單元 23: 柴比雪夫定理
(課本 §3.11)

提供隨機變數 $Y \in (\mu - k\sigma, \mu + k\sigma)$ 的機率的下界.

定理 3.14. (Chebyshev's Theorem). 設隨機變數 Y 的期望值爲 μ 且有限變異數爲 σ^2, 則對任意常數 $k > 0$,

$$P(|Y - \mu| < k\sigma) \geq 1 - \frac{1}{k^2}$$

或

$$\begin{aligned} P(|Y - \mu| \geq k\sigma) &= 1 - P(|Y - \mu| < k\sigma) \\ &\leq 1 - \left(1 - \frac{1}{k^2}\right) = \frac{1}{k^2} \end{aligned}$$

<證> (略), 延至 §4.10.

註 1. Chebyshev's Theorem 適用於任意的隨機分布, 不論機率直方圖是否爲鐘型 (bell-shaped).

注意: 經驗法則僅適用於直方圖爲鐘型的機率分布, 且彼此之間不矛盾.

註 2. Chebyshev's Theorem 的結論是相當保守的下界, 亦即 $Y \in (\mu - k\sigma, \mu + k\sigma)$ 的眞正機率經常超過下界 $1 - \frac{1}{k^2}$; 但對任一 $k > 1$, 均可製造出一對應的機率分布, 使其落入 $(\mu - k\sigma, \mu + k\sigma)$ (針對此一特定的 k) 的機率剛好就是下界 $1 - \frac{1}{k^2}$ (亦即, 眞實地達到下界).

例 1. 設經過長期觀察後, 每天的顧客數 Y 爲一期望值爲 20 且標準差 (standard deviation, 簡稱 std) 爲 2 的隨機變數 (但其機率分布未知). 試問明天的顧客數大於 16 且小於 24 的機率爲何?

<解> 因爲分布未知, 僅可採用 Chebyshev's Theorem. 又

$$\frac{24 - 16}{2} = 4 = 2\sigma$$

以及

$$\frac{16 + 24}{2} = 20 = \mu$$

故,

$$\begin{aligned}
\text{所求} &= P(16 < Y < 24) \\
&= P(20 - 2(2) < Y < 20 + 2(2)) \\
&\geq 1 - \frac{1}{2^2} \text{ (因爲 } k = 2) \\
&= \frac{3}{4}
\end{aligned}$$

註. 若 $\sigma = 1$, 則

$$\frac{24-16}{2} = 4 = 4\sigma$$

故,

$$\begin{aligned} P(16 < Y < 24) &\geq 1 - \frac{1}{4^2} \text{ (因爲 } k = 4) \\ &= \frac{15}{16} \end{aligned}$$

因此, σ 的值有相當的影響力 (效用).