隨機法則

黃文璋教授
國立高雄大學應用數學系

1 律法

自古以來,人們的行爲就受著律法的約束。例如,在舊約聖經出埃及記第十九章,耶和華降臨西奈山頂上,召摩西(Moses)上山頂,要他向以色列人頒佈十誡(The Ten Commandments),以及立下各種典章(These are the laws you are to set before them),讓以色列人遵循。在第二十四章裡記載,耶和華對摩西說"你上山到我這裡來,住在這裡,我要將石版並我所寫的律法和誡命賜給你,使你可以教訓百姓。"摩西和他的幫手約書亞(Joshua)起來,上了神的山。另外,在史記漢高祖本記,劉邦於入咸陽後,與父老約法三章耳"殺人者死,傷人及盜抵罪"。又佛教裡有所謂菩薩十戒,沙彌十戒等。國有國法,家有家法,幫有幫規。以目前台灣而言,除了憲法及各種法律外,學校裡學生手册上還列有各種章則。這種種的法律及規定,都是要人們遵守,是一種人爲的限制,若不遵守,可能就會有一些處罰。

大家可能也聽過所謂叢林法則。在叢林裡,各種動物間,長久以來,也自然形成一些各自須遵守的規則,而達到一種平衡的狀態。

科學裡的各種律法也不少。如在實數中有交換律(commutative law): $x+y=y+x$, $xy=yx$; 結合律(associative law): $x+(y+z)=(x+y)+z$, $x(yz)=(xy)z$;分配律:$x(y+z)=xy+xz$ 等。至於物理中的萬有引力定律,熱力學定律等,都是大家耳熟能詳的。

人所定出來的以及動物間自然形成的律法,這些不計好了,因隨著環境的變遷及想法的改變,這些律法有可能被更改。至於數學及自然科學中的各種律法,則皆屬必然性,是毫無意外的成立。本文所要討論的則是隨機中的律法。這是並非人爲所訂出來的,而是自然產生的律法。人的貢獻只是

將這些律法找出,並證實這些律法的確是對的。

在數學裡講究的是不變,放諸四海皆準的律法,因爲那是一個必然性的世界。如$0.2+0.6=0.8$, $0.6>0.2$ 等。沒有人會對這類式子存疑。但在隨機世界裡,就不是這麼簡單了。我們可能會問,眞是0.2嗎?如銅板出現正面的機率究竟是否眞爲0.2? 兩個事件的機率,也不要冒冒失失地相加,而要留意其含義。如可不可以相加?相加後的值又到底代表什麼?再如銅板 A 出現正面機率爲0.6,銅板 B 出現正面機率爲0.2, 0.6雖大於0.2,但若分別丟此二銅板,是否能保證銅板 A 會較銅板 B 先出現一正面?學過機率論的人當然知道不一定。那 $0.6>0.2$ 在此處的意義爲何呢?歸根究底,就牽涉到機率究竟是什麼?了解其意義後,才能理解機率值較大的含義。

在隨機世界裡,要遵循的是隨機法則,必然世界裡的那些數學中的律法,當然還是要藉助的。但若僅以數學律法來看隨機世界的問題,那學到的還是數學,不是隨機了。由於學生初次接觸機率統計,往往是在中學數學課程,所以對隨機性的內涵,常未能充分了解,只徒然以爲機率統計爲一類簡單的數學。因此雖唸過不少機率統計,但若都是透過數學的眼光來學習,效果即使不是零也很接近零了。

2 機率的意義

機率論裡的一主要工作,是探討隨機中的律法,將隨機中的必然行爲找出來。

隨機中還有必然性?司馬遷在"報任少卿書"一文(見古文觀止),說他致力完成史記,乃是"亦欲以究天人之際,通古今之變,成一家之言"。我們也可以這麼說,學機率論的目的,是爲了探究隨機中的必然性,掌握各種隨機現象的變化,而成爲一能隨機思考的人。

有許多現象,其發生的情況,看起來似乎是毫無規律的(所謂隨機現象)。譬如說,有一號稱是公正的銅板,你丟了10次,得到7個正面3個反面,難免懷疑其公正性。但另一人丟,卻得到2個正面8個反面。公正到底是什麼?你感到很納悶。有些學生不太用功,卻振振有辭地說"用功有什麼用,你

看xxx唸了那麼多,也考不好”。在棒球比賽裡,以往打擊率很高的球員,有時在接連幾場比賽中,一支安打也沒有。於是每逢他出場打擊,觀眾給他不少噓聲。

上述這類現象,可說是我們日常生活裡,屢屢會遇到的。我們當然可含混地將這種不規律性歸之於隨機現象的特性。那談機率又有何意義呢?

報紙每日刊登各地當日的降雨機率,我們並不清楚那些機率值是如何產生,也少有人去討論如何判定其正確性。但當一機率學家說“某銅板出現正面的機率爲 p”,你當然可理直氣壯地問他,機率爲 p 的含義是什麼?

對機率的解釋有一些不同的方式。譬如有所謂“主觀的觀點”(subjective point of view),可以由過去的經驗,或不管三七二十一地認定某事件的機率就是 p。另一種是所謂“客觀的解釋”(objective interpretation),以多次重覆實驗後,一事件出現的相對頻率(relative frequency)來表示機率,又稱爲頻率對機率的解釋(frequentistic interpretation of probability)。

頻率對機率的解釋乃是基於大數法則(Law of Large Numbers)。

大數法則爲機率論中一很基本且重要的定理,此問題之歷史如下。在十七世紀末及十八世紀初,瑞士數學家James Bernoulli(1654-1705)證明了一掛他名字的定理。此Bernoulli 定理於西元1713年,他死後才出版的機率論最早的著作: Ars Conjectandi(The Art of Conjecturing),才首度爲世人所知。Bernoulli考慮的是所謂伯努力(Bernoulli)隨機數列 $\{X_n, n \geq 1\}$,即 X_n 只取0,1兩個值。這是大數法則之一最簡單的特例。在十九世紀初,法國數學家Poisson(1781-1840),在較一般的條件下(即允許 $P(X_n = 1)$ 對不同的 n 可以不同),證出了一類似的定理。然後一直到西元1866年,俄國數學家Chebyshev(1821-1894),用他發明的不等式證出了他的版本。稍後俄國數學家Markov(1856-1922),發現Chebyshev的結果可更一般些。

這以後大數法則便沒什麼突破,直到西元1926年,年輕的俄國數學家Kolmogorov(1903-1987),得到獨立隨機變數會遵循大數法則的充要條件。西元1928年, Khinchine(1894-1959,爲俄國最重要的機率學家之一)證明對於獨立且有共同分佈(簡稱i.i.d.)之隨機變數,只要期望值存在,便可適用大數法則,此即底下的定理2。

假設重覆地做一試驗,並觀測其中一特定的事件 A。譬如說持續地丟一骰子,並觀測偶數是否發生,則 A 即爲出現的結果爲偶數。再定義隨機變數 $X_n, n \geq 1$, $X_n = 1$ 或0,就依第 n 次試驗的結果是 A 或 A^c(A 之餘集,表 A 未發生), $\{X_n, n \geq 1\}$ 即爲一組i.i.d.的伯努力隨機變數。則 n 次試驗後,事件 A 發生之相對頻率爲

$$\frac{n_A}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i, \tag{1}$$

其中 n_A 表 A 發生之總次數。由實際的經驗中,我們注意到當 n 很大時, n_A/n 似乎逐漸"趨近"某一定值。若以該定值,當作事件 A 之機率,便是頻率對機率的解釋。而等價地,人們等於觀察到對一組i.i.d.的伯努力隨機變數,其樣本平均(sample mean)$\sum_{i=1}^{n} X_i/n$,會隨著n之增大,而逐漸"趨近"某一定值。此定值剛好是 A 發生的的機率 $P(A)$, 又等於 X_i 之期望值 $E(X_i) = 1 \cdot P(A) + 0 \cdot P(A^c)$。由於只要樣本數 n 夠大,便有此性質,所以才稱爲大數法則。

以這種相對頻率的觀點來解釋機率,是廣爲機率學家所接受的一種方式。我們不是看一次的結果,不爭一時,但爭千秋。只要次數 n 夠大, n_A/n 就會"接近"$P(A)$。若 n 很大了, n_A/n 仍離所宣稱之 $P(A)$ 很遠,那便可能懷疑 $P(A)$ 是否正確?

在第1節裡我們提到出現正面機率分別爲0.6及0.2之二銅板 A 及 B。若只丟幾次,銅板 A 出現的正面數 n_A,不一定比銅板 B 出現的正面數 n_B 多,但因 n 夠大時, $n_A/n, n_B/n$ 分別會"接近"0.6及0.2,因此只要丟得夠多次,我們若要猜那一個銅板得到較多的正面數,當然要猜 A。這是機率較大的意義。

對一離散型的隨機變數 X,當我們說 X 遵循某一機率法則,譬如說 X 取值在 $\{a_k, k \geq 1\}$, 且設 $P(X = a_k) = p_k$,這是什麼意思呢?可以如下地解釋:設 $X_1, X_2, \cdots, X_n$ 爲i.i.d.之隨機變數,且對 $\forall i \geq 1$, $P(X_i = a_k) = p_k, k \geq 1$。則對$\forall k \geq 1$,

$$\frac{[X_i = a_k]\text{之}i\text{的個數}}{n}$$

當 n 很大時,會"接近"p_k。還是用頻率來解釋。當 n 不太大時,會有起伏的結果,但只要 n 夠大,規律性便產生了,這是機率裡的法則,是在大樣本下的結果。只要樣本數夠多,它必然有這種結果。

對一任給的隨機變數 X,因對一 $a \in R$, $P(X=a)$ 可能爲0,此時不妨考慮一機率值爲正的區間 $A=[c,d]$, $P(A)>0$,則便可有類似的機率法則。

對任一隨機變數 X,所謂 X 之法則(the law of X,或說 X 之機率法則),以$\mathcal{L}(X)$ 表之,表 X 之分佈函數(distribution function),或機率密度函數(probability density function,簡稱p.d.f.),或特徵函數(characteristic function)等可以唯一決定 X 之分佈者(見Tucker(1967)p.99)。例如,若說 $\mathcal{L}(X)$ 爲$\mathcal{N}(\mu,\sigma^2)$,表 X 有 $\mathcal{N}(\mu,\sigma^2)$ 分佈。大數法則,便對 X 是否遵循某一法則(即是否有某一分佈),提供理論依據。

3 大數法則

在上節中,我們屢屢提到接近及趨近。希望讀者警覺性夠高,會問如何接近?以及如何趨近?要知我們是在討論機率,所以接近及趨近,都是在機率之下。但這中間還是有些細節要特別留意,不是光一個機率之下便可一筆帶過。

我們先給常用的Chebyshev不等式(Chebyshev's Inequality)。

定理1.設 X 爲一隨機變數,且設 $\mu = E(X), \sigma^2 = \mathrm{Var}(X)$ 皆存在,其中 $\sigma > 0$。則對 $\forall k > 0$,

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

證明.對一事件 A,令 I_A 表 A 之指示函數(indicator function),顯然$I_A + I_{A^c} = 1$。取 $A = \{|X-\mu| \geq k\sigma\}$,則

$$\begin{aligned}\sigma^2 = E(X-\mu)^2 &= E(|X-\mu|^2 I_A) + E(|X-\mu|^2 I_{A^c}) \\ &\geq E(|X-\mu|^2 I_A) \geq k^2\sigma^2 E(I_A)\end{aligned}$$

$$= k^2\sigma^2 P(|X-\mu| \geq k\sigma) \text{ 。}$$

得證。

我們知道,對一隨機變數 X,期望值 μ 表 X 之集中處,標準差 σ 則爲量測X 與 μ 差距的大小。(2)式給出 X 偏離期望值 k 個標準差之機率的一上界。此不等式之特點在其普遍性(適用變異數存在且不爲0之每一隨機變數),而非精確性((2)式右側有時比左側大很多)。

例1.設 $X_1, X_2, \cdots$ 爲i.i.d.之隨機變數,且設 $E(X_1) = \mu$, $\mathrm{Var}(X_1) = \sigma^2$ 皆存在。令 $S_n = \sum_{i=1}^n X_i$, $\overline{X}_n = S_n/n$,則 $E(\overline{X}_n) = \mu$, $\mathrm{Var}(\overline{X}_n) = \sigma^2/n$。由Chebyshev不等式得

$$P(|\overline{X}_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2} \text{ 。} \tag{3}$$

故對 $\forall\varepsilon > 0$,

$$\lim_{n\to\infty} P(|\overline{X}_n - \mu| \geq \varepsilon) = 0 \text{ 。} \tag{4}$$

(4)式指出,在例1的條件下, $n \to \infty$ 時, S_n/n 機率收斂(converge in probability,以符號 P 表之)至 μ。又(4)式可如下地解釋: ε 可視爲欲以樣本平均 $\overline{X}_n$ 來估計 μ 所要求之精確度, (4)式保證不論 ε 選的多小, $\overline{X}_n$ 與 μ 之差距會落在此精確度內之機率,即 $P(|\overline{X}_n - \mu| < \varepsilon)$,當觀測數 n 逐漸增大,會趨近至1。直觀上此結果是對的,因對i.i.d.之隨機數列 $\{X_n, n \geq 1\}$,其和S_n 之期望值及變異數,分別爲 X_1 之期望值及變異數的 n 倍,即 $E(S_n) = nE(X_1)$, $\mathrm{Var}(S_n) = nE(X_1)$。$\overline{X}_n$ 之期望值與 X_1 之期望相同,但變異數卻成爲 $1/n$ 倍,隨著 n 之增大而減小且趨近至0。期望值爲 $E(X_1) = \mu$,且變異數趨近至0,表 n 很大時, $\overline{X}_n$ 不會太偏離 μ,而這就是(4)式的意思。此現象稱爲弱大數法則(Weak Law of Large Numbers)。我們等於是在 X_1 之變異數存在的假設下,證出弱大數法則。事實上此條件是多餘的,只要 X_1 之期望值存在即可。我們敘述此定理如下,證明則略去,可參考黃文璋(1994)pp.254-255。

定理2.(弱大數法則).設 $X_1, X_2, \cdots$ 爲i.i.d.之隨機變數,以 μ 爲其共同期望值。則 $\overline{X}_n$ 機率收斂至 μ。即

$$\overline{X}_n \xrightarrow[n \to \infty]{P} \mu \text{。}$$

只要 X_1 之期望值存在, (4)式便成立。但若變異數也存在,則(3)式成立。(3)式爲一較明確之敘述,因對 $P(|\overline{X}_n - \mu| \geq \varepsilon)$ 給出一以 n 表示之上界。

弱大數法則對頻率對機率的解釋,提供了一理論依據。當 $n \to \infty$ 時, (1) 式右側會機率收斂至 $E(X_1) = p = P(A)$。故 n_A/n 機率收斂至 $P(A)$。因此 n 很大時,可放心地以 n_A/n 來估計(estimate)$P(A)$。

弱大數法則爲大數法則之一較弱的形式。我們常說當 n 很大時, $\overline{X}_n$ 會以某種方式接近 μ,現在總算知道"某種方式"之一了。由於 $\overline{X}_n$ 爲隨機變數,所以我們不能如數列中的極限寫成

$$\lim_{n \to \infty} \overline{X}_n = \mu,$$

或說

$$\lim_{n \to \infty} |\overline{X}_n - \mu| = 0 \text{。}$$

而是說$|\overline{X}_n - \mu| \geq \varepsilon$ 的機率會趨近至0。

隨機變數 $\{X_n, n \geq 1\}$ 不獨立時,弱大數法則亦可能成立,見黃文璋(1994)第5.3節。也就是弱大數法則之適用性還要更廣些。由於本文只是初步的介紹,我們僅列出一較簡單且常見的弱大數法則的版本。

例2. 設某人參與一賭局, 第n 次之淨所得以X_n 表之, $n \geq 1$, 又設$\{X_n, n \geq 1\}$ 爲i.i.d.。令$S_n = \sum_{i=1}^{n} X_i, n \geq 1$, 表$n$ 次後之總所得, $\overline{X}_n = S_n/n$ 表n 次後之平均所得。以$\mu = E(X_1)$ 表每次之期望所得。若$\mu > 0$, 由定理2得, $n \to \infty$ 時,

$$P(S_n \geq n\mu/2) \geq P(|\overline{X}_n - \mu| \leq \mu/2) \to 1 \text{。}$$

也就是若進行賭局的次數夠多(n 很大), 則會淨贏一任意大的數目(至少$n\mu/2$) 的機率將很大。若$\mu < 0$, 則情況剛好反過來。當$\mu = 0$, 通常便稱此爲一公正的賭局。但是"公正"這個用語有時卻沒那麼恰當, 尤其是當$\mathrm{Var}(X_1) = \infty$ 的時候。見黃文璋(1994)pp.264-265之說明。

大數法則有一更強的形式。法國數學家Borel(1871-1956,爲現代機率論的創始者之一),在西元1909年,首先證出i.i.d.伯努力隨機變數的結果。一般教科書常見的版本如下,證明見Chung(1974)Theorem 5.4.2。其他形式的強大數法則(Strong Law of Law Numbers)亦見Chung(1974)。在此幾乎確實地收斂(convergence almost surely(簡稱a.s.),又稱convergence almost everywhere(a.e.), convergence with probability one(w.p.1)。

定理3.(Kolmogorov強大數法則).設 $\{X_n, n \geq 1\}$ 爲i.i.d.之隨機變數,令 $\overline{X}_n = \sum_{i=1}^{n} /n, n \geq 1$。又設 $E(X_1) = \mu$ 存在。則 $\overline{X}_n$ 幾乎確實地收斂至 $E(X_1)$,即

$$\overline{X}_n \xrightarrow[n \to \infty]{a.s.} E(X_1)\text{。} \tag{5}$$

粗略地講,定理3指出,對"幾乎所有"的樣本數列 $X_1, X_2, \cdots$,隨著n之增大,前 n 個樣本的平均 $\overline{X}_n$,愈來愈接近 $E(X_1)$。也就是"大多數的時候",隨機變數 $X_1, X_2, \cdots$ 之觀測值 $x_1, x_2, \cdots$,會滿足只要 n 夠大,則

$$E(X_1) - \varepsilon < \bar{x}_n < E(X_1) + \varepsilon, \tag{6}$$

其中 $\bar{x}_n = \sum_{i=1}^{n} x_i/n$,而 ε 爲一任給的正數。

對未學過較高等的機率論的讀者,欲使其了解"幾乎所有"或"大多數的時候"的概念,是有些困難。幸好在統計裡,通常只要懂弱大數法則便足夠了。(5)式是說

$$P(\lim_{n\to\infty} \overline{X}_n = E(X_1)) = 1\text{。} \tag{7}$$

即會使 $\{\overline{X}_n, n \geq 1\}$ 趨近至 $E(X_1)$ 之樣本的集合之機率爲1,而不會使$\{\overline{X}_n, n \geq 1\}$ 趨近至 $E(X_1)$ 之樣本的集合之機率爲0。也許只要注意到

$$\lim_{n\to\infty} \overline{X}_n = E(X_1) \tag{8}$$

爲一事件,有時成立有時不成立, (7)式就是要求(8)中之事件發生之機率爲1。

利用大數法則,對一可重覆觀測的事件,其發生之機率,可由發生次數之相對頻率來估計。所謂一事件發生之機率,不再只是一理論上的值,而是可經由多次的觀測而看出來。在統計上,大數法則成爲藉助數據分析來估計一未知參數(parameter)之基礎。

4 中央極限定理

與大數法則並列爲機率論中兩個最重要的結果,就是中央極限定理(Central Limit Theorem)。這是隨機世界裡另一影響深遠的法則。

二項分佈(binomial distribution)是機率論早期發展裡最常遇到的一個分佈。除了源自於丟銅板及丟骰子等,如在大數法則中的討論,有時我們關心某特定的事件 A 是否發生,也就是我們看到的結果只有 A 發生或 A^c 發生。則在 n 次觀測後, A 發生的次數 S_n 有二項分佈 $\mathcal{B}(n,p)$,其中 $p=P(A)$。若以 X_i 表第 i 次之結果, $X_i=1$ 或0, 依在第 i 次觀測中, A 是否發生。則 $S_n=\sum_{i=1}^n X_i$。大數法則說

$$\frac{S_n}{n} \xrightarrow[n\to\infty]{a.s.} p\text{。} \tag{9}$$

或者可以說(這要用到機率論裡關於收斂的結果)

$$\frac{S_n-np}{n} \xrightarrow[n\to\infty]{a.s.} 0\text{。} \tag{10}$$

有時我們並非只對平均 S_n/n 之值有興趣,而是想知道 S_n/n(或 S_n)會落在某一區間之機率。

由於 $\mathcal{B}(n,p)$ 分佈之期望值爲 np,所以我們知道 S_n 大約是在 np 上下。但對較大的 n,在計算機率時,常倍感吃力。例如,丟一公正的銅板1,000次,則所得之正面數 $S_{1,000}$ 有 $\mathcal{B}(1{,}000,1/2)$ 分佈。故 $S_{1,000}$ 介於490至515之機率爲

$$P(490\le S_{1,000}\le 515)$$

$$= \sum_{i=490}^{515} \binom{1,000}{i} (\frac{1}{2})^{1,000} 。$$

對計算 $\binom{1,000}{490}$,相信便不會是一太令人高興的事了。

那些令人敬佩的機率學家,發現對任二非負整數 k 及 j,只要 n 夠大,則

$$\begin{aligned}
(11) \qquad & P(j \le S_n \le k) \\
& = P(j - 0.5 \le S_n \le k + 0.5) \\
& = P(\frac{j - np - 0.5}{\sqrt{np(1-p)}} \le \frac{S_n - np}{\sqrt{np(1-p)}} \le \frac{k - np + 0.5}{\sqrt{np(1-p)}}) \\
& \doteq \Phi\left(\frac{k - np + 0.5}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{j - np - 0.5}{\sqrt{np(1-p)}}\right),
\end{aligned}$$

其中對 $\forall x \in R$,

$$(12) \qquad \Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du$$

表標準常態分佈(standard normal distribution,爲法國的Poincaŕe(1854-1912)所取的名稱。Poincaŕe被認爲是最後一位興趣廣泛(universalist)的數學家) $\mathcal{N}(0,1)$ 之分佈函數。也就是以標準常態分佈來做爲二項分佈機率之估計,讓求二項分佈的機率,變成不再是負擔了。又在(11)中,之所以會有 $+0.5$ 與 -0.5,是爲了使此近似更精確,當 n 很大時,兩個0.5略去並無甚影響。

上述二項分佈趨近至常態分佈的結果,爲De Moivre (1667-1754)及Laplace (1749-1827)所先後證出(又稱De Moivre-Laplace Theorem)。De Moivre,這位被視爲牛頓(Newton,1642-1727) 之後繼者,給出 $p = 1/2$ 的結果於西元1714年出版的Doctrine of Chances一書中。而有法國牛頓之稱的Laplace,在其於西元1812年所出版之不朽的著作Théorie Analytique des Probabilités,推廣De Moivre之結果至一般的 p。Laplace也了解到此結果之重要性。這可說是底下我們要討論的中央極限定理(此名稱爲Pólya (1887-1985,當代著名分析學家,在機率論及組合學亦有很多重要成果)於西元1920年所取的)之最早被知道之特例。常態分佈也是這樣被引進機率理論中的。

De Moivre-Laplace版的中央極限定理之推廣,爲機率論中一重要的課題。Chebyshev利用動差法(method of moments),於西元1887年,給出對一般的隨機變數成立之嚴密證明。俄國數學家Liapounov(1857-1918),於西元1900及1901年,在較一般的條件下,利用特徵函數給出一現代的證明。之後當然尚有繼續的推廣。底下爲一今日最常見的版本,適用於任何變異數存在之i.i.d.的隨機數列。

定理4.設 $\{X_n, n \geq 1\}$ 爲i.i.d.之隨機變數,且設 $E(X_1) = \mu, \mathrm{Var}(X_1) = \sigma^2 > 0$ 皆存在。令 $S_n = \sum_{i=1}^{n} X_i$,則$(S_n - n\mu)/(\sigma\sqrt{n})$ 會分佈收斂(converge in distribution)至標準常態分佈,以

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow[n \to \infty]{d} \mathcal{N}(0,1)$$

表之。即對 $\forall b > a$,

$$\lim_{n\to\infty} P(a < \frac{S_n - n\mu}{\sigma\sqrt{n}} \leq b) = \Phi(b) - \Phi(c)\text{。} \tag{13}$$

定理4也可表示成樣本平均的結果,即(13)式等價於

$$\lim_{n\to\infty} P(a < \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \leq b) = \Phi(b) - \Phi(c), \tag{14}$$

其中 $\overline{X}_n = S_n/n$ 。

定理4指出,將 S_n 減去其期望值,除以標準差(此過程稱爲將 S_n 常態化(normalized)),其分佈可以標準常態分佈來逼近。

定理4由於極限分佈爲常態分佈, 所以若稱爲常態收斂定理(Normal Convergence Theorem) 或許更適合。中央極限定理是用在處理獨立的隨機變數之和, 這些變數原先可能是某項觀測所產生之誤差。此定理指出,在適當的條件下,常態化誤差總和之極限,遵循常態分佈。而定理4有人也把它當做是一種自然的法則(Law of Nature)。此定理可有許多不同方向的推廣, 但若假設任意地改變, 則中央極限定理可能就不再成立了。事

實上有許多隨機現象,其極限分佈並非常態分佈,稀有事件法則(Law of Rare Events)即爲一例。

自(13)式得到對 $\forall\, b \in R$,

$$(15)\quad P(S_n \le b) = P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \le \frac{b - n\mu}{\sigma\sqrt{n}}\right) \doteq \Phi((b - n\mu)/(\sigma\sqrt{n}))\text{。}$$

因此我們大約可以說,當 n 很大時, S_n 之分佈近似於期望值爲$n\mu$, 變異數爲$n\sigma^2$ 之常態分佈。亦即S_n 之近似分佈只與$X_1, X_2, \cdots, X_n$ 之共同期望值μ 及變異數σ^2 有關, 而與它們的分佈無關。這是中央極限定理一特殊之處。另外我們也可給出中央極限定理的收斂速率(rate of convergence),此即Berry-Essen Theorem, 其證明則略去。

定理5. 除了定理4之假設外, 再令$\gamma = E(|X_1^3|) < \infty$。則對$\forall b \in R$ 及$n \ge 1$,

$$(16)\qquad \left|P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \le b\right) - \Phi(b)\right| \le \frac{5\gamma}{\sigma^3\sqrt{n}}\text{。}$$

並沒有一通用的規則告訴我們樣本數多大時,中央極限定理才會對S_n(或 $\overline{X}_n$)提供一很好的估計。不過若 $X_1, X_2, \cdots, X_n$ 之共同分佈不要過度"偏斜"(skewed,即不對稱),且是單峰的(unimodal,即p.d.f.只有一極大值),則經驗顯示,當樣本數$n \ge 30$,以常態分佈來估計關於 S_n(或 $\overline{X}_n$)之機率值時,通常至少到小數一位精確。在統計裡常要做各種估計,估計的精確性是很重要的。定理5指出,以中央極限定理來估計關於 S_n 之機率,其誤差成長的速率之階數(order)不超過 $n^{-1/2}$。

底下給一中央極限定理之應用。

例3. 假設有n 個正數, 每一數若其小數部分大於或等於0.5, 則進位至下一整數, 若小於0.5, 則捨去小數部分。例如$7.5 \to 8, 3.472 \to 3$。經過這樣概算後,我們想看總和會產生的誤差有多大?

解.令X_i 表第i 個數產生之誤差, 且設$X_1, X_2, \cdots, X_n$ 爲i.i.d. $\mathcal{U}[-1/2, 1/2)$ 之隨機變數。則$E(X_1) = \mu = 0, \mathrm{Var}(X_1) = \sigma^2 = 1/12$。設$n$夠大, 則可利

用中央極限定理來估計誤差和$S_n = \sum_{i=1}^{n} X_i$。例如, 若$n = 12$,則

$$P(-1 < S_n \leq 1) = P(-1 < \frac{S_n - n\mu}{\sigma\sqrt{n}} \leq 1) \doteq \Phi(1) - \Phi(-1) \doteq 0.6827 \text{。}$$

一般而言,若n 夠大,則對 $\forall \alpha > 0$,

$$P\left(-\alpha\sqrt{\frac{n}{12}} < S_n \leq \alpha\sqrt{\frac{n}{12}}\right) \doteq \Phi(\alpha) - \Phi(-\alpha) \text{。} \tag{17}$$

買東西在結帳時,若買的物品多些,有時也弄不清楚店員是否算對,爲了風度,也不好意思拿出計算機來按,但自己心算又沒那麼好。一個方法是求其大概的和。上例是將小數部分四捨五入,視實際情況,也可將個位四捨五入,或末兩位四捨五入。但如此一來便造成一些誤差。誤差大小是否合理呢?上例便是利用中央極限定理來估計這種概算法的誤差。例如, $n = 12$ 時,若誤差不超過1,大約可以接受。若誤差超過2,由於此機率約爲

$$1 - (\Phi(2) - \Phi(-2)) \doteq 1 - 0.9545 = 0.0455,$$

算是很小,便可能要懷疑總和有誤了。

例4.高斯誤差理論(Gauss theory of errors). 重複做一觀測, 譬如說量測一物體之長度, 則即使一切條件不變, 每次所得之值可能不盡相同。事實上所得之觀測值散佈在期望值$E(X)$ 附近。高斯稱觀測值與期望值之差$X - E(X)$ 爲隨機誤差(random error)。期望值$E(X)$ 當然不見得等於眞實的值, 因測量說不定會產生有系統的(systematic)誤差。此種問題我們不去討論, 我們只考慮隨機誤差$X - E(X)$ 。在誤差理論中, 通常假設隨機變數X 之期望值及變異數均有限, 有時對誤差的分佈也做一些假設, 不過此處我們不假設知道誤差的分佈。

誤差理論是高斯對機率論的主要貢獻, 因研究此理論他並發展出最小平方法(method of least squares)。統計裡一主要的工作便是估計,既是估計便有誤差,而掌握誤差的大小當然很重要。底下我們便看看中央極限定理在此理論的應用。

設重複做一觀測n 次, 而得到觀測值$X_1, X_2, \cdots, X_n$, 並設這些是i.i.d. 之隨機變數, $E(X_1) = \mu, \mathrm{Var}(X_1) = \sigma^2 < \infty$ 。若令

$$X_i = \mu + \varepsilon_i, i = 1, 2, \cdots, n, \tag{18}$$

則$\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n$ 亦爲i.i.d.,且$E(\varepsilon_1) = 0, \mathrm{Var}(\varepsilon_1) = \sigma^2$ 。此處μ 可想成觀測量的實際值, 而ε_i 即爲第i 次觀測之誤差 。

我們通常用觀測的平均值 $\overline{X}_n = \sum_{i=1}^n X_i/n$,當做$\mu$ 之估計值 。則由弱大數法則

$$\overline{X}_n \xrightarrow[n \to \infty]{P} \mu \text{ 。}$$

但是$\overline{X}_n$ 當然並不剛好是μ, 因此我們想知$\overline{X}_n$ 會偏離μ 的情況 。例如, 我們可能想估計誤差至少會是a 之機率值,$a > 0$ 。即想估計

$$P(|\overline{X}_n - \mu| \geq a) \text{ 。} \tag{19}$$

由中央極限定理, 當n 夠大,便有

$$P(|\overline{X}_n - \mu| \geq a) \doteq 2(1 - \Phi(a\sqrt{n}/\sigma)) \text{ 。} \tag{20}$$

另外, 有時我們會問, 要觀測幾次, 才可使(19) 中之機率小於α? $1 - \alpha$ 在統計上稱爲顯著水準(significance level), 或信賴係數(confidence coefficient),此因

$$P(|\overline{X}_n - \mu| \leq a) \geq 1 - \alpha \text{ 。} \tag{21}$$

由(20), 令

$$2(1 - \Phi(a\sqrt{n}/\sigma)) = \alpha,$$

再取

$$n \geq \frac{\sigma^2}{a^2} z_{1-\alpha/2}^2, \tag{22}$$

即得所需之最少觀測次數,其中

$$z_\alpha = \Phi^{-1}(\alpha),\ 0 < \alpha < 1, \tag{23}$$

即 $\Phi(z_\alpha) = \alpha$。舉例而言, 若$a = 0.1\sigma, \alpha = 0.01$, 則$z_{1-\alpha/2} = z_{0.995} \doteq 2.576$,因此$n \geq 100 z_{0.995}^2 \doteq 676$。

在給下例之前, 我們先介紹信賴區間(confidence interval) 的概念。設X 爲一期望值爲μ, 變異數爲σ^2 之隨機變數。一個得到μ 的方法爲進行一數列之獨立的觀測, 且得到n 個i.i.d. 的隨機變數$X_1, X_2, \cdots, X_n$, 均與X 有相同的分佈。再以樣本平均$\overline{X}_n$ 來估計μ。先固定一信賴係數$1-\alpha$, α 通常取得很接近0, 再找出a 之值, 使得

$$P(|\overline{X}_n - \mu| \leq a) = 1 - \alpha \text{。} \tag{24}$$

由上式, μ 會落在區間$[\overline{X}_n - a, \overline{X}_n + a]$ 之機率爲$1-\alpha$, 故此區間稱爲信賴係數爲$1-\alpha$ 之μ 的一信賴區間。有時即使(24)式左側機率並非剛好等於$1-\alpha$,而是大於$1-\alpha$,亦說此區間之信賴係數爲$1-\alpha$。a 之找法仍可藉助(20)式。由(20)得

$$P(|\overline{X}_n - \mu| \leq a) \doteq 1 - 2(1 - \Phi(a\sqrt{n}/\sigma)) \geq 1 - \alpha \text{。}$$

因此a 要滿足

$$a\sqrt{n}/\sigma \geq z_{1-\alpha/2} \text{。}$$

例5. 設X 只取0 或1 兩個值, 且設$P(X = 1) = p$。令μ 及σ 分別表X 之期望值及標準差。則$\mu = p, \sigma = \sqrt{p(1-p)} \leq 1/2$。若取$\alpha = 0.01$, 因$z_{1-\alpha/2} = z_{0.995} \doteq 2.576$,故$a$ 要滿足

$$a\sqrt{n}/\sigma \geq 2.576 \text{。}$$

雖然σ 爲未知,因$\sigma \leq 1/2$, 故取$a \geq 1.288/\sqrt{n}$ 即可。因此(近似的) 0.99 信賴區間爲

$$[\overline{X}_n - 1.288/\sqrt{n}, \overline{X}_n + 1.288/\sqrt{n}] \text{。}$$

例如, 若丟一銅板10,000 次, 且得正面4,950 次, 則正面出現的機率p之0.99 的信賴區間約爲$[0.495 - 0.013, 0.495 + 0.013]$, 即$[0.482, 0.508]$。

例6. 設有一銅板出現正面的機率p 爲未知, 我們想估計p 之值。此問題與下述這些問題類似: 欲估計高雄市的選民中, 支持某特定市長候選人的比例p; 欲估計某新藥能治癒某病之機率p。對於銅板, 可依下述方式估計p。丟一銅板n 次, 以S_n 表所得正面數, 又令$\overline{X}_n = S_n/n$ 表正面數之相對頻率。由大數法則知, 以$\overline{X}_n$ 來估計p 似乎是合理的。我們當然希望$\overline{X}_n$ 與實際的p 值不要差太遠, 譬如說$|\overline{X}_n - p| \le \varepsilon$, 其中$\varepsilon > 0$ 爲一給定之常數。

明顯可看出, 不論樣本數n 取多大, 都不一定保證$|\overline{X}_n - p| \le \varepsilon$ 必成立, 因就是有可能n 次全得到正面(或反面)。我們頂多能做的是, 給一$\alpha, 0 < \alpha < 1$, 使得

$$P(|\overline{X}_n - p| \le \varepsilon) \ge 1 - \alpha \text{。} \tag{25}$$

在上式中, ε 表以 $\overline{X}_n$ 來估計 p 之準確性, $1-\alpha$ 則表對此準確性之"信心"。當 n 夠大時,利用中央極限定理得

$$\begin{aligned} P(|\overline{X}_n - p| \le \varepsilon) &= P(|S_n - np|/\sqrt{np(1-p)} \le n\varepsilon/\sqrt{np(1-p)}) \\ &\doteq 2\Phi(n\varepsilon/\sqrt{np(1-p)}) - 1 \\ &= 2\Phi(\sqrt{n}\varepsilon/\sqrt{p(1-p)}) - 1 \text{。} \end{aligned}$$

我們便是要看n 多大時, $2\Phi(\sqrt{n}\varepsilon/\sqrt{p(1-p)}) - 1 \ge 1 - \alpha$ 才會成立。即 n 要滿足

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

因$p(1-p) \le 1/4$, 故

$$n \ge n_0 = \frac{1}{4\varepsilon^2} z_{1-\alpha/2}^2 \tag{26}$$

便滿足所需。

例如, 若$\varepsilon = 0.005$ 且$\alpha = 0.05$, 則$z_{1-\alpha/2} = z_{0.975} \doteq 1.960$, 此時取

$$n \ge n_0 = 10,000 \cdot (1.960)^2 \doteq 38{,}416,$$

便得

$$P(|\overline{X}_n - p| \le 0.005) \ge 0.95 \text{。}$$

我們常在報上看到底下這類敘述: xx民調公司,於8月5日 、6日,對二十歲以上成人進行電話抽樣訪問,有效樣本共1,130個,在百分之九十五信心水準下,誤差在正負2.9個百分點之間。調查結果顯示有32% 的民眾支持 $\cdots$。

這其中百分之九十五,表(25)式中 $1-\alpha=0.95$,通常是一事先給定的值; 1,130表訪問成功的樣本數,每次不盡相同,即(25)式中的 n 之值; 2.9個百分點,表(25)式中之 $\varepsilon=0.029$; 32% 表(25)式中 $\overline{X}_n=0.32$。此處爲取樣後不放回(sampling without replacement),所以對應的 $X_1, X_2, \cdots, X_n$ 其實不獨立。但假設選民總數夠多,因而可忽略取樣後不放回,與取樣後放回(sampling with replacement)間之差異。將 $1-\alpha$ 及 n 代入(26)式,並取等號,便可解出誤差 ε 之近似值:

$$\begin{aligned}\varepsilon &\doteq \frac{1}{2\sqrt{n}} z_{1-\alpha/2} \\ &\doteq \frac{1.96}{2\sqrt{1{,}130}} \\ &\doteq 0.02915 = 2.915\% \text{。}\end{aligned}$$

通常取 $\alpha=0.05$,且 ε 不超過3%,將此二值代入(26),得

$$n \geq \frac{1.96^2}{4 \cdot 0.03^2} \doteq 1{,}067 \text{。}$$

實際運作時,訪問成功的樣本要在1,067個左右,再將訪問後,所得之成功的樣本數 n ,代入(26)式,以得誤差 ε 之近似值。

當候選人間之支持率很接近時,就要取更大的 n,以減小 ε 之值,才能提高判斷究竟誰領先之正確性。譬如說,訪問結果甲得到支持率30%,乙得到支持率29%,但若抽樣誤差是3%,便只能說調查結果,甲乙二人的支持率,在統計上是無法分出差異的。由(26) 式,在同樣的 α 之下,若誤差 ε 要由3%降至1%,則樣本數 n 要變成9倍,在訪問上困難增加很大。

除了上述這些應用外, 中央極限定理在建立模式方面也很有用。由此定理得知, 若一隨機變數爲許多變異數爲有限之獨立且有共同分佈之隨

機變數之和, 其分佈便近似於常態分佈。尤有進者, 這些隨機變數也不一定要有相同的分佈, 只要每一個的值都很小即可(見Chung(1974)Theorem 7.21中央極限定理之一推廣)。許多自然界的現象均有此性質, 即爲許多很小的獨立的隨機變數之和。例如,人的身高或智商是由許多獨立(或近乎獨立)的基因或環境的因素所決定, 每一因素都造成一些微小的影響。其他如測量所產生的誤差也往往可視爲許多獨立的小誤差之和。

假設我們想建立一個屬於上述這種型式的現象之模式, 我們可能會對此現象的分佈先做一些假設。則由中央極限定理, 常態分佈是一個很自然的假設。譬如在前面提過的這些例子中, 身高、智商或測量產生的誤差, 往往可假設有常態分佈。而這一切,都是基於此一適用性如此之廣的上天賦與我們的法則—中央極限定理。

另外,中央極限定理也可當做大數法則之一補充。對一組i.i.d.的隨機變數 $\{X_n, n \geq 1\}$, 令 $S_n = \sum_{i=1}^{n} X_i$, $\overline{X}_n = S_n/n, n \geq 1$。乍看之下,中央極限定理比弱大數法則有用,因後者只是說 $\overline{X}_n$ 會"收斂"至期望值 $\mu = E(X_1)$, 而前者卻可對 $\overline{X}_n$ 所落在的範圍,給出機率之估計值。但不要忘記,中央極限定理要在變異數 $\sigma^2 = \mathrm{Var}(X_1)$ 存在之假設下,才能適用。大數法則卻只要求期望值存在。要知在機率統計裡,變異數存在比期望值存在之假設強很多。在較強的假設下,得到較好的結果,是合理且不足爲奇的。

在大數法則中(見(10)式),

$$\frac{S_n - n\mu}{n} \xrightarrow[n \to \infty]{a.s.} 0,$$

如果將上式左側分母減小些,只除以 $\sqrt{n}$,或者精確地講,除以 $\sigma\sqrt{n}$,如此左側變大,於是左側不再往0跑,而是會"接近"(分佈收斂)一非退化的隨機變數(且有 $\mathcal{N}(0,1)$ 分佈)。又在變異數亦存在的假設下,由

$$\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow[n \to \infty]{d} \mathcal{N}(0,1),$$

因左側分母(爲一常數數列 $\{\sigma/\sqrt{n}, n \geq 1\}$)趨近至0,所以分子會"趨近"至0(事實上是機率收斂至0,這要用到一些機率方面的理論)。所以

由中央極限定理,可直觀上看出大數法則成立。

5 大數下的迷思

我們先看底下一段報導:

綠色基金會以台北市一個月的車禍率爲十萬分之一爲例,並非每輛車會出十萬分之一的車禍,而是台北市三百萬輛車有三十輛會在一個月內出事。時間與母體群大小在這種風險機率計算中非常重要,一般來說只有大母體才能使用,例如車禍就是大母體計算很好的例子,但核電廠數目屈指可數,似乎不宜適用大母體計算。(83年6月21日聯合報第3版,記者李彥甫)

該記者認爲在大母體下,車禍率十萬分之一,導致三百萬輛車有三十輛會在一個月內出事。這可能是該記者在有些機率的概念下,所做出之推論。有趣的是報導的標題是"數據需要科學解釋"。對大數法則有這樣的理解,並非不尋常,我們再看一例。

有一陣子大家樂很風行, 賭徒藉著台灣銀行每期開出的愛國獎券中獎號碼來對獎。經常有沒簽中的賭徒對愛國獎券之搖獎機率產生懷疑。報章雜誌遂刊登一些教導大家"正確"了解機率學的文章。我們引用下面一段話, 供大家參考, 可能更有助於提醒我們須正確了解大數法則的含義(見趙慕嵩等(1987))。原文一字不易:

事實上, 所謂失常的機率只是在機率學中必然性的短暫現象,其實還是正常的。譬如一顆六面的正方形骰子, 上面有一到六的點數, 理論上每擲一次就應該使得每個數字各有六分之一的出現機會, 那麼連擲六次, 是否1, 2, 3, 4, 5, 6 等數字剛好各出現一次? 當然不會, 可是如果連擲六億次, 每個點數出現的次數就非常接近一億次, 而滿足於六分之一的理論機率。

如果按照理論的機率, 愛券連開十五萬次獎, 開出105 萬組號碼, 那麼00 到99 這一百組號碼就有可能各出現一萬零五百次, 而接近於理論機率。

大數法則是說 n 很大時,樣本平均 $\overline{X}_n$ 便"很接近"期望值 $E(X_1)$,有時會以樣本和 S_n 來表示此法則,即 $(S_n - nE(X_1))/n$ 會"很接近"0。很多人便誤以爲 n 愈大時, S_n 便愈接近 $nE(X_1)$。這當然是不正確的概念。要知即使對常數數列,若 $n \to \infty$ 時, $a_n/b_n \to 0$,並不導致$a_n \to 0$。取 $a_n = \sqrt{n}, b_n = n$ 即可得知。在此例中, n 愈大時, a_n 也愈大,但二者之比值 $a_n/b_n = 1/\sqrt{n}$ 卻隨著 n 之增大而減小。在大數法則裡也有類似的情景。由下例便可理解。

例7. 連續投擲一出現正面機率爲p 之銅板。令$X_n = 1$ 或0 就依第n 次出現正面或反面, 則$S_n = \sum_{i=1}^{n} X_i$ 表投擲n 次後共得之正面數。因$E(X_1) = p$, 由大數法則, 當n 很大時, S_n/n 會很接近p 之機率很大。次由定理4,

$$P(|S_n - np| \leq \alpha\sqrt{np(1-p)}) \doteq \Phi(\alpha) - \Phi(-\alpha)\text{。}$$

例如, 取$p = 1/2, \alpha = 1$。則若$n = 10^4$,

$$P(4{,}950 \leq S_n \leq 5{,}050) \doteq 0.6827\text{。}$$

若$n = 10^6$,

$$P(499{,}500 \leq S_n \leq 500{,}500) \doteq 0.6827\text{。}$$

不難看出對同樣的α 值, 當n 愈大, S_n 會落在的區間其半徑愈大。這可能與不少人的直觀不合,以爲n 愈大, S_n 會愈接近$nE(X_1)$。事實上n 愈大, S_n 的變異會變大(如上述第一個區間的半徑爲50, 第二個爲500), 但與n 相比(即$\sqrt{np(1-p)}/n = \sqrt{p(1-p)}/\sqrt{n}$), 變異卻是相對地減小。大數法則告訴我們的是$S_n/n$ "接近"$E(X_1)$, 而非S_n "接近"$nE(X_n)$! 甚至, 不論n 多大, 均不能百分之百的保證S_n/n 會必等於$E(X_1)$, 否則這便不是一隨機的現象了。

由上例也可看出,在大樣本之下,大數法則及中央極限定理所各自扮演的角色。

切記大數法則主要是針對平均,而非和。想通了這點,對諸如底下的敘述,便能知道其中之問題所在。

1.丟一公正銅板50次,且皆得正面,則反面應快出現了。

2.丟一公正銅板1,000次,得正面數較反面數多20,則在下1,000次投擲中,反面數會較正面數多20的機會很大。

3.若平均8個月有一次飛機失事,而在過去7個月中皆無飛機失事,因此本月會有飛機失事的機會不小。

不過下列敘述倒是對的。

丟一公正銅板1,000 次,得正面數較反面數多20,繼續丟1,000 次,則此2,000 次中,共得正面數之期望值較共得反面數之期望值多20。

最後,對幾個不同的 n,我們來看丟一公正銅板 $2n$ 次,得正面數與反面數相同的機率 $p_{2n} = \binom{2n}{n}2^{-2n}$ 之大小。

$$p_2 = \binom{2}{1}(\frac{1}{2})^2 = \frac{1}{2},$$
$$p_4 = \binom{4}{2}(\frac{1}{2})^4 = \frac{6}{16},$$
$$p_{10} = \binom{10}{5}(\frac{1}{2})^{10} = \frac{252}{1,024},$$

p_{100} 則約爲0.08。隨著 n 之增大, p_{2n} 變小,也就是愈來愈不容易有恰好半數是正面。事實上(見黃文璋(1994)第二章例9.1)

$$p_{2n} \sim \frac{1}{\sqrt{\pi n}},$$

其中符號"$\sim$"表 $n \to \infty$ 時,兩側之商趨近至1,即所謂近似相等(asymptotically equal)。

6 其他法則

隨機世界裡尚有一些其他法則各項住一片天。例如,稀有事件法則或稱小數法則(Law of Small Numbers)。由此法則引出了波松分佈(Poisson distribution)。另外,所謂無記憶性質(Memoryless Property)引出指數分佈(exponential distribution) 及幾何分佈(geometric distribution) 。極限裡的結果尚有疊對數法則(Law of the Iterated Logarithm) 等。當然還有

很多法則、定理及性質等,如同希臘神話中的眾神,各有其管轄範圍。因此隨機世界中雖然多變,但卻不亂,有大大小小的規範必須遵循。了解各種隨機法則,對此世界,就會由見山不是山見水不是水,再度走入見山是山見水是水的體認。

附錄:極限的概念

極限是數學裡一很重要的題材,經由極限,可以得到不少有意思的結果。對於極限的定義及定理,如果是關於常數函數(數列),可查一般的微積分的書,黃文璋(1999)第九章也可參考,如果是關於隨機變數,可查一般機率論的書(如黃文璋(1994)第五章)。爲了便於讀者閱讀,本節列出本文提到的一些極限的定義。對於欲深入了解機率及統計理論的讀者,這些知識當然是不夠的。

首先,談到極限,就牽涉到無限,此一連詩詞中都會出現的名詞(如唐詩三百首中,韋應物的賦得暮雨送李胄詩中有"相送情無限"),要眞正了解其意義可能要花點功夫。

其實不要說極限的概念不易弄清,極限中所常涉及的很大及很小,也是一不易掌握的概念。我們常說積少成多。引用此成語時,多少都有打氣或勉勵的意思。類似的成語尚有集腋成裘、聚沙成塔等。甚至愚公移山的精神,也是基於積少成多。

當然如果你學過微積分,便知道由少要至多,其中尚有一些微妙的關係。例如,比較下述三等式,其中等式左側均有 n 項:

$$\frac{1}{n^2}+\frac{1}{n^2}+\cdots+\frac{1}{n^2}=\frac{1}{n},$$
$$\frac{1}{n}+\frac{1}{n}+\cdots+\frac{1}{n}=1,$$
$$\frac{1}{\sqrt{n}}+\frac{1}{\sqrt{n}}+\cdots+\frac{1}{\sqrt{n}}=\sqrt{n}\text{。}$$

當 n 很大時,三個等式的左側,都是很多個很小的量相加。但在第一式中,和仍然很小($1/n$),第二式中的和爲一普通大的數(1),至於第三式中的和($\sqrt{n}$)則會很大。

所以,如果你問一位文學家積少成多對不對,他很可能答對,但如果問一數學家,他很可能會小心地反問你,少是多少?如何積法?以及怎樣叫多?

事實上積少後會如何,遠比前述幾個例子複雜多了。在微積分裡,這是關於級數收斂發散的題材。簡單地講,當 n 很大時, $n^{-1/2}$ 及 n^{-1} 皆往0跑,但速度有快慢: n^{-1} 較快, $n^{-1/2}$ 較慢。同樣地, n 很大時, $n^{1/2}, n^1, n^2$ 皆往無限大跑,但 n^2 的速度最快, $n^{1/2}$ 最慢。所以少是有等級的,大也是有等級的。與 n^2 相比, n 並不算大($n/n^2 = 1/n$),但若與 $n^{1/2}$ 相比, n 便很大了。$n^{-1/2}, n^{-1}, n^{-2}$ 之關係亦類似。就好像學然後知不足。當一個人拿到博士學位後,由於有博士學位的人依學問來排,大致成一金字塔,大部分的人在底下,而有些絕頂聰明的人在最上面,與下層的距離是如此遙遠。若常抬頭向上看,當然覺得仰之彌高,因此有時亦要向下看看,會發現仍有不少不如自己的。顯然學問的好壞也是相對比較的。在論語學而篇裡,孔子說"無友不如己者"。這句話應該不是如字面上所說,要人交朋友不要交到不如自己的,否則不要說窒礙難行,就算有人眞做到了,每個朋友都比自己好的人生,也不見得是快樂的。

史記李斯傳"太山不讓土壤故能成其大,河海不擇細流故能就其深"。細流聚集後,有時成河有時成海。河海亦各有大小。成河或成海,端視細流多細,及聚集多少細流。

經過長期的千錘百鍊,終於能以較簡潔的方式,來描述極限的概念。我們先對一般的"(常數)數列 $\{a_n, n \geq 1\}$ 之極限爲 a"給一較明確之定義。怎樣才是

$$\lim_{n\to\infty} a_n = a? \tag{27}$$

也就是 $n \to \infty$ 時, $a_n \to a$,其中符號"$\to$"表趨近。當 n 很大時,a_n 與 a 的差距將很小,即 $|a_n - a|$ 要很小。要多小才算很小呢?可以給一正數ε 當做 $|a_n - a|$ 之上界,而因並非所有 $|a_n - a|$ 皆須很小,只要 n 很大時,$|a_n - a|$ 很小即可。換句話說,只要可以找到一正整數n_0,使得 $n \geq n_0$ 時, $|a_n - a| < \varepsilon$ 就夠了。若覺得此 ε 不夠小,可以再換一個。最後,若不論給那一個 $\varepsilon > 0$,皆可辦到能找到一 n_0 (此 n_0 可與 ε 有關,即對不同的 ε 可找到不同的n_0),使

得 $n \geq n_0$ 時(即 n 夠大), $|a_n - a| < \varepsilon$,則我們便同意(27)成立。

我們倒不是說 n 很大時, a_n 要等於 a,即 $a_n - a = 0$,而是說只要 n 夠大, $|a_n - a|$ 須小於任意一正數。而 $|a_n - a| \geq 0$ 當然永遠成立,但一非負的數要能小於每一正數,就只有0了。所以說此時(27)成立。

任給一 $\varepsilon > 0$,要說(27)成立,就須找到一正整數 n_0,使得 $n \geq n_0$ 時,

$$|a_n - a| < \varepsilon,$$

即 $a_n \in (a - \varepsilon, a + \varepsilon)$。如果 ε 給得愈小,通常n_0 便要愈大才行。但若一直可以辦到,就不得不接受(27)的結果了。我們將上述想法寫成一定義如下。

定義1.設有一數列 $\{a_n, n \geq 1\}$。對一 $a \in R$,若 $\forall \varepsilon > 0$, 存在一 $n_0 \geq 1$,使得 $n \geq n_0$ 時, $|a_n - a| < \varepsilon$,則稱 $\lim_{n\to\infty} a_n = a$。

不少初學者對定義1並不容易接受,在邏輯上,這是一個不算容易的命題。此因其中包含四個敘述: (1) $\forall \varepsilon > 0$, (2)存在一 $n_0 \geq 1$, (3)使得 $n \geq n_0$ 時, (4) $|a_n - a| < \varepsilon$。要弄清楚其中的因果關係,是要經過一段時間的。我們的目的是要讓 $|a_n - a| < \varepsilon$,在什麼情況下?只要 $n \geq n_0$ 即可,對前面的 $a_1, a_2, \cdots, a_{n_0-1}$ 可以不用理會。而 n_0 又是什麼?只要找到一個即可,是可以隨著所給的 ε 而不同的。

例8.設 $a_n = (2n^2+3)/(n^2+2n), n \geq 1$,底下我們證明$\lim_{n\to\infty} a_n = 2$。欲使

$$\begin{aligned} |a_n - 2| &= \left|\frac{2n^2+3}{n^2+2n} - 2\right| = \left|\frac{4n-3}{n^2+2n}\right| \\ &< \frac{4n}{n^2+2n} = \frac{4}{n+2} < \frac{4}{n} < \varepsilon, \end{aligned}$$

只要 $n > 4/\varepsilon$ 即可。故若取 $n_0 = [4/\varepsilon] + 1$,其中 $[x]$ 表小於或等於 x 之最大整數,則依定義即得證 $\lim_{n\to\infty} a_n = 2$。

至於 $\lim_{n\to\infty} a_n$ 不存在又是什麼意思?也就是我們要能展示 $\forall a \in R$,

$$\lim_{n\to\infty} a_n \neq a。 \tag{28}$$

而上式即 $\lim_{n\to\infty} a_n = a$ 不成立,即要推翻

(29) $\forall\varepsilon > 0$, 存在一$n_0 \geq 1$, 使得$n \geq n_0$時, $|a_n - a| < \varepsilon$ 。

我們依序來看:要說 $\forall\varepsilon > 0, \cdots$ 不成立,只要找到一$\varepsilon > 0$,使得 $\cdots$ 不成立即可。在此 $\cdots$,即(29)之後三個敘述。而欲使存在一 $n_0 \geq 1$, $\triangle\triangle$ 不成立,便須對 $\forall n_0 \geq 1$, $\triangle\triangle$ 不成立。在此 $\triangle\triangle$,即(29)之後二敘述。最後欲 $n \geq n_0$ 時, $|a_n - a| < \varepsilon$,不成立,便只要找到一 $n \geq n_0$,使得 $|a_n - a| \geq \varepsilon$。總結如下:對一$a \in R$,若存在一 $\varepsilon > 0$,使得對 $\forall n_0 \geq 1$,必存在一 $n \geq n_0$,使 $|a_n - a| \geq \varepsilon$,則(28)成立。而若 $\forall a \in R$, 皆能辦到此事,則 $\lim_{n\to\infty} a_n$ 便不存在,或說 $n \to \infty$ 時, a_n 之極限不存在,即 $\{a_n, n \geq 1\}$ 發散。

例9.試證 $\lim_{n\to\infty}(-1)^n$ 不存在。

證明.令 $a_n = (-1)^n$。要證明 $\forall a \in R$, $\lim_{n\to\infty}(-1)^n \neq a$。先設 $a \geq 0$。取 $\varepsilon = 1/2$,則當 n 爲奇數時,

$$|a_n - a| = |-1-a| = a+1 > \varepsilon,$$

即對 $\forall n_0 \geq 1$,必存在一奇數 $n \geq n_0$,使得 $|a_n - a| \geq \varepsilon$。

其次設 $a < 0$。仍取 $\varepsilon = 1/2$,則當 n 爲偶數時

$$|a_n - a| = |1-a| = 1-a > \varepsilon,$$

即仍有對 $\forall n_0 \geq 1$,存在一偶數 $n \geq n_0$,使得 $|a_n - a| \geq \varepsilon$。

得證 $\lim_{n\to\infty}(-1)^n$ 不存在。

當 $n \to \infty$,數列 $\{a_n, n \geq 1\}$ 之極限存在,表隨著 n 之增大,數列趨向穩定,即 a_n 之值逐漸不會有太大的變化,而收斂到某一定值 a。對一隨機數列 $\{X_n, n \geq 1\}$, $n \to \infty$ 時,此數列之極限存在,有一些不同的意義。換句話說,隨機數列有不同的收斂方式。我們列出本文所提到的幾種。在底下定義2及3中, $\{X_n, n \geq 1\}$ 爲一數列之隨機變數,與隨機變數 X 定義在同一機率空間 $(\Omega, \mathcal{F}, P)$。

定義2.設存在一機率爲0之事件 N(即 $N \in \mathcal{F}$,且 $P(N)=0$),使得

$$\lim_{n\to\infty} X_n(\omega) = X(\omega),\ \forall \omega \in \Omega \backslash N, \tag{30}$$

則稱 $n \to \infty$ 時, $\{X_n, n \geq 1\}$ 幾乎確實地收斂至 X,且以

$$X_n \xrightarrow[n\to\infty]{a.s.} X$$

表之。即 $n \to \infty$ 時, $X_n \to X$, a.s.。

幾乎確實地收斂,類似於一函數數列(series of functions)$\{f_n(x), n \geq 1\}$ 之逐點收斂(pointwise convergence)。不過此處收斂允許對一機率爲0之集合不必成立(因此才稱幾乎確實)。也就是對每一 $\omega \in \Omega$,我們檢驗數列$\{X_n(\omega), n \geq 1\}$ 是否趨近至 $X(\omega)$,若不收斂的 ω 之集合N 之機率爲0,便稱此爲幾乎確實地收斂。這是一種很自然的收斂方式,有時又稱爲強收斂(strong convergence)。(30)又可以

$$P(\lim_{n\to\infty} X_n = X) = 1 \tag{31}$$

表之。亦即

$$P(\lim_{n\to\infty} |X_n - X| < \varepsilon) = 1,\ \forall \varepsilon > 0\text{。} \tag{32}$$

(31)式指出

$$X_n \xrightarrow[n\to\infty]{a.s.} X$$

若且唯若 $P(M)=1$,其中 $M = \Omega \backslash N$,表 Ω 中使

$$\lim_{n\to\infty} X_n(\omega) = X(\omega)$$

之所有 ω 所形成之集合。

對一事件 A,若 $P(A) = 1$,則稱此爲一幾乎確實的事件。如 $X \geq 0$, a.s.,即表$P(X \geq 0) = 1$。若 $P(A) = 0$,便稱此爲一不可能的事件(impossible event),或零集(null set)。譬如說,自區間 $[0,1]$ 中任取一點,並以 X 表之,則因 $P(X=a)=0$, $\forall a \in R$,因此$[X=a]$ 爲一不可能的事件。

例10.令 $\Omega=\{0,1\}$, $\mathcal{F}$ 包含 Ω 之所有子集合,$P(\{0\})=P(\{1\})=1/2$。則 $(\Omega,\mathcal{F},P)$ 構成一機率空間。當 n 爲偶數時,令 $X_n(0)=5, X_n(1)=10$,當n 爲奇數時,令 $X_n(0)=10, X_n(1)=5$。則數列 $X_1(0), X_2(0), X_3(0), \cdots$ 不收斂, $X_1(1), X_2(1), X_3(1), \cdots$ 亦不收斂。故使數列 $\{X_n(\omega), n\geq 1\}$ 收斂之 ω 的集合爲空集合,即 $\{X_n, n\geq 1\}$ 收斂之機率爲0。

例11.令 $\Omega=\{1,2,3,\cdots\}$, $\mathcal{F}$ 包含 Ω 之所有子集合, $P(\{k\})=2^{-k}, k\geq 1$,則 $(\Omega,\mathcal{F},P)$ 構成一機率空間。令

$$X_n(\omega)=\begin{cases} 2^n & ,\omega\geq n+1, \\ 0 & ,1\leq\omega\leq n。\end{cases}$$

對 $\forall\omega\in\Omega$,只要 $n\geq\omega$,則 $X_n(\omega)=0$,故$n\to\infty$ 時 $X_n(\omega)\to 0$。即得

$$X_n \xrightarrow[n\to\infty]{a.s.} X,$$

其中 $P(X=0)=1$。

定義3.若

$$\lim_{n\to\infty} P(|X_n-X|<\varepsilon)=1,\ \forall\varepsilon>0, \tag{33}$$

則稱 $n\to\infty$ 時, $\{X_n, n\geq 1\}$ 機率收斂至 X,以

$$X_n \xrightarrow[n\to\infty]{P} X$$

表之。

(33)式與

$$\lim_{n\to\infty} P(|X_n-X|\geq\varepsilon)=0,\ \forall\varepsilon>0, \tag{34}$$

等價。又依極限之定義, (34)式成立,若且唯若對 $\forall\varepsilon>0$ 及$\delta>0$,存在一 $n_0\geq 1$,使得

$$P(|X_n-X|\geq\varepsilon)\leq\delta,\ \forall n\geq n_0。 \tag{35}$$

另外,不難看出(33), (34)二式分別等價於

$$\lim_{n\to\infty} P(|X_n - X| \le \varepsilon) = 1,\ \forall\varepsilon > 0,$$
$$\lim_{n\to\infty} P(|X_n - X| > \varepsilon) = 0,\ \forall\varepsilon > 0\ 。$$

例12.令 $\{Z_n, n \ge 1\}$ 爲i.i.d.之隨機變數,以 $\mathcal{U}(0,1)$ 爲其共同分佈,又令 $Y_n = \max\{Z_1, Z_2, \cdots, Z_n\}$ 。則因 $P(Y_n \ge 1) = 0$,故對$\forall 0 < \varepsilon < 1$,當 $n \to \infty$ 時,

$$P(|Y_n - 1| \ge \varepsilon) = P(Y_n < 1 - \varepsilon) = (1-\varepsilon)^n \to 0\ 。$$

故

$$Y_n \xrightarrow[n\to\infty]{P} 1\ 。$$

定義4.設 X_n 之分佈函數爲 $F_n, n \ge 1$, X 之分佈函數爲 F 。若對每一 F 之連續點 x,

(36) $$\lim_{n\to\infty} F_n(x) = F(x),$$

則稱 $n \to \infty$ 時, $\{F_n, n \ge 1\}$ 弱收斂(converge weakly)至 F,且以

$$F_n \xrightarrow[n\to\infty]{w} F$$

表之 。此時亦稱 $\{X_n, n \ge 1\}$ 分佈收斂至 X,以

$$X_n \xrightarrow[n\to\infty]{d} X$$

表之 。

故若

(37) $$\lim_{n\to\infty} P(X_n \le x) = P(X \le x),\ \forall x\text{使得}P(X = x) = 0,$$

則 $\{X_n, n \ge 1\}$ 分佈收斂至 X 。分佈收斂事實上是分佈函數的收斂,而非隨機變數之收斂,這是與機率收斂及幾乎確實地收斂不同的地方 。因此

稱 $\{F_n, n \geq 1\}$ 弱收斂至 F,比稱 $\{X_n, n \geq 1\}$ 分佈收斂至 X 恰當。又分佈收斂也常被稱爲法則收斂(convergence in law),顧名思義,是指其機率法則(即分佈函數)收斂。

例13.對 $\forall n \geq 1$,設隨機變數 X_n 之p.d.f.爲

$$f_n(x) = \begin{cases} 1/2 & \text{,若 } x = 1 - 1/n \text{ 或 } 1 + 1/n, \\ 0 & \text{,其他。} \end{cases}$$

則 X_n 之分佈函數 F_n 爲

$$F_n(x) = \begin{cases} 0 & , x < 1 - 1/n, \\ 1/2 & , 1 - 1/n \leq x < 1 + 1/n, \\ 1 & , x \geq 1 + 1/n, \end{cases}$$

且除了 $x = 1$ 之外, $n \to \infty$ 時, $F_n(x) \to F(x)$,其中

$$F(x) = \begin{cases} 0, & x < 1, \\ 1, & x \geq 1, \end{cases}$$

爲一只在 $x = 1$ 不連續之分佈函數。故 $\{F_n, n \geq 1\}$ 弱收斂至 F。若取X 滿足 $P(X = 1) = 1$,則 X 之分佈函數爲 F,且

$$X_n \xrightarrow[n \to \infty]{d} X \text{。}$$

上述三種收斂有如下關係:

$$X_n \xrightarrow[n \to \infty]{a.s.} X \text{ 導致 } X_n \xrightarrow[n \to \infty]{P} X,$$

$$X_n \xrightarrow[n \to \infty]{P} X \text{ 導致 } X_n \xrightarrow[n \to \infty]{d} X \text{。}$$

若 X 爲一退化的隨機變數,即存在一常數 c,使得 $P(X = c) = 1$,則此時

$$X_n \xrightarrow[n \to \infty]{P} X \text{ 與 } X_n \xrightarrow[n \to \infty]{d} X$$

等價。

習　題

1. 設$\{X_n, n \geq 1\}$ 爲i.i.d.之隨機變數,以Cauchy 分佈 $\mathcal{C}(\theta, a)$ 爲其共同分佈。試問此時大數法則是否適用?

2. 欲抽樣調查市民中支持某政策之比例p, 且要求估計值與p 之差不超過0.02的機率, 大於或等於0.95。在下述二情況下分別估計此時樣本數至少要若干?

(i) 已知$p \leq 0.2$;

(ii) 對p 值毫無概念。

3. 投擲一公正的銅板400次,求所得之正面數介於190與210間之機率的近似值。

4. 投擲一公正的骰子12,000 次。

(i) 估計點數6 出現的次數介於1,800 至2,200 次間之機率;

(ii)若點數6 出現2,500 次, 你是否會懷疑這不是一公正的骰子? 試說明理由。

5. 投擲一公正的銅板1,000 次, 求k 之值使得出現正面的次數介於490 與k 間的機率約爲0.5。

6. 某選舉之候選人想估計選民對其支持的比例p, 問須成功地訪問多少選民, 才有約95 % 的把握誤差不超過4.5 %。另外,若成功訪問1,092位選民,則有95 % 的把握誤差約不超過若干?

7. 欲估計市民對某項政策之支持度p,須成功地訪問多少市民,才有約99%的把握誤差不超過1%。

8. 假設澎湖與高雄間有兩家船運公司各有客輪往返二地。兩家公司之客輪的品質及一切狀況假設均類似, 設有n 個客人獨立且隨機地自二家公司擇一搭乘。令X_1, X_2分別表此二家公司每日會被選中搭乘的人數,則X_1, X_2皆有$\mathcal{B}(n, 1/2)$ 分佈。若每家公司每日只提供s 個船位, $s < n$, 則有一正的機率$f(s) = P(X_1 > s)$ 有客人搭不上。
(i)以中央極限定理估計 $f(s)$;
(ii)試證 $s > (n + z_{1-\alpha}\sqrt{n})/2$ 時, $f(s) < \alpha$ (近似)成立,其中$0 < \alpha < 1$; (iii)利用(ii),當 n =1,000, 求使 $f(s) < 0.01$ 之最小的 s;

9. 假設某廠牌之燈泡壽命有期望值爲10 天之指數分佈。在某年初裝上一新燈泡, 一旦用壞便重裝一新的。試估計在一年中使用的燈泡超過50 個之機率。

10. 設$\{X_n, n \geq 1\}$ 有i.i.d.之$\mathcal{N}(0,1)$ 分佈。試以函數Φ 來估計n 很大時, $P(X_1^2 + X_2^2 + \cdots + X_n^2 \leq x)$ 之值, 其中Φ 爲$\mathcal{N}(0,1)$ 之分佈函數。

11. 設$\{X_n, n \geq 1\}$ 爲i.i.d.之隨機變數, $E(X_1) = \mu$, $\mathrm{Var}(X_1) = 9$。試分別以Chebyshev 不等式及中央極限定理求出a 之值(或近似值), 使得$P(|\overline{X}_n - \mu| < a) \geq 0.9$。

12. 設隨機變數 X 有$\Gamma(n, \lambda)$ 分佈, n 爲一正整數。試利用中央極限定理估計n 很大時, $P(X \leq x)$ 之值,$x \in R$。

13. 投擲一公正的銅板, 直至出現100 個正面才停止, 試估計至少要投226 次的機率。

14. 假設某校有55% 的學生帶眼鏡, 隨機抽取100 個學生, 求其中帶眼鏡者不超過50 個之機率的近似值。

15. 投擲一公正的骰子100 次, 求所得點數和介於330 與380 間之機率的近似值。

16. 設某工廠生產之零件中有3 % 是不良品, 每1,000 個零件裝一盒。任取一盒, 求其中不良品數不超過5 % 之機率的近似值。

17. 某有線電視公司讓500個家庭免費試看一個月,根據過去經驗,試看結束後,每一家庭有6成的機會,願意每月付費600元收看。試估計此試看案結束後,該公司每月可增加180,000元收入之機率。

18. 某投幣式電動玩具店,以稱重量來估計每周所收之10元銅板個數。假設1公斤中, 10元銅板個數 X 之期望值爲142,標準差爲0.25。現設某周共收400公斤的10元銅板。

(i)求10元銅板個數之期望值;

(ii)求最小的 a,使得此400公斤中,至少有 a 個10元銅板之機率不超過0.001。

19. 設 $\{X_n, n \geq 1\}$ 爲i.i.d.之 $\mathcal{B}er(p)$ 隨機變數。取 $\varepsilon = 0.01$。欲求 n_0,使得 $n \geq n_0$ 時,

$$P(|\overline{X}_n - p| \leq \varepsilon) \geq 0.95 \text{。}$$

(i)利用Chebyshev不等式求之;

(ii)利用中央極限定理求之;

(iii)比較(i)及(ii)中所得之 n_0,說明何以其中之一大很多。

20. 某人自區間 $[0,1]$ 中任取一點,設取中某點 c。他感到很疑惑,問道"取中的點爲 c 不是一不可能的事件嗎?那怎會發生呢?"試解釋此一不可能卻會發生的現象,並因此說明爲何不可能的事件可以一再發生(第一次取中 c_1,第二次取中 $c_2, \cdots$)。

參考文獻

1. 黃文璋(1994).機率論講義。全華科技圖書股份有限公司,台北。

2. 黃文璋(1999).數學欣賞。華泰文化事業股份有限公司,台北。

3. 趙慕嵩、張文輝及胡不歸(1987).羅盤、神明牌、風水、獎券科。時報周刊民國76年10月,第504期, 22-28。

4. Chung, K. L. (1974). *A Course in Probability Theory*, 2nd ed. Academic Press, New York.

5. Tucker, H. G. (1967). *A Graduate Course in Probability.* Academic Press, New York.