單元 16: 二項機率分布
(課本 §3.4)

定義 1. 二項實驗擁有下列性質:

(1) 實驗 (experiment) 由固定的 n 個相同的試驗 (identical trial) 所組成.

(2) 每種試驗會產生兩種結果中的一種, 稱其中一種爲成功 (S), 另一種爲失敗 (F).

(3) 單一試驗的成功機率爲一數值 p, 且對所有的試驗均相同. 失敗的機率爲 $q = 1 - p$.

(4) 試驗間是相互獨立的.

(5) 感興趣 (欲探討) 的隨機變數 $Y \stackrel{\text{def}}{=} n$ 次試驗中, 觀察到的成功次數.

註. “成功” 並不一定是要求日常用語中所謂 “好的” 事情, 僅代表一試驗中二種可能結果中的一個即可.

例 1. 一用來偵測飛行器的早期預警系統是由 4 個相同且獨立運作的雷達所組成. 設每一雷達能偵測到入侵飛行器的機率爲 0.95, 且感興趣的隨機變數 Y 爲無法偵測到飛行器的雷達數. 試問此爲一二項實驗嗎?

<解> 定義 “成功” = “無法偵測到”. 檢查 5 種性質如下:

(1) 4 種相同的試驗, 每種爲一特定雷達是否偵測出飛行器.

(2) 每種試驗產生: “成功” = “無法偵測出” 或 “失敗” = “偵測出”.

(3) 所有試驗的

$$\begin{aligned} p &= P(\text{成功 (S)}) \\ &= P(\text{無法偵測出}) \\ &= 1 - 0.95 \\ &= 0.05 \end{aligned}$$

(4) 因爲是獨立運作, 故試驗間是相互獨立的.

(5) $Y = 4$ 個試驗中成功的次數.

所以, 此爲一 $n = 4$, $p = 0.05$, 且 $q = 0.95$ 的二項實驗.

例 2. 設一大母體中某候選人的支持率爲 40%. 隨機選取一大小爲 $n = 10$ 的樣本, 且令 $Y =$ 觀察到的支持數. 試問此爲一二項實驗嗎?

<解> 因爲是以 “不放回的方式” 取樣, 首先

$$P(\text{選出的第一人支持}) = 0.4$$

但 (條件機率, conditional probability)

$$\begin{aligned} & P(\text{選出的第二人支持}|\text{選出的第一人支持}) \\ = \ & \frac{|A| - 1}{|S| - 1} \\ \neq \ & 0.4 \end{aligned}$$

此乃因爲設 $|A| =$ 支持的人數, 則 $|A|/|S| = 0.4$. 故試驗間並不相互獨立. 因此, 不爲一二項實驗.

註 1. 令事件 $S_1 =$ 選出的第一人支持, 且 $S_2 =$ 選出的第二人支持, 則根據全機率律, 得無條件機率

(unconditional probability)

$$\begin{aligned}
& P(\text{選出的第二人支持}) \\
&= P(S_1)P(S_2|S_1) + P(\overline{S_1})P(S_2|\overline{S_1}) \\
&= (0.4)\frac{|A|-1}{|S|-1} + (0.6)\frac{|A|}{|S|-1} \\
&= \frac{|A|-0.4}{|S|-1} \\
&= \frac{(0.4)|S|-0.4}{|S|-1} \ (\text{因爲 } |A|/|S| = 0.4) \\
&= (0.4)\frac{|S|-1}{|S|-1} \\
&= 0.4
\end{aligned}$$

事實上, 對任意 k, 無條件機率

$$P(\text{選出的第 } k \text{ 人支持}) = 0.4$$

想想看!

註 2. 一般而言, 若母體夠大且樣本相對地小, 後來試驗爲成功的條件機率會近似地相同, 而無關於前面試驗的結果 (因爲在大母體, 小樣本之下, 移除一人或少數人並不會實質地改變支持率). 因此試驗間爲近似地獨立, 且得一近似的二項實驗.

註 3. 事件 $(Y = y)$ 相當於事件

$$A \stackrel{\text{def}}{=} \{\text{所有 } n \text{ 元組} \,|\, n \text{ 個位置中有 } y \text{ 個放 } S, \text{剩下的 } n-y \text{ 個位置放 } F\}$$

又對於 A 中的每一個 n 元組, 其機率爲

$$p^y(1-p)^{n-y}$$

(根據獨立性與乘法交換率), 且

$$|A| = \binom{n}{y}$$

所以,

$$p(y) = P(A) = \binom{n}{y} p^y(1-p)^{n-y}$$

定義 2. 稱隨機變數 Y 具有 "根據 n 次試驗且成功機率爲 p 的二項分布" 若且爲若

$$p(y) = \binom{n}{y} p^y(1-p)^{n-y}$$

其中 $y = 0, 1, 2, \ldots, n$ 且 $0 < p < 1$, 並以

$$Y \sim b(n, p)$$

表示之. 也就是說,

$$Y \sim b(n,p) \;\Leftrightarrow\; p(y) = \binom{n}{y} p^y (1-p)^{n-y}$$

例 3. 設一捆 5000 個保險絲中, 有 5% 是有瑕疵的. 若抽 5 個檢測, 試求至少有一個瑕疵的機率.

<解> 因爲是大母體 (N=5000), 小樣本 ($n = 5$), 故可假設隨機變數

$$Y \overset{\text{def}}{=} \text{瑕疵數} \sim b(5, 0.05)$$

所以,

$$\begin{aligned} P(\text{至少有一個瑕疵}) &= 1 - P(Y = 0) \\ &= 1 - \binom{5}{0}(0.05)^0(0.95)^5 \\ &= 1 - (0.95)^5 \\ &= 1 - 0.774 \\ &= 0.226 \end{aligned}$$

是一個相當大的機會可看到至少有一個是瑕疵的, 即使樣本是如此的小.

例 4. 經驗顯示感染某病的人中, 會有 30% 會自動康復. 藥廠研發出一種新藥, 並隨機選取 10 位病人, 予以

服用後, 得 9 位康復. 假設新藥完全無效, 試求服藥的 10 位病人中, 至少有 9 位康復的機率.

<解> 令 Y = 服藥的10 位病人中, 康復的病人數. 則

$$Y \sim b(10, 0.3)$$

此乃因爲在無藥效的假設下, 自動康復率爲 0.3. 因此,

$$\begin{aligned}
\text{所求} &= P(Y \geq 9) \\
&= P(Y = 9) + P(Y = 10) \\
&= \binom{10}{9}(0.3)^9(0.7) + \\
&\qquad \binom{10}{10}(0.3)^{10}(0.7)^0 \\
&= 0.000138 + 0.000006 \\
&= 0.000144
\end{aligned}$$

爲一極低的機率值.

接著探討出現 9 位康復的可能原因:

(1) 藥確實無效, 此乃一純屬巧合的稀有事件.

(2) 藥有效.

推論: 傾向採用第 (2) 個原因, 因爲一極低發生率 (不可能發生) 的事件居然會發生, 顯現是由於假設錯誤而造成的極小機率值.

例 5. (接續例 3.) 若抽 20 個檢測, 試求至少有 4 個瑕疵的機率.

<解> 因爲樣本大小 $n = 20 \ll$ 母體大小 $N = 5000$, 可設

$$Y \sim b(20, 0.05)$$

故,

$$\begin{aligned} \text{所求} &= P(Y \geq 4) \\ &= 1 - P(Y \leq 3) \\ &= 1 - 0.984 \text{ (查表1)} \\ &= 0.016 \end{aligned}$$

一相當小的值. 所以, 若檢測 20 個保險絲時, 確實觀察到 3 個以上 (不含) 的瑕疵品, 則<u>可懷疑</u> 5% 的瑕疵率是錯誤的.

定理 3.7. 設 $Y \sim b(n, p)$. 則

$$\mu = E(Y) = np$$

且

$$\sigma^2 = \mathsf{Var}(Y) = npq,\ \ q = 1 - p$$

<證> 根據期望值的定義,

$$\begin{aligned}
& E(Y) \\
= \ & \sum_{y=0}^{n} y \cdot \binom{n}{y} p^y (1-p)^{n-y} \\
= \ & \sum_{y=1}^{n} y \cdot \frac{n!}{y!(n-y)!} p^y (1-p)^{n-y} \\
= \ & np \sum_{y=1}^{n} \frac{(n-1)!}{(y-1)!(n-y)!} p^{y-1} (1-p)^{n-y}
\end{aligned}$$

接著經由如下的變數變換: 令

$$z = y - 1$$

以及

$$\begin{aligned}
y = 1 &\ \Rightarrow\ z = 0 \\
y = n &\ \Rightarrow\ z = n - 1
\end{aligned}$$

可由上式得出

$$E(Y) = np \sum_{z=0}^{n-1} \frac{(n-1)!}{z!(n-1-z)!} p^z (1-p)^{n-1-z}$$

又上式累加符號內的式子剛好是 $b(n-1,p)$ 的 pmf, 故根據 pmf 的定義, 完整的和 (亦即, 從 $z=0$ 加到 $z=n-1$) 爲 1. 因此,

$$E(Y) = np(1) = np$$

註. 針對離散隨機變數 Y, 機率函數 (probability function) $p(y)$ 又稱作機率質量函數 (probability mass function), 並簡記成 pmf.

根據變異數的公式

$$\sigma^2 = E(Y^2) - [E(Y)]^2$$

故需先求

$$E(Y^2) = E[Y(Y-1)] + E(Y)$$

而

$$\begin{aligned}
&E[Y(Y-1)] \\
&= \sum_{y=0}^{n} y(y-1)\frac{n!}{y!(n-y)!}p^y(1-p)^{n-y} \\
&= n(n-1)p^2 \cdot \\
&\qquad \sum_{y=2}^{n} \frac{(n-2)!}{(y-2)!(n-y)!}p^{y-2}(1-p)^{n-y}
\end{aligned}$$

接著經由類似的變數變換 $z = y - 2$, 可由上式得出

$$E[Y(Y-1)] = n(n-1)p^2 \cdot \sum_{z=0}^{n-2} \frac{(n-2)!}{z!(n-2-z)!} p^z (1-p)^{n-2-z}$$

同樣可觀察出上式累加符號內的式子爲 $b(n-2, p)$ 的 pmf, 故完整的和爲 1. 所以,

$$E[Y(Y-1)] = n(n-1)p^2(1) = n(n-1)p^2$$

因此, 得

$$\begin{aligned} E(Y^2) &= E[Y(Y-1)] + E(Y) \\ &= n(n-1)p^2 + np \end{aligned}$$

且

$$\begin{aligned} \sigma^2 &= E(Y^2) - [E(Y)]^2 \\ &= n(n-1)p^2 + np - n^2p^2 \\ &= n^2p^2 - np^2 + np - n^2p^2 \\ &= np(1-p) \\ &= npq \end{aligned}$$

例 6. 調查一大公司中 20 位員工有關是否贊同某項政策. 若得 6 位支持, 試估計 “眞正但未知” 的支持率 p.

<解> 因爲大母體 (公司), 小樣本, 故可設

$$Y \stackrel{\text{def}}{=} 20 \text{ 人中贊同的人數} \sim b(20, p)$$

則

$$P(Y=6) = \binom{20}{6} p^6 (1-p)^{14}$$

應是一個大的值, 因爲 $(Y=6)$ 確實發生了. 所以, 可採用使得 $P(Y=6)$ (一個 p 的函數) 爲最大的 p 值, 作爲眞正的 p 值的估計值. 又

$$\text{最大化 } P(Y=6) \quad \Leftrightarrow \quad \text{最大化 } p^6(1-p)^{14}$$

因爲常數 $\binom{20}{6} > 0$. 此亦相當於

$$\text{最大化 } g(p) \stackrel{\text{def}}{=} 6 \ln p + 14 \ln(1-p)$$

因爲 $\ln(\cdot)$ 爲一遞增函數.

最大化的標準過程如下:

(1) 求臨界點 (critical points): 令

$$\frac{d}{dp}[6 \ln p + 14 \ln(1-p)] = \frac{6}{p} - \frac{14}{1-p} = 0$$

由此導出

$$6(1-p) - 14p = 0 \quad \Leftrightarrow \quad 6 - 20p = 0$$

因此,

$$p = \frac{6}{20} \in (0, 1)$$

(2) 驗證: 繪出 $g'(p)$ 的符號圖 (sign chart) 如下.

故, 在

$$p = \frac{6}{20}$$

時有最大值.

因此, 以 $\frac{6}{20}$ 估計 p, 與直觀相符 (因爲 p 表示任一試驗成功的機率, 含有相對頻率的意味, 故以 20 次中 6 次成功的比例 $\frac{6}{20}$ 估計 p 是相當合理的).

註. 此種估計 p 的方法稱作最大概似法 (method of maximum likelihood). 得到的估計值稱作最大概似估計值 (maximum likelihood estimate, 簡稱 mle).