大數法則

黃文璋
國立高雄大學應用數學系

1 數大便是美

徐志摩的散文,很多人中學時代都讀過,膾炙人口的作品實在不少。他的有些句子後來常被引用,“數大便是美”是其中之一。我們節錄原來文章如下:

數大便是美

徐志摩日記西湖記

> 數大便是美, 碧綠的山坡前幾千隻綿羊, 挨成一片的雪絨, 是美; 一天的繁星, 千萬隻閃亮的眼神, 從無極的藍光中下窺大地, 是美;泰山頂上的雲海, 巨萬的雲峰在晨光裏靜定著, 是美; $\cdots$; 數大便是美。數大了似乎按照著一種自然律, 自然的會有一種特別的排列, 一種特別的節奏, 一種特殊的式樣, 激動我們審美的本能, 激發我們審美的情緒。
>
> 所以, 西湖的蘆荻, 與花塢的竹林, 也無非是種數大的美, 不是智力可以分析的, $\cdots$。

徐志摩舉了好幾個自然界的景象, 來描述數大便是美, 引起了很多人的共鳴。曾在一部落格(blog)中, 看到一篇標題為“星巴克小熊之數大就是美”的文章, 其中有底下一段話:

> 為何說數大就是美呢? 這是因為並不是每一隻小熊都可愛。有時當我買熊時, 旁邊的人都說這隻好醜喔! 但是我還是不偏心的買回家, 我相信蒐集起來一堆熊, 一定很有成就感。現在第七隻的加入後, 果然聲勢浩大, 就像以前課本寫的"數大就是美", 果然沒錯。

徐志摩的數大便是美, 常被寫成"數大就是美"。很多時候數量一大, 的確會讓人產生某種特別的看法。即使連7隻醜小熊, 都有人可感受到數大的美。徐志摩認為"數大了似乎按照著一種自然律"。自古以來, 人們便常注意到, 宇宙裡各種隨機現象, 雖看起來像是雜亂無章, 但卻似乎遵循某些規律在運轉。雖然這不是"數大便是美"所能貼切地描述, 但的確是如徐志摩所說的, 數大後常會按照某一種自然律。在隨機現象中, 我們可稱這種自然律為**隨機法則**。就是因有這些法則, 讓我們對隨機現象, 仍能有所掌握, 仍能預測。這其中最該一提的, 就是**大數法則**(law of large numbers)。我們將分別介紹**巨數法則**(law of truly large numbers), **弱大數法則**(weak law of large numbers), 及**強大數法則**(strong law of large numbers)。

2 極限的定義

懂點機率統計的人, 常開口閉口大數法則。大數法則究竟是什麼? 討論大數法則, 無可避免的, 會涉及極限。只是極限可不是一簡單的概念。但弄懂極限, 是進入較高深數學的第一步。本節我們便稍微介紹極限。

任給一數列$\{a_n, n \geq 1\}$, 當$n \to \infty$時, a_n不見得會趨近至某一定值。例如, $a_n = (-1)^n, n \geq 1$, 則此數列為$-1, 1, -1, 1, \cdots$, 不論n多大, 數列都是$1, -1$交錯著, 不會接近任一定值。但若$a_n = (1/2)^n, n \geq 1$, 則隨著n之增大, a_n愈來愈小, 一直往0接近。我們以

$$n \to \infty \text{時}, a_n \to 0$$

表之, 也可寫成

$$\lim_{n\to\infty} a_n = 0 \text{ 。}$$

極限是微積分的基礎, 但極限並不是一簡單的概念 。想要弄懂極限, 當然要找本微積分書, 仔細唸一唸 。黃文璋(1999)第九章“極限的概念”, 為一討論極限之通俗性文章 。

不過凡是牽涉到無限大, 就要謹慎對待 。男孩對女孩說“我每天(每一第n 天)都愛你”, 這是否就隱含“海枯石爛, 永不變心”呢? 倒也未必, 海枯石爛表示時間n趨近至∞ 。但$n \to \infty$時會如何, 不能以每個n會如何來推測 。舉一簡單的例子 。令$a_n = 1/n, n \geq 1$, 則對每一$n \geq 1$, a_n皆為正數 。但

$$\lim_{n\to\infty} a_n = 0$$

就非正數了 。中學時代, 你應學過**數學歸納法** 。令A_n表一與n有關的命題, $n \geq 1$ 。在有些情況下, 可以數學歸納法證明每一命題A_n皆成立 。其原理就是先證出A_1成立, 然後再證明A_n成立, 可導致A_{n+1}成立, 則每一命題A_n便皆成立了, $n \geq 1$ 。但這只證出每一命題A_n為眞, 並未導致$n \to \infty$時, 命題仍成立 。

對於$n \to \infty$時, $a_n \to a$, 是說當n不斷地增大, a_n可任意接近a 。但任意接近a是什麼意思? 此表a_n與a之差距$|a_n - a|$可任意小 。而怎樣是任意小? 由你定好了, 0.01算小嗎? 還是0.0001才算小?, 就任給一差距, 以ε表之, 即要求$|a_n - a|$須小於ε 。但何時$|a_n - a| < \varepsilon$?我們可沒說對每一$n \geq 1$, $|a_n - a| < \varepsilon$都要成立, 而是說n要很大很大時, $|a_n - a| < \varepsilon$才須成立 。但怎樣是很大很大? 如果從某一項n_0開始, $|a_n - a| < \varepsilon$皆成立, 你大約便只好服氣了 。於是便產生了下述極限的定義: 設有數列$\{a_n, \geq 1\}$, 及實數a,

若每一$\varepsilon > 0$, 存在一$n_0 \geq 1$, 使得$|a_n - a| < \varepsilon$, 當$n \geq n_0$,
則稱$n \to \infty$時, $a_n \to a$, 以$\lim_{n\to\infty} a_n = a$表之 。

這個定義並非微積分一開始發展便有的, 而是經過一百多年, 數學家才想出以此方式來定義極限 。要知數學上一個概念, 常是經過長期的演變, 才

能以簡潔、嚴密、一般、且抽象的方式呈現。後來的學習者,自然得花一點功夫, 才能領悟此概念之內涵。

3 巨數法則

巨數法則, 雖然英文名為law of truly large numbers, 但其實與law of large numbers 並不太相干。在一般正規的機率論書籍中, 不會提到此法則。它主要是出現在通俗性的文章中, 有時也被稱為law of large numbers。我們實在不願稱它為"**眞大數法則**", 只好含混地稱它為**巨數法則**。在Diaconis and Mosteller (1989) 一文中, 對此法則給出如下定義:

> With a large enough sample, any outrageous thing is likely to happen(當樣本數夠大, 任何聳人聽聞的事, 都可能發生)。

數學家Littlewood (1953) 認為一件事若發生的機率為百萬分之一, 便可稱之為**令人驚奇** (surprising)。若採用此定義, 各種令人驚奇的事, 可說在世界各地, 經常在發生。譬如說49取6的樂透彩, 一張彩券會中頭獎的機率很小, 僅13,983,816分之1。對一特定的人, 要中頭獎當然很難, 因他不太會買很多很多張。但"有人"中頭獎, 就是一常會發生的事件。又如一家庭中有4人生日相同, 這當然不容易。若忽略閏年, 假設一年有365天, 且每人在1年中任一天出生之機率皆為1/365。則任4人生日為365天中某一特定的日子之機率為$1/365^4$。因此4人生日相同的機率為$1/365^3$ =1/48,627,125, 實在很小。但世界人口已超過69.27億(至2011年6月), 若有某一家庭中有4人生日相同(注意, 這又比限制是父母及2小孩生日相同更容易發生多了), 並不該太令人驚訝。

雖與大數法則不同, 但由於都涉及大樣本, 有些人遂將二者混在一起。如Shermer (2004, 科學人2004年9月號"奇蹟? 機率?"一文為其中文翻譯(姚若潔譯))一文提到:

> A principle of probability called the Law of Large Numbers shows that an event with a low probability of occurrence in a small number of trials has a high probability of occurrence

in a large number of trials. Events with million-to-one odds happen 295 times a day in America.

發生機率百萬分之一的奇怪事件, 在美國每天可發生295次, 這是基於美國有2.95億(295百萬)的人。又順便一提, 根據**美國人口普查局**(U.S.Census Bureau)網頁, 至2011年6月, 美國人口超過3.11億。

薛莫(Michael Shermer) 為Scientific American 的專欄作家。上述他那篇文章雖然有趣, 但對機率的描述, 卻不夠準確, 恐易引起誤會。首先, 他所引用的不是大數法則, 而是巨數法則。其次,他說"在數量樣本較少時, 機率很小的事件, 在數量樣本較大時, 其發生的機率會變高"(英文原意如此, 這是科學人的中譯), 這樣講是不對的。應該是說"發生機率很小的事件, 若試驗數(或說樣本數)較少時, 會有這種事件發生的機率不高;若試驗數(或說樣本數)較大時, 會有這種事件發生的機率會變高"。至於單一事件發生的機率, 不會隨試驗數之多寡而改變。另外, 最後一句話, 改為"在一個人身上每天發生機率為百萬分之一的怪事件, 在美國平均一天可發生295次"較恰當。

一事件發生的機率p雖然很小, 重複觀測n次, 假設各事件相互獨立, 則n次皆未發生之機率$(1-p)^n$, 隨著n之增大, 此機率愈來愈接近0, 而至少發生一次之機率$1-(1-p)^n$則逐漸接近1。這也可以解釋, 只要觀測數夠多, 一聳人聽聞的事件, 其發生就不該令人驚訝。更何況世上各種千奇百怪的事, 實在不少, 要發生某一令人匪夷所思的事, 就更容易了。譬如說, 假設美國有一研究機構, 養了一批猴子, 每隻發一台打字機。他們每日就是在那兒亂敲。終於有一天, 發現某隻猴子敲出一串從華盛頓開始的美國歷任總統的名子。雖然上了頭版新聞, 只是機率學家卻不會感到驚訝。讀者可能也會明白了, 何以通俗性的文章中, 會稱此為law of truly large numbers。由於閱讀對象為較一般的民眾, 當看到一件很不尋常的事, 以"眞是夠大的觀測數"來解釋, 可能會較易使人覺得原來如此。巨數法則可用來解釋何以生活上處處有巧合(coincidence)。關於巧合事件之無所不在, 可參考黃文璋(2003)第四章"純屬巧合"一文。幾年前曾引起一陣風波的**聖經密碼**(The Bible Code)事件, 也可以用巨數法則解釋。可

參考Shermer(2003, 科學人2003年7月號“聖經裡眞的有密碼?”為其中文翻譯(姚若潔譯))一文。

巨數法則雖不難理解, 但仍常有人無法正確使用。在“看守台灣季刊”第四卷第二期(2002年夏季號)的“編後語”中有底下一段文字:

> 就在完成本期彙編時, 網路上看到一則新聞(7/9), 英國有一對白人夫婦接受人工授精, 結果孕婦生出一對黑人雙胞胎, 成為英國首件人工授精搞烏龍案例。
>
> 為避免出錯, 人工授精的過程十分嚴格, 每一個步驟都會重複檢查, 理論上可能會出錯的機率只有百萬分之一。這次不僅出錯, 甚至是黑白搞烏龍! 這樣的機率可能更低。如此低的機率, 怎麼可能發生?
>
> 按台北市銀行的公告, 樂透彩頭獎的簽中機率大約是525萬分之一。這個簽中機率更低, 低於人工授精出錯機率的五分之一。然而, 總會有人簽中! 按台電公告的核電反應爐每年發生重大災變的機率是十萬分之一。這個機率是人工授精的出錯機率的10倍, 更是簽中頭獎的50多倍。怎麼可能保證不會發生?

該如何解釋呢? 首先, 發生機率再小的事件, 只要機率為正, 便都可能發生, 並無法保證不會發生。但如前所述, 每天有各種事物在進行, 發生一些令人覺得離譜、怪誕不經的事, 可以說是必然。對於北銀42取6的樂透彩, 一週開獎兩期, 每期簽注人數多達幾百萬甚至有千萬, 有人簽中頭獎, 自然不稀奇。那天若有人第二次中頭獎, 也不用太驚訝。至於台電公告的核電反應爐, 每年發生重大災變的機率是十萬分之一, 注意是“每年”。又假設此機率是針對“1座”核電反應爐, 而台灣目前只有3座核電廠。不提每年3座與一年開獎104期, 且每期有幾百萬人簽注之比, 而只提525萬(實際是5,245,786)與10萬之比, 眞是明察秋毫而不見輿薪。沒人可保證核電反應爐不會發生災變, 雖我們並不清楚十萬分之一的機率是如何求出, 及此機率值是否正確。但如果接受台電的數據, 則因核電反應爐也有其使用年

限, 因此在未來10年, 雖會繼續產生不少樂透彩頭獎得主, 以及發生各種烏龍事件, 但台灣3座核電反應爐, 在10年內會至少有1座發生災變的機率, 仍是不太高 。

4 弱大數法則

大數法則又稱**大數率**, 或**平均法則**(law of averages) 。由於有大數法則, 使得在**不確定性**(uncertainty)中, 我們仍能掌握一些**確定性**(certainty) ; 在**混亂**(chaos)中, 仍有其**秩序**(order) 。大數法則是說, 若一實驗(或觀測), 能持續且重複地進行, 則觀測值之平均, 將任意接近期望的結果 。比較正式一點的說, 就是隨機所產生樣本之平均, 當樣本數很大, 將有很大的機率, 接近母體之平均 。

機率論早期的發展, 常對某事件是否發生有興趣 。如投擲銅板是否出現正面, 玩撲克牌是否得到3條等 。換句話說, 對只有兩個結果的觀測有興趣 。以$X_i = 1$, 表第i次觀測該事件發生, $X_i = 0$, 表第i次觀測該事件未發生 。如此觀測到一串0, 1的數列 。這種數列, 今日稱為**伯努力數列**(Bernoulli sequence) 。而這類只有兩個結果的實驗, 便稱為**伯努力試驗**(Bernoulli trial) 。這是因瑞士數學家**伯努力**(Jacob Bernoulli, 1654-1705), 最先探討而得名 。在他死後8年, 1713年, 他姪兒**尼古拉斯伯努力**(Nicholas Bernoulli, 1687-1759), 替他出版那本可說是機率論最早的書籍Ars Conjectandi (原文為拉丁文, 英文書名為The Art of Conjecturing) 。在這本書中, 伯努力證明了一以他的姓命名的定理, 即**伯努力法則**(Bernoulli law):

> 獨立且重複地觀測一發生機率為p之事件A, 當觀測次數趨近至∞, 事件發生之相對頻率接近p之機率, 將趨近1 。

註1. Bernoulli 一家可說是數學史上相當顯赫的一家 。在三代裡, 至少有8位, 對數學 、機率或統計有貢獻 。其中有5位算是相當傑出 。在數學 、機率及統計裡, 常會見到Bernoulli, 不見得是同一人 。我們列出伯努力四代的部分成員於圖1 。西方人的姓(last name)較少相同, 名(first

name)則很容易相同。在同一領域出名且姓相同的, 往往並不太多。所以我們通常只以姓稱呼西方的科學家、文學家及藝術家等名人。如牛頓、高斯及歐拉等, 都是姓, 而幾百年來也沒產生第二個著名科學家是牛頓、高斯或歐拉。若真有同姓的, 再附上名, 便可區隔。只是這套慣用的稱呼法, 對Bernoulli 就不適用了。由於這是一數學大家族, 不但你見到的Bernoulli 可能不是同一人, 有幾位的名也相同。中國人對於取名, 常會避長者諱。外國人則有時為小孩取某長者之名, 以為紀念。這點在Bernoulli家族很明顯, 還有取與自己同名的。圖1中有11個人, 卻只有4個不同的名, 有時以二世(II), 三世(III)來區隔, 有時也不區隔。另一個Bernoulli 家族的名會令人產生困擾的是, 有幾位曾在歐洲幾個國家居住, 而在不同的地方會有不同的名。以我們所提到的這位Bernoulli 家族第一位數學家為例, Jacob 有時寫成Jakob;他又稱為Jacques(有時寫成Jaques);有時又稱為James。為了提昇機率與數理統計的發展, 1975年還成立了一個**伯努力協會**(Bernoulli Society)的國際組織。可見伯努力家族在機率與統計界的被重視。

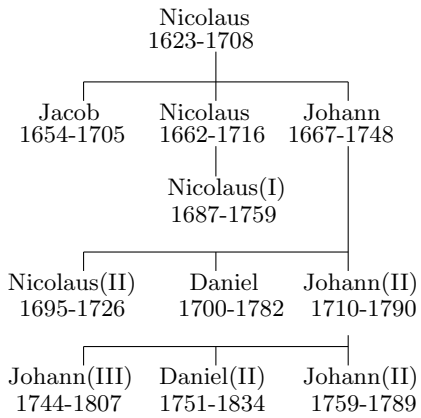

圖1. Bernoulli 數學大家族

伯努力所指的“趨近”是什麼意思?

令$n(A)$表觀測n次, 事件A所發生之次數。又以$p_{n,k}$表$n(A) = k$之機

率。則

(1) $$p_{n,k} = P(n(A) = k) = \binom{n}{k} p^k (1-p)^{n-k},\ k = 0, 1, \cdots, n\ 。$$

當n很大時, 事件A發生之相對頻率$n(A)/n$與p之差距不應太大。但$n(A)$可能每回觀測都不盡相同, 畢竟這是一隨機現象。有時$n(A) = n$, 有時$n(A) = 0$。我們上一節才討論巨數法則。投擲一公正銅板100次, 100次全出現正面的機率當然很低, 僅$1/2^{100}$。但若一直重複做這件事(每回投擲100個銅板), 譬如說做了2^{100}回, 則其中出現一回100個全是正面, 就不用太奇怪。如果回數再多些, 譬如做2^{110}回, 就更容易出現好幾回100個全是正面了。因平均可出現$2^{110}/2^{100} = 2^{10} = 1,024$(次)。

註2.要完成2^{100}回投擲銅板100次, 其實並非易事。假設以電腦模擬, 且1秒鐘可模擬1萬兆($= 2^{16}$)回, 夠快了吧! 1天有86,400秒, 1年365天約有$3.1536 \cdot 10^7$秒。因此一年約可模擬$3.1536 \cdot 10^{23}$回。又$2^{100} \doteq 1.2676506 \cdot 10^{30}$, 二者相除, 得到約要模擬$4.01969 \cdot 10^6$年。要四百多萬年才能模擬完, 中華民族才號稱有五千年文化。野史裡偶有投擲出100個正面的記載, 如宋朝名將狄青曾辦到, 見黃文璋(2003)p.71, 那些銅板當然都是特製的。如果銅板為公正, 你現在知道了, 投擲100個, 是很難出現100個全是正面。若想靠多投擲幾回而得, 那回數之多, 乃遠超乎我們所能想像。

對於隨機現象的種種解釋, 自然須以機率為依歸。伯努力是認為$n(A)/n$與p之差是"不太可能"過大。即只要n夠大, $|n(A)/n - p| > \varepsilon$之機率應很小, 其中$\varepsilon$為任一正數。而此機率為

(2) $$\sum_{|k/n-p|>\varepsilon} p_{n,k} = \sum_{|k/n-p|>\varepsilon} \binom{n}{k} p^k (1-p)^{n-k}\ 。$$

伯努力辛苦地證明

(3) $$\lim_{n\to\infty} \sum_{|k/n-p|>\varepsilon} \binom{n}{k} p^k (1-p)^{n-k} = 0\ 。$$

其中的求和當然不會是容易的工作。今日已少有人看過(可能也不想看, 甚至不知道)他的證法。有興趣的讀者, 可參考Rényi(1970)pp.195-

196。事實上, 利用**柴比雪夫不等式**(Chebyshev inequality, Pafnuty L. Chebyshev, 1821-1894, 為俄國著名數學家), 可輕易地證出比伯努力更一般的結果, 可參考黃文璋(2010)p.124。

伯努力所指的事件發生之相對頻率, 其實就是

$$f_n(A) = \frac{n(A)}{n} = \frac{X_1 + \cdots + X_n}{n}, n \geq 1,$$

其中$X_1, \cdots, X_n$為獨立且有相同分佈(independent and identically distributed, 簡稱iid)之隨機變數, X_i取值0或1, 且$P(X_i = 1) = p, i \geq 1$。隨機變數只取值0或1, 且取值1的機率為p, 便稱為有參數p之**伯努力分佈**(Bernoulli distribution), 以$Ber(p)$表之。伯努力即指出, iid的隨機變數$X_i, i \geq 1$, 以$Ber(p)$為共同分佈時, 當$n \to \infty$時, $(X_1 + \cdots + X_n)/n$ 會偏離p"太遠"(如差距超過ε)的機率趨近至0。

讀者可看出大數法則與巨數法則, 是不同的兩個法則。曾在網路上一篇文章中, 看到底下一段話:

> 記得保險業流行一條"大數法則", 意思是只要大量接觸客人, 十位或一百位或一千位, 總有一位客人投保。換句話說, 你失敗, 只是個人不夠勤力, 接觸的客人不夠多之故。

其實這應是巨數法則, 而非大數法則。對一事件A當觀測數n很大, 巨數法則關心A是否發生, 大數法則則是說事件A發生的總次數$n(A)$, "約"是np, 其中p為A發生的機率。也可以這麼說, 大數法則是比巨數法則更精確的法則。它指出當觀測數n很大, 一特定事件大約發生多少次。因n很大時, 事件A發生約np次, 即使p很小, 只要$np > 1$, 則看到事件A發生, 就不足為奇, 而這就是巨數法則所指出的現象。

大數法則可以支持**頻率對機率的解釋**。設有一銅板, 出現正面的機率為p(或一事件發生之機率為p), 只投擲一次, 不是正面就是反面, 無法感受p的意義。但只要投擲數夠大, 銅板出現正面的相對頻率, 就很可能會接近p了。伯努力之後, 機率學家們, 繼續探討大數法則。1928年, 俄國機率學家**辛欽**(Aleksandr, Y. Khinchin, 1894-1959), 證明對iid的隨機

變數, 只要期望值存在, 不論分佈為何, 大數法則便成立。即對$\forall n \geq 1$, 設$X_1, \cdots, X_n$為iid之隨機變數, 且設$E(X_1)$存在(即$-\infty < E(X_1) < \infty$), 則**樣本平均**(sample mean)

$$\overline{X}_n = \frac{X_1 + \cdots + X_n}{n}, n \geq 1,$$

當$n \to \infty$時, 會**機率收斂**(converges in probability)至$E(X_1)$, 以

$$\overline{X}_n \xrightarrow[n \to \infty]{P} E(X_1) \tag{4}$$

表之。在此機率收斂的定義如下。

定義1. 設有一數列之隨機變數$\{Y_n, n \geq 1\}$, 及一隨機變數Y, 若

$$\lim_{n \to \infty} P(|Y_n - Y| > \varepsilon) = 0, \forall \varepsilon > 0, \tag{5}$$

則稱$n \to \infty$時, $Y_n, n \geq 1$, 機率收斂至Y, 且以$Y_n \xrightarrow[n \to \infty]{P} Y$表之。

(5)式是說, 對$\forall \varepsilon > 0$, 當$n \to \infty$時, $|Y_n - Y| > \varepsilon$之機率趨近至0。亦即$|Y_n - Y| \leq \varepsilon$之機率趨近至1:

$$\lim_{n \to \infty} P(|Y_n - Y| \leq \varepsilon) = 1, \forall \varepsilon > 0 \text{。} \tag{6}$$

辛欽所證出的結果, 後來被稱為**弱大數法則**。當iid的隨機數列$X_n, n \geq 1$, 只取$0, 1$兩個值, 且$P(X_1 = 1) = p$, 則$E(X_1) = p$, 這時弱大數法則就回到伯努力的版本。弱大數法則並無法保證n很大時, 樣本平均$\overline{X}_n$與隨機變數之期望值$E(X_1)$"必會"很接近, 這與常數數列的收斂大不相同。對於隨機現象, 所有的保證, 都是機率式的。即弱大數法則保證n很大時, 樣本平均$\overline{X}_n$, 與期望值$E(X_1)$很接近的機率很大。不論$\overline{X}_n$與$E(X_1)$之差距要多小(譬如說$|\overline{X}_n - E(X_1)| \leq \varepsilon$, 其中$\varepsilon$可為一任意正數), 此機率要多大(譬如說$P(|\overline{X}_n - E(X_1)| \leq \varepsilon)$大於$1 - \delta$, 其中$\delta$可為一任意正數), 皆能辦到, 只要樣本數$n$夠大。即對每一$\varepsilon > 0$, 及$0 < \delta < 1$, 存在一$n_0 \geq 1$, 使得

$$P(|\overline{X}_n - E(X_1)| \leq \varepsilon) > 1 - \delta, \text{ 當} n \geq n_0, \tag{7}$$

或等價地說

$$\lim_{n\to\infty} P(|\overline{X}_n - E(X_1)| \le \varepsilon) = 1,\ \forall \varepsilon > 0 \text{ 。} \tag{8}$$

弱大數法則並沒有說, n很大時, X_n會等於$E(X_1)$ 。很多初學者會誤以為如此, 但這是錯的 。即使n很大時, $|\overline{X}_n - E(X_1)|$必很小, 此說法也是錯的 。這些皆非機率式的說法 。正確的說法是n很大時, $|\overline{X}_n - E(X_1)|$很小的機率會很大 。以數學式子表示就是(7)式, 或(8)式 。對於隨機現象, 我們通常只能做機率式的保證 。初學者或許會以為, 我們的保證好似都有些保留, 不像數學中一向斬釘截鐵式的說法有權威 。但這就是機率與數學之別: 看似不太確定, 其實是更可靠的保證 。只要想甲醫生肯定地說這病人活不過3個月, 乙醫生說病人活不過3個月的機率為0.99, 你覺得哪一種講法較精準?

弱大數法則不只針對iid的隨機變數才成立, 條件可放寬些 。歷來機率學家, 給出不同條件下的弱大數法則, 使其適用性更廣 。

大數法則針對期望值存在的隨機變數, 期望值若不存在, 像是**柯西分佈**(Cauchy distribution), 就不適用了 。圖2給出當$X_n, n \ge 1$, 為iid之隨機變數, 且以$Ber(1/2)$為共同分佈, $\overline{X}_n$之一模擬圖形, $1 \le n \le 1,000$ 。可看出除了n不太大時, $\overline{X}_n$與0.5有較大差距, 之後就與0.5很接近了 。對參數2.5及1之柯西分佈, 以$\mathcal{C}(2.5, 1)$表之, 即機率密度函數為

$$f(x) = \frac{1}{\pi(1 + (x - 2.5)^2)}, x \in R,$$

圖3給出$\overline{X}_n$之模擬圖形 。由於此分佈之期望值不存在, 大數法則不適用, 圖形顯示$\overline{X}_n$之振盪頗大 。

在統計裡, 很多不錯的統計方法, 都基於大數法則 。著名的**動差估計法**(method of moment estimator)為一例 。設隨機變數X有參數λ之**波松分佈**(Poisson distribution), 以$X \sim P(\lambda)$表之, $\lambda > 0$ 。即X之機率密度函數為

$$f(x) = \frac{e^{-\lambda}\lambda^x}{x!}, x = 0, 1, \cdots \text{ 。}$$

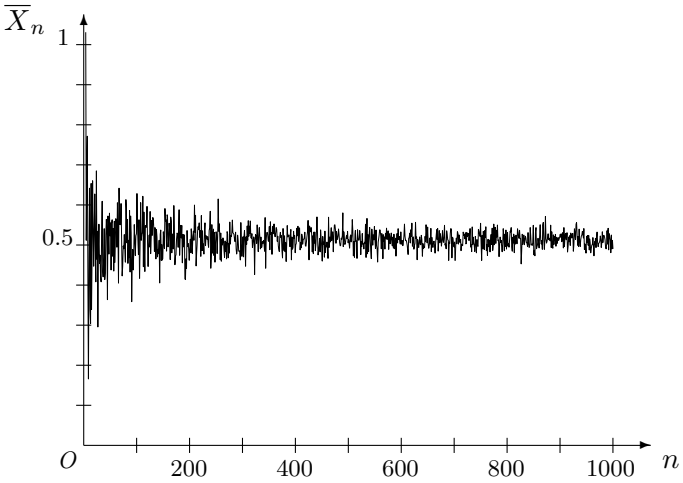

圖2. $Ber(1/2)$分佈$\overline{X}_n$, $1 \le n \le 1,000$, 之模擬圖

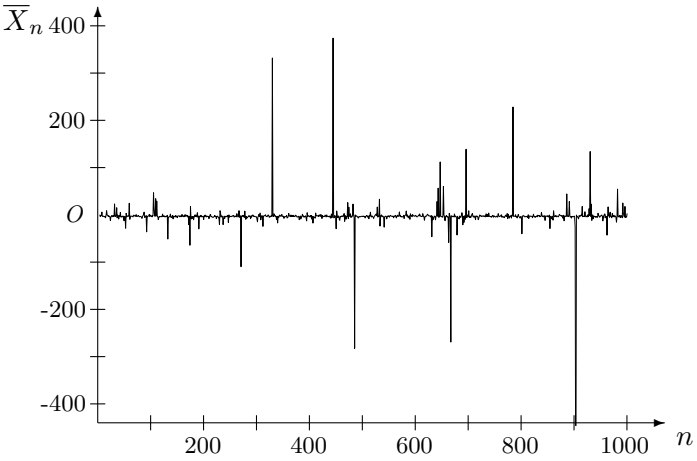

圖3. $\mathcal{C}(2.5,1)$分佈$\overline{X}_n$, $1 \le n \le 1,000$, 之模擬圖

我們想要估計λ, 獨立且重複地觀測n次, 得到樣本$X_1, \cdots, X_n$。則可利用樣本平均$\overline{X}_n$作為λ之估計量。這種因$E(X) = \lambda$, 就以樣本平均$\overline{X}_n$取代期望值$E(X)$, 作為λ之估計量, 即為動差估計法。由於$E(X) = \lambda$恰為X之期望值, 此略可說明何以動差估計量通常是不錯的估計量。因至少我們確定取樣夠多時, 它接近欲估計的參數之機率便很大。

我們常會接觸到機率為0之事件。自區間$[0,1]$隨機地取一個點, 會取

中0.3的機率為何? 0, 你可能會不假思索地回答。你還知道任何一點被取中之機率皆為0。如果令X表取中之點, 則X為一連續型的隨機變數。因此X會等於$[0,1]$間任一數之機率皆為0。修過機率及統計的課程, 使你具備基本的機率知識。但有人一取之下, 取中0.729, 他問你機率不是為0嗎? 怎麼卻發生了?

要知**機率為0的事件, 不代表不會發生**, 此現象常讓初學者感到迷惑。事實上, 事件之"不會發生"、"會發生", 及"一定發生"等, 皆非專業的說法。專業的說法分別是:機率為0, 機率為正, 及機率為1。有人取中0.729, 他懷疑0.729被取中的機率其實大於0。如何跟他解釋? 此時大數法則便可派上用場。你請他再取1次, 第3次取, 繼續取, 依序取了n次, 如此得到隨機數列$X_1, \cdots, X_n$, 其中$X_i = 1$, 表第i次取中0.729, $X_i = 0$, 表第i次未取中0.729, $i \geq 1$。然後算出0.729出現次數的相對頻率, 即求

$$\overline{X}_n = \frac{X_1 + \cdots + X_n}{n}, n \geq 1 \text{。}$$

雖然$X_1 = 1$, 但自X_2起, 應都是0了。隨著n之增大, $\overline{X}_n$將愈來愈接近0。何以我們那麼"確定"? 因弱大數法則告訴我們$n \to \infty$時, $\overline{X}_n \xrightarrow[n \to \infty]{P} E(X_1) = 0$ 。

由於上述這種困擾, 有時機率學家, 傾向以**幾乎不可能**(almost impossible), 及**幾乎確定**(almost certain), 來取代"不可能"(不會), 及"確定"(必會)。附帶一提, 當有人告訴你, 他自區間$[0,1]$隨機地取一個點, 取中0.729, 你可大膽的請問他, 是否真的是0.729? 你的量尺夠精確嗎? 為何你能如此有信心, 該數很可能是因量測不夠精確, 才以為是0.729, 其實並不是0.729? 此因$[0,1]$間有理數的集合是**可數的**(countable), 而可數集合之**勒貝格測度**(Lebesgue measure)為0, [0, 1]間無理數之勒貝格測度為1。因此隨機取一點, 會取中無理數之機率為1。我們當然寧可認為是因他沒量精準, 才以為得到的是有理數。

最後, 隨機變數的不獨立, 或分佈不同, 在適當條件下, 弱大數法則仍會成立。可參考黃文璋(2010)5.3節。

5 強大數法則

對於上一節事件A發生的相對頻率$f_n(A) = n(A)/n$, 我們想知道$n \to \infty$時, 其極限行為。直觀上由弱大數法則, 會認為

$$\lim_{n\to\infty} f_n(A) = p, \tag{9}$$

其中$p = P(A)$。但我們知道此並不正確。因極限不見得要存在, 且即使存在, 也不見得是p。對$n \geq 1$, $f_n(A)$有時恆為0, 有時恆為1。前者的極限為0, 後者的極限為1。我們最多可以問的是:是否對"幾乎所有"(almost all)回的觀測, (9)式皆成立? 在此, 一回觀測, 指的是第1次, 第2次, $\cdots$, 第n次, $\cdots$, 一直無止盡的進行。我們想知道觀測到(9)式之機率是否為1? 即是否有

$$P(\lim_{n\to\infty} f_n(A) = p) = 1, \tag{10}$$

或等價地問

$$P(\lim_{n\to\infty} f_n(A) \neq p) = 0 \tag{11}$$

成立否? 只是(11)式不見得好驗證。我們得先找出$\lim_{n\to\infty} f_n(A) \neq p$的觀測之集合$N$, 再看$P(N)$是否為0。除非$p = 1$, 否則1個$A$一直發生的觀測, 當然使$f_n(A) \nrightarrow p$。故此觀測屬於$N$。我們知道第1至第$n$次觀測, A皆發生的機率為p^n, $n \geq 1$。而$n \to \infty$時, p^n趨近至0。但N中可是包含無限多個觀測, 是否可數呢? 不能再說下去了, 讀者大約已搞糊塗了。總之, N中有那些元素, 不是那麼容易描述, 因此要驗證$P(N) = 0$並非易事。

由(9)式, 後來就衍生出更一般的**強大數法則**。

弱大數法則與強大數法則, 本質上並不是兩個完全不同的法則, 而是描述樣本平均以不同的方式收斂至母體平均。弱大數法則是說, 當樣本數無止盡的增大, 樣本平均與母體平均的差距, 可以小於任意正數之機率, 將無止盡的接近1。而強大數法則是說, 當樣本數無止盡的增大, 樣本平均的極限與母體平均相等的機率, 將無止盡的接近1。即設

對$\forall n \geq 1$, $X_1, \cdots, X_n$為iid之隨機變數, 且設$E(X_1)$存在, 則樣本平均$\overline{X}_n$, 當$n \to \infty$時, 會**幾乎確實地收斂**(converges almost surely)至$E(X_1)$, 以

$$\overline{X}_n \xrightarrow[n \to \infty]{a.s.} E(X_1) \tag{12}$$

表之。在此幾乎確實地收斂之定義如下。

定義2.設有一數列之隨機變數$\{Y_n, n \geq 1\}$, 及一隨機變數Y, 若

$$P(\lim_{n\to\infty} Y_n = Y) = 1, \tag{13}$$

則稱$n \to \infty$時, $\{Y_n, n \geq 1\}$幾乎確實地收斂至Y, 且以$Y_n \xrightarrow[n \to \infty]{a.s.} Y$表之。

(13)式等價於

$$P(\lim_{n\to\infty} |Y_n - Y| \leq \varepsilon) = 1, \forall \varepsilon > 0 \text{。} \tag{14}$$

因此(12)式與底下三式皆等價:

$$P(\lim_{n\to\infty} \overline{X}_n = E(X_1)), \tag{15}$$

$$P(\lim_{n\to\infty} |\overline{X}_n - E(X_1)| \leq \varepsilon) = 1, \forall \varepsilon > 0, \tag{16}$$

$$P(\lim_{n\to\infty} |\overline{X}_n - E(X_1)| > \varepsilon) = 0, \forall \varepsilon > 0 \text{。} \tag{17}$$

不少初學者對於(13)式的涵義感到困惑, 我們略微說明如下。首先不論定義1或定義2中, $\{Y_n, n \geq 1\}$要機率收斂, 或幾乎確實地收斂至Y, $\{Y_n, n \geq 1\}$與Y, 皆要定義在同一機率空間。現假設$\{Y_n, n \geq 1\}$與Y, 皆定義在同一機率空間$(\Omega, \mathcal{F}, P)$, 又注意到這些隨機變數皆是由Ω映至R。則(13)式就是

$$P(\{\omega | \omega \in \Omega, \lim_{n\to\infty} Y_n(\omega) = Y(\omega)\}) = 1 \text{。} \tag{18}$$

在這類式子中, 我們常省略“$\omega \in \Omega$”, 即只寫成

$$P(\{\omega | \lim_{n\to\infty} Y_n(\omega) = Y(\omega)\}) = 1, \tag{19}$$

甚至簡單地以(13)式表之。大家在微積分中學過函數數列。設有一數列之函數$f_n, n \geq 1$, 且設這些函數有相同的定義域。對每一定義域中的x, 可得一數列$f_n(x), n \geq 1$。有些x會使$\lim_{n\to\infty} f_n(x)$存在, 有些則不會。例如, 取

$$f_n(x) = x^n + n^2(1-x^2)^n + nx^{2n} + (\cos(x-0.5))^n, x \in [-1,1], n \geq 1,$$

$$f(x) = 0, x \in [-1,1]。$$

則滿足$\lim_{n\to\infty} f_n(x) = f(x)$之$x$的集合為$[-1,1]\backslash\{-1,0,0.5,1\}$。不收斂點僅有$-1, 0, 0.5,$及1等4點, 其勒貝格測度為0, 因此$\{f_n, n \geq 1\}$幾乎確實地收斂至$f$。因如前所述, 實數上可數個點之勒貝格測度為0, 故若在$[-1,1]$中可數個點改變函數f之值, 仍會有$\{f_n, n \geq 1\}$幾乎確實地收斂至f。例如,取$g(x) = 1$, 若x為$[-1,1]$中之有理數; $g(x) = 0$, 若x為$[-1,1]$中之無理數。則$\{f_n, n \geq 1\}$幾乎確實地收斂至g。機率函數亦為一測度, $\{Y_n, n \geq 1\}$是否幾乎確實地收斂至Y, 就要看$\{\omega | \lim_{n\to\infty} Y_n(\omega) = Y(\omega)\}$之機率是否為1, 或等價地說不收斂的$\omega$之集合$\{\omega | \lim_{n\to\infty} Y_n(\omega) \neq Y(\omega)\}$, 其機率是否為0。要注意的是, 與實數系統不同, 有時候一個ω的機率便為正, 特別是若機率空間為離散型。在機率論裡, 關於幾乎確實地收斂, 學生往往對其意義感到很茫然。其實只要與實數裡函數數列的收斂相對照, 就不難理解了。

當iid的$\{X_i, i \geq 1\}$以$Ber(p)$, 為共同分佈, 此時強大數法則之證明, 可參考黃文璋(2010)pp.324-326。一般分佈下的證明, 見Chung(2001) Theorem 5.4.2。又強大數法則亦有非iid隨機變數的版本。

在機率論裡會證明, 幾乎確實地收斂會導致機率收斂。因此前者是較強的收斂, 後者是較弱的收斂。這是何以會各命名為強大數法則及弱大數法則。我們以一簡單的例子, 來略微說明機率收斂與幾乎確實地收斂之別。

假設科學家研究複製人的技術逐漸進步。第k次複製時, 將人體分成k個區域, 而複製出的k個產品, 每一皆恰有一區域與人體不符, 但與人體不符的區域, k個複製品皆不一樣。將全部複製品依序編號。顯然至第k次複製, 共有$\sum_{i=1}^{k} i = k(k+1)/2$個產品。機率收斂就對應這個情況: 編號愈大的複製品, 與人體的差異愈來愈小。但任選一區域, 不論編號多大, 皆有很多複製品, 與此區域不符。這就對應不幾乎確實地收斂。

現在看另一情況。每次複製一個, 第k次複製, 將人體分成的k個區域, 是由頭頂依序至腳底。而那一複製品, 只有一個區域與人體不符, 且是最靠近腳底的那個區域。複製品依序編號, 則隨編號之增大, 複製與人體逐漸可重疊, 幾乎不分軒輊。這就對應幾乎確實地收斂。

最後, 強大數法則亦有隨機變數非iid的版本, 如見黃文璋(2010)定理7.8。此處只是初步的介紹, 不多討論。

參考文獻

1. 黃文璋(1999). 數學欣賞。華泰文化事業股份有限公司, 台北。
2. 黃文璋(2003). 隨機思考論。華泰文化事業股份有限公司, 台北。
3. 黃文璋(2010). 機率論, 第二版。華泰文化事業股份有限公司, 台北。
4. Chung, K.L.(2001). *A Course in Probability Theory*, 3rd ed. Academic Press, New York.
5. Diaconis, P. and Mosteller, F.(1989). Methods for studying coincidences. *Journal of the American Statistical Association*, 84, 853-861.
6. Littlewood, J.E.(1953). *A Mathematican's Miscellany.* Methuen, London.
7. Rényi, A.(1970). *Foundations of Probability.* Holden-Day, Inc., San Francisco.

8. Shermer, M.(2003). Codified claptrap. *Scientific American*, June , 288(6), 35.

9. Shermer, M.(2004). Miracle on probability street. *Scientific American*, August, 291(2), 32.