選舉問題 (Election Problem)

- 某日電視台進行選舉民調，在某州進行民調樣本數是 2000 人，結果顯示 52% 的人支持候選人 A 並容許 3% 的誤差，換句話說，即大約有 49%~55% 的選民支持候選人 A
  - 這樣的結果可信嗎？
  - 要調查的人數是如何決定的？
- 設 p = 州內支持候選人 A 的比率，即 p = 任意找一個投票人他是支持 A 的機率
- 設 n 是母體的樣本數，同時 X = 樣本中支持 A 的人數

選舉問題 (Election Problem) (Cont.)

- 因此， X/n = 樣本中支持 A 的比率
- 民調專家想要用樣本比率 X/n 來估計全面性的比率 p ，問題來了，多大的 n 可以讓人們覺得這樣的民調是可信的？
  - 100% 可信 → 調查州裡的每位投票人即
  - 可接受誤差 (例：誤差正負 3%) → $\frac{X}{n}-0.03 \le p \le \frac{X}{n}+0.03$
  - 換另外一種想法 → 我想要看到至少 95% 可信程度，即以 95% 求出應有的 n

$$P\left(\frac{X}{n}-0.03 \le p \le \frac{X}{n}+0.03\right) \ge 0.95$$

選舉問題 (Election Problem) (Cont.)

- 也就是說，你要找到一個合適的 n 以使得你至少有 95% 的可信度確定，調查出來的支持 A 的百分比在誤差 3% 的情況是多少
- 因此可將其表示為一個區間 $\left[\frac{X}{n}-0.03, \frac{X}{n}+0.03\right]$
- 這樣的區間被稱之為**信賴區間** (Confidence interval)
- 95% 則被稱之為**信賴水準** (Confidence level)
- 信賴區間小且信賴水準高 $\rightarrow$ n 要越大
- $P(np-0.03n \le X \le np+0.03n) \ge 0.95$
- 實際要解的式子是 $P(np-0.03n \le X \le np+0.03n) = 0.95$

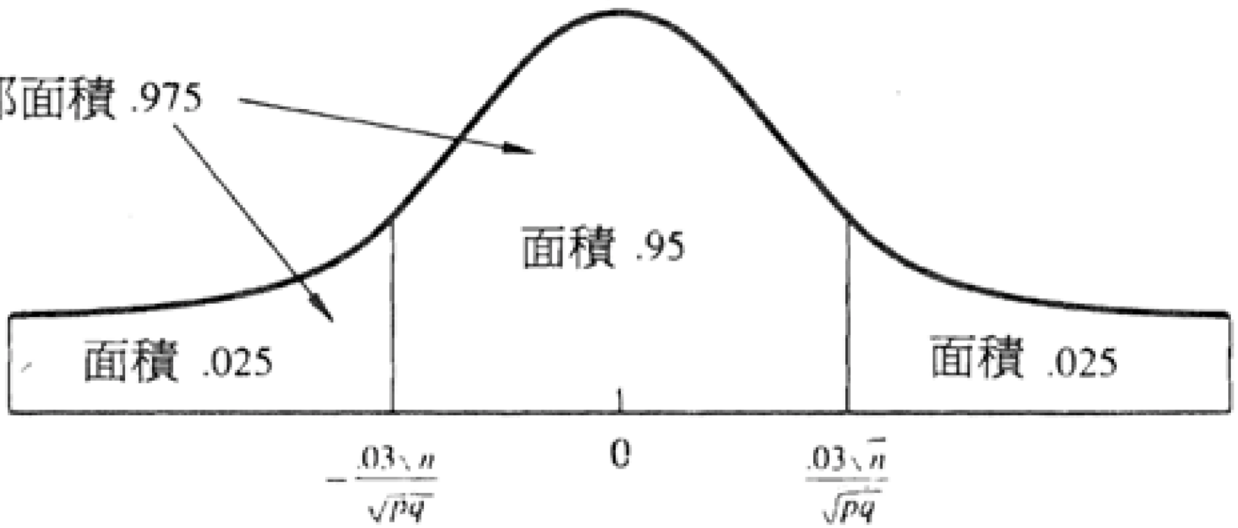
選舉問題 (Election Problem) (Cont.)

- 在大母體中進行不放回取樣 $\rightarrow$ 獨立試驗
- X 是具有參數 n 與 p 的二項式分佈
- 當 n 很大時 $\rightarrow$ 二項式近似常態分佈，所以 X 是近似常態且具參數 $\mu = np, \sigma^2 = npq$
- 將式子 $P(np - 0.03n \le X \le np + 0.03n) = 0.95$ 減去 μ 並除以 σ
- 得到 $$P\left(-\frac{0.03n}{\sqrt{npq}} \le \frac{X-\mu}{\sigma} \le \frac{0.03n}{\sqrt{npq}}\right) = 0.95$$

$$P\left(-\frac{0.03\sqrt{n}}{\sqrt{pq}} \le X^* \le \frac{0.03\sqrt{n}}{\sqrt{pq}}\right) = 0.95$$

選舉問題 (Election Problem) (Cont.)

- 我們知道，在單位常態下方面積是 95% 是介於 $-0.03\sqrt{n}/\sqrt{pq}$ 與 $0.03\sqrt{n}/\sqrt{pq}$ 之間
- 則面積的 97.5% 是落在 $0.03\sqrt{n}/\sqrt{pq}$ 左邊
- 從常態列表可查出 $F^*(1.96)=0.975$



選舉問題 (Election Problem) (Cont.)

- 所以 $\frac{0.03\sqrt{n}}{\sqrt{pq}}=1.96 \longrightarrow \sqrt{n}=\frac{1.96\sqrt{pq}}{0.03} \longrightarrow n=4268\,pq$
- 但我們不知 p 或 q，我們以最差的情況來計算，即當 $pq = ¼$ (即 $p = ½$ 與 $q = ½$)，則

$$n \geq 4268 \times \frac{1}{4} = 1067$$

- 使用這樣的 n 做民調可得到 95% 的可信度