中央極限定理

黃文璋

國立高雄大學應用數學系

1 平均法則

台灣的樂透彩發行以來, 提供了一個一夕致富的管道。只要花50元買張彩券, 就有機會成為百萬甚至上億的富翁。只是要中頭獎實在不容易, 49取6的最難, 共有13,983,816種組合, 42取6的有5,245,786種組合。即使是38取6的, 仍有2,760,681種組合。這麼多組合要簽哪一組呢? 有心人會統計出熱門(常出現的)及冷門(較少出現)的號碼, 以供簽注者參考。那究竟該簽熱門還是冷門號碼?

認為該簽熱門號碼, 與認為該簽冷門號碼, 乃是基於不同的想法。前者大抵是認為各號碼出現的機率可能不同, 有些號碼氣較旺較易出現。至於後者, 大約還是認為各號碼出現的機率相同。因此過去幾期出現較少的號碼, 未來幾期該較容易出現, 他們認為這是**平均法則**(law of average)。既然平均要相同, 和當然也要相同! 數字會自己設法彌補過去幾期少出現的次數。在他們眼中, 各數字彷彿有生命, 會有記憶。

大數法則(law of large numbers)有兩個版本。對一數列之獨立且有共同分佈(independent and identically distributed, 簡稱iid)之隨機變數$X_1, X_2, \cdots$, 只要$-\infty < E(X_1) < \infty$, 則$n$很大時, **樣本平均**(sample mean)$\overline{X}_n = (X_1 + \cdots + X_n)/n$, 會很接近$E(X_1)$的機率很大, 此即**弱大數**

法則(weak law of large numbers)。也就是說

$$(1) \qquad \lim_{n\to\infty} P(|\overline{X}_n - E(X_1)| \le \varepsilon) = 1,\ \forall \varepsilon > 0。$$

但並沒有說n很大時, $\overline{X}_n = E(X_1)$。事實上, 除了少數較特殊的情況(如存在常數c, 使得$P(X_1 = c) = 1$), 否則n愈大時, $P(\overline{X}_n \ne E(X_1))$愈接近1。

大數法則由於涉及樣本平均, 故又稱**平均法則**。它是指(1)成立, 或更強的, 指

$$(2) \qquad P(\lim_{n\to\infty} |\overline{X}_n - E(X_1)| \le \varepsilon) = 1,\ \forall \varepsilon > 0,$$

此即**強大數法則**(strong law of large numbers)。但再度絕對沒有說n很大時, $\overline{X}_n = E(X_1)$。或以和表之, 並非n很大時, $X_1 + \cdots + X_n = nE(X_1)$一定成立。

對於樂透彩, 假設是公正的, 即每次開獎, 每一種組合出現的機率皆相同, 則開愈多期, 各數字出現的累積頻率, 差異通常頗大, 不會愈來愈接近。機率相同的後果常是頻率不同。很容易想像, 如果做芝麻餅, 就不能閉著眼睛隨機地灑芝麻, 這樣灑的後果, 芝麻的散佈會很不均勻, 有些地方厚, 有些地方稀疏。機率裡的**均勻分佈**(uniform distribution), 與生活中之散佈的很均勻, 可是大異其趣的。

再以投擲銅板為例, 假設持續地投擲一公正的銅板。投擲若干次後, 得到正面較反面多100次。那在未來的投擲, 反面數是否有趕上的傾向? 有些人認為會, 所依據的理由自然是平均法則: 正面數與反面數最終該相等。有些人認為不會, 因銅板是沒有記憶的, 每次銅板出現正、反面的機率皆為$1/2$。因此正、反面不必有趨於平手的傾向。

這類問題很多, 如果資料顯示, 飛機平均每起降300次, 發生1次意外, 如今已起降290次皆平安無事, 你會不會認為該有1次意外發生了, 否則就達不到300次起降的1次意外? 如果某項醫學手術成功率為10分之1, 而前9個病人的手術皆失敗, 你是第10位, 是否可以很安心地上手術台?

你現在知道了, 這類問題的答案, 或者說在我們的假設下(每次獨立地投擲銅板, 飛機每次起飛的情況與其他次獨立), 我們所觀測的過程都是無記憶的。

不過對於投擲公正銅板, 我們仍要弄清平手的意義。雖然目前正面領先100次, 對未來的出現正面或反面並無影響。但在未來有"某一時刻", 正反面會達到平手的機率, 卻為1。只是呢, 我們得補充一句, 在目前正面領先100次的情況下, 未來有某一時刻, 正面領先數達到100萬次的機率, 亦為1。只要投擲數夠多, 各種樣式(pattern)都可能出現。例如, 連續10個正面, 一串3正面、3反面、3正面、3反面。前者機率為1/1,024, 後者機率為1/4,096。各種有趣的樣式很多, 細心觀察, 將會發現不少, 不用太驚訝。一直1正、1反、1正、1反, $\cdots$, 有人以為這是最容易的, 其實很難出現。即使正、反面數經常性地維持在各半附近, 也很不容易。

仍以投擲一公正的銅板為例。先看投擲數$n = 10$。但會恰好得到5個正面之機率並不大, 只有

$$\binom{10}{5}(\frac{1}{2})^{10} = \frac{252}{1{,}024} \doteq 0.246\,,$$

小於1/4的機率。就算不要求5正面5反面, 所得正面數總該介於4與6(5的前後各10%的數)間吧! 此機率為

$$\binom{10}{4}(\frac{1}{2})^{10} + \binom{10}{5}(\frac{1}{2})^{10} + \binom{10}{6}(\frac{1}{2})^{10} = \frac{672}{1{,}024} = 0.65625 < \frac{2}{3}\text{。}$$

即有大於1/3的機率, 所得正面數小於4或大於6。如果投擲1,000,000次, 則期望會得到500,000次正面。但會恰好得到500,000個正面之機率更小了,

$$\binom{1{,}000{,}000}{500{,}000}(\frac{1}{2})^{1{,}000{,}000}\text{。}$$

不過這回我們就有極大的把握(機率極度接近1), 所得正面數介於400,000與600,000(500,000的前後各10%的數)間。甚至對短很多的區間, 如半徑500的區間[499,500, 500,500], 所得正面數會落在此區間的機率, 就已約0.6826(見下一節)。半徑為500, 比半徑為1的區間$[4, 6]$, 前者自然長很多。但前者之相對半徑(即除以投擲數n)為500/1,000,000=0.0005, 遠小於後者之相對半徑$1/10 = 0.1$。正面數不會剛好等於投擲數之半, 而是在投擲數之半的左右波動。波動區間會隨著投擲數之變大而變長, 但**相對半**

徑會變短。這些我們下一節再說明。

<u>**註1.**</u> 二數列$\{a_n,\, n \geq 1\}$及$\{b_n,\, n \geq 1\}$, 稱為**近似相等**(asymptotically equal), 並以$a_n \sim b_n$表之, 其意義為$n \to \infty$時, $a_n/b_n \to 1$。要注意的是, 二數近似相等, 其差有可能會很大。如$a_n = n + \sqrt{n}$, $b_n = n$, 則$a_n \sim b_n$, 但$n \to \infty$時, $a_n - b_n = \sqrt{n} \to \infty$。又$n!$之計算並無快速的方法。即使僅利用掌上型科學用計算機, 經由取對數, 我們可以立即知道$2^{1{,}000{,}000}$的大小。此數有301,030位, 約為$9.900656658 \cdot 10^{301{,}029}$。但計算機上雖有階乘鍵, 通常最大只能表示到$9.999999999 \cdot 10^{99}$, 而$69! \doteq 1.711 \cdot 10^{98}$, 故70!就已太大而求不出了, 更不要說100!, 甚至1,000,000!了。一個著名的估計$n!$之**位階**(order)的方法, 是利用**史得林公式**(Stirling formula)

$$n! \sim \sqrt{2\pi}\; n^{n+1/2}\, e^{-n} \text{。} \tag{3}$$

由此得(見黃文璋(2010) p.128)投擲一公正銅板$2n$次, 恰好出現n個正面之機率

$$\binom{2n}{n}(\frac{1}{2})^{2n} \sim \frac{1}{\sqrt{\pi n}} \text{。} \tag{4}$$

$1/\sqrt{\pi n}$隨著n之增大而漸減, 下降速度之位階為$n^{-1/2}$。舉例來看。利用(4)式, 投擲1公正銅板100次, 會得到恰好50個正面之機率約為

$$\frac{1}{\sqrt{50\pi}} \doteq 0.08;$$

若投擲1,000,000次, 會得到恰好500,000個正面之機率約為

$$\frac{1}{\sqrt{500{,}000\pi}} \doteq 0.0008 \text{。}$$

投擲數成為10,000倍, 但機率減為1/100。

正、反面數各半固然不易, 有某一面長期領先, 卻稀鬆平常。令$S_n = X_1 + \cdots + X_n$, $n \geq 1$, 其中$X_i = 1$, 表第i次投擲得正面, $X_i = -1$, 則表得反面。由於是公正銅板, $P(X_i = 1) = P(X_i = -1) = 1/2$。利用**布朗運**

動(Brownian motion)的結果來估計(見黃文璋(1995)第七章),

$$(5)\qquad P(\max_{1\le n\le 10{,}000} S_n \le a) \doteq P(\max_{0\le s\le 1} X(s) \le \frac{a}{100})$$
$$= P(|Z| \le \frac{a}{100}) \doteq \begin{cases} 0.0796, & a = 10, \\ 0.1586, & a = 20, \\ 0.3830, & a = 50, \\ 0.6826, & a = 100, \end{cases}$$

其中Z有**標準常態分佈**(standard normal distribution), 以$\mathcal{N}(0,1)$表之, $\{X(t), t\ge 0\}$表一**標準布朗運動**。S_n為正, 表投擲至第n次, 正面數領先, S_n為負, 則表反面數領先, S_n為0則表正反面平手。(5)式指出當投擲總數從第1次至第10,000次裡, 正面數領先都不超過10次的機率約為0.0796, 但正面數領先不超過100次的機率約為0.6826。也就是S_n在0的附近上下波動, 即正、反面數領先幅度均不太大較不易發生, 但有一面領先較多次反易發生。

仍利用布朗運動來估計, 由**反正弦法則**(arc-sine law), 可得

$$(6)\qquad P\left(\frac{\text{正面領先次數}}{\text{總投擲次數}} \ge x\right) \doteq 1 - \frac{2}{\pi}\arcsin\sqrt{x}\ 。$$

當$x = 0.993$, 此機率約為

$$1 - \frac{2}{\pi}\arcsin\sqrt{0.993} \doteq 0.0533\ 。$$

因此投擲10,000次中, 有一面領先次數至少9,930次(另一面僅領先不超過70次)之機率約為$0.0533\cdot 2 = 0.1066$。超過10分之1的機率, 可不算太小。因此看到這種現象, 並不用太驚訝。

我們列出4個S_n, $1\le n\le 10{,}000$, 之模擬圖於圖1, 以讓讀者了解S_n的波動情況。

2 中央極限定理

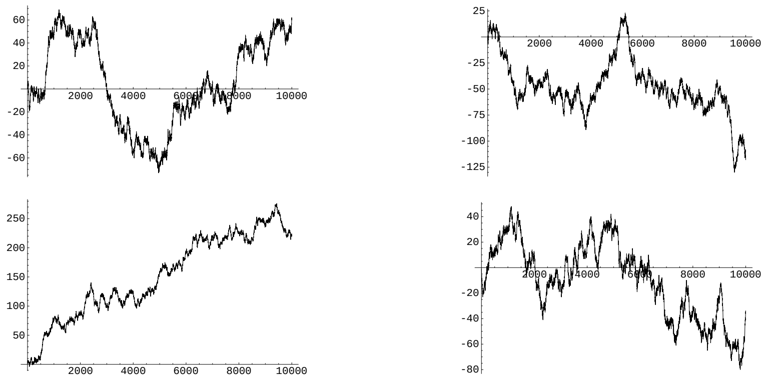

圖1. 4個$\{S_n, 1 \leq n \leq 10{,}000\}$的模擬圖

掌握誤差是很重要的。人類登陸月球, 有預計登陸的地點。誤差若太大, 太空船本來要降落平地, 結果落在一峽谷中就不太妙。登月使命完成後, 要回來地球, 須穿越大氣層, 落在某處海中, 誤差若太大, 落在喜馬拉雅山頂上, 要如何救回太空人, 就很傷腦筋。現代科學中, 微積分是一重要的工具。在微積分裡, 我們處處會見到近似的影子, 常常在證明誤差很小。舉例而言,大家都知道三角函數的重要, 但如何求其函數值呢? 對於$\sin x$, 我們可用多項式來逼近: 對$\forall x \in R$,

$$(7)\quad \sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots + (-1)^{n-1}\frac{x^{2n-1}}{(2n-1)!} + R_{2n}(x),$$

其中餘項

$$(8)\qquad |R_{2n}(x)| \leq \frac{|x|^{2n+1}}{(2n+1)!}\text{。}$$

如果$x \in [0,1]$, 且取$n = 4$, 則

$$\sin x = x - \frac{1}{6}x^3 + \frac{1}{120}x^5 - \frac{1}{5{,}040}x^7 + R_8(x),$$

且

$$|R_8(x)| \leq \frac{x^{2n+1}}{9!} \leq \frac{1}{362{,}880} \leq 2.755732 \cdot 10^{-6}\text{。}$$

故在$x \in [0,1]$, 以多項式

$$f(x) = x - \frac{1}{6}x^3 + \frac{1}{120}x^5 - \frac{1}{5{,}040}x^7$$

做為$\sin x$之近似，誤差不超過$2.755732 \cdot 10^{-6}$。如果你覺得此誤差不夠小，取更大的n即可。

對一數列之iid的隨機變數$\{X_n, n \geq 1\}$，且$E(X_1)$存在，大數法則指出，$n \to \infty$時，$\overline{X}_n$會以某種方式收斂至$E(X_1)$。但我們也已指出，不論n多大，除非這些隨機變數為常數，否則並無法保證$\overline{X}_n = E(X_1)$。那有沒有什麼方式，可以如(7)式中，知道多項式$\sum_{i=1}^{n}(-1)^{i-1}x^{2i-1}/(2i-1)!$與$\sin x$之差異$|R_{2n}|$究竟有多大？此問題若能回答，才能更有信心地以$\overline{X}_n$來估計$E(X_1)$。

中央極限定理(central limit theorem)，差不多可以說是機率論與統計學中，最重要的一個定理。此定理是說，若有夠多的獨立的隨機變數之和(或平均)，將可以常態分佈做為近似。設對$\forall n \geq 1$, $X_1, \cdots, X_n$為iid之隨機變數，且$\mu = E(X_1)$及$\sigma^2 = \mathrm{Var}(X_1)$皆存在。中央極限定理指出：對任二實數$a < b$,

$$\lim_{n\to\infty} P\left(a \leq \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \leq b\right) = \Phi(b) - \Phi(a), \tag{9}$$

其中$\Phi(x)$表$\mathcal{N}(0,1)$之分佈函數，即

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\, du,\ x \in R\,。 \tag{10}$$

由此得**標準化**後的隨機變數$(X_1 + \cdots + X_n - n\mu)/(\sigma\sqrt{n})$, n較大時，可以標準常態分佈來近似。證明可參考黃文璋(2010)第四章定理5.2。我們知道當隨機變數W有$\mathcal{N}(0,1)$分佈，則$\alpha W + \beta$有$\mathcal{N}(\beta, \alpha^2)$分佈。因此由中央極限定理知：當$X_1, \cdots, X_n$為iid之隨機變數，且$\mu = E(X_1)$及$\sigma^2 = \mathrm{Var}(X_1)$皆存在，則$n$較大時，$X_1 + \cdots + X_n$可以$\mathcal{N}(n\mu, n\sigma^2)$分佈來近似。若以樣本平均來表示，便有$n$較大時，$\overline{X}_n$可以$\mathcal{N}(\mu, \sigma^2/n)$分佈來近似。事實上(9)導致

$$\lim_{n\to\infty} P\left(a \leq \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \leq b\right) = \Phi(b) - \Phi(a),\ b > a\,。 \tag{11}$$

又$\Phi(x)$滿足

$$\Phi(x) + \Phi(-x) = 1,\ \forall x \in R\,。 \tag{12}$$

由(9)式得(取$b=c,\ a=-c$)n很大時,

$$(13)\qquad P\left(\left|\frac{X_1+\cdots+X_n-n\mu}{\sigma\sqrt{n}}\right|\le c\right)\doteq 2\Phi(c)-1,\ c\ge 0\text{ 。}$$

此處用到(12)式, $\Phi(-c)=1-\Phi(c)$ 。由(13)式得

$$(14)\qquad P\left(|\overline{X}_n-\mu|\le\frac{c\,\sigma}{\sqrt{n}}\right)\doteq 2\Phi(c)-1,\ c\ge 0,$$

及

$$(15)\qquad P\left(|\overline{X}_n-\mu|>\frac{c\,\sigma}{\sqrt{n}}\right)\doteq 2(1-\Phi(c)),\ c\ge 0\text{ 。}$$

因此又有(令$c\sigma/\sqrt{n}=\varepsilon$)

$$(16)\qquad P(|\overline{X}_n-\mu|>\varepsilon)\doteq 2(1-\Phi(\varepsilon\sqrt{n}/\sigma)),\ \varepsilon\ge 0\text{ 。}$$

弱大數法則指出$\forall\varepsilon>0$, n很大時, $P(|\overline{X}_n-\mu|>\varepsilon)$很小 。但究竟多小? 中央極限定理便給出$n$很大時, $P(|\overline{X}_n-\mu|>\varepsilon)$約為$2(1-\Phi(\varepsilon\sqrt{n}/\sigma))$ 。不論$\{X_n, n\ge 1\}$之共同分佈為何, 都有此結果, 這是中央極限定理的威力所在 。雖然無法保證n很大時, $|\overline{X}_n-\mu|$不會超過ε 。但此機率$P(|\overline{X}_n-\mu|>\varepsilon)$的大小, 卻可掌握, 即約$2(1-\Phi(\varepsilon\sqrt{n}/\sigma))$ 。此值隨著n之趨近∞($n\to\infty$時, $\Phi(\varepsilon\sqrt{n}/\sigma)\to 1$), 而趨近至0 。另外, 若(16)式成立, 即得$n\to\infty$時, $P(|\overline{X}_n-\mu|>\varepsilon)\to 0$ 。因此看起來中央極限定理是一較弱大數法則強的定理 。不過前者要假設變異數存在, 後者不用 。在較強的假設下, 是可能得到較強的結果 。

在中央極限定理裡, 不論隨機變數之共同分佈為何, 只需期望值及變異數皆存在便適用 。甚至它有更一般的型式 。獨立性及分佈相同的假設皆可放寬 。當然此時分佈有要滿足的條件, 可參考黃文璋(2010)第五章定理5.3 。由於有此推廣的版本, 更可以解釋何以不少隨機現象, 皆可以常態分佈來近似 。諸如人的智商 、身高 、膽固醇含量, 這些隨機的量, 可視為很多微小的獨立隨機效應所造成的, 因此皆能用常態分佈來描述 。19世紀時, 比利時統計學家**奎特雷**(Adolphe Quetelet, 1796-1874)及英國統計學家**高頓**(Francis Galton, 1822-1911), 發現很多生物裡的特徵, 其分佈皆符

合常態分佈。這也是此分佈被稱為**常態**的一個主因, 也隱含若是其他分佈, 就不算常態。常態分佈就是如此被推崇。

中央極限定理的歷史, 可參考Tijms(2004)p.169。最早的版本, 是出生於法國, 後來移民至英國的數學家**棣美弗**(Abraham de Moivre, 1667-1754)所提出。在一篇於1733年出版的文章中, 他利用常態分佈, 做為投擲一公正的銅板許多次, 出現正面數之分佈的近似。這項發現遠超過當時代的人對機率論的理解, 因此並未引起重視。就在幾乎已被遺忘時, 著名的法國數學家**拉普拉斯**(Pierre-Simon, Marquis de Laplace, 1749-1827), 把此結果從塵封中救了出來。在拉普拉斯1812年出版的那篇機率史上不朽的論文Théorie Analytique des Probabilités裡, 推廣棣美弗的結果。他以**常態分佈**來近似**二項分佈**(見底下註2)。但如同棣美弗, 拉普拉斯的結果, 仍未引起同時代的人太多注意。直到19世紀結束後, 中央極限定理的重要性, 才被完全認識。1901年, 俄國數學家**里阿普那夫**(Aleksandr M. Lyapunov, 1857-1918), 給出比定理較一般的敘述, 及嚴密的證明。之後一直到今日, 如同central(中央、中心)一詞, 找出不同的條件, 使得大量的和, 可以常態分佈來近似, 一直是機率論裡, 極核心的研究題材之一。

由於前述這一段淵源, 中央極限定理又稱**棣美弗-拉普拉斯定理**。

常態分佈又稱**高斯分佈**(Gaussian distribution)。德國今日用歐元, 往昔德國10馬克, 是以大數學家**高斯**(Carl F. Gauss, 1777-1855)為人像, 人像左側伴隨高斯的, 不是他數學上其他成就, 而是一常態分佈的機率密度函數及其圖形。**誤差理論**是高斯對機率論的主要貢獻, 而中央極限定理, 能對此理論有很好的解釋, 我們第6節再討論。常態分佈機率密度函數的圖形, 原先稱為**高斯曲線**(Gaussian curve), 或**高斯誤差曲線**(Gaussian error curve)。不過自著名的英國統計學家**卡爾・皮爾生**(Karl Pearson, 1857-1936)之後, 人們開始稱此為**常態曲線**(normal curve)。另外, "中央極限定理"的名稱, 則是1920年, 由出生於匈牙利的數學家**波里亞**(George Pólya, 1887-1985)所取的。時至今日, 機率論裡的定理雖然很多, 但中央極限定理的地位, 仍是屹立不搖。

<u>**註2.**</u> 設$X_1, \cdots, X_n,\ n \geq 1$, 為iid之隨機變數, 以參數$p$之**伯努力分佈**(Ber-noulli distribution)為共同分佈, 此分佈以$Ber(p)$表之, $0 < p < 1$ 。即有$P(X_i = 1) = p = 1 - P(X_i = 0),\ i = 1, \cdots, n$ 。令$S_n = X_1 + \cdots + X_n,\ n \geq 1$ 。則S_n有參數n, p之**二項分佈**(binomial distribution), 以$\mathcal{B}(n,p)$表之 。即S_n之機率密度函數為

$$P(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k},\ k = 0, 1, \cdots, n\ 。 \tag{17}$$

拉普拉斯就是將棣美弗的結果, 推廣至投擲的銅板不一定是公正的情況 。

由於$\mu = E(S_n) = np,\ \sigma^2 = \mathrm{Var}(S_n) = np(1-p)$ 。故對S_n有$\mathcal{B}(n,p)$分佈, 當n夠大,

$$P\left(a \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq b\right) \doteq \Phi(b) - \Phi(a),\ a < b\ 。$$

這就是所謂二項分佈$\mathcal{B}(n,p)$趨近至常態分佈, 為中央極限定理之一特別的情況 。對二項分佈, 若採**連續性的更正**(continuity correction), 可更精確地表示上式 。即對任二非負整數k及j, 當n夠大時,

$$\begin{aligned} & P(j \leq S_n \leq k) \\ & = P(j - 1/2 \leq S_n \leq k + 1/2) \\ & = P\left(\frac{j - np - 1/2}{\sqrt{np(1-p)}} \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq \frac{k - np + 1/2}{\sqrt{np(1-p)}}\right) \\ & \doteq \Phi\left(\frac{k - np + 1/2}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{j - np - 1/2}{\sqrt{np(1-p)}}\right)\ 。 \end{aligned} \tag{18}$$

當然, 如果n實在很大, 使得$1/2$與$\sqrt{np(1-p)}$相比極小, 則上式中之$1/2$可忽略 。不過如果n不是特別地大, 有$1/2$這項, 會使近似較精確 。

對iid的隨機變數$X_1, \cdots, X_n,\ n \geq 1$, 且設$E(X_1) = \mu$, $\mathrm{Var}(X_1) = \sigma^2$皆存在, 則當$n$夠大時, 由(13)式得

$$P(|X_1 + \cdots + X_n - n\mu| > c\,\sigma\sqrt{n}) \doteq 2(1 - \Phi(c)),\ c > 0\ 。 \tag{19}$$

由於當$x \to \infty$, $\Phi(x)$趨近至1的速度很快, 因此$1-\Phi(x)$及$2(1-\Phi(x))$趨近至0的速度亦很快。例如, $1-\Phi(4)$就已約是$3.16712 \cdot 10^{-5}$了。表1給出一些$1-\Phi(x)$及$2(1-\Phi(x))$之值。

表1. 一些$1-\Phi(x)$及$2(1-\Phi(x))$之值

x	$1-\Phi(x)$	$2(1-\Phi(x))$
1	0.1586552539	0.3173105079
2	0.0227501320	0.0455002639
3	0.0013498980	0.0026997961
4	0.0000316712	0.0000633425
5	$2.866515719 \cdot 10^{-7}$	$5.733031438 \cdot 10^{-7}$
6	$9.865876450 \cdot 10^{-10}$	$1.973175290 \cdot 10^{-9}$
7	$1.279812544 \cdot 10^{-12}$	$2.559625088 \cdot 10^{-12}$
8	$6.220960574 \cdot 10^{-16}$	$1.244192115 \cdot 10^{-15}$
9	$1.128588406 \cdot 10^{-19}$	$2.257176812 \cdot 10^{-19}$
10	$7.619853024 \cdot 10^{-24}$	$1.523970605 \cdot 10^{-23}$

例1. 你買了一批20個電池, 一次用一個。假設電池壽命為一隨機變數, 每一個平均可用14天, 標準差則為3天。一旦沒電立即更新。則這批電池共可用超過9個月且少於10個月之機率約為若干?

解. 本例即求$P(270 < X_1 + \cdots + X_{20} < 300)$, 而$X_1, \cdots, X_{20}$為iid, $\mu = E(X_1) = 14$, $\sigma = (\mathrm{Var}(X_1))^{1/2} = 3$。因$n\mu = 20 \cdot 14 = 280$, $\sigma\sqrt{n} = 3\sqrt{20}$, 由中央極限定理,

$$
\begin{aligned}
&P(270 < X_1 + \cdots + X_{20} < 300) \\
&= P\left(\frac{270-280}{3\sqrt{20}} < \frac{X_1 + \cdots + X_{20} - 280}{3\sqrt{20}} < \frac{300-280}{3\sqrt{20}}\right) \\
&\doteq \Phi(1.491) - \Phi(-0.745) \doteq 0.9320 - 0.2281 = 0.7039 \text{。}
\end{aligned}
$$

例2. 如果有人告訴你, 他投擲1個公正的銅板10,000次, 得到5,250個正面, 你相信嗎?

解. 由大數法則, 你知道約會得到5,000個正面。但由史得林公式, 會恰好得到5,000個正面的機率, 僅約0.008。你不能要求一定要出現恰好5,000個正面, 才相信此銅板為公正。但一個合理的要求是, 出現的正面數該在5,000附近。如果離5,000過遠, 就懷疑銅板不是公正。更進一步, 如果得到的正面數過多, 可能會傾向相信銅板出現正面的機率大於1/2; 如果得到的正面數過少, 可能會傾向相信銅板出現正面的機率小於1/2。

如果銅板為公正, 要得到至少5,250個正面之機率為

$$\sum_{i=5,250}^{10,000} \binom{10,000}{i} (\frac{1}{2})^i,$$

因有$\binom{10,000}{i}$這項, 除非i很接近10,000, 否則上述和中的任一項, 都不太好求出, 更不要說求其和了。這時中央極限定理便可派上用場了。

對本例中投擲公正銅板的情況。令$X_i = 1$, 表第i次得正面, $X_i = 0$, 表第i次得反面, $i = 1, 2, \cdots$,10,000。則$P(X_i = 1) = P(X_i = 0) = 1/2$。因此$\mu = E(X_i) = 1/2,\ \sigma^2 = \mathrm{Var}(X_i) = 1/4,\ \sigma = 1/2$。由此得

$$\frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} = \frac{5{,}250 - 5{,}000}{1/2 \cdot 100} = 5\text{。}$$

而由表1, $P(X_1 + \cdots + X_n - n\mu > 5\sigma\sqrt{n})$之機率約為$2.8665 \cdot 10^{-7}$(我們要的其實是$P(X_1 + \cdots + X_n - n\mu \geq 5\sigma\sqrt{n})$, 不過二者差異極有限)。此機率小至夠讓我們強烈懷疑此銅板之公正性, 我們推測正面出現的機率很可能大於1/2。

另一方面, 取$c = 3$, 得對投擲一公正的銅板10,000次, 出現的正面數落在$[4{,}850, 5{,}150]$間之機率, 約為0.9974。此機率在通常的情況下, 便已被認為夠大了。換句話說, 所得正面數不太容易落在此區間外。

由(13)式, 且令$\mu = 1/2,\ \sigma = 1/2$,

$$(20)\quad P(X_1 + \cdots + X_n \in [\frac{n}{2} - \frac{c\sqrt{n}}{2}, \frac{n}{2} + \frac{c\sqrt{n}}{2}]) \doteq 2\Phi(c) - 1,\ c \geq 0\text{。}$$

n次投擲一公正的銅板, 所得正面數落在一以$n/2$為中心之對稱區間的機率若給定$(2\Phi(c) - 1)$, 則區間半徑$c\sqrt{n}/2$隨著n之增大而增大, 但成長速度

的位階為$n^{1/2}$, 若除以投擲數n, 得$n^{-1/2}$。也就是區間的相對半徑為漸減, 以$n^{-1/2}$的位階下降。對同一機率, 隨著投擲數n之增大, 所得正面數落在的範圍愈來愈大。但相對而言, 乃落在一愈來愈窄的區間。

考慮相對區間是有道理的。對投擲100次, 區間半徑5, 佔投擲數之5%; 投擲1,000,000次, 區間半徑成為500, 區間雖然變長, 但半徑卻只佔投擲數之0.05%。相對於投擲數, 後者是一較無關痛癢的區間。

註3. 我們已指出, 投擲一公正的銅板, 投擲數愈多, 愈不容易出現正反面各有半數。諸如樂透彩的開獎, 投擲骰子都有類似的結果。以骰子為例。投擲一公正的骰子愈多次, 愈難得到每一面都出現1/6的投擲數。譬如說投擲6億次, 不要預期每一面都出現1億次。每一面出現的次數偏離1億有幾千是很正常的。事實上, 令$X_i=1$, 表第i次投擲, 點數1出現, $X_i=0$, 表點數1未出現, $i=1,2,\cdots,6\cdot 10^8$。對應(19)式, $n=6\cdot 10^8$, $\mu=1/6$, $\sigma=\sqrt{5}/6$, 取$c=1$, 則$n\mu=10^8$, $c\sigma\sqrt{n}=(\sqrt{30}/6)\cdot 10^4\doteq$9,129,

$$P(|X_1+\cdots+X_n-10^8|>9{,}129)\doteq 2(1-\Phi(1))\doteq 0.3174\text{ 。}$$

即出現1點的次數, 偏離1億次超過9,129次的機率, 約有0.3174, 此乃一不算小的機率。而6個面中, 會有任何一個面出現的次數, 與1億偏離超過9,129次的機率就更大了。

再看一個例子。

例3. 某次考試題目全為選擇題共50題, 每題有4個選項。某生宣稱他都沒唸書, 因此全部都用猜的。成績發下後, 他對了21題。對此你有何評論。

解. 如果隨機地猜, 答對的題目數, 應在期望值$50\cdot 1/4=12.5$附近。若偏離此數過遠, 我們會懷疑是否眞隨機地猜。若答對的題目若過多, 顯示該生可能並非完全不會。如今該生對了21題, 21題夠不夠多呢? 我們來分析看看。令$X_i=1$, 表第i題答對, $X_i=0$, 表第i題答錯, $i=1,\cdots,50$。則$P(X_i=1)=1/4$, $P(X_i=0)=3/4$。因此$\mu=E(X_i)=1/4$, $\sigma^2=\mathrm{Var}(X_i)=3/16$, $\sigma=\sqrt{3}/4$。由此得

$$\frac{X_1+\cdots+X_n-n\mu}{\sigma\sqrt{n}}=\frac{21-50\cdot 1/4}{\sqrt{3/4}\cdot\sqrt{50}}\doteq 2.776\text{ 。}$$

又$1-\Phi(2.776)\doteq 0.0027$。也就是如果是隨機地猜, 則會猜中至少21題之機率約為0.0027。此機率夠小, 足以讓我們得到該生很可能並非完全沒唸書之推論。

例2及例3這類問題, 皆屬統計裡**假設檢定**(hypothesis testing)的範疇。在數學裡我們常在證明, 有人以為學數學就是不斷地在證明, 證明一個個的命題是否成立。但在科學上或生活中, 一件事是否為眞, 常只有天曉得。我們只能依發生機率之大小, 來做推論。不論是例2中的投擲銅板, 或例3中的考試, 即使銅板為公正, 是可能投擲出5,250個正面; 考試全用猜的, 也是有可能答對21題, 甚至50題全猜對。因為此二事件之機率皆為正。而機率為正的事件, 當然就有可能發生。由於眞相為何, 常就是無法判定, 所以採用"假設"一詞。事實上, 我們以為的眞相, 難道就是眞相嗎? 一切都是假設。而得到的推論是**接受**某一假設, 或**拒絕**某一假設。假設檢定裡, 設計出一套做推論的程序。發生機率之大小, 為做推論之主要依據。而當樣本數夠大, 往往中央極限定理便可用上, 可簡化計算。

例4. 欲估計某地選民對某候選人之支持度p。經由簡單隨機抽樣, 得到n個樣本$X_1,\cdots,X_n$, 其中$X_i=1$, 表第i個人支持該候選人, $X_i=0$, 表不支持。以$\overline{X}_n$來估計p似乎是合理的。這樣估計的誤差為何?

解. 假設選民中有p的比例支持該候選人。但$X_1,\cdots,X_n$並不獨立, 因這是屬於取出後不放回, 你總不能同一個人問他兩次同樣的問題。不過通常選民數很多, 所以取出後不放回的誤差, 就忽略好了。假設$P(X_i=1)=p$, $P(X_i=0)=1-p$, 因$\mu=E(X_i)=p$, $\sigma^2=\mathrm{Var}(X_i)=p(1-p)$, 利用(16)式得

$$P(|\overline{X}_n-p|>\varepsilon)\doteq 2(1-\Phi(\varepsilon\sqrt{n}/(p(1-p)))),\ \varepsilon\geq 0\text{ 。}\tag{21}$$

上式給出$\overline{X}_n$與p之差異大於ε之機率的近似值。問題是上式左側與p有關, 而我們就是因不知p, 才要取樣來估計p。如果n夠大, $\overline{X}_n$應會夠接近p,

則(21)式中右側的p以$\overline{X}_n$取代, 誤差便不致於太大。即得

$$(22) \qquad P(|\overline{X}_n - p| > \varepsilon) \doteq 2(1 - \Phi(\varepsilon\sqrt{n}/(\overline{X}_n(1 - \overline{X}_n))))\text{。}$$

如果事先給定一α值, $0 < \alpha < 1$, α通常是一較小的的值, 而要求

$$(23) \qquad P(|\overline{X}_n - p| > \varepsilon) < \alpha,$$

或等價地,

$$(24) \qquad P(|\overline{X}_n - p| \leq \varepsilon) \geq 1 - \alpha,$$

則n要多大呢? 由(23)式,

$$(25) \qquad P(|\overline{X}_n - p| \leq \varepsilon) \doteq 2\Phi(\varepsilon\sqrt{n}/(p(1 - p))) - 1 \geq 1 - \alpha\text{。}$$

對$\forall\, 0 < y < 1$, 令z_y滿足

$$(26) \qquad \Phi(z_y) = y\text{。}$$

這是辦得到的, 因$\Phi(x)$為一嚴格漸增的連續函數, 且$\lim_{x \to -\infty} \Phi(x) = 0$, $\lim_{x \to \infty} \Phi(x) = 1$。則由(25)式得

$$\frac{\varepsilon\sqrt{n}}{p(1 - p)} \geq z_{1-\alpha/2}\text{。}$$

即

$$n \geq \frac{p(1 - p)z^2_{1-\alpha/2}}{\varepsilon^2}\text{。}$$

由於$p(1 - p) \leq 1/4$, $0 \leq p \leq 1$, 故樣本數

$$(27) \qquad n \geq \frac{z^2_{1-\alpha/2}}{4\varepsilon^2}$$

即可。當然由於(21)式為一近似的式子, 所以此n亦為一近似值。

例如, 若$\varepsilon = 0.03$, $\alpha = 0.05$(這是一般做民調所取的ε及α), 則因$z_{1-\alpha/2} = z_{0.975} \doteq 1.960$, 而$1.960^2/(4 \cdot 0.03^2) \doteq 1{,}067.1$。故取樣本數$n \geq 1{,}068$。

你可能覺得以$1/4$取代$p(1-p)$有時會不夠精準。此點還好, 因這樣會造成得取較大的n, 並不會減小精確性。又(21)式也只是一近似公式而已。但受訪者不見得能誠實回答問題, 民意也隨時會變。因此民調的結果只能供參考。在一些細節過度在意精確度, 可能會見樹不見林。例如, 誤差$\varepsilon = 0.03$看似蠻大的, 可否降低, 如取$\varepsilon = 0.01$? 由(27)式, ε由0.03降至0.01, 樣本數要成為9倍, 達到9,604。對民調而言, 困難度增加很大, 說不定因此造成更大的誤差。以更好的調查設計, 包含更好的問卷題目, 更好的訪員訓練等, 使樣本更隨機, 並設法減小拒訪率, 以得到更接近眞實的意見, 才是較重要的。

表2給出一些常用的z_p之值。至於一般的p, 就要藉助查表或使用計算機而得。

表2. 常用之z_p值

p	0.90	0.95	0.975	0.99	0.995	0.999
z_p	1.282	1.645	1.960	2.326	2.576	3.090

對iid的隨機變數$X_1, X_2, \cdots$, 且設$\mu = E(X_1)$, $\sigma^2 = \text{Var}(X_1)$皆存在, 中央極限定理給出$n$很大時,

$$P\left(\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \le b\right) \doteq \Phi(b),\ b \in R\ 。\tag{28}$$

由上式可得n很大時, $\overline{X}_n$與μ會相距某值以上之機率(即$P(|\overline{X}_n - \mu| > \varepsilon)$, $\varepsilon > 0$)的近似值。這是比大數法則更進一步的結果。但上式左右二機率值的差異有多大呢? 畢竟實際應用時, 我們只會見到有限的n。機率中有一Berry-Essen Theorem, 它指出如果$\gamma = E(|X_1|^3)$亦存在, 則對$\forall b \in R$,

$$\left|P\left(\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \le b\right) - \Phi(b)\right| \le \frac{5\gamma}{\sigma^3\sqrt{n}}\ 。\tag{29}$$

上式給出了(28)式左右二機率差異之一上界。

例5. 設$P(X_1 = 1) = p = 1 - P(X_1 = 0)$。則$\mu = p$, $\sigma^2 = p(1-p)$, $\gamma =$

$E(|X_1|^3) = E(X_1^3) = p$。因此

$$\left| P\left(\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \le b \right) - \Phi(b) \right| \le \frac{5p}{(p(1-p))^{3/2}\sqrt{n}} = \frac{5}{p^{1/2}(1-p)^{3/2}\sqrt{n}}\text{。}$$

當$p = 1/2$, 上式右側成為$20/\sqrt{n}$。$n = 100$時, 此值為2, 這樣的誤差上界自然毫無參考價值。所以對n不是很大時, 不等式(29)用途並不大。不過數學中的不等式常是如此, 普遍性重於精確性。不論分佈為何, 只要三次動差存在, (29)式便適用。普遍性導致(29)式左側之上界有時無法夠精準。但由(29)式亦可得到$n \to \infty$時, $P((\overline{X}_n - \mu)/(\sigma/\sqrt{n}) \le b) \to \Phi(b)$。

3 中央極限定理之圖示

一隨機變數X有$\mathcal{N}(0,1)$分佈, 表其機率密度函數為

$$f_X(x) = \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2},\ x \in R, \tag{30}$$

而分佈函數$\Phi(x) = P(X \le x)$如(10)式。若隨機變數Y之機率密度函數為

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma}\, e^{-(y-\mu)^2/(2\sigma^2)},\ y \in R, \tag{31}$$

便稱有參數μ, σ^2之常態分佈, 以$\mathcal{N}(\mu, \sigma^2)$表之。$\mu$及$\sigma^2$分別為$\mathcal{N}(\mu, \sigma^2)$之期望值及變異數, $\mu \in R$, $\sigma^2 > 0$。經標準化後, $(Y - \mu)/\sigma$有$\mathcal{N}(0,1)$分佈。因此Y之分佈函數亦可以$\Phi(x)$表之。即

$$P(Y \le y) = P\left(\frac{Y-\mu}{\sigma} \le \frac{y-\mu}{\sigma} \right) = \Phi\left(\frac{y-\mu}{\sigma} \right)\text{。} \tag{32}$$

即不論參數為何, 常態分佈的機率值皆可以$\Phi(x)$表之。這是常態分佈方便之處。圖2給出$\mathcal{N}(0,1)$分佈及$\mathcal{N}(\mu, \sigma^2)$分佈之機率密度函數圖形。可看出對一般的$\mathcal{N}(\mu, \sigma^2)$分佈, 其機率密度函數之圖形, 與$\mathcal{N}(0,1)$分佈的機率密度函數之圖形, 形狀是一樣的, 只是經平移與尺度的改變而已。這種圖形為**鐘形曲線**(bell-shaped curve), 且兩側下降至0的速度極快。

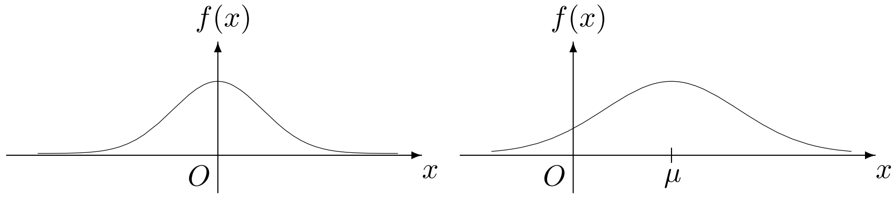

圖2. $\mathcal{N}(0,1)$及$\mathcal{N}(\mu,\sigma^2)$分佈機率密度函數之圖形

很多教科書說樣本數$n \geq 30$, 中央極限定理便適用。即若有iid的隨機變數$X_1, \cdots, X_n$, 且$\mu = E(X_1)$, $\sigma^2 = \mathrm{Var}(X_1)$皆存在, 則當$n \geq 30$, $(\overline{X}_n - \mu)/(\sigma/\sqrt{n})$, 或$(X_1 + \cdots + X_n - n\mu)/(\sigma\sqrt{n})$, 皆可以$\mathcal{N}(0,1)$分佈來近似。若$X_i$只取0, 1二值, 且$P(X_i = 1) = p$, 書上更說, 根據**經驗法則**(a rule of thumb), 只要np及$n(1-p)$皆大於5, 中央極限定理便適用。亦即當$p = 0.2$時, $n \geq 26$即可; 而當$p = 0.5$時, n更只要少少的11以上便可。眞有這麼神奇嗎? 極限下的結果, 卻只要不用太大的樣本數, 便可適用? 事實上, 若X_i's共同分佈之機率密度函數, 對稱於其期望值, n往往不用太大, 中央極限定理便已適用。否則n便要較大些。以p_i表骰子出現i點之機率, $i = 1, \cdots, 6$。於三組不同的p_i's下, 圖3給出$n = 5$, 及$n = 10$, $X_1 + \cdots + X_n$機率密度函數之圖形。若p_i's皆為1/6, 則即使$n = 5$, 此機率密度函數之圖形, 便已有常態分佈的約略形狀。p_i's若為對稱(即$p_1 = p_6$, $p_2 = p_5$, $p_3 = p_4$), $n = 10$時, 此機率密度函數之圖形, 已頗有常態分佈的形狀。而可看出即使p_i's較偏斜, $n = 10$時, 機率密度函數之圖形狀雖仍有些鋸齒狀, 但已大致有常態分佈之形狀。

但如果讓n持續增大, 可能便無法如此圖示中央極限定理了。因X_n或$\overline{X}_n$, 都是要標準化後, 才會近似$\mathcal{N}(0,1)$分佈。詳細討論見黃文璋(2011)一文。

4 信賴區間

中央極限定理在建立**信賴區間**(confidence interval)時, 常也能發揮功能。底下給一情況來說明。

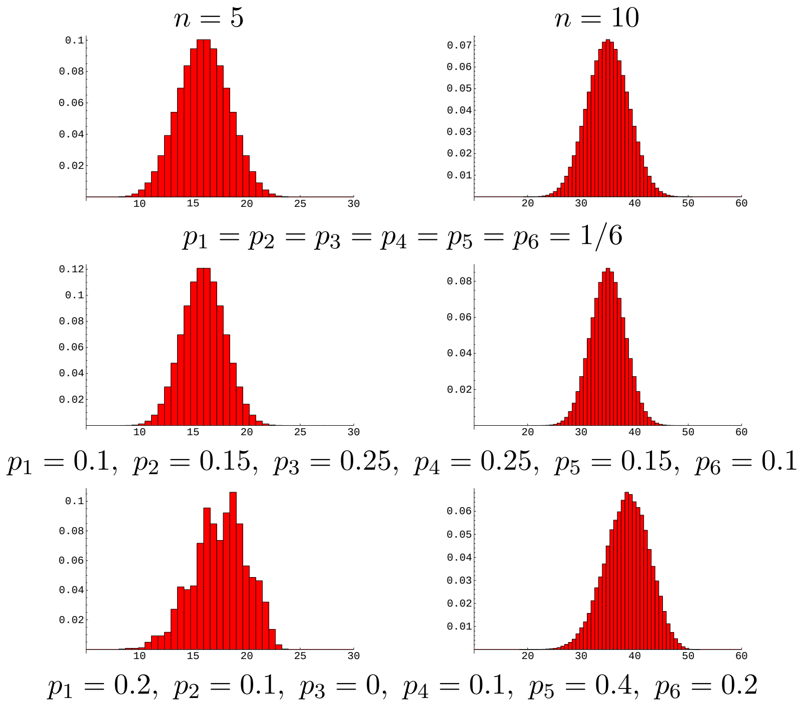

圖3. 各面出現機率不同的3個骰子, 分別投擲5次及10次,
所得點數和機率密度函數之圖形

假設$X_1,\cdots,X_n$為iid之隨機變數, 且$\mu=E(X_1)$, $\sigma^2=\mathrm{Var}(X_1)$皆存在。對於$\mu$, 我們知道常以樣本平均$\overline{X}_n$來估計。有時會想以一區間來估計$\mu$, 這是所謂**區間估計**。即使不知$X_i$'s之共同分佈為何, 但$\sigma$已知, 利用(14)式, 且取$c=z_{1-\alpha/2}$(因此$2\,\Phi(c)-1=2(1-\alpha/2)-1=1-\alpha$), 其中$0<\alpha<1$, 得

$$(33)\quad P\left(-z_{1-\alpha/2}\cdot\sigma/\sqrt{n}\le\overline{X}_n-\mu\le z_{1-\alpha/2}\cdot\sigma/\sqrt{n}\right)\doteq 1-\alpha\text{ 。}$$

上式等價於

$$(34)\quad P(\overline{X}_n-z_{1-\alpha/2}\cdot\sigma/\sqrt{n}\le\mu\le\overline{X}_n+z_{1-\alpha/2}\cdot\sigma/\sqrt{n})\doteq 1-\alpha\text{ 。}$$

當σ已知, 區間

$$(35)\qquad [\overline{X}_n-z_{1-\alpha/2}\cdot\sigma/\sqrt{n},\ \overline{X}_n+z_{1-\alpha/2}\cdot\sigma/\sqrt{n}],$$

便稱參數μ之近似的("近似的"三字常被省略)$1-\alpha$信賴區間。此區間有時以$\overline{X}_n \pm z_{1-\alpha/2}\cdot\sigma/\sqrt{n}$表之。通常$\alpha$為一較小的正數, 如0.01, 0.05, 0.10等。如果σ未知, 以一區間來估計μ, 該區間當然不能有另一未知的參數σ。在此情況下, 常以**樣本變異數**(sample variance)S_n^2之正的平方根S_n取代σ, 其中

$$S_n^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X}_n)^2,\ n\geq 2\ 。\tag{36}$$

S_n^2滿足$E(S_n^2)=\sigma^2$。則μ之近似的$1-\alpha$信賴區間成為

$$[\,\overline{X}_n-z_{1-\alpha/2}\cdot S_n/\sqrt{n},\ \overline{X}_n+z_{1-\alpha/2}\cdot S_n/\sqrt{n}\,]\ 。\tag{37}$$

$1-\alpha$, 或$100(1-\alpha)\%$, 稱為上述信賴區間之信心水準。利用信賴區間的概念, 在例4關於民調的例子裡, 當$\varepsilon=0.03$, $\alpha=0.05$, $n=1{,}068$, 給出支持度p之估計值$\overline{X}_n$後, 在民調裡常說"在95%之信心水準下, 抽樣誤差不超過3%"。這其實就等同於說p之95%信賴區間為$[\overline{X}_n-0.03,\overline{X}_n+0.03]$。

<u>**例6.**</u> 對於一發生機率為p之事件, 經重複觀測, 則可以該事件出現之相對頻率來估計p。我們亦可給出p之信賴區間, 作法類似例4。即設有隨機變數$X_1,\cdots,X_n$, 其中$X_i=1$, 表第i次觀測事件發生, $X_i=0$, 表未發生。則$X_1,\cdots,X_n$為iid, 且$P(X_i=1)=p$, $P(X_i=0)=1-p$。因此$\mu=E(X_i)=p$, $\sigma^2=\mathrm{Var}(X_i)=p(1-p)$。則由(35)式, 且將其中的$\sigma$以$\sqrt{\overline{X}_n(1-\overline{X}_n)}$取代, 得$p$之近似的$100(1-\alpha)\%$信賴區間為

$$\left[\overline{X}_n-z_{1-\alpha/2}\sqrt{\overline{X}_n(1-\overline{X}_n)/n},\ \overline{X}_n+z_{1-\alpha/2}\sqrt{\overline{X}_n(1-\overline{X}_n)/n}\right]\ 。\tag{38}$$

以$\sqrt{\overline{X}_n(1-\overline{X}_n)}$取代$\sigma$, 較以$S_n$取代$\sigma$, 形式較簡單。

對42取6的樂透彩, 開出的頭獎號碼, 連號(含各種可能之連號, 如6連號, 一個3連號一個2連號, 及三個2連號等)發生之機率不小, 大於1/2, 高達0.5568($=$2,921,002/5,245,786)。此利用排列組合便可求出, 並不困難。現假設有某人並未去計算此機率, 他統計了400期, 發現其中有219期

開出的頭獎號碼有連號。因$219/400 = 0.5475$, 且$z_{0.975} \doteq 1.960$, 故由所得數據, 得到連號出現機率之95%信賴區間為

$$\left[0.5475 - 1.960\sqrt{0.5475 \cdot \frac{0.4525}{400}},\ 0.5475 + 1.960\sqrt{0.5475 \cdot \frac{0.4525}{400}}\right]$$
$$\doteq\ [0.4987,\ 0.5963]\ 。$$

實際的連號機率0.5568的確落在此區間。

信賴區間的涵意, 一般人習焉而不察。我們必須略微闡釋些, 也可參考黃文璋(2006)一文。在取樣前(35)及(37)式, 皆為隨機區間。取樣後, 得到$\overline{X}_n = \overline{x}_n$, 代入(35)或(37)式, 則得一常數區間, 如

$$[\,\overline{x}_n - z_{1-\alpha/2} \cdot \sigma/\sqrt{n},\ \overline{x}_n + z_{1-\alpha/2} \cdot \sigma/\sqrt{n}\,]\ 。$$

此區間要嘛包含μ, 要嘛不包含, 說有$1-\alpha$的機率會包含μ, 是不正確的講法。有如一公正的銅板, 出現正、反面之機率各為1/2, 一旦投擲得到正面, 就不能再說此面為正面之機率為1/2。

$1-\alpha$就是機率。如果做很多次實驗, 得到很多信賴區間, 長度皆為$2z_{1-\alpha/2} \cdot \sigma/\sqrt{n}$(或$2z_{1-\alpha/2} \cdot S_n/\sqrt{n}$), 但起始點跟終點可能不同。由機率的意義, 這些區間中大約有$1-\alpha$的比率, 會涵蓋μ。例如, 若$\alpha = 0.05$, 且得到100個信賴區間, 則其中有95個左右會涵蓋μ。實際上有幾個涵蓋μ當然不一定。但在取樣前, 我們知道(忽略這些95%信賴區間只是"近似的"), 會涵蓋μ之區間數, 此隨機變數有$\mathcal{B}(100, 0.95)$分佈。可能95個, 也可能85個(此機率很小, 利用中央極限定理, 此機率小於$\Phi(-4.588) < \Phi(-4) \doteq 3.164 \cdot 10^{-5}$, 更精確值為$\Phi(-4.588) \doteq 2.238 \cdot 10^{-6}$), 或其他值, 但平均有95個。

5 風險

網路上有一"圈中人保險網", 其中有"大數法則"一詞, 其解釋為:

又稱"大數定律"或"平均法則", 是概率論主要法則之

一。此法則的意義是: 在隨機現象的大量重複出現中, 呈現幾乎必然的規律, 這類規律就是大數法則。大數法則是近代保險業賴以建立的數理基礎。根據大數法則的定律, 保的危險單位愈多, 損失概率的偏差愈小。反之, 承保的危險單位愈少, 損失概率的偏差愈大。因此, 保險人運用法則就可以比較精確地預測危險, 合理地釐定保險費率。

上段話在說些什麼呢?

大數法則是保險的基本原理之一。很多有關保險的文獻中, 都會提到大數法則。但其中關於大數法則的敘述, 常不太正確。如有某一人壽公司的壽險百科中, 在一標題為"保險基本原理"的文章中, 總共四個原理, 第二個原理便是大數法則。其說明如下:

> 大數法則是指一件事重覆發生的次數很多時, 其發生的機率就會接近眞實的情形。
>
> 我們用擲骰子來說明"大數法則", 大家都知道骰子擲1, 2, 3, 4, 5, 6點的機率各是六分之一, 可是實際上擲6次卻很難得到1, 2, 3, 4, 5, 6點各一次, 那這個機率到底是如何得來的呢? 以前有位西方數學家, 擲了一萬次, 得出來各點的機率不是等於六分之一, 他又繼續擲, 擲了五萬次, $\cdots$,六萬次, $\cdots$,十萬次, 發現得到1, 2, 3, 4, 5, 6點的機率愈來愈平均, 也就是六分之一。
>
> 大數法則運用在保險上面最常見的就是死亡率, 壽險業利用"台灣壽險業第三回經驗生命表"來計算保費的基準, "台灣壽險業第三回經驗生命表"是以一千萬人為基準的。

事實上對公正的骰子。擲6次, 是遠比擲6萬次, 更容易各面均出現投擲數的6分之1次。希望讀者已能了解。

至於本節一開始所引的那段關於大數法則的解釋, 所提到的"損失概率的偏差愈小", 讀者應已知道, 偏差的大小, 並無法由大數法則得到, 而是必須仰賴中央極限定理。但有關保險的文獻中, 往往只提及大數法則。可

能是"中央極限定理", 光從字面上, 就未讓人覺得這是太淺顯的概念。不像"大數法則"四字, "數大便是美", 好像很淺顯, 令人以為可以輕易地去解釋。

保險, 對保户而言, 當然不是一**公正賽局**(fair game)。以意外險為例, 通常只有死亡或全身殘障才理賠。一年幾千元的保費, 賠償高達數百萬元, 保户似乎頗划得來。保費的計算, 乃根據過去死亡的資料訂出。以期望淨所得而言, 保險公司是不會吃虧的。對保户來說, 保險的目的, 並非為了賺錢。這一年平安無事, 幾千元的保費消失了, 應感慶幸, 他可不想被理賠。

只要保户愈多, 保險公司的風險就愈小。為了簡化, 一切營運成本都不計好了, 且假設保險公司無任何準備金。則若保險公司只有一個保户, 且是保意外險, 則一旦該保户死亡, 保險公司便破產了。例如, 設任何一人, 一年會意外死亡的機率為0.0007。保500萬, 一年保費設為5,000元。對保户, 期望淨所得為$5\cdot 10^6\cdot 0.0007-5{,}000=-1{,}500$(元), 是划不來。但有0.0007的機率, 保險公司會破產。不過實際上不會只有1人投保, 只要保户愈多, 保險公司破產的機率便愈小。保險由於每人保的金額、項目不盡相同; 因工作性質、年齡等原因, 每人保費也可能不同。底下我們以一簡單的例子來顯示"保户愈多, 保險公司破產的機率便愈小"之現象。

例7. 假設一次賭1元, 贏的話得1元, 輸的話, 1元便為賭場收走。設每次賭客贏的機率為p, $0<p<1$, 且賭了n次, $n\geq 1$。即設有iid的隨機變數$X_1,\cdots,X_n$, $P(X_i=1)=p$, $P(X_i=-1)=1-p$。賭客之淨所得為$X_1+\cdots+X_n$。我們想求$P(\sum_{i=1}^n X_i\geq 1)$, 此為賭$n$次後賭客之淨所得為正之機率。由於$\mu=E(X_i)=2p-1$, $\sigma^2=\mathrm{Var}(X_i)=4p(1-p)$, 利用中央極限定理得, 當$n$夠大時,

$$(39)\quad P(\sum_{i=1}^n X_i\geq 1)=P\left(\frac{\sum_{i=1}^n X_i-n(2p-1)}{2\sqrt{p(1-p)}\sqrt{n}}\geq\frac{1-n(2p-1)}{2\sqrt{p(1-p)}\sqrt{n}}\right)$$
$$\doteq 1-\Phi\left(\frac{(n+1)/2-np}{\sqrt{np(1-p)}}\right)\ 。$$

輪盤的歷史悠久, 可說是最古老的賭戲之一。這是源自於17世紀, 法國數學家**巴斯卡**(Blaise Pascal, 1623-1662)所發明的一種遊戲。現假設賭的是**歐式輪盤**(European roulette), 即有1至36的數字, 奇數為紅色, 偶數為黑色, 加上一綠色的0(**美式輪盤**(American roulette)則尚有一00), 0出現算莊家贏。押紅黑中的某一顏色, 賠率為1賠1, 則$p = 18/37$, 約為0.4865, 僅略小於1/2, 每賭1次, 莊家平均只淨贏$1/37 \doteq 0.027$(元), 賭客並未太吃虧(美式輪盤則對賭客更不利), 賭客總不能期望賭場裡提供"公正賽局"。對一些不同的n, 表3給出賭客淨所得為正, 亦即賭場虧錢之機率。若賭的次數不多, 如只賭1次, 賭場有約0.4865的機率會虧, 賭100次, 賭場虧的機率都還有約0.3555。但隨著賭的次數持續增加, 賭場幾乎可說穩賺不賠。

表3. 歐式輪盤賭場虧之機率

賭的次數n	10^2	10^3	10^4	10^5
賭場虧之機率	0.3555	0.1876	$3.327 \cdot 10^{-3}$	$5.997 \cdot 10^{-18}$

由上例, 讀者應可接受只要保戶愈多, 保險公司可說就不用擔心有破產的問題。當然這裡面有一點要留意, 就是保險公司要有相當夠的資金才行。因並非所有保戶同時加入保險, 且保費一次收齊。若才收了幾份保費, 就有人要理賠, 便付不出理賠金了。

例8. 承上例。賭場想知道有99%的信心, 它至少可淨賺多少元?

解. 賭場淨賺至少a元, 即表賭客淨所得少於$-a$元。現要求$P(\sum_{i=1}^n X_i \le -a)$。令

$$\begin{aligned}
&P(\sum_{i=1}^n X_i \le -a)\\
&= P\left(\frac{\sum_{i=1}^n X_i - n(2p-1)}{2\sqrt{p(1-p)}\sqrt{n}} \le \frac{-a-n(2p-1)}{2\sqrt{p(1-p)}\sqrt{n}}\right)\\
&\doteq \Phi\left(\frac{-a-n(2p-1)}{2\sqrt{np(1-p)}}\right) = 0.99\text{。}
\end{aligned}$$

因$z_{0.99} \doteq 2.326$, 由上式得

$$-a - n(2p-1) \doteq 2.326\sqrt{np(1-p)}\ 。$$

故

$$a \doteq n(1-2p) - 2.326\sqrt{np(1-p)}\ 。 \tag{40}$$

即有99%的信心, 賭場約至少可淨贏$(n(1-2p) - 2.326\sqrt{np(1-p)})$元 。

對$p = 18/37$, 及一些不同的n, 表4給出a之近似值 。通常$p < 1/2$, 由(40)式看出, 只要n夠大, a皆為正 。對於$p = 18/37$, $n \to \infty$時, $a \sim n/37$ 。

表4. 歐式輪盤賭場有99%的信心至少可淨贏的金額

賭的次數n	10^2	10^3	10^4	10^5	10^6
賭場	-8.9	-9.7	154.0	2,335.1	25,864.5

6 高斯誤差理論

重覆做一物理實驗, 譬如某天文學家量測兩星球間之距離, 則即使用同樣的儀器及技術, 由於大氣條件的改變等因素, 每次所量得之值, 可能不盡相同 。那該取那一值呢? 假設n次量測後, 得到觀測值$x_1, \cdots, x_n$, 則人們常以這些觀測值之平均, 當做眞實值μ之估計 。即以

$$\overline{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{41}$$

估計μ, 這是一很自然的作法 。此一估計值有下述二特性, 其中$\hat{\mu}$表μ之一估計值 。

(i)滿足估計之誤差和$\sum_{i=1}^{n}(x_i - \hat{\mu})$為0之$\hat{\mu}$即為$\overline{\mu}$;

(ii)滿足估計之誤差平方和$\sum_{i=1}^{n}(x_i - \hat{\mu})^2$最小之$\hat{\mu}$亦為$\overline{\mu}$ 。

估計有時會高估, 有時會低估, 但總的來說, 不應傾向高估或低估, 所以誤

差和應為0。而唯一會使誤差和為0之估計值, 就是$\overline{\mu}$。另外, "絕對誤差之和"要愈小愈好。但絕對誤差之和$\sum_{i=1}^{n}|x_i-\hat{\mu}|$, 處理上較麻煩。因此我們常考慮誤差平方和$\sum_{i=1}^{n}(x_i-\hat{\mu})^2$, 其平方根的位階, 就與$\sum_{i=1}^{n}|x_i-\hat{\mu}|$一樣了。而$\overline{\mu}$也是唯一使$\sum_{i=1}^{n}(x_i-\hat{\mu})^2$最小之估計值。由於有此二特性, 高斯認為觀測值之平均$\overline{\mu}$, 應為眞實值μ之一最佳的估計值。但什麼樣的誤差, 會使$\overline{\mu}$成為一最佳的μ之估計值? 看到一現象, 去想究竟是什麼因, 才會造成這樣的果。科學上除了探討由某個因可得到那些果, 也常由所看到的某個果, 去推測產生的因。

令$X_1,\cdots,X_n$表n次觀測後所得之樣本。假設這些是iid的隨機變數。若令

$$X_i=\mu+Y_i,\ i=1,\cdots,n, \tag{42}$$

此處μ為觀測的眞實值, 為某一定值, 而Y_i即為第i次觀測之誤差, 即

$$Y_i=X_i-\mu,\ i=1,\cdots,n\text{ 。}$$

由於$X_1,\cdots,X_n$為iid, 故$Y_1,\cdots,Y_n$亦為iid。在(i)及(ii)二條件下, 我們想看誤差的分佈會是什麼? 在機率裡, 稱此為分佈之**刻畫**(characterization of distribution)。

令f表$Y_1,\cdots,Y_n$共同的機率密度函數。你應可看出除非我們做一些合理的假設, 否則當然無法決定f。由於量測的特性, 很接近的兩個誤差值, 可合理的假設其發生之可能性應也很接近。為了簡便, 我們乾脆就假設$f(x)>0,\ \forall x\in R$, 且f之一階導數存在且連續(滿足這種性質的函數稱為**平滑的**(smooth))。

這樣還是無法決定f, 你可找到無限多個平滑的機率密度函數。諸如$f(x)=(\pi(1+(x-\theta)^2))^{-1},\ x\in R$, 其中$\theta\in R$為一常數, 皆是。

高斯以如下的手法決定f。

首先$Y_1,\cdots,Y_n$之聯合機率密度函數為$f(y_1)\cdots f(y_n)$。至於觀測到的樣本$x_1,\cdots,x_n$之**概似函數**(likelihood function)為

$$L(\mu|x_1,\cdots,x_n) \tag{43}$$

$$= f(y_1)\cdots f(y_n)$$
$$= f(x_1-\mu)f(x_2-\mu)\cdots f(x_n-\mu),$$

其中$x_i, y_i \in R,\ i=1,\cdots,n$。函數$L$就是將聯合機率密度函數$\prod_{i=1}^n f(y_i)$, 視為一$\mu$之函數, 而將$x_1,\cdots,x_n$固定。此時已並不將$L$當作機率密度函數, 而是$x_1,\cdots,x_n$給定, 且讓$\mu$可以變動。

高斯的想法是, 既然

$$X_1 = x_1, \cdots, X_n = x_n$$

此一事件發生, 而$\overline{\mu}$又被認為是μ之最佳估計值。所以當$\mu=\overline{\mu}$, 應使概似函數L達到最大值。也就說當$\mu=\overline{\mu}$時, 恰會使觀測值$x_1,\cdots,x_n$最易產生。此想法是基於此估計值既是最佳, 那就要$x_1,\cdots,x_n$是最可能被觀測到。一百多年後, 現代統計學的創始者之一**費雪**(Ronald A. Fisher, 1890-1962), 所提出的**最大概似法**(method of maximum likelihood), 就是源自高斯的想法。多了此一假設, 高斯在1809年, 如下導出誤差有常態分佈。

首先L在$\mu=\overline{\mu}$達到極大值, 等價於$\log L$在$\mu=\overline{\mu}$達到極大值。因此

(44) $$\frac{d}{d\mu}\log L(\mu|x_1,\cdots,x_n)\bigg|_{\mu=\overline{\mu}} = 0 \text{。}$$

即

(45) $$\sum_{i=1}^n \frac{f'(x_i-\overline{\mu})}{f(x_i-\overline{\mu})} = 0 \text{。}$$

所以f亦須滿足(44)式。先令$u_i = x_i - \overline{\mu}$, 再令$g(u)=\log f(u)$, 則(44)式成為

(46) $$\sum_{i=1}^n g'(u_i) = 0 \text{。}$$

上式並非對所有$u_1,\cdots,u_n$皆須成立, 而是對$\forall n\geq 1$, 及滿足

(47) $$\sum_{i=1}^n u_i = 0$$

之$u_1, \cdots, u_n$成立即可(因$\sum_{i=1}^n u_i = \sum_{i=1}^n (x_i - \overline{\mu}) = n(\sum_{i=1}^n x_i/n - \overline{\mu})$, 故(41)式與(46)式等價)。要注意的是, 一函數的極值其實不一定發生在微分等於0處。不可微的點(由於假設f可微, 所以上述L沒有不可微的點), 定義域之端點, 皆可能是極值發生處。所以這裡隱含地假設L之極值不會發生在端點。

在(45)式中, 取$n = 2$, 及$u_1 = u$, $u_2 = -u$, 得

$$g'(u) + g'(-u) = 0,\ \forall u \in R \text{。}$$

即g滿足

$$g'(-u) = -g'(u),\ \forall u \in R \text{。} \tag{48}$$

再於(45)式中, 取$n = 3$, 及$u_1 = a + b$, $u_2 = -a$, $u_3 = -b$, 得

$$g'(a+b) + g'(-a) + g'(-b) = 0,\ \forall a, b \in R \text{。}$$

利用(47)式, 上式成為

$$g'(a+b) = g'(a) + g'(b),\ \forall a, b \in R \text{。} \tag{49}$$

滿足(48)式之連續函數(由於假設f'為連續, 所以$g' = f'/f$亦為連續函數)為線性函數(見底下註4)。即

$$g'(u) = c_1 u,\ u \in R, \tag{50}$$

其中c_1為一常數。因此

$$g(u) = \frac{1}{2} c_1 u^2 + c_2,\ u \in R,$$

其中c_2亦為常數。故

$$f(u) = e^{g(u)} = e^{c_2} e^{c_1 u^2/2},\ u \in R \text{。}$$

c_1必須是負的, 否則$\int_{-\infty}^{\infty} f(u)du$會等於$\infty$(之前已假設$f(x) > 0$, $\forall x \in R$)。令$c_1 = -1/\sigma^2$, 其中$\sigma > 0$為一常數。又由$\int_{-\infty}^{\infty} f(u)du = 1$, 可解出$e^{c_2} = (\sqrt{2\pi}\sigma)^{-1}$。即得

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/(2\sigma^2)},\ x \in R \text{。} \tag{51}$$

可驗證對於上述f,

$$L(\mu|x_1, \cdots, x_n) = \frac{1}{(\sigma\sqrt{2\pi})^n} e^{-\sum_{i=1}^{n}(x_i-\mu)^2/(2\sigma^2)},$$

的確在$\mu = \overline{\mu} = \sum_{i=1}^{n} x_i/n$達到極大值。故(51)式所給的$f$, 確實是我們所要的解。

在一些不太強的假設下, 高斯巧妙地導出量測的誤差有期望值為0的常態分佈。這就是**高斯的誤差理論**(Gauss theory of errors), 又稱**誤差法則**(error law), 或**誤差律**。

註4. 設g為R上之一連續函數, 且滿足

$$g(x+y) = g(x) + g(y),\ \forall x, y \in R \text{。} \tag{52}$$

則$g(x) = g(1)x$, $\forall x \in R$。證明如下。

設存在$x_0 \neq 0$, 使得$g(x_0) \neq g(1)x_0$。令$h(x) = g(x) - g(1)x$, 則h亦滿足$h(x+y) = h(x)+h(y)$, $\forall x, y \in R$, 且$h(x_0) \neq 0$。又由$h(nx) = nh(x)$可得$\lim_{n\to\infty} |h(nx_0)| = \infty$。但此乃不可能, 因可如下地證明$h$為有界。易知$h(1) = 0$, 故$h(x+1) = h(x)$, $\forall x \in R$。即h為一週期函數。而h又為$[0,1]$上之一連續函數, 因此在$[0,1]$上必為有界。週期性又導至h在R上為有界, 得證。

另外, 如果g為一定義在$[0,\infty)$上之連續函數, 且滿足$g(x+y) = g(x) + g(y)$, $\forall x, y \geq 0$, 則仿上述推導, 可得$g(x) = g(1)x$, $\forall x \geq 0$。

誤差有期望值為0之常態分佈是否合理? 對於一項量測, 中間有很多步驟, 每一步驟, 都可能產生一些微小的誤差。如果量測沒有**系統性的偏**

差(systematic error), 每一項小誤差都可視為對稱於0。中央極限定理便導致總誤差有期望值為0之常態分佈。高斯的誤差理論是1809年提出, 拉普拉斯則在1812年, 提出以常態分佈來近似二項分佈。所以中央極限定理與誤差理論原本是各自發展。我們現在看到, 二者其實可以相互呼應的。

由於有中央極限定理, 再加上高斯的誤差理論, 使得在隨機現象中, 常態分佈常扮演極重要的角色。有很長一段時間, 常態分佈做為描述隨機現象, 佔著獨尊的地位。有些人甚至誤以為任何隨機的量皆有常態分佈。

7 結語

中央極限定理用途廣泛, 第1節中所提到的布朗運動, 其中之一些理論的推導, 也要用到中央極限定理。此定理與大數法則互相搭配, 使得機率與統計, 更能發揮其功能。

我們有時說某位運動選手在某比賽中表現失常, 或說大失常態。一個人的某項表現, 很可能是隨機的, 在某一眞實值附近波動。中央極限定理, 或高斯的誤差理論, 導致實際觀測值常會有常態分佈, 以$\mathcal{N}(\mu,\sigma^2)$表之, 其中$\mu$可視為表現的眞實值, 或平均值, σ則為表現的標準差。對於常態分佈, 觀測值與μ之差異, 超過一個標準差σ, 機率約0.3173, 算是很容易發生; 超過兩個標準差, 機率就不太大, 才約0.0455; 超過三個標準差, 機率僅約0.0027, 應是夠小了。視不同的情況, 表現在μ之若干個標準差之外, 便稱為**失常**、不正常, 或說**不是常態**。至於表現在μ附近, 當然就稱為常態, 或說正常。很多服飾、家具及物品的設計, 主要都是針對常態(正常)的人。以鞋子為例, 若腳之大小非常態, 過小或過大, 店裡找不到適合的, 只好調貨或訂做。鞋店視其庫存的大小, 只會備有幾個標準差之內的鞋。你的腳大小非常態, 只好委曲些。一個人全身上下, 有許多的特徵。例如, 身高、體重、眼睛大小、英文能力、數學能力、跑步快慢、血糖高低等。仔細列舉, 總可達1,000項, 或更多。某一項特徵落在三個標準差之外(例如, 身高比平均矮三個標準差)的機率, 就以0.0027計。由於特徵多達1,000項, 所以每個人有兩、三項($1{,}000 \cdot 0.0027 = 2.7$)極不

正常的特徵, 可說稀鬆平常。由表1, 落在平均值5個標準差之外的機率, 雖僅約$5.733 \cdot 10^{-7}$, 微乎其微, 但全世界有超過65億($= 6.5 \cdot 10^{9}$)人口, 而$(6.5 \cdot 10^{9}) \cdot (5.733 \cdot 10^{-7}) \doteq 3{,}726$。所以不論那一項特徵, 世界上都有許多"極不正常"(在平均5個標準差之外)的人。當然不正常不一定是壞事, 尤其是那些正向的不正常。音樂神童、圍棋天才、記憶超人、梵谷(Vincent Van Gogh, 1853-1890, 荷蘭著名畫家), 及伍茲(Tiger Woods, 1975-, 美國著名高爾夫球手)等, 都屬在某方面極不正常的人。我們若不想僅當個平庸的人(譬如說, 樣樣表現都在平均的一個標準差內), 就要找出自己那些正向的不正常特徵, 設法突顯, 並往那些特徵發揮。

參考文獻

1. 黄文璋(1995). 隨機過程。華泰文化事業股份有限公司, 台北。
2. 黄文璋(2006). 統計裡的信賴。數學傳播季刊, 30(4), 48-61。
3. 黄文璋(2010). 機率論, 第二版。華泰文化事業股份有限公司, 台北。
4. 黄文璋(2011). 關於中央極限定理之圖示。高中數學電子報第56期 (2011年6月27日)。
5. Tijms, H.(2004). *Understanding Probability*: *Chance Rules in Everyday Life*. Cambridge University Press, Cambridge.