k cells) 中的一個).

(3) 單一試驗的結果爲第 i 類的 (或落入盒 i) 的機率爲 p_i, $i = 1, 2, \ldots, k$, 且對所有的試驗, 此現象均成立.

註. 因爲試驗結果必爲 k 種中的一種, 故

$$p_1 + p_2 + \cdots + p_k = 1$$

(4) 試驗間是相互獨立的.

(5) 令隨機變數 $Y_i = n$ 次試驗中, 結果爲第 i 類 (或落入盒 i) 的試驗數, $i = 1, 2, \ldots, k$.

註. 因爲有 n 個試驗結果且必爲 k 種中的一種, 故

$$Y_1 + Y_2 + \cdots + Y_k = n$$

則對於 $i = 1, 2, \ldots, k$, $0 \le y_i \le n$ 且 $\sum_{i=1}^{k} y_i = n$,

$$\begin{aligned} & P(Y_1 = y_1, Y_2 = y_2, \ldots, Y_k = y_k) \\ = & \ \frac{n!}{y_1! y_2! \cdots y_k!} p_1^{y_1} p_2^{y_2} \cdots p_k^{y_k} \\ \stackrel{\text{或}}{=} & \ \binom{n}{y_1\, y_2\, \cdots\, y_k} p_1^{y_1} p_2^{y_2} \cdots p_k^{y_k} \end{aligned}$$

並稱隨機變數 $(Y_1, Y_2, \ldots, Y_k)$ 具有參數爲 n, p_1, p_2, $\ldots$, p_k 的多項分布, 且表示成

$$(Y_1, Y_2, \ldots, Y_k) \sim \text{multinomial}(n, p_1, p_2, \ldots, p_k)$$

對於 $i = 1, 2, \ldots, k$, $0 \le y_i \le n$ 且 $\sum_{i=1}^{k} y_i = n$,它的 joint pmf

$$\begin{aligned} p(y_1, y_2, \ldots, y_k) & = \frac{n!}{y_1! y_2! \cdots y_k!} p_1^{y_1} p_2^{y_2} \cdots p_k^{y_k} \\ & \stackrel{\text{或}}{=} \binom{n}{y_1\, y_2\, \cdots\, y_k} p_1^{y_1} p_2^{y_2} \cdots p_k^{y_k} \end{aligned}$$

例 1. 根據人口普查資料, 美國成年人口的分類比例如下表:

類別	年齡	比例
1	18-24	.18
2	25-34	.23
3	35-44	.16
4	45-64	.27
5	65 ↑	.16

隨機選出 5 位成人, 試問其中有 1 位介於 18-24, 2 位介於 25-34, 2 位介於 45-64 的機率爲何?

<解> 因爲選出的人數 5 遠小於全國總成人數, 故可視爲放回, 獨立, 相同的試驗, 重複 5 次, 而得

$$(Y_1, Y_2, Y_3, Y_4, Y_5) \sim \text{multinomial}(5, .18, .23, .16, .27, .16)$$

且所求爲

$$\begin{aligned} & p(1,2,0,2,0) \\ = & \frac{5!}{1!2!0!2!0!}(.18)^1(.23)^2(.16)^0(.27)^2(.16)^0 \\ = & 30(.18)(.23)^2(.27)^2 \\ = & .0208 \end{aligned}$$

定理 5.13. 設多項分布隨機變數

$$(Y_1, Y_2, \ldots, Y_k) \sim \text{multinomial}(n, p_1, p_2, \ldots, p_k)$$

則 (1) 對於 $i = 1, 2, \ldots, k$,

$$E(Y_i) = np_i,\ \ \text{Var}(Y_i) = np_i q_i$$

其中 $q_i = 1 - p_i$.

(2) 若 $s \neq t$,

$$\text{Cov}(Y_s, Y_t) = -np_s p_t$$

<證> (1) 針對 Y_i 而言, k 類多項實驗相當於如下的二項實驗: 以 p_i 的機率產生第 i 類結果, $1 - p_i$ 的機率得到不是第 i 類的結果, 亦即, 將其它的結果合併爲不是第 i 類, 則對於 $1 \leq i \leq k$,

$$Y_i \sim b(n, p_i)$$

故對於 $i = 1, 2, \ldots, k$,

$$E(Y_i) = np_i \text{ 且 } \text{Var}(Y_i) = np_i q_i$$

其中 $q_i = 1 - p_i$.

(2) 可將 k 類多項實驗視爲一串 n 個相同獨立的試驗, 且針對 $s \neq t$, 定義

$$U_i = \begin{cases} 1, & \text{若第 } i \text{ 次試驗產生第 } s \text{ 類結果} \\ 0, & \text{其它} \end{cases}$$

且

$$W_i = \begin{cases} 1, & \text{若第 } i \text{ 次試驗產生第 } t \text{ 類結果} \\ 0, & \text{其它} \end{cases}$$

其中 $i = 1, 2, \ldots, n$. 則

$$Y_s = \sum_{i=1}^{n} U_i$$

且

$$Y_t = \sum_{j=1}^{n} W_j$$

又第 i 次試驗不可能同時產生第 s 類與第 t 類結果, 所以 U_iW_i 恆為 0, 故對於 $i = 1, 2, \ldots, n$,

$$E(U_iW_i) = 0$$

另對於 $i = 1, 2, \ldots, n$,

$$E(U_i) = p_s$$

且

$$E(W_i) = p_t$$

所以, 對於 $i = 1, 2, \ldots, n$,

$$\begin{aligned} \mathrm{Cov}(U_i, W_i) &= E(U_iW_i) - E(U_i)E(W_i) \\ &= 0 - p_sp_t = -p_sp_t \end{aligned} \tag{1}$$

當 $1 \le i \ne j \le n$ 時, 第 i 次試驗與第 j 次試驗是相互獨立的, 故

$$\begin{aligned}
&E(U_iW_j)\\
&= 1 \cdot P(U_iW_j = 1) + 0 \cdot P(U_iW_j = 0)\\
&= P(U_i = 1, W_j = 1)\\
&= P(U_i = 1)P(W_j = 1)\\
&= p_sp_t
\end{aligned}$$

其中第三個等號成立乃因爲獨立試驗乃相當於它們的相關事件 $(U_i = 1)$ 與 $(W_j = 1)$ 爲獨立的事件, 因而獨立的交集事件機率等於各自事件機率的乘積所致; 並由此可導出

$$\begin{aligned}
\mathrm{Cov}(U_i, W_j) &= E(U_iW_j) - E(U_i)E(W_j)\\
&= p_sp_t - p_sp_t = 0 \qquad (2)
\end{aligned}$$

所以, 當 $s \ne t$ 時, 根據線性組合的共變異數公式, (1) 式, 與 (2) 式,

$$\begin{aligned}
&\mathrm{Cov}(Y_s, Y_t)\\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} \mathrm{Cov}(U_i, W_j)\\
&= \sum_{i=1}^{n} \mathrm{Cov}(U_i, W_i) + \sum_{1\le i\ne j\le n} \mathrm{Cov}(U_i, W_j)\\
&= \sum_{i=1}^{n} (-p_sp_t) + \sum_{1\le i\ne j\le n} (0)\\
&= -np_sp_t
\end{aligned}$$

註. Y_s 與 Y_t 爲負相關, 與直觀相符, 因爲第 s 類的增加會迫使第 t 類的減少.