

# 統計裡的信賴

黃文璋

## 1. 前言

民國94年10月9日聯合報有一則引人注意的新聞：

最近高捷案與中鋼董事長年薪等事件嚴重打擊民進黨形象，陳總統下重手要求調查，引起外界對府院關係緊張的揣測。陳總統昨天公開說，「相信」謝院長，謝長廷則在澎湖隔空「感恩」，表示他「謝謝總統的信賴。」

高雄捷運 BOT 案，因泰勞暴動事件，而揭露一連串的問題，檢調單位調查其中的弊案，陳水扁總統亦下令代理高雄市長的葉菊蘭，成立高捷案調查小組，將行政院長謝長廷也列為調查對象。有些人解讀謝長廷已失去陳水扁對他的“政治信任”。上一則新聞就是這樣來的。

我們常提到相信、信心、信任、信賴等。在不同的情境下，有不同的涵義，像是信任還有政治信任。只是政壇上曾肝膽相照或情同父子，最後形同陌路者屢見不鮮。所以政治人物在說相信或信賴時，內心中究竟有多相信或多信賴，是很令人存疑的。至於風雨生信心，這個信心又有多大呢？新約聖經約翰一書第五章第4節，還有“使我們勝了世界的，就是我們的信心。”我國行政法中也有“信賴保護原則”。這些我們不擬討論，我們有興趣的是統計學裡的意義。統計學裡有信賴區間(confidence interval)及信心水準(confidence level)。信心水準又稱信賴水準，或信賴係數(confidence coefficient)。底下為二常可見到的例子。

1. 區間 (2.8, 3.2) 是參數  $\mu$  的 95% 信賴區間。
2. 這次調查於10月7日至8日晚間舉行，成功訪問了 1,068 位成人，另492人拒訪。在百分之九十五的信心水準下，抽樣誤差在正負三個百分點以內。調查是以台灣地區住宅電話為母體，作尾數兩位隨機抽樣。

對於上述二段中的信賴區間及信心水準，大家是否了解其意義呢？信賴區間與信心水準的解讀，這是預計民國95年開始實施的“普通高級中學必修科目「數學」課程綱要”中，第二學年“機率與統計 (I)”主題裡，所包含的內容。舊的課程綱要中，並無此主題。顯然負責制訂課程綱要的數學專家們，體會到此主題在現今社會的重要，認為高中生也該熟悉，才將此主題加入新

的數學課程綱要中。屬於高中的題材，理應不太深奧，我們也的確常在報章雜誌上看到此二名詞。在前述綱要“實施方法”中的“教材編選”，列出一些一般性的建議，如“教材應注意到銜接、統整和連結”等，只有對此主題特別寫著：

機率與統計 (I) 新增信賴區間與信心水準的解讀一節，其相關的教學活動建議由全班每一位同學各自以亂數表模擬丟銅板的過程，代入銅板正面機率信賴區間的算式  $\sqrt{\frac{p(1-p)}{n}}$  來得到各自所得的信賴區間，並察覺大多數同學所得的信賴區間會涵蓋銅板正面機率的真實值。

我們引用上段文字，一字不改，連標點符號皆未更動。課程綱要是給編寫教科書的作者看，依據此綱要，才知道教科書中，要包含那些題材，以及編寫時該注意那些事項。那上述這段對作者的建議，是否講得清楚明白呢？對此，我們的信心水準可能不是很大。

在充滿數據的今日，處處要用到統計，統計學中的很多名詞，早已常出現在生活中。在數學裡，傳統的代數及幾何，皆已有兩千多年的歷史，很多名詞都已千錘百鍊，因此較易為人所接受。但統計學發展的時日尚不是很久，現代統計學的發展，大約只有100年。時至今日，很多數學的名詞，其意義大家已很清楚，望辭生義，並早就融入我們的語言中。如平行、垂直、比例、拋物線、黃金分割等。又像是最大公約數，近年來為許多政治人物所愛引用，如“中華民國是朝野的最大公約數”。而統計學則是從原有的語言中借名詞，比如估計、充分性、顯著性、假設檢定、信賴、信心等皆是。語言中不是很清晰的概念，自然不易對應到統計學中，有嚴格定義的名詞。甚至連極基本的期望值及變異數，雖期望、變異皆是很普通的用語，但有不少人學完一年統計學後，對期望值及變異數的意義，仍是一知半解。所以雖然信賴區間及信心水準，出現於很多統計分析的報告中，且因民意調查的普遍，此二名詞更常出現於報章媒體，但真正了解其意義的人，恐怕少之又少。本文便是要闡釋統計學裡信賴及信心的涵義。

## 2. 信賴區間

我們常在做估計，估計這個綠燈是否過得去，估計這次考試得幾分，估計追上這個女孩子的機率。比較嚴肅一點的，則有估計明年的經濟成長率，估計高雄捷運10年後，每日載客量等。

估計的方法很多，以銅板出現正面的機率  $p$  為例。直觀上投擲  $n$  次，假設出現  $k$  次正面，則以  $k/n$  來估計  $p$ 。對涉及可以重複觀測的實驗，如估計某種新藥之治癒率，估計袋中白球所佔比率，估計選民對某一候選人之支持率等，都可採用此簡單的方法。只是在很多情況下，觀測  $n$  次，不見得能做到像投擲銅板一樣，結果是獨立且有共同分佈(independent and identically distributed, 簡稱 iid)。必須要有好的實驗設計，收集到的數據，品質才會較佳。統計學裡已證明，前述以  $k/n$  估計  $p$ ，雖然簡單，但在某種評比下卻為“最佳”(這是一致最小變異不偏估計

量(uniformly minimum variance unbiased estimator, 簡稱 UMVUE))。雖是最佳, 有時我們對只給出一估計值, 亦即只給出點估計量(point estimator), 並不太滿意, 因這通常不是  $p$  的真實值。或者說我們認為點估計量恰好是真實值的機率, 通常是微乎其微。退而求其次, 我們想知道  $p$  大約是多少, 即  $p$  大約落在那一範圍。例如, 假設投擲一銅板 100 次, 得 58 個正面, 因此以 0.58 來估計銅板出現正面的機率  $p$ 。現有某人投擲, 假設得 59 個正面,  $k/n = 0.59$ , 他有些懷疑。0.58, 0.59 差不多, 你這樣說明。那 60 次正面呢? 54 次呢? 因投擲不見得就會剛好得到  $k/n = 0.58$ , 所以有時你會覺得, 與其只給個估計值, 還不如給一  $p$  可能會落在的範圍, 說不定還好些。再看一例。對一只有二候選人之選舉, 若其中候選人 A 民調結果之支持率, 領先候選人 B 1.2%, 候選人 A 是否就放心了呢? 可能未必。調查會有誤差, 甚至民意也會變, 1.2% 的領先不見得夠安全, 因此給出支持率大致的範圍, 也就是 區間估計量(interval estimator), 此處應是必要的。

常態分佈是一很重要的分佈, 在很多不同的隨機現象中, 可描述其中之機率模型。符號  $X \sim \mathcal{N}(\mu, \sigma^2)$ , 即表隨機變數  $X$  有期望值  $\mu$ , 及變異數  $\sigma^2$  的常態分佈, 其中  $\mu \in R$ ,  $\sigma > 0$ 。  $X$  之機率密度函數為

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}, \quad x \in R.$$

我們列出一些常態分佈的性質如下:

- (i) 設  $X \sim \mathcal{N}(\mu, \sigma^2)$ 。令  $a, b$  為二實常數且  $a \neq 0$ , 則  $aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$ 。
- (ii) 設  $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ ,  $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$ , 且  $X$  與  $Y$  獨立, 則  $X + Y \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$ 。
- (iii) 設  $X \sim \mathcal{N}(\mu, \sigma^2)$ , 則  $Z = (X - \mu)/\sigma \sim \mathcal{N}(0, 1)$ 。  $\mathcal{N}(0, 1)$  即標準常態分佈, 其機率密度函數之圖形如圖 1。

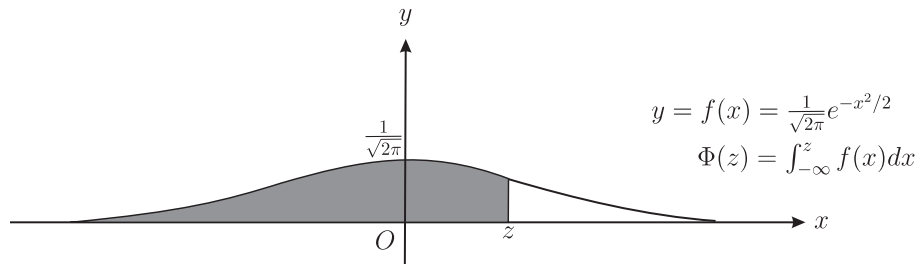


圖 1.  $\mathcal{N}(0, 1)$  分佈的機率密度函數圖形, 斜線部分的面積 =  $\Phi(z)$ ,  $z \in R$ 。

性質 (i) 是說經過線性變換, 原來的常態分佈還是保有。

性質 (ii) 也是常態分佈處處可見的原因之一, 因二獨立的常態隨機變數之和, 仍有常態分佈。

由於性質 (iii), 使得常態分佈的機率值表, 僅需給  $\mathcal{N}(0, 1)$  的便夠了。因若  $X \sim \mathcal{N}(\mu, \sigma^2)$ , 則

$$P(X \leq x) = P\left(\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right) = P\left(Z \leq \frac{x - \mu}{\sigma}\right).$$

其中  $Z$  有  $\mathcal{N}(0, 1)$  分佈。所以只要給出  $\Phi(z)$  的機率值表, 便可求出  $P(X \leq x)$  之值, 其中

$$\Phi(z) = P(Z \leq z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx, \quad z \in R. \quad (1)$$

由上式且利用圖 1 中,  $y = f(x)$  的圖形對稱於  $y$  軸, 可得當  $Z \sim \mathcal{N}(0, 1)$ ,

$$P(|Z| \leq z) = 2\Phi(z) - 1, \quad z > 0. \quad (2)$$

當然我們不會忘記中央極限定理(central limit theorem)。對  $\forall n \geq 1$ , 設  $X_1, \dots, X_n$  為 iid 之隨機變數, 又設  $\mu = E(X_1)$ ,  $\sigma = \sqrt{\text{Var}(X_1)} > 0$  皆存在。令  $\bar{X}_n = \sum_{i=1}^n X_i/n$ ,  $n \geq 1$ , 表樣本平均(sample mean)。則

$$\lim_{n \rightarrow \infty} P\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq z\right) = \Phi(z), \quad z \in R. \quad (3)$$

上式指出  $n$  很大時,  $(\bar{X}_n - \mu)/(\sigma/\sqrt{n})$  有近似的  $\mathcal{N}(0, 1)$  分佈。由 (3) 式又得

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| \leq z\sigma/\sqrt{n}) = 2\Phi(z) - 1, \quad z > 0. \quad (4)$$

假設  $X_1, \dots, X_n$  為 iid 之隨機變數, 以  $\mathcal{N}(\mu, \sigma^2)$  分佈為其共同分佈, 設  $\sigma^2$  已知。由常態分佈的性質 (i) 及 (ii), 可得  $\bar{X}_n \sim \mathcal{N}(\mu, \sigma^2/n)$ 。即樣本平均仍有常態分佈, 期望值不變, 變異數變小, 為原來的  $1/n$  倍。 $\bar{X}_n$  為  $\mu$  之一很好的點估計量, 但  $\mu$  之區間估計量呢? 來看底下的討論。

由性質 (iii),  $(\bar{X}_n - \mu)/(\sigma/\sqrt{n}) \sim \mathcal{N}(0, 1)$ 。又由 (2) 式,

$$P\left(\left|\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}\right| \leq z\right) = 2\Phi(z) - 1, \quad z > 0. \quad (5)$$

由上式又得

$$P(\bar{X}_n - \sigma z/\sqrt{n} \leq \mu \leq \bar{X}_n + \sigma z/\sqrt{n}) = 2\Phi(z) - 1, \quad z > 0. \quad (6)$$

上式等價於

$$P(\mu \in [\bar{X}_n - \sigma z/\sqrt{n}, \bar{X}_n + \sigma z/\sqrt{n}]) = 2\Phi(z) - 1, \quad z > 0. \quad (7)$$

我們便藉 (7) 式以求出  $\mu$  之一信賴區間。

首先對  $\forall 0 < y < 1$ , 令  $z_y$  滿足

$$\Phi(z_y) = P(Z \leq z_y) = y, \quad (8)$$

其中  $Z \sim \mathcal{N}(0, 1)$ 。由於  $\Phi(z)$  為嚴格漸增函數, 且  $\lim_{z \rightarrow -\infty} \Phi(z) = 0$ ,  $\lim_{z \rightarrow \infty} \Phi(z) = 1$ , 所以對  $\forall 0 < y < 1$ ,  $z_y$  唯一存在。若  $\Phi(x) = y$ , 則  $x = z_y$ 。即  $x = z_y$  為  $\Phi(x) = y$  之反函數。在圖 1 中, 若斜線部分的面積為  $y$ , 則有唯一的  $z_y$  與其對應。

給定一  $\alpha > 0$ ,  $\alpha$  通常是一較小的數, 我們想給  $\mu$  之一  $1 - \alpha$  信賴區間。令  $2\Phi(z) - 1 = 1 - \alpha$ , 即  $\Phi(z) = 1 - \alpha/2$ , 故得  $z = z_{1-\alpha/2}$ 。因此

$$\mu \in [\bar{X}_n - \sigma z_{1-\alpha/2}/\sqrt{n}, \bar{X}_n + \sigma z_{1-\alpha/2}/\sqrt{n}]$$

之機率為  $1 - \alpha$ 。區間

$$I = [\bar{X}_n - \sigma z_{1-\alpha/2}/\sqrt{n}, \bar{X}_n + \sigma z_{1-\alpha/2}/\sqrt{n}] \quad (9)$$

便為  $\mu$  之一信心水準為  $1 - \alpha$  之信賴區間, 或只簡單地說  $1 - \alpha$  信賴區間, 或用百分比, 稱為  $\mu$  之一  $100(1 - \alpha)\%$  信賴區間。當  $\alpha = 0.1, 0.05$  及  $0.01$  時 (這是幾個常取的  $\alpha$  值), 即得 90% 信賴區間, 95% 信賴區間, 99% 信賴區間。此時  $z_{1-\alpha/2}$  之近似值分別為  $z_{0.95} \doteq 1.645$ ,  $z_{0.975} \doteq 1.960$ , 及  $z_{0.995} \doteq 2.576$ 。

在 (6) 式中, 依序取  $z = 1, 2, 3$ , 則有

$$P(\bar{X}_n - \sigma/\sqrt{n} \leq \mu \leq \bar{X}_n + \sigma/\sqrt{n}) \doteq 0.6826,$$

$$P(\bar{X}_n - 2\sigma/\sqrt{n} \leq \mu \leq \bar{X}_n + 2\sigma/\sqrt{n}) \doteq 0.9545,$$

$$P(\bar{X}_n - 3\sigma/\sqrt{n} \leq \mu \leq \bar{X}_n + 3\sigma/\sqrt{n}) \doteq 0.9974.$$

由於  $\bar{X}_n$  之標準差為  $\sqrt{\sigma^2/n} = \sigma/\sqrt{n}$ ,  $[\bar{X}_n - \sigma/\sqrt{n}, \bar{X}_n + \sigma/\sqrt{n}]$ ,  $[\bar{X}_n - 2\sigma/\sqrt{n}, \bar{X}_n + 2\sigma/\sqrt{n}]$ , 及  $[\bar{X}_n - 3\sigma/\sqrt{n}, \bar{X}_n + 3\sigma/\sqrt{n}]$  等三區間, 分別為距  $\bar{X}_n$  不超過一個標準差, 兩個標準差, 及三個標準差之信賴區間, 信心水準分別為 68.26%, 95.45%, 99.74%。在高中數學綱要中, 對“信賴區間與信心水準的解讀”之說明為

常態分配及 68 - 95 - 99.7 規律。僅需處理二元資料, 不必引進機率模型, 以教學活動瞭解信賴區間與信心水準的解讀。

其中提及的“68 - 95 - 99.7 規律”就來自於此。即若  $X \sim \mathcal{N}(\mu, \sigma^2)$ , 則當  $i = 1, 2, 3$ ,  $P(|X - \mu| \leq i\sigma)$  分別約等於 0.68, 0.95 及 0.997。只是“規律”二字的涵義卻不知為何。因為此三數字加上百分比, 便不過是機率, 而非什麼規律。又所謂 68 - 95 - 99.7, 也只適用常態分佈 (或最多

加上近似的常態分佈), 其他分佈通常不適用。例如, 設  $X$  有指數分佈, 參數為  $\lambda$ , 其中  $\lambda > 0$ 。即  $X$  之機率密度函數為  $f(x) = \lambda e^{-\lambda x}$ ,  $x > 0$ 。又  $X$  之期望值  $\mu = E(X) = 1/\lambda$ , 且標準差  $\sigma = \sqrt{\text{Var}(X)} = 1/\lambda$ 。則當  $i = 1, 2, 3$ ,  $P(|X - \mu| \leq i\sigma)$  分別等於

$$\begin{aligned}\int_0^{2/\lambda} \lambda e^{-\lambda x} dx &= 1 - e^{-2} \doteq 0.8647, \\ \int_0^{3/\lambda} \lambda e^{-\lambda x} dx &= 1 - e^{-3} \doteq 0.9502, \\ \int_0^{4/\lambda} \lambda e^{-\lambda x} dx &= 1 - e^{-4} \doteq 0.9817.\end{aligned}$$

即得到“86 – 95 – 98”, 而非“68 – 95 – 99.7”。

你會不會想問, 為何不取  $\alpha = 0$ , 而得一 100% 的信賴區間? 百分之百的信賴, 不是最好嗎? 由於  $z_0 = -\infty$ ,  $z_1 = \infty$ , 代入 (9) 式, 得  $\mu$  之百分之百的信賴區間為  $(-\infty, \infty)$ , 即整個實數集合。這個區間大的變成毫無意義。這也足以說明,  $\alpha$  過小, 如 0.0001 等, 不見得就是那麼適合。一般而言, 信心水準太高, 信賴區間往往便過長, 反失去估計的目的。所以我們通常不會追求過高的信心水準, 更不會有百分之百的信賴區間。

底下給一簡單的例子。

例 1: 某工廠生產某種飲料, 假設飲料的容量有 (單位為 cc) 常態分佈, 標準差為 1.2cc。從某日的產品隨機抽取 10 瓶, 容量分別為 330.1, 329.9, 330.9, 331.2, 329.1, 330.8, 328.9, 329.5, 330.2, 329.8。試給出容量期望值  $\mu$  之一 95% 信賴區間。

解: 在此  $X_1, \dots, X_{10}$  為 iid 之  $\mathcal{N}(\mu, 1.2^2)$  隨機變數。因  $\bar{X}_n = 330.04$ ,  $\alpha = 0.05$ , 由 (9) 式得  $\mu$  之一 95% 信賴區間為

$$\begin{aligned}& [330.04 - 1.2 \cdot 1.96/\sqrt{10}, 330.04 + 1.2 \cdot 1.96/\sqrt{10}] \\ & \doteq [329.27, 330.78].\end{aligned}$$

讀者是否已看出, 前述對常態分佈的討論中, 給定一  $\alpha$ , 其中  $0 < \alpha < 1$ , 則  $1 - \alpha$  的信賴區間並不唯一。諸如

$$\begin{aligned}& [\bar{X}_n - \sigma z_{1-\alpha/3}/\sqrt{n}, \bar{X}_n + \sigma z_{1-2\alpha/3}/\sqrt{n}], \\ & [\bar{X}_n - \sigma z_{1-2\alpha/5}/\sqrt{n}, \bar{X}_n + \sigma z_{1-3\alpha/5}/\sqrt{n}],\end{aligned}$$

甚至對任意  $\alpha_1, \alpha_2 > 0$ , 只要  $\alpha_1 + \alpha_2 = \alpha$ , 則

$$[\bar{X}_n - \sigma z_{1-\alpha_2}/\sqrt{n}, \bar{X}_n + \sigma z_{1-\alpha_1}/\sqrt{n}],$$

皆是。不過可以證明(如果你的幾何知識夠好,由圖1便可看出),在同樣的 $\sigma$ ,  $n$ 及 $\alpha$ 之下,(9)式給的 $1 - \alpha$ 信賴區間為最短。而在同樣的信心水準下,我們傾向取區間長度較短者,愈短表估計愈精準。

在 $\mathcal{N}(\mu, \sigma^2)$ 分佈中,若 $\sigma^2$ 未知,也有辦法求 $\mu$ 之信賴區間,但需以樣本變異數(sample variance)

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

取代 $\sigma^2$ 。另外,當然也可求 $\sigma^2$ 之信賴區間,也是分 $\mu$ 已知或 $\mu$ 未知兩情況。還有單邊信賴區間。分佈不是常態,也可對其中的參數給信賴區間。一般數理統計的書,在這方面都有很多討論,在此不贅言。

不是常態分佈,參數的信賴區間,有時較不易求得。不過如果樣本數夠多,便可藉助中央極限定理,以求取參數之信賴區間,見下列。

例2: 設 $X_1, \dots, X_n$ 為iid, 且

$$P(X_1 = 1) = p, P(X_1 = 0) = 1 - p, 0 < p < 1.$$

隨機變數只取0, 1兩個值,稱為有伯努力分佈(Bernoulli distribution)。因 $E(X_1) = p$ ,  $\text{Var}(X_1) = p(1 - p)$ , 由中央極限定理,  $n \rightarrow \infty$ 時,

$$P\left(\frac{\bar{X}_n - p}{\sqrt{p(1-p)}/\sqrt{n}} \leq z\right) \rightarrow \Phi(z), z \in R. \quad (10)$$

因此 $n$ 較大時,

$$P\left(\left|\frac{\sqrt{n}(\bar{X}_n - p)}{\sqrt{p(1-p)}}\right| \leq z_{1-\alpha/2}\right) \doteq 1 - \alpha. \quad (11)$$

即得 $p$ 之一近似的 $1 - \alpha$ 信賴區間

$$[L, U], \quad (12)$$

其中

$$L = \frac{2n\bar{X}_n + z_{1-\alpha/2}^2 - z_{1-\alpha/2}(4n\bar{X}_n(1 - \bar{X}_n) + z_{1-\alpha/2}^2)^{1/2}}{2(n + z_{1-\alpha/2}^2)},$$

$$U = \frac{2n\bar{X}_n + z_{1-\alpha/2}^2 + z_{1-\alpha/2}(4n\bar{X}_n(1 - \bar{X}_n) + z_{1-\alpha/2}^2)^{1/2}}{2(n + z_{1-\alpha/2}^2)}.$$

雖然是利用中央極限定理, $n$ 要很大才行。實際應用時,往往只要 $np$ 及 $n(1 - p)$ 皆大於5,上述 $[L, U]$ 便給出一還不錯的 $p$ 之近似的 $1 - \alpha$ 信賴區間。你可能會問, $p$ 不是未知嗎? 那

如何確定  $np$  及  $n(1-p)$  是否大於 5? 要知有時我們會對  $p$  有些了解, 如知道  $0.3 \leq p \leq 0.7$ , 這時  $n \geq 17$  便可以了。

前述信賴區間有些複雜, 既然是近似, 不妨用更簡單的形式。首先由弱大數法則(weak law of large numbers),  $n \rightarrow \infty$  時, 樣本平均  $\bar{X}_n$  會機率收斂(converges in probability) 至  $E(X_1) = p$ 。即  $n \rightarrow \infty$  時, 對  $\forall \varepsilon > 0$ ,

$$P(|\bar{X}_n - p| \leq \varepsilon) \longrightarrow 1. \quad (13)$$

因此利用機率論裡的結果,  $n$  很大時, (11) 式中的  $p(1-p)$  可以  $\bar{X}_n(1-\bar{X}_n)$  取代, 而得

$$P\left(\left|\frac{\sqrt{n}(\bar{X}_n - p)}{\sqrt{\bar{X}_n(1-\bar{X}_n)}}\right| \leq z_{1-\alpha/2}\right) \doteq 1 - \alpha. \quad (14)$$

由此求出  $p$  之近似的  $1 - \alpha$  信賴區間

$$\left[\bar{X}_n - z_{1-\alpha/2}\sqrt{\bar{X}_n(1-\bar{X}_n)/n}, \bar{X}_n + z_{1-\alpha/2}\sqrt{\bar{X}_n(1-\bar{X}_n)/n}\right], \quad (15)$$

計算上較簡易。也可寫成  $\bar{X}_n \pm z_{1-\alpha/2}\sqrt{\bar{X}_n(1-\bar{X}_n)/n}$  為  $p$  之一近似的  $1 - \alpha$  信賴區間。對照前言所引高中數學實施方法那一段, 我們猜想“信賴區間的算式  $\sqrt{p(1-p)/n}$ ”就是指 (15) 式中的  $\sqrt{\bar{X}_n(1-\bar{X}_n)/n}$ 。不過 (12) 式或 (15) 式, 才真正是近似的所謂“信賴區間的算式”。還必須一提,  $p$  為一未知的參數, 要給  $p$  的信賴區間, 其中怎能含有  $p$ ?

舉例來比較。設  $n = 400$ ,  $\bar{X}_n = 0.29$ , 且  $\alpha = 0.05$ 。則利用 (12) 及 (15) 式所得之 0.95 信賴區間, 分別為  $[0.2477, 0.3363]$ , 及  $[0.2455, 0.3345]$ , 差別並不太大。

對於本例, 我們亦可事先決定樣本數  $n$ , 以達到所需之精確度。在有關對某候選人之支持率, 對行政院某部會之施政滿意度, 或對某產品之喜愛度的調查, 常會產生伯努力分佈。此時對所欲估計的  $p$  (表支持率、滿意度等), 會得到一如 (12) 或 (15) 式之  $p$  的  $1 - \alpha$  信賴區間。我們當然希望此區間長度短些, 如此估計才夠精準。這時就有賴提高樣本數  $n$  了。

設採用 (15) 式當做  $p$  之一近似的  $1 - \alpha$  信賴區間, 且希望此區間的半徑, 或說估計之誤差, 不超過  $d$ 。即要有

$$z_{1-\alpha/2}\sqrt{\bar{X}_n(1-\bar{X}_n)/n} \leq d. \quad (16)$$

因  $0 \leq \bar{X}_n \leq 1$ , 故  $\bar{X}_n(1-\bar{X}_n) \leq 1/4$ , 因此使 (16) 式成立之一充分條件為

$$\frac{z_{1-\alpha/2}^2}{4n} \leq d^2.$$

即得樣本數之一下界:

$$n \geq \frac{z_{1-\alpha/2}^2}{4d^2}. \quad (17)$$



若取  $\alpha = 0.05$ ,  $d = 0.03$ , 即得

$$n \geq \frac{z_{0.975}^2}{4 \cdot 0.03^2} \doteq \frac{1.96^2}{0.0036} \doteq 1,067.1.$$

即需有  $n \geq 1,068$ 。一般在做民調時, 若在百分之九十五信心水準下, 抽樣誤差要在正負三個百分點以內, 則有效樣本數至少要有 1,068, 這個 1,068 就是這樣算出來的。實際在調查時, 由於會有拒訪等失敗的樣本, 所以一開始設定的抽樣數要多些。但最後成功的樣本數當然也不一定剛好是 1,068。假設最後成功的樣本數為  $n$ , 則由 (15) 式, 抽樣誤差 (即區間半徑)

$$d = z_{1-\alpha/2} \sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}} \leq \frac{z_{1-\alpha/2}}{2\sqrt{n}}. \quad (18)$$

例如, 設  $\alpha = 0.05$ ,  $n = 900$ , 因  $z_{0.975} \doteq 1.96$ , 得  $d \leq 1.96/60 \doteq 0.0327$ 。此時可以說在百分之九十五信心水準下, 抽樣誤差在正負 3.27 個百分點以內。

由 (17) 式可看出, 若想減小抽樣誤差, 則樣本數會快速增加。例如, 在  $\alpha = 0.05$  下, 若要  $d = 0.01$ , 則

$$n \geq \frac{z_{0.975}^2}{4 \cdot 0.01^2} \doteq \frac{1.96^2}{0.0004} = 9,604.$$

所需之樣本數為  $d = 0.03$  時之 9 倍。一般而言, 抽樣誤差要降為  $1/k$  倍, 所需之樣本數要成為  $k^2$  倍。反之, 若樣本數增為  $k$  倍, 則抽樣誤差僅降為  $1/\sqrt{k}$  倍。若要提高信心水準呢? 例如, 設  $\alpha$  由 0.05 降為 0.01,  $d$  仍設為 0.03, 則因  $z_{0.995} \doteq 2.576$ , 故所需之樣本數至少要有

$$n = \frac{2.576^2}{4 \cdot 0.03^2} \doteq 1,843.3,$$

即至少要有樣本數 1,844。

在估計時, 給出信賴區間是常有必要的。可以這麼說, 點估計就像醫生的鐵口直斷: 醫生向病人家屬說, 病人只能再活一年。實際的情況是, 可能 8 個月後病人便死了, 也可能又活了 2 年 6 個月。大約少有剛好一年後死的。若能給一區間, 如 6 個月至 3 年, 雖不再是那麼明確, 使醫生好像不夠權威, 但換得一些保證, 或說一些信心, 使醫生講的話較可能是對的 (即病人再活的時間長度, 會落在醫生所講的區間中)。這是採用區間估計量的目的之一。

最後必須一提的是, 在上例中, 將  $\bar{X}_n(1-\bar{X}_n)$  以  $1/4$  取代, 這樣其實有些粗糙 (只要看  $\bar{X}_n = 0.1$  的情況即知), 造成的誤差遠比 (12) 式與 (15) 式間的差異大多了。只有當  $p$  較接近 0.5 時, 這種取代才不致有太離譜的影響。如果事先知道  $p$  的範圍, 如  $0.1 \leq p \leq 0.3$ , 則  $p(1-p) \leq 0.21$ , 可對  $p(1-p)$  得到比  $1/4$  更精確的推估。此時 (17) 式成為  $n \geq 0.21z_{1-\alpha/2}^2/d^2$ , 而 (18) 式成為  $d \leq \sqrt{0.21}z_{1-\alpha/2}/\sqrt{n}$ 。

以上討論顯示，在勢均力敵的情況下，民調領先，選舉結果卻逆轉是很平常的。例如，設候選人A的支持率為 32.3%，候選人B的支持率為 31.1%。由於大部分的民調是設定  $\alpha = 0.05$ ， $d = 0.03$ ，在此情況下，候選人A支持率的 95%信賴區間為 [29.3%, 35.3%]，候選人B支持率的 95%信賴區間為 [28.1%.34.1%]。二區間有很大的重疊，因此選舉結果逆轉發生並不足為奇。這也是為何信賴區間的長度要愈短愈好的原因之一。不過我們已指出，欲減小抽樣誤差  $d$ ，樣本數  $n$  得大幅度增加。在實際做民調時，困難度將增大不少。相較之下，更好的抽樣方法，更謹慎的規劃，反而是較可行的。另外，民調多半採用電話訪問，先天已有一些不易掌控的因素，又這些信賴區間也只是“近似的”。所以選舉時所做民調的信賴區間，信心水準等，有時不能過度採信。

### 3. 信賴的意義

前言裡提到，政治裡的信賴，可能並不太能信。那統計學裡的信賴呢？它的意義為何？真可信賴嗎？

當你看到有人指著一區間說，這是  $\mu$  的 95% 信賴區間，他是在信賴什麼呢？95% 的意義又是什麼呢？對於 (9) 式，我們通常說有  $100(1 - \alpha)\%$  的信心  $\mu$  會屬於區間  $I$ 。但對於例 1，我們是否可說  $\mu$  會落在區間 [329.27, 330.78] 之機率約為 0.95？不少人以為此答案是肯定的。事實上，對於例 1，敘述  $P(\mu \in [329.27, 330.78]) \doteq 0.95$  並不正確。(9) 式為一隨機區間，在取樣前，的確有  $1 - \alpha$  的機率，此區間會包含  $\mu$ 。但是一旦取得一組樣本  $x_1, x_2, \dots, x_n$ ，且將 (9) 式中之  $\bar{X}_n$  以  $\bar{x}_n = \sum_{i=1}^n x_i/n$  取代，則所有隨機性便消失了，而是得到一固定的區間。又因  $\mu$  為一常數（只是不知其值為何）， $\mu$  要嘛落在此區間，要嘛不落在此區間，說  $P(\mu \in [\bar{x}_n - \sigma z_{1-\alpha/2}/\sqrt{n}, \bar{x}_n + \sigma z_{1-\alpha/2}/\sqrt{n}]) = 1 - \alpha$  自然不對。例如，在例 1 中，假設該工廠有一資深員工，知道實際的  $\mu$  之值應很接近 331.0。則若你告訴他  $P(\mu \in [329.27, 330.78]) = 0.95$ ，他一定斥為無稽，覺得統計人員，不過是書生，書生之見，只能聽聽。此正如設一袋中有 1 個紅球 9 個白球，某人隨機地取一球，假設取中紅球。明明手上拿著紅球，這時你告訴他此球為白球之機率為 0.9，他必覺得你不知所云。正確的說法是，在同樣的  $\alpha$  及  $n$  之下，若我們持續地取樣，每次各得一信賴區間，則這些信賴區間中，“約有”  $100(1 - \alpha)\%$  個會涵蓋  $\mu$  值。以弱大數法則來解釋，涵蓋  $\mu$  之信賴區間數，比上全部的信賴區間數，當後者趨近至  $\infty$  時，會機率收斂至  $1 - \alpha$ 。對於  $1 - \alpha$  的信心水準，其中  $1 - \alpha$  所指的，就是機率而已。

藉圖形來說明。圖 2 為  $\mathcal{N}(\mu, \sigma^2)$  分佈之樣本平均  $\bar{X}_n$  之機率密度函數的圖形，圖 3 為在  $\sigma^2$  已知之下，依序取樣 14 次（每次皆取  $n$  個樣本），所得  $\mu$  之 14 個 95% 信賴區間。若  $\bar{X}_n$  介

於  $\mu - 1.96\sigma/\sqrt{n}$  及  $\mu + 1.96\sigma/\sqrt{n}$  之間，則得到的信賴區間會包含  $\mu$ 。由於圖2中機率密度函數的圖形介於  $\mu - 1.96\sigma/\sqrt{n}$  與  $\mu + 1.96\sigma/\sqrt{n}$  間之面積約為0.95， $\bar{X}_n$  會落在此範圍內之機率便也約為0.95。圖3中之信賴區間，第9個並未包含  $\mu$ 。即有  $13/14 \doteq 92.86\%$  的區間包含  $\mu$ 。弱大數法則告訴我們，若取樣夠多次（因此得到很多  $\bar{X}_n$  之實際值  $\bar{x}_n$ ），則其中“約有”95% 左右個（口語有時講20次中有19次）信賴區間會包含  $\mu$ 。至於對任一特別的區間，說其有 95% 的機率會包含  $\mu$ ，則是沒有意義的。因這一特定的區間已非隨機區間，常數（但未知） $\mu$  會落在此區間的機率，不是1便是0。

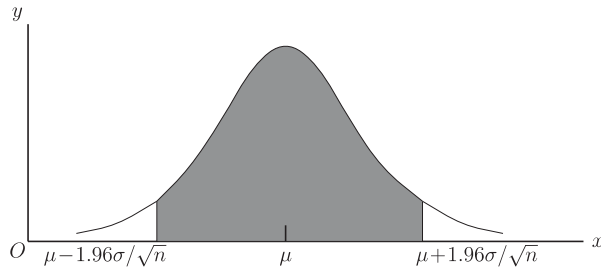


圖2.  $y = f(x) = \frac{1}{(\sigma/\sqrt{n})\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2/n)}, x \in R.$

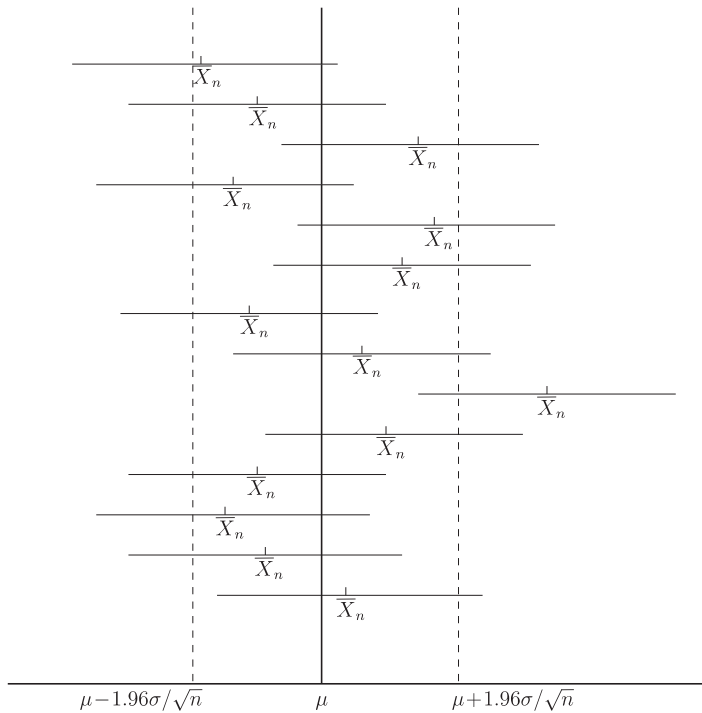


圖3. 對  $\mathcal{N}(\mu, \sigma^2/n)$  分佈經由重複取樣所得  $\mu$  之14個 95% 信賴區間。

仍要提醒讀者，事件發生的機率0.95，並非每20次“必會”發生19次。運氣不好，100次有可能1次也未發生。我們來看前言中所引用的那段教材編選。首先“由全班每一位同學各自以亂數表模擬丟銅板的過程”，這是什麼意思呢？假設  $p = 0.6$ ，則可將亂數表上 0, 1, 2, 3, 4, 5 皆當做1，表出現正面；6, 7, 8, 9 皆當做0，表出現反面。 $n$  如果是400，就從亂數表上取400個數字，因而得到一串 1, 0 的數列  $X_1, \dots, X_{400}$ 。求出  $\bar{X}_n$ ，再給一  $\alpha$ ，便得到一  $p$  之  $1 - \alpha$  信賴區間。對相同的  $p, n, \alpha$ ，每位同學各自從亂數表上某一數字開始模擬，並各得一  $p$  的  $1 - \alpha$  信賴區間。但接著的“並察覺大多數同學所得的信賴區間會涵蓋銅板正面機率的真實值”，就不是正確的說法。事實上，學校中只要班級數夠多，就很可能有的班級，全班學生們所得的信賴區間，其中有好幾個沒有涵蓋銅板出現正面的機率  $p$ 。見下例。

例 3: 設一班有40位同學，老師給定一  $p$ ，其中  $0 < p < 1$  (譬如  $p$  就是0.6)。每位同學各自以亂數表，模擬投擲銅板的過程，而各得一銅板出現正面機率  $p$  之 95% 信賴區間。則全班所得的40個信賴區間，共有幾個會涵蓋  $p$  呢？

解：會涵蓋  $p$  的區間數以  $X$  表之。取樣前，每一信賴區間有0.95之機率會涵蓋  $p$ 。全班所得的40個信賴區間，要看有幾個涵蓋  $p$ ，相當於獨立地做某實驗40次，每次成功機率皆為0.95，要看成功幾次。故 (見底下註1)  $X$  有二項分佈(binomial distribution)，參數為40, 0.95。即  $X \sim \mathcal{B}(40, 0.95)$ 。令

$$F(x) = P(X \leq x), \quad x \in R.$$

$F(x)$  即為  $X$  不超過  $x$  之機率。表1給出  $\mathcal{B}(40, 0.95)$  及  $\mathcal{B}(40, 0.90)$  分佈之一些  $F(x)$  值。

由表1知，全班所得之40個 95% 信賴區間中，至少有4個 (即  $X \leq 36$ ) 不涵蓋  $p$  之機率約為0.13815。假設某校高二有25個班，每班假設都有40人。若每班皆做此模擬實驗，則全年級全班所得之信賴區間，至少有4個不涵蓋  $p$  之班級數，約有  $\mathcal{B}(25, 0.13815)$  分佈。平均則約有

$$25 \times 0.13815 \doteq 3.45375 \text{ (班)}.$$

假設全台灣普通高級中學的高二共有2,500班 (每班仍設為40人)，每班都做此實驗。則雖然  $P(X \leq 33) \doteq 0.00339$  並不太大，但因

$$2,500 \times 0.00339 \doteq 8.475 \text{ (班)},$$

即全台灣平均約有8個班，全班所得之信賴區間，至少有  $7 (= 40 - 33)$  個不涵蓋  $p$ 。有高達7個信賴區間不涵蓋  $p$ ，這幾個班級的學生，大約不會認為，大多數同學所得的信賴區間會涵蓋  $p$ 。

其次來看如果是求  $p$  之 90% 的信賴區間會如何？

由表1知，對  $\mathcal{B}(40, 0.90)$  分佈， $P(X \leq 32) \doteq 0.04190$ 。故一校之全年級 (仍設有25班)，約有

表1.  $\mathcal{B}(40, 0.95)$  及  $\mathcal{B}(40, 0.90)$  分佈之一些機率值

$\mathcal{B}(40, 0.95)$ 分佈		$\mathcal{B}(40, 0.90)$ 分佈	
$X$	$F(x)$	$X$	$F(x)$
40	1.00000	40	1.00000
39	0.87149	39	0.98522
38	0.60094	38	0.91953
37	0.32326	37	0.77719
36	0.13815	36	0.57687
35	0.04803	35	0.37098
34	0.01388	34	0.20627
33	0.00339	33	0.09952
32	0.00071	32	0.04190
31	0.00013	31	0.01550
30	0.00002	30	0.00506

$$25 \times 0.04190 \doteq 1.0475 \text{ (班)},$$

至少有8個信賴區間沒有涵蓋  $p$ 。全台灣就約有104班了。由於前述“實施方法”，並未講明是求多少百分比之信賴區間，因此未留後路地說：

並察覺大多數同學所得的信賴區間會涵蓋銅板正面機率的真實值，

顯然並不恰當，易引起教師及學生之一些困擾。

註1: 假設獨立地做一實驗  $n$  次，每次成功的機率為  $p$ ，其中  $n$  為一正整數， $0 < p < 1$ 。則總共之成功數  $X$ ，便有  $\mathcal{B}(n, p)$  分佈。 $X$  之機率密度函數為

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, \dots, n,$$

又  $E(X) = np$ ,  $\text{Var}(X) = np(1-p)$ 。

例3顯示，本欲藉模擬實驗，讓學生了解信賴區間之涵義，卻因敘述過於簡略，既未交待清楚隨機的本質，也未留意到樣本數夠多後，小機率事件的發生很容易（對95%，全年級1,000個學生，所得之信賴區間，平均約有50個不會涵蓋  $p$ ）。統計的教學有其難度，若所涉及機率的涵義未說明白，有時立意良善的教學方式，不見得會有好效果。

## 4. 信賴區間與品質管制

信賴區間與假設檢定有一些對應關係，我們藉品質管制來說明。仍以常態分佈為例。有些工廠的品質管制（以下簡稱品管）人員，便以 (9) 式中之信賴區間，作為品管之依據。假設某產品之規格（如長度、重量等）須為  $\mu$ 。經隨機抽取  $n$  個樣本後，得到一如 (9) 式之 95% 信賴區間 ( $\sigma^2$  假設為已知)。則若  $\mu$  落在此區間，便認為該批產品為合格，否則認為不合格。偶而有品管人員心存疑惑，取樣愈多 ( $n$  愈大)，則信賴區間的長度愈短 ((9) 式中之區間長度為  $2\sigma z_{1-\alpha/2}/\sqrt{n}$ ，隨著  $n$  增大而變小)，因此愈不容易包含  $\mu$ ，如此一來，不是產品愈容易不合規格嗎？更何況取樣多本來就已較麻煩，所以他們對取樣較多存有抗拒之心。你認為他們的抗拒合理嗎？檢視圖 3，區間長度若較短，則會包含實際  $\mu$  值之機會的確是較小，品管人員排斥較大的  $n$  似乎是有道理的。其實不然。

若採用 (9) 式做為信賴區間，對一固定產品（因此  $\sigma$  相同），在同一  $\alpha$  值之下，信賴區間隨著  $n$  之增大而變短。但不要忘記， $\alpha$  沒有改變，換句話說，對這些或長或短的隨機區間，我們皆有相同的  $100(1-\alpha)\%$  信心，認為  $\mu$  會落在其中。

由弱大數法則， $n$  愈大時， $\bar{X}_n$  有愈靠近  $\mu$  之傾向。因此  $n$  愈大時，以  $\bar{X}_n$  為中心，只需要較小的半徑  $\sigma z_{1-\alpha/2}/\sqrt{n}$ ，該區間仍有相同的機率  $1-\alpha$  會涵蓋  $\mu$ 。有點像若飛彈射得愈準，則雖爆破半徑較小，對目標物仍可有相同的摧毀效果。

即使涵蓋  $\mu$  的機率相同，如前所述，我們仍較偏好區間長度較短者。原因很簡單，給出估計的區間長度愈短，表示推論愈精準。A 先生給出  $\mu$  之一長度為 10 之 95% 信賴區間，B 先生給出  $\mu$  之一長度為 5 之 95% 信賴區間（但二者之中心點可能不同），一般人當然覺得後者較準。這就是辛苦地取樣較多 ( $n$  較大)，所換得之代價。

## 5. 結語

我們常提到機率，一事件的機率為  $p$ ，花些功夫，機率二字的涵義，逐漸可以明白。對於信心水準  $1-\alpha$ ，其中的  $1-\alpha$  乃指機率。以一統計量來估計某一參數，稱做點估計，以一隨機區間來估計參數，稱做區間估計。此區間表示參數可能的活動範圍；參數會落在此區間的機率，稱做信心水準。由於一點會等於參數之機率微乎其微，甚至是 0，因此我們對點估計是否為參數之真實值，“信心”自然不夠。就估計的觀點而言，區間較一點更可信賴，稱此區間為信賴區間，有其道理。只是一旦做完實驗，取到一組樣本，所得之信賴區間，為一確定的區間，這時區間會包含參數的機率不是 1 就是 0，而不再是  $1-\alpha$  了。給出一區間明確的“信心水準”，而不像政治人物常掛在口中，可信度不高的“百分之百信賴”，或“完全有信心”，使此區間更可信賴。至於有多信賴？或說信賴程度為何？信心水準就是給出信賴程度。

—本文作者任教於國立高雄大學應用數學系—