# 單元 19: 超幾何機率分布
(課本 §3.7)

設母體中有 $N$ 個元素, 其中 $r$ 個爲第一類 (如, 紅色), 剩下的 $(N-r)$ 個爲第二類 (如, 黑色), 如圖所示. 若實驗爲由 $N$ 個元素中, 隨機選出 $n$ 個. 令

$$Y = \text{選出的第一類 (或紅色) 元素的個數}$$

則 $Y$ 的可能值爲 0, 1, 2, ..., $\min(r, n)$, 且

$$\begin{aligned} P(Y=y) &= (\text{選出 } y \text{ 個第一類的方法數}) \cdot \\ &\quad (\text{選出 } (n-y) \text{ 個第二類的方法數}) / \\ &\quad (\text{選出 } n \text{ 個元素的總方法數}) \\ &= \frac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}} \end{aligned}$$

其中 $y = 0, 1, 2, \ldots, \min(r, n)$ 且 $n - y \le N - r$. 並稱隨機變數 $Y$ 有超幾何機率分布 (hypergeometric probability distribution). 正式定義如下:

**定義 1.** $Y \sim$ hypergeometric$(N, r, n)$ 若且爲若

$$\begin{aligned} p(y) &= P(Y=y) \\ &= \frac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}} \end{aligned}$$

其中 $y = 0, 1, 2, \ldots, \min(r, n)$ 且 $n - y \leq N - r$.

**定理 3.10.** 設 $Y \sim \text{hypergeometric}(N, r, n)$.
則

$$\mu = E(Y) = n\left(\frac{r}{N}\right)$$

且

$$\sigma^2 = \text{Var}(Y) = n\left(\frac{r}{N}\right)\left(1 - \frac{r}{N}\right)\left(\frac{N-n}{N-1}\right)$$

<證> (略) 第五章有簡易的證法.

**註 1.** 令

$$p = \frac{r}{N} \text{ 且 } q = 1 - p = 1 - \frac{r}{N}$$

則

$$E(Y) = np$$

且

$$\text{Var}(Y) = npq\left(\frac{N-n}{N-1}\right)$$

與二項分布的期望值和變異數相似, 並稱因子

$$\frac{N-n}{N-1}$$

爲一調整 (adjustment) 因子 (這是適當的, 當 $n$ 相對於 $N$ 是大的時候). 另外, 當 $n$ 固定時, 若 $N \to \infty$ (亦即, 大母體, 小樣本), 則

$$\frac{N-n}{N-1} \to 1$$

更明顯地顯示出此種相似.

例 1. 設一捆 20 個產品中, 有 4 個瑕疵品. 隨機選取 5 個檢查, 若有超過一個壞的, 則退貨. 問退貨的機率爲何? 另求大小爲 5 的樣本中, 瑕疵產品數的期望值與變異數.

<解> 令 $Y$ = 檢查出的瑕疵數. 則

$$Y \sim \text{hypergeometric}(20, 4, 5)$$

又

$$\begin{aligned}
P(\text{退貨}) &= 1 - P(Y \le 1) \\
&= 1 - P(Y=0) - P(Y=1) \\
&= 1 - \frac{\binom{4}{0}\binom{16}{5}}{\binom{20}{5}} - \frac{\binom{4}{1}\binom{16}{4}}{\binom{20}{5}} \\
&= 1 - 0.2817 - 0.4696 \\
&= 0.2487
\end{aligned}$$

且期望值

$$E(Y) = 5\left(\frac{4}{20}\right) = 1$$

變異數

$$\begin{aligned}\text{Var}(Y) &= 5\left(\frac{4}{20}\right)\left(1-\frac{4}{20}\right)\left(\frac{20-5}{20-1}\right)\\ &= \frac{16}{20}\cdot\frac{15}{19}\\ &= \frac{48}{76}\\ &= 0.632\end{aligned}$$

註 2. 當 $N \to \infty$ 且 $n$, $\frac{r}{N} \stackrel{\text{def}}{=} p$ 為固定時 (亦即, 大母體, 小樣本), 針對每一固定的 $y$,

$$\lim_{N\to\infty} \frac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}} = \binom{n}{y} p^y (1-p)^{n-y}$$

亦即, hypergeometric($N, r, n$) 的 pmf 的極限剛好等於 $b(n,p)$ 的 pmf, 此乃從 pmf 的角度反映出超幾何機率分布與二項機率分布的相似性.

<證> 自行閱讀如下的證明. 經由展開二項係數, 化簡, 求極限 (在 $y$, $n$ 固定且 $\frac{r}{N} = p$ 也固定下) 的過程可證

得. 首先, 展開二項係數, 得

$$\frac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}}$$

$$= \frac{n!}{y!(n-y)!}\left[\frac{r(r-1)\cdots(r-y+1)}{N(N-1)\cdots(N-y+1)}\right]\cdot \left[\frac{(N-r)(N-r-1)\cdots}{(N-y)(N-y-1)\cdots}\right. \left.\frac{(N-r-(n-y-1))}{(N-y-(n-y-1))}\right]$$

接著化簡, 得

上式

$$= \binom{n}{y}\left(\frac{r}{N}\right)\left[\frac{r}{N}\left(\frac{1-\frac{1}{r}}{1-\frac{1}{N}}\right)\right]\left[\frac{r}{N}\left(\frac{1-\frac{2}{r}}{1-\frac{2}{N}}\right)\right]\cdots \left[\frac{r}{N}\left(\frac{1-\frac{y-1}{r}}{1-\frac{y-1}{N}}\right)\right]\cdot\left[1-\frac{r-y}{N-y}\right]\cdot \left[1-\frac{r-y}{N-y-1}\right]\cdots \left[1-\frac{r-y}{N-y-(n-y-1)}\right] \qquad (1)$$

在 $\frac{r}{N}=p$ 固定, $y$ 固定, 以及 $N, r\to\infty$ 下, 由

$$\left(\frac{r}{N}\right)$$

開始至

$$\left[\frac{r}{N}\left(\frac{1-\frac{y-1}{r}}{1-\frac{y-1}{N}}\right)\right]$$

的 $y$ 項均收斂到

$$\frac{r}{N}(1)=p$$

而且由

$$\left[1-\frac{r-y}{N-y}\right]$$

到

$$\left[1-\frac{r-y}{N-y-(n-y-1)}\right]$$

的 $(n-y)$ 項均可分別表示成

$$\left[1-\frac{r}{N}\left(\frac{1-\frac{y}{r}}{1-\frac{y}{N}}\right)\right]$$

以及

$$\left[1-\frac{r}{N}\left(\frac{1-\frac{y}{r}}{1-\frac{y+(n-y-1)}{N}}\right)\right]$$

而且在 $n$ 也固定下, 均收斂到

$$1-\frac{r}{N}(1)=1-p$$

最後, 將上述求得的極限合併 (相乘) 並根據 (1) 式, 得

$$\frac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}} \to \binom{n}{y} p^y (1-p)^{n-y}$$