

單元 32: 柴比雪夫定理

(課本 §4.10)

定理 4.13. (柴比雪夫定理). 設隨機變數 Y 的期望值 μ 及變異數 σ^2 均有限, 則對任意的 $k > 0$,

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

或

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

<證> 僅證明連續隨機變數的情況, 同理可證離散的情形. 設 $f(y)$ 為 Y 的 pdf, 則根據變異數的定義及隨機變數函數的期望值公式,

$$\begin{aligned}\text{Var}(Y) &= \sigma^2 \\ &= \int_{-\infty}^{\infty} (y - \mu)^2 f(y) dy \\ &= \int_{-\infty}^{\mu - k\sigma} (y - \mu)^2 f(y) dy + \\ &\quad \int_{\mu - k\sigma}^{\mu + k\sigma} (y - \mu)^2 f(y) dy + \\ &\quad \int_{\mu + k\sigma}^{\infty} (y - \mu)^2 f(y) dy \quad (1)\end{aligned}$$

因爲 (1) 式中中間積分的被積函數 $(y - \mu)^2 f(y) \geq 0$, 故中間積分大於或等於 0, 而得

$$\sigma^2 \geq \int_{-\infty}^{\mu - k\sigma} (y - \mu)^2 f(y) dy + \int_{\mu + k\sigma}^{\infty} (y - \mu)^2 f(y) dy \quad (2)$$

接著當 $y \leq \mu - k\sigma$ 時, $y - \mu \leq -k\sigma < 0$ 並導出

$$(y - \mu)^2 \geq k^2 \sigma^2$$

以及當 $y \geq \mu + k\sigma$ 時, $y - \mu > k\sigma > 0$ 亦導出

$$(y - \mu)^2 \geq k^2 \sigma^2$$

又因爲 (2) 式中兩項積分的範圍分別爲 $y \leq \mu - k\sigma$ 以及 $y \geq \mu + k\sigma$, 故在此積分範圍內, 此二項積分的被積函數

$$(y - \mu)^2 f(y) \geq k^2 \sigma^2 f(y)$$

而可導出

$$\begin{aligned} \sigma^2 &\geq \int_{-\infty}^{\mu - k\sigma} k^2 \sigma^2 f(y) dy + \int_{\mu + k\sigma}^{\infty} k^2 \sigma^2 f(y) dy \\ &= k^2 \sigma^2 \left[\int_{-\infty}^{\mu - k\sigma} f(y) dy + \int_{\mu + k\sigma}^{\infty} f(y) dy \right] \end{aligned}$$

最後根據一個重要的已知結論

$$P(a \leq Y \leq b) = \int_a^b f(y)dy$$

上式相當於

$$\begin{aligned}\sigma^2 &\geq k^2 \sigma^2 [P(Y \leq \mu - k\sigma) + P(Y \geq \mu + k\sigma)] \\ &= k^2 \sigma^2 P(|Y - \mu| \geq k\sigma)\end{aligned}$$

兩邊同除 $k^2 \sigma^2$, 得

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

亦相當於

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

例 1. 根據經驗, 某機器的維修時間

$$Y \sim \text{gamma}(3.1, 2)$$

若一新手需時 22.5 分才完成維修, 試問與過去的經驗相符嗎?

<解> 根據經驗以及 $\text{gamma}(\alpha, \beta)$ 隨機變數的性質, 期望值

$$\mu = \alpha\beta = (3.1)(2) = 6.2$$

且變異數

$$\sigma^2 = \alpha\beta^2 = (3.1)(4) = 12.4$$

故

$$\sigma = \sqrt{12.4} = 3.52$$

又

$$\frac{22.5 - 6.2}{3.52} = 4.63$$

亦即，此新手的維修時間 22.5 分是偏離期望值 4.63 個標準差。故根據柴比雪夫定理，得

$$\begin{aligned} P(Y \geq 22.5) &\leq P(|Y - \mu| \geq 4.63\sigma) \\ &\leq \frac{1}{(4.63)^2} = 0.0466 \end{aligned}$$

為一個小的值。因此，可推論新手的動作是慢了些。

註。 $P(Y \geq 22.5)$ 的真實值為

$$\int_{22.5}^{\infty} \frac{y^{2.1} e^{-y/2}}{2^{3.1} \Gamma(3.1)} dy$$

是一不易求得的值，但如上例，可透過柴比雪夫定理繞過繁雜的計算，而得到所要的推論。