

單元 29: Beta 機率分布

(課本 §4.7)

常用於模型化：比率，如化學產品中不純度的比率，機器處在維修狀態所占的比率，...，故其可能值 $\in [0, 1]$ 。正式定義如下：

定義 1. $Y \sim \text{beta}(\alpha, \beta)$, $\alpha > 0$, $\beta > 0$, 若且為若 Y 的 pdf

$$f(y) = \begin{cases} \frac{y^{\alpha-1}(1-y)^{\beta-1}}{B(\alpha, \beta)}, & 0 \leq y \leq 1 \\ 0, & \text{其它} \end{cases}$$

其中

$$B(\alpha, \beta) = \int_0^1 y^{\alpha-1}(1-y)^{\beta-1} dy = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

註 1. 雖然 beta 的 pdf 定義在 $0 \leq y \leq 1$ ，但透過轉換

$$y^* = \frac{y-c}{d-c}$$

可將 beta 的 pdf 推廣至任意區間： $c \leq y \leq d$ (結果待證)。

註 2. Beta 隨機變數的 cdf

$$\begin{aligned} F(y) &= \int_0^y \frac{t^{\alpha-1}(1-t)^{\beta-1}}{B(\alpha, \beta)} dt \\ &\stackrel{\text{表成}}{=} I_y(\alpha, \beta) \end{aligned}$$

稱作非完整 beta 函數 (incomplete beta function), 無明確的完整型 (closed-form), 但當 α, β 均為整數時, 由分部積分 (integration by parts), 可得

$$\begin{aligned} I_y(\alpha, \beta) &= \int_0^y \frac{t^{\alpha-1}(1-t)^{\beta-1}}{B(\alpha, \beta)} dt \\ &= \sum_{i=\alpha}^n \binom{n}{i} y^i (1-y)^{n-i} \end{aligned}$$

其中 $0 < y < 1$, $n = \alpha + \beta - 1$. (一個與二項分布的 pmf 之間的關係.)

定理 4.11. 若 $Y \sim \text{beta}(\alpha, \beta)$, $\alpha > 0$, $\beta > 0$, 則期望值

$$\mu = E(Y) = \frac{\alpha}{\alpha + \beta}$$

且變異數

$$\sigma^2 = \text{Var}(Y) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

<證> 根據期望值的定義,

$$\begin{aligned}
 E(Y) &= \int_{-\infty}^{\infty} y f(y) dy \\
 &= \int_0^1 y \cdot \left(\frac{y^{\alpha-1} (1-y)^{\beta-1}}{B(\alpha, \beta)} \right) dy \\
 &= \frac{1}{B(\alpha, \beta)} \int_0^1 y^{\alpha} (1-y)^{\beta-1} dy \\
 &= \frac{1}{B(\alpha, \beta)} \cdot B(\alpha+1, \beta) \\
 &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+\beta+1)} \\
 &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \frac{\alpha\Gamma(\alpha)\Gamma(\beta)}{(\alpha+\beta)\Gamma(\alpha+\beta)} \\
 &= \frac{\alpha}{\alpha+\beta}
 \end{aligned}$$

其中第五及第六個等號成立乃根據 $B(\alpha, \beta)$ 與 $\Gamma(\alpha)$ 的定義所致，並經由化簡得最後一個等號；另外，第四個等號成立乃因為 $\text{beta}(\alpha, \beta)$ 的 pdf 的完整型積分爲 1，亦即，

$$\int_0^1 \frac{y^{\alpha-1} (1-y)^{\beta-1}}{B(\alpha, \beta)} dy = 1$$

由此可導出，對於 $\alpha > 0, \beta > 0$,

$$\int_0^1 y^{\alpha-1} (1-y)^{\beta-1} dy = B(\alpha, \beta) \quad (1)$$

故根據 (1) 式, 得

$$\begin{aligned}
 \int_0^1 y^\alpha (1-y)^{\beta-1} dy \\
 &= \int_0^1 y^{(\alpha+1)-1} (1-y)^{\beta-1} dy \\
 &= B(\alpha+1, \beta)
 \end{aligned}$$

同理, 根據隨機變數的函數期望值的公式, 以及合併被積函數中的 y 項, 且經由整理, 由 (1) 式得

$$\begin{aligned}
 E(Y^2) &= \int_0^1 y^2 \left(\frac{y^{\alpha-1} (1-y)^{\beta-1}}{B(\alpha, \beta)} \right) dy \\
 &= \frac{1}{B(\alpha, \beta)} \int_0^1 y^{(\alpha+2)-1} (1-y)^{\beta-1} dy \\
 &= \frac{1}{B(\alpha, \beta)} \cdot B(\alpha+2, \beta) \\
 &\quad \text{(根據 (1) 式)} \\
 &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \frac{\Gamma(\alpha+2)\Gamma(\beta)}{\Gamma(\alpha+\beta+2)} \\
 &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \frac{(\alpha+1)\alpha\Gamma(\alpha)\Gamma(\beta)}{(\alpha+\beta+1)(\alpha+\beta)\Gamma(\alpha+\beta)} \\
 &= \frac{(\alpha+1)\alpha}{(\alpha+\beta+1)(\alpha+\beta)}
 \end{aligned}$$

所以,

$$\begin{aligned}\text{Var}(Y) &= E(Y^2) - [E(Y)]^2 \\&= \frac{(\alpha + 1)\alpha}{(\alpha + \beta + 1)(\alpha + \beta)} - \left(\frac{\alpha}{\alpha + \beta}\right)^2 \\&= \frac{\alpha}{\alpha + \beta} \left[\frac{\alpha + 1}{\alpha + \beta + 1} - \frac{\alpha}{\alpha + \beta} \right] \\&= \frac{\alpha}{\alpha + \beta} \cdot \frac{\alpha(\alpha + \beta) + \alpha + \beta - \alpha(\alpha + \beta) - \alpha}{(\alpha + \beta)(\alpha + \beta + 1)} \\&= \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}\end{aligned}$$

例 1. 石油批發商一週的銷售率可用 $\alpha = 4$, $\beta = 2$ 的 beta 分布模型化. 試求至少銷售庫存的 90% 的機率.

<解> 由題意, 銷售率 $Y \sim \text{beta}(4, 2)$, 且其機率密度函數 pdf

$$\begin{aligned}f(y) &= \begin{cases} \frac{\Gamma(6)}{\Gamma(4)\Gamma(2)}y^3(1-y), & 0 \leq y \leq 1 \\ 0, & \text{其它} \end{cases} \\&= \begin{cases} 20y^3(1-y), & 0 \leq y \leq 1 \\ 0, & \text{其它} \end{cases}\end{aligned}$$

所以，銷售率至少為 90% 的機率

$$\begin{aligned}P(Y \geq .9) &= \int_{0.9}^1 20(y^3 - y^4) dy \\&= 20 \left(\frac{y^4}{4} - \frac{y^5}{5} \right) \Big|_{0.9}^1 \\&= 5y^4 - 4y^5 \Big|_{0.9}^1 \\&= 1 - (0.9)^4(5 - 3.6) \\&= 1 - 0.91854 \\&= 0.08146\end{aligned}$$