

單元 17: 幾何 (等比) 機率分布 (課本 §3.5)

獨立地重複相同的試驗. 若 $P(\text{成功}) = p$, 且令

$Y =$ 第一次成功出現時, 所需的試驗次數

則 Y 的可能值為 $1, 2, 3, \dots$ 且針對 $y = 1, 2, 3, \dots$,

$$\begin{aligned} P(Y = y) &= P(\text{前面 } (y - 1) \text{ 次試驗均失敗且第 } y \text{ 次成功}) \\ &= (1 - p)^{y-1} p \quad (\text{由獨立性}) \\ &= q^{y-1} p, \quad q = 1 - p \end{aligned}$$

並稱隨機變數 Y 有幾何 (等比) 機率分布 (geometric probability distribution). 正式定義如下:

定義 1. $Y \sim \text{geometric}(p)$ 若且為若

$$\begin{aligned} p(y) &= P(Y = y) \\ &= q^{y-1} p \end{aligned}$$

其中 $y = 1, 2, 3, \dots$, $0 \leq p \leq 1$.

註. 幾何 (等比) 機率分布常用於 “等候時間的長度” 的模型建構.

例 1. 設一引擎在任何 “單一小時” 內壞掉的機率為 0.02. 試求 $P(\text{引擎存活 } 2 \text{ 小時})$.

<解> 令 $Y =$ 到第一次壞掉時, 所在的 “單一小時” 數, 則由題意得,

$$Y \sim \text{geometric}(0.02)$$

故,

$$\begin{aligned} P(\text{引擎存活 } 2 \text{ 小時}) &= P(Y \geq 3) \\ &= 1 - P(Y = 1) - P(Y = 2) \\ &= 1 - (0.02) - (0.98)(0.02) \\ &= 0.9604 \end{aligned}$$

定理 3.8. 設 $Y \sim \text{geometric}(p)$. 則

$$\mu = E(Y) = \frac{1}{p}$$

且

$$\text{Var}(Y) = \frac{1-p}{p^2}$$

註. 平均值與 p 成反比.

<證> 根據期望值的定義,

$$\begin{aligned} E(Y) &= \sum_{y=1}^{\infty} y \cdot q^{y-1} p = p \cdot \sum_{y=1}^{\infty} y q^{y-1} \\ &= p \cdot \frac{d}{dq} \left[\sum_{y=0}^{\infty} q^y \right] \\ &\quad \text{(因爲此時微分與加法的互換是成立的)} \\ &= p \cdot \frac{d}{dq} \left[\frac{1}{1-q} \right] = p \cdot \frac{1}{(1-q)^2} = \frac{1}{p} \end{aligned}$$

又

$$\begin{aligned} E[Y(Y-1)] &= \sum_{y=1}^{\infty} y(y-1) q^{y-1} p \\ &= qp \sum_{y=1}^{\infty} y(y-1) q^{y-2} \\ &= qp \cdot \frac{d^2}{dq^2} \left[\sum_{y=0}^{\infty} q^y \right] = qp \cdot \frac{d^2}{dq^2} \left[\frac{1}{1-q} \right] \\ &= qp \cdot \frac{2}{(1-q)^3} = \frac{2q}{p^2} \end{aligned}$$

故,

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

且

$$\begin{aligned}\text{Var}(Y) &= E(Y^2) - [E(Y)]^2 \\ &= \frac{2q}{p^2} + \frac{1}{p} - \frac{1}{p^2} \\ &= \frac{2q + p - 1}{p^2} = \frac{1 - p}{p^2}\end{aligned}$$

例 2. 承接例 1, 得平均等候時間

$$E(Y) = \frac{1}{p} = \frac{1}{0.02} = 50 \text{ 小時}$$

且變異數

$$\text{Var}(Y) = \frac{1 - 0.02}{(0.02)^2} = \frac{0.98}{0.0004} = 2450$$

以及標準差

$$\sigma = \sqrt{2450} = 49.497 \text{ 小時}$$

例 3. 在一大公司中調查員工是否贊同某項政策, 直到出現第一位贊同者為止. 若第 5 位受訪者是第一位贊同者, 試估計 “真正但未知” 支持率 p .

<解> 因爲大母體 (大公司), 故可設

$$Y \stackrel{\text{def}}{=} \text{直到第一位贊同的受訪人數} \sim \text{geometric}(p)$$

則

$$P(Y = 5) = (1 - p)^4 p$$

應是一個大的值, 因為 $(Y = 5)$ 確實發生了. 所以, 可採用使得 $P(Y = 5)$ (p 的函數) 為最大值的 p 值, 做為真正的 p 值的估計值. 又

$$\text{最大化 } (1 - p)^4 p$$

相當於

$$\text{最大化 } g(p) \stackrel{\text{def}}{=} 4 \ln(1 - p) + \ln p$$

過程如下:

(1) 求臨界值 (critical points): 令

$$g'(p) = \frac{-4}{1 - p} + \frac{1}{p} = 0$$

得

$$\frac{1 - 5p}{p(1 - p)} = 0$$

因此,

$$p = \frac{1}{5} \in (0, 1)$$

(2) 驗證: 繪 $g'(p)$ 的符號圖如下:

故, 在 $p = \frac{1}{5}$ 時有最大值.

因此, 以 $\frac{1}{5}$ 估計 p , 與直觀相符 (因為第 5 位受訪者才是第一位支持者, 故以相對頻率的角度來說, 以 $\frac{1}{5}$ 估計是合理的).

註. 此種估計 p 的方法稱作最大概似法 (method of maximum likelihood).