單元 15: 隨機變數或其函數的期望值
(課本 §3.3)

定義 1. 令 Y 爲一離散隨機變數, 且其機率函數爲 $p(y)$ (簡記成 $Y \sim p(y)$). 則 Y 的期望值 (expected value 或 mean)

$$\mu \overset{\text{或}}{=} E(Y) \overset{\text{def}}{=} \sum_y yp(y)$$

上式中的累加符號式 $\sum$ 是涵蓋所有可能的 y 值.

定理 3.2. 令 $Y \sim p(y)$ 且 $g(Y)$ 爲 Y 的一實值函數. 則

$$E[g(Y)] = \sum_y g(y)p(y)$$

<證> 僅考慮 Y 取有限個值

$$y_1, y_2, \cdots, y_n$$

的情況. 同理可證 Y 取可數無窮個值的情況.

因爲 $g(y)$ 不一定是 1-1 (一對一), 故假設 $g(Y)$ 取值

$$g_1, g_2, \cdots, g_m$$

其中 $m \le n$. 則 $g(Y)$ 也是一隨機變數, 且對於 $i = 1, 2, \ldots, m$, 其機率函數

$$
\begin{aligned}
P[g(Y) = g_i] &= \sum_{\substack{\text{所有 } y_j \text{ 使得} \\ g(y_j) = g_i}} p(y_j) \\
&\overset{\text{表成}}{=} p^*(g_i)
\end{aligned}
$$

根據期望值的定義,

$$
\begin{aligned}
E[g(Y)] &\overset{\text{def}}{=} \sum_{i=1}^{m} g_i \cdot p^*(g_i) \\
&= \sum_{i=1}^{m} g_i \cdot \{ \sum_{\substack{\text{所有 } y_j \text{ 使得} \\ g(y_j) = g_i}} p(y_j) \} \\
&= \sum_{i=1}^{m} \sum_{\substack{\text{所有 } y_j \text{ 使得} \\ g(y_j) = g_i}} g_i p(y_j) \\
&= \sum_{j=1}^{n} g(y_j) p(y_j)
\end{aligned}
$$

其中最後一個等號成立乃因爲

$$
\bigcup_{i=1}^{m} \{ y_j | g(y_j) = g_i \} = \{ y_1, y_2, \ldots, y_n \}
$$

且

$$g_i = g(y_j)$$

註. 求 $E[g(Y)]$ 時, 不需要根據定義先求出隨機變數 $g(Y)$ 的機率函數 (分布), 而可直接使用已知的 Y 的機率函數 $p(y)$ 以求得

$$E[g(Y)] = \sum_y g(y)p(y)$$

定義. 隨機變數 Y 的變異數

$$\sigma^2 \overset{\text{或}}{=} \mathrm{Var}(Y) \overset{\text{def}}{=} E[(Y-\mu)^2]$$

三個與期望值有關的常用定理, 如下述.

定理 3.3. 令 $Y \sim p(y)$ 且 c 爲一常數, 則

$$E(c) = c$$

<證> 定義常數函數 $g(Y) = c$, 則根據定理 3.2, 以及 $g(y) = c$,

$$E(c) = E(g(Y)) = \sum_y g(y)p(y) = \sum_y cp(y)$$

將 c 提出, 並根據 $\sum_y p(y) = 1$, 上式可化簡為

$$E(c) = c\sum_y p(y) = c \cdot 1 = c$$

定理 3.4. 設 $Y \sim p(y)$, $g(Y)$ 為一 Y 的函數且 c 為一常數, 則

$$E[cg(Y)] = cE[g(Y)]$$

<證> 由定理 3.2, 因為 $cg(Y)$ 亦為一 Y 的函數, 故

$$\begin{aligned} E[cg(Y)] &= \sum_y cg(y)p(y) \\ &= c\sum_y g(y)p(y) \\ &= cE[g(Y)] \end{aligned}$$

定理 3.5. 設 $Y \sim p(y)$ 且

$$g_1(Y), g_2(Y), \ldots, g_k(Y)$$

為 k 個 Y 的函數, 則

$$\begin{aligned} &E[g_1(Y) + g_2(Y) + \cdots + g_k(Y)] \\ &\quad = E[g_1(Y)] + E[g_2(Y)] + \cdots + E[g_k(Y)] \end{aligned}$$

<證> 僅證明 $k=2$ 的情況 (同理可證明任意 k 的情況). 因爲可視 $g_1(Y)+g_2(Y)$ 爲一 Y 的函數, 故根據定理 3.2,

$$\begin{aligned} & E[g_1(Y)+g_2(Y)] \\ &= \sum_y [g_1(y)+g_2(y)]p(y) \\ &= \sum_y g_1(y)p(y) + \sum_y g_2(y)p(y) \\ &= E[g_1(Y)] + E[g_2(Y)] \end{aligned}$$

註. 定理 3.4 與定理 3.5 合併稱作期望值的線性性質 (linearity), 亦即, 對任意 Y 的函數 $g_1(Y), g_2(Y), \ldots, g_k(Y)$ 及常數 $a_1, a_2, \ldots, a_k$,

$$\begin{aligned} & E[a_1 g_1(Y) + a_2 g_2(Y) + \cdots + a_k g_k(Y)] \\ &= a_1 E[g_1(Y)] + a_2 E[g_2(Y)] + \cdots \\ &\qquad + a_k E[g_k(Y)] \end{aligned}$$

定理 3.6. 設 $Y \sim p(y)$, 則

$$\begin{aligned} \mathrm{Var}(Y) &= \sigma^2 \\ &= E[(Y-\mu)^2] \\ &= E(Y^2) - [E(Y)]^2 \\ &\overset{\text{或}}{=} E(Y^2) - \mu^2 \end{aligned}$$

<證> 根據變異數的定義及期望值的線性性質,

$$\begin{aligned}
\sigma^2 &= E[(Y-\mu)^2] \\
&= E[Y^2 - 2Y\mu + \mu^2] \\
&= E(Y^2) - 2\mu E(Y) + \mu^2 \\
&\quad (\text{因為 } \mu,\ \mu^2 \text{ 均為常數}) \\
&= E(Y^2) - 2\mu\cdot\mu + \mu^2 \\
&= E(Y^2) - \mu^2
\end{aligned}$$

例 1. 設 $Y \sim \begin{array}{c|c} y & p(y) \\ \hline 0 & 1/8 \\ 1 & 1/4 \\ 2 & 3/8 \\ 3 & 1/4 \end{array}$. 試求 Y 的期望值 μ, 變異數 σ^2, 與標準差 σ.

<解> 根據期望值的定義,

$$\begin{aligned}
\mu &= E(Y) \\
&= \sum_y y p(y) \\
&= 0\cdot\frac{1}{8} + 1\cdot\frac{1}{4} + 2\cdot\frac{3}{8} + 3\cdot\frac{1}{4} \\
&= \frac{1+3+3}{4} \\
&= 1.75
\end{aligned}$$

根據變異數公式 (定理 3.6) 以及隨機變數的函數期望值公式 (定理 3.2),

$$\begin{aligned}\sigma^2 &= E(Y^2) - [E(Y)]^2 \\ &= \sum_y y^2 p(y) - \mu^2 \\ &= 0^2 \cdot \frac{1}{8} + 1^2 \cdot \frac{1}{4} + 2^2 \cdot \frac{3}{8} + 3^2 \cdot \frac{1}{4} - \left(\frac{7}{4}\right)^2 \\ &= \frac{1+6+9}{4} - \frac{49}{16} \\ &= \frac{15}{16} = .9375\end{aligned}$$

最後, 標準差

$$\sigma = \sqrt{\sigma^2} = \sqrt{.9375} = .97$$

例 2. 設機器 A 在一天運作 t 時下的維修次數爲一期望值與變異數均爲 0.1t 的隨機變數 Y_1, 且運作成本爲

$$C_A(t) = 10t + 30Y_1^2$$

另機器 B 在一天運作 t 時下的維修次數爲一期望值與變異數均爲 0.12t 的隨機變數 Y_2, 且運作成本爲

$$C_B(t) = 8t + 30Y_2^2$$

又設維修時間可忽略, 且每晚將機器調整以致於隔天機器的運作如新的一樣. 試問何者的每天平均成本較低, 若一工作天爲 (a) 10 小時及 (b) 20 小時?

<解> 機器 A 的每天平均運作成本爲

$$\begin{aligned} E[C_A(t)] &= E\left[10t + 30Y_1^2\right] \\ &= 10t + 30E\left(Y_1^2\right) \qquad (1) \end{aligned}$$

此乃因爲 $10t$, 30 爲常數. 又根據

$$\text{Var}(Y) = E(Y^2) - \mu^2$$

得

$$E(Y^2) = \text{Var}(Y) + \mu^2$$

故延續 (1) 式的計算得

$$\begin{aligned} E[C_A(t)] &= 10t + 30\left[\text{Var}(Y_1) + (E(Y_1))^2\right] \\ &= 10t + 30\left[0.1t + (0.1t)^2\right] \\ &= 13t + 0.3t^2 \end{aligned}$$

同理, 機器 B 的每天平均運作成本爲

$$\begin{aligned} E[C_B(t)] &= 8t + 30E\left(Y_2^2\right) \\ &= 8t + 30\left[\text{Var}(Y_2) + (E(Y_2))^2\right] \\ &= 8t + 30\left[0.12t + (0.12t)^2\right] \\ &= 11.6t + 0.432t^2 \end{aligned}$$

故, (a) 當一工作天爲 10 小時,

$$E[C_A(10)] = 13(10) + (0.3)(10)^2 = 160$$

且

$$\begin{aligned} E[C_B(10)] &= 11.6(10) + (0.432)(10)^2 \\ &= 159.2 \end{aligned}$$

得機器 B 較經濟 (選機器 B). 一個直觀的理由是, 在短期內, 機器 B 較低的成本會反應出效果.

(b) 當一工作天爲 20 小時,

$$E[C_A(20)] = 13(20) + (0.3)(20)^2 = 380$$

且

$$\begin{aligned} E[C_B(20)] &= 11.6(20) + (0.432)(20)^2 \\ &= 404.8 \end{aligned}$$

得機器 A 較經濟 (選機器 A). 一個直觀的理由爲, 在長期下, 機器 A 較小及少變異的維修次數會反應出效果.