單元 5: 二項分布的常態近似
(課本 §7.5)

若

$$Y \sim \text{binomial}(n, p)$$

則

$$Y = \sum_{i=1}^{n} X_i$$

其中

$$X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Bernoulli}(p)$$

且

$$E(X_i) = p \quad \text{以及} \quad \text{Var}(X_i) = pq,\ q = 1 - p$$

因此,

1. 由中央極限定理, 當 $n \to \infty$ 時,

$$\sqrt{n}\left(\frac{\frac{Y}{n} - p}{\sqrt{pq}}\right) = \frac{Y - np}{\sqrt{npq}} \xrightarrow{\mathcal{D}} N(0, 1)$$

由此導出,

$$Y \approx N(np, npq)$$

2. 以常態分布隨機變數近似二項分布隨機變數的準則.
令隨機變數 $W \sim N(np, npq)$.

首先, 由圖示,

$$
\begin{aligned}
&P(Y \le 3) \\
&\quad = \text{由 0 至 3, 四個 } Y \text{ 的 pdf 長方條面積和} \\
&\quad \approx P(W \le 3.5) \\
&\quad = P\left(\frac{W - np}{\sqrt{npq}} \le \frac{3.5 - np}{\sqrt{npq}}\right) \\
&\quad = P\left(Z \le \frac{3.5 - np}{\sqrt{npq}}\right)
\end{aligned}
$$

一個較好的估計, 因為取 $W \le 3$, 少了最後的半條長方條, 會導致較大的誤差, 如圖示.

對於整數 $a < b$,

$$
\begin{aligned}
&P(a \le Y \le b) \\
&\quad = \text{由 } a \text{ 至 } b \text{ 的 } Y \text{ 的 pdf 長方條面積和} \\
&\quad \approx P\left(a - \frac{1}{2} \le W \le b + \frac{1}{2}\right) \\
&\quad = P\left(\frac{a - \frac{1}{2} - np}{\sqrt{npq}} \le Z \le \frac{b + \frac{1}{2} - np}{\sqrt{npq}}\right)
\end{aligned}
$$

一個較好的估計, 因為若取 $a \le W \le b$, 則少了頭與尾的兩個半條長方條, 如圖示.

對於整數 c,

$$\begin{aligned}
&P(Y=c)\\
&= \ Y \text{ 在 } c \text{ 的 pdf 長方條面積}\\
&\approx \ P\left(c-\frac{1}{2}\le W\le c+\frac{1}{2}\right)\\
&= \ P\left(\frac{c-\frac{1}{2}-np}{\sqrt{npq}}\le Z\le\frac{c+\frac{1}{2}-np}{\sqrt{npq}}\right)
\end{aligned}$$

一個較好的估計, 因爲 $P(W=c)=0$, 明顯不對, 故需前後各延伸 $\frac{1}{2}$.