庶民中央極限定理

黃文璋
國立高雄大學應用數學系

1 前言

德國馬克是以前西德的法定貨幣, 東西德統一後, 繼續使用至2002年歐元流通為止。有段時間10馬克紙幣上, 放的是**高斯**(Carl F. Gauss, 1777-1855)之肖像。大數學家高斯, 在許多領域, 如統計、代數, 及分析等, 皆有極開創性的突破。甚至還有人認為, 不少高斯的研究成果, 至今仍能深深影響人類生活, 其貢獻較其他舉世知名的科學家, 都廣泛許多, 可參考黃文璋(2008)一文。但被挑出來放在紙幣上與他長相左右, 不是別的, 而是**常態分佈**的**機率密度函數**及其圖形。由此一方面可看出常態分佈的重要, 一方面大約可理解, 德國這個國家, 其國民的數學素養怎會不好? 他們鈔票上有數學家, 還有一個大有來頭的函數。這函數絕非簡單, 其中有兩個數學中極重要的常數π及e, 連同幾個數學中常用的符號$\sqrt{\ }$及希臘字母μ與σ。

近年來可能是鑑於各行各業常在做民調, 民調已融入生活, 負責制訂高中數學課程綱要的委員, 因此覺得有必要讓一般人民了解其中常出現的**信賴區間**。於是在民國95年版的高中數學課程綱要裡, 開始引進信賴區間。民國99年版的課程綱要中, 仍保有此題材。雖僅討論**二項分佈**$\mathcal{B}(n,p)$裡參數p之估計, 但為了近似, 也引進**中央極限定理**, 及隨之而來

的常態分佈。德國在鈔票上有常態分佈, 我們在高中裡介紹中央極限定理, 算是與他們互別苗頭。只是鈔票能用就好, 其上的圖案, 有何深遠的含意, 懂不懂都沒關係。但既然將這些題材擺進高中數學, 顯然是想提高國民的統計素養, 且還不得不擠掉一些原本有的題材, 因此總要能傳遞正確的概念。否則不但無法達到統計深耕的目的, 並將有礙日後學習進階的統計。

大家可能也聽過有些人是怎麼看統計的:

> 有三種謊言: 謊言, 可惡的謊言, 統計。

時至今日, 統計早已成為做決策不可缺少的工具。但若未運用得當, 將比不採任何統計方法還要糟。統計工作者, 當然得避免提出的分析報告, 被與謊言歸於同一類, 棄之如敝屣。只是統計裡的一些基本想法, 所謂**統計思維**, 與數學是大異其趣的, 可參考黃文璋(2009)一文。統計裡強調**隨機性**, 與一向著重**必然性**的數學, 本質上並不易相容。因此要在中學裡教統計, 比較好的作法, 是讓統計成為單獨的一科, 而非像目前放進數學課程中。要知信賴區間與中央極限定理, 這些連在大學機率統計課程裡, 都不算簡單的題材, 即使循序漸進地跨入, 學生都難免一知半解, 如今卻堂而皇之地進入高中數學。在低估此題材之深度下, 缺乏足夠鋪陳, 便快速進入一雖非邈然不可攀, 但也有些難度的境界, 不論在教學或學習, 引起很多困擾, 乃可預期。

有如憲法是國家的根本大法, 教科書乃依課程綱要而編寫。為何我們說制訂時, 低估信賴區間等題材之深度, 甚至因此還訂出這套缺乏一言而為天下法的課程綱要呢? 在課程綱要"常態分布、信賴區間與信心水準的解讀"那項下寫著:

> 高中程度的統計推論只做隨機變數期望值的估計, 它的背後理論是中央極限定理。要介紹中央極限定理, 就需要引入常態分布。此部分僅做通識性的介紹, 以活動方式建立學生對於中央極限定理的直觀。
>
> 對一固定的信心水準, 給出信賴區間公式, 再讓學生以亂數

表模擬或實驗投擲正面出現機率為p的銅板n次, 代入信賴區間公式, 以說明信心水準的意涵; 並以此解讀, 何以大多數的學生所得的信賴區間都會涵蓋p?

首先, "隨機變數期望值的估計, 它的背後理論是中央極限定理", 此講法不知從何而生。要知**參數估計**, 通常分**點估**計及**區間估計**。中央極限定理, 對於點估計(不管是否估計期望值), 根本用不到; 而對區間估計, 不少情況下也不需要, 遑論成為背後理論? 以此做為引進中央極限定理的理由, 高斯若地下有知, 恐亦會感到莫名所以。其次"此部分僅做通識性的介紹", 雖說得輕鬆, 只是中央極限定理的概念, 豈有那麼容易? 而且究竟"通識性"是指什麼? 難道高中課程裡, 有非通識性的嗎? 由於課程綱要中特別強調通識性, 有些教科書可能因企圖落實通識性, 將氣勢磅礴的中央極限定理, 寫得花拳繡腿。這還不打緊, 有時還因此忽略嚴謹性, 甚至將此定理詮釋錯誤。

主要由於隨機性的本質, 再加上不同分佈間的差異, 對參數估計, 往往有不一而足的方法, 由於各有優點(可參考黃文璋(2007)一文), 因此能長期並存。沒有放之四海而皆準的估計, 乃統計的特色之一。參數區間估計的方法不會唯一, 便也很顯然。因此有關統計的教材裡, 通常不會出現有如昔日的**建構式數學**, "給出"信賴區間公式, 這種定於一尊的說法。

至於對銅板正面出現的機率p, 課程綱要中提議以實驗來顯示"大多數"學生得到的信賴區間皆會涵蓋p。此講法完全缺乏隨機的概念, 其不恰當, 雖早在黃文璋(2006)一文中便已指出, 只是顯然未被理會。

另外, 有幾本教科書裡, 皆曾出現類似"本書僅討論離散型的隨機變數"之句子, 頗令人納悶。不是有重頭戲常態分佈嗎? 這難道不是連續型? 原來課程綱要裡有一句"高中課程只處理離散型的隨機變數"。編撰者只顧擷取課程綱要中的句子, 也不管究竟正不正確。

絕學無憂, 重重問題固然令人憂心忡忡。只是這些題材都已經放進高中數學了, 一時也無可奈何。務實地看, 到底有沒有什麼方式, 可以略微提高學習成效? 雖云道可道非常道, 本文仍嘗試對現行高中數學裡, 環繞中央極限定理的幾個概念, 做些闡釋。

2 隨機現象

宇宙的運轉, 有**必然性**和**隨機性**。銅板落地所需時間可求出, 這是必然性, 但那一面將朝上卻不知, 這是隨機性。對事先不能預知結果的觀測或試驗, 便稱為**隨機現象**(random phenomenon)。意思本已很明白, 但曾有教科書又畫蛇添足, 加以如下說明:

> 一般而言, 如果一個試驗具有隨機現象, 可在相同條件之下重複進行; 雖然無法預知其結果, 但是我們可以知道所有可能出現的結果。

這段文字不太通順就不去管它, 但必須一提的是, 以這樣的方式解釋隨機現象, 是不夠"一般"的。

諸如A先生是否今年底前結婚, B籃球隊是否得年度冠軍, 都屬隨機現象。但"今年"只有一次; 至於球隊是可能在未來某年獲冠軍, 只是每年每隊的成員都不盡相同, 連參賽隊伍也不見得一樣, 因此每年的球賽, 並不能算是在相同條件下重複進行。

另外, "可以知道所有可能出現的結果", 此句也有些爭議。首先, 我們得釐清什麼是可能, 不可能。有時人們遇到較困難的事, 會說"這是不可能的任務", 但再如何不可能, 有些最後仍完成。考完大學學測, 某人問你的總級分可能多少? 你正為考得不太好而傷心, 實在沒興趣回答此問題, 便敷衍地說0至75分都有可能。問的人覺得你不可能考到75分, 也不可能考0分。所以可能, 不可能, 不見得有那麼嚴格的界限。因此將"知道所有可能出現的結果", 特別放進對隨機現象的解釋, 彷彿此為隨機現象須具備的條件, 並不恰當。

你想知道今天出門, 所遇到最高的人, 身高可能是多少? 你其實連世界身高最高紀錄為何, 都不太清楚。但你認為將"所有可能出現的結果", 取成介於150公分至300公分間, 便萬無一失。換句話說, 不像投擲一公正的骰子一次, 並觀測所得點數的情況, 在討論隨機現象時, 那些才屬於可能出現的結果, 有時不太確定。這時將所有可能出現的結果之集合, 取得較大些, 乃無關緊要。亦即雖列在可能出現的結果中, 有些其實根本"不可

能”出現。因此, 在不少情況下, 所有可能出現的結果之集合, 不見得有一致的看法。此集合有時是為了不至於掛一漏萬而取的, 倒不見得是我們一開始便知道那些結果眞會出現。讀者也不難想通, 不能以發生的機率不為0, 來界定那些屬於可能出現的結果。例如, 由0至1的區間中, 隨機地取1點, 則整個區間中的點, 當然都屬所有可能出現的結果。但顯然該區間中任何一點, 被取中的機率皆為0。

對一隨機現象, 所有可能出現的結果之集合(可以是觀測所得, 或指定), 稱為**樣本空間**。我們已指出, 前述集合常非固定。因此如Feller(1968) pp.13-14所說, 樣本空間的概念, 有如幾何學裡的點、線、面, 並無嚴格定義, 指定樣本空間是較明確的作法。投擲一銅板兩次, 如果不指定, 卻問樣本空間為何? 這是不清不楚的問題。若觀測兩次會出現正、反面的情況, 則樣本空間可取為{(正, 正), (正, 反), (反, 正), (反, 反)}, 其中(正, 反)表第一次得正面且第二次得反面, 餘類推。若觀測兩次共出現的正面數, 則樣本空間可取為{0, 1, 2}。若觀測兩次共出現的正面數減去共出現的反面數, 則樣本空間可取為{-2, 0, 2}。因觀測的不同, 尚可有其他樣本空間。整個機率的架構是很有彈性的, 萬不可以為對投擲一銅板兩次, 樣本空間一定是{(正, 正), (正, 反), (反, 正), (反, 反)}。會這樣誤解, 原因大約是根深蒂固古典機率的想法所造成。以為銅板必公正, 各次投擲必相互獨立等。趁早跳出此桎梏, 對往後靈活且正確的學習機率, 其實是較好的。通常若確定觀測什麼現象, 便可寫出一樣本空間。甚至可更一般, 連觀測都不必, 直接說“取”, 或“令”樣本空間為何, 然後繼續之後的討論。

為何我們強調上述那幾點呢? 對於數學, 人們通常較能嚴肅以對, 因從小在數學課中, 我們一再被要求須精準。但對涉及隨機現象的問題, 以為反正也不一定會出現什麼結果, 因此不管提出什麼結論, 都不見得立即能被判定對或錯, 何必如數學中的嚴謹? 在一知半解下, 隨機遂轉化為隨便, 對事物的描述, 便常不覺得須很明確。

今日大家對機率一詞, 皆能朗朗上口, 但究竟其意義為何, 一般人卻可能不甚了了。在機率教科書上, 通常可見到下述三種對機率的解釋:

以相同的可能性來定義機率, 此為**古典的定義**。

在多次相同條件下重複試驗後, 以一事件出現的相對頻率, 來定義機率, 此為**統計的定義**, 或說**客觀的解釋**, 或說**頻率對機率的解釋** 。

以觀察者對一事件之相信程度, 來定義機率, 此即**主觀的觀點**, 或說**對機率的主觀解釋** 。

由前述三種解釋, 得到啓示, 二十世紀的三十年代, 產生了以公理化的方式, 來定義機率 。定義中包含三基本要件:

樣本空間, **事件的集合**, 及**機率函數** 。

各基本要件, 皆有一些條件 。如樣本空間不可為空集合, 這是很顯然的; 又事件為樣本空間之子集合, 其機率須介於0與1之間等, 這符合長久以來人們對機率的認知 。前述三基本要件, 便構成**機率空間**, 想進一步了解的讀者, 可參考黃文璋(2011a)一文 。

你可能會好奇, 我們不是早就經常在求機率, 為什麼還需要機率空間? 此有如原本只是一群人分兩隊在操場踢球, 發展夠久後, 總會想訂個足球的比賽規則 。此規則當然會吸收長期以來, 在球場上形成的默契, 但會考慮的更周詳, 以因應各可能發生的情況 。

在機率空間裡, 樣本空間可以是虛擬的, 不必真的在執行什麼試驗, 或做什麼觀測 。就算真有試驗, 如前已指出, 樣本空間可以是指定的, 而試驗也不必是可以重複的 。對實際情況, 不能重複的試驗, 所涉及的機率, 往往是主觀的 。但主觀機率, 常也是根據過去一些客觀的資料而產生, 並非都毫無憑據 。只是由於觀測僅有一次, 在一翻兩瞪眼下, 難免被質疑所給的機率之正確性 。比如說, 我們想知道最近形成的那一個颱風, 是否會登陸台灣 。這怎能重複觀測? 氣象局預測登陸機率低, 颱風卻帶來強風豪雨; 預測登陸機率高, 颱風卻不見蹤影, 屢有發生 。

在機率空間裡, 集合不論有限或無限, 離散或連續, 都可以用來當做樣本空間 。不像在古典機率裡, 樣本空間只能是有限集合, 而且由於發生機率相同之故, 須知樣本空間中有幾個元素, 才能給定機率, 因此對那些能放進樣本空間, 便很在意 。事件的集合, 更是可大可小 。可小到只包含空

集合及樣本空間兩個元素; 也可大到包含樣本空間之所有子集合 。在古典的機率裡, 以及一般較初等的討論裡, 就是取這最大的樣本空間 。至於機率, 只對事件才給, 而機率函數也僅需滿足幾個特定的條件即可 。在看似有約束的機率空間中討論機率, 但一切可說彈性極大 。

你不妨這樣想, 訂出一很有彈性的足球比賽規則, 其中每隊人數 、場地大小, 及比賽時間等皆不固定, 但都得滿足某些條件 。訂規則者可不管地球上是否存在能供幾十億, 甚至更多的人比賽的場地 。換句話說, 所有對機率的討論, 可以僅是紙上談兵 。但這看起來抽象的機率空間, 卻非常一般, 無所不包 。那些原本常見的投擲骰子, 及撲克牌遊戲等, 所產生的機率情境, 不過都是機率空間之簡單實例而已 。

自引進機率空間的架構後, 機率理論便開始長足發展, 一切概念遂逐漸清晰起來 。大部分的情況下還無妨, 即使機率空間究竟為何, 並沒有很清楚, 也能討論機率 。只是若遇到有爭議, 或處理較細膩的問題時, 可能便得先釐清, 到底當下的機率空間為何? 有如必要時, 就得將平常束之高閣的憲法搬出 。

你是否信心十足, 覺得自己並非一般人, 了解機率的意義有何困難? 在奇摩的網頁上, 曾有人提出下述問題:

> 我先生的爺爺本身是雙胞胎, 我外婆也生過雙胞胎, 但公婆及我父母都沒有生下雙胞胎, 請問我們會不會生下雙胞胎? 如果會, 那機率會是多少?

有一個可能是醫學院的學生, 給出如下的答覆:

> 你的問題讓我一看就很想回答, 因為很像教授考試的題目, 大部分教授考的題目, 在臨床上都沒什麼用途 。
>
> 一般而言, 孕婦生雙胞胎機率是1/89 。如果家族中有雙胞胎, 這機率比1/89一定還要大 。你問我; 你會不會生下雙胞胎? 答案是會, 但是不保證, 只是機率比較大, 就好像你問我; 走在路上會不會被廣告看板砸下來, 打到頭? 我的答案也是會, 但是機率比較小 。結論是: 有問好像沒問 。

> 機率是統計學上騙人的東西, 許多事情要重複做100次才有機率可言, 懷孕不可能100次, 每次懷孕生雙胞胎機率是1/89, 但單次懷孕生雙胞胎機率若不是0%, 就是100%, 就好像問我, 50元銅幣丟到地上一次, 是蘭花機率有多少? 事實上, 50元銅幣丟到地上, 不是總統府, 就是蘭花, 如果丟到地上100次, 那麼機率就會接近50%。如果丟到地上一次, 蘭花的機率若不是0%, 就是100%。

前述回答者對機率的解釋, 雖然不正確, 但可能是不少人的看法。要知只要尚未確定前, 皆可問生出雙胞胎之機率。只是一旦"你"確定後, 對你而言, 此機率便不是0就是1了。投擲銅板的情況也類似, 並非要投擲很多次, 僅投擲1次就能談機率。而投擲100次, 也非(正面出現的)機率就會接近50%。而是正面出現次數的相對頻率, 在"某種意義"下, 會接近0.5。至於究竟如何接近, 以及對於隨機現象, 到底我們能否保證什麼? 這些問題較難些, 我們留待第5節再討論。

最後必須一提的是, 依機率空間的定義, 樣本空間中允許僅有一個元素。例如, 假設投擲一兩面皆為正之銅板, 則會出現那一面? 毫無疑問, 當然只可能是正面, 即樣本空間內只有正面一個元素。這原本並非隨機現象, 但我們仍將此視為隨機現象。或更明確地說, 視此為一**退化的**(degenerate)隨機現象。此兩正面銅板一擲之下, 出現正面之機率為1, 出現反面之機率為0。將退化的情況亦納入(其實是不得不納入), 河海不擇細流故能就其深, 所有觀測及試驗, 便都是隨機現象了。兩千多年前, **畢氏**(**畢達哥拉斯**, Pythagoras, 約西元前580-500年)學派有"萬物皆數"(All is number)。我們則說"萬物皆隨機"(All is random)。處在這隨機世界裡, 我們怎能不將隨機以及機率的概念, 都弄得更清楚些?

3 隨機變數

不時會聽到有些具統計博士學位的學者坦誠, 他們是進了研究所後, 才眞弄清楚隨機變數的意義。看來隨機變數的概念並不簡單。但不必訝

異, 這是我們高中數學裡便有的題材 。

我們知道, 樣本空間裡的元素, 不一定是數字 。例如, 做一民意調查, 想了解民眾對某議題的看法, 選項可能有極力贊成 、贊成 、不贊成 、極不贊成, 及沒意見等 。由於我們一向較慣於處理數字的問題, 因此可將極力贊成視為5, 贊成視為4, 不贊成視為3, 餘此類推 。也就是將觀測的結果數值化 。又如袋中有5紅球 、2白球及3藍球, 每次隨機取一球, 取出後放回, 連取10次, 想看能取中幾個紅球 。則可將紅球視為1, 白球及藍球皆視為0 。即使原先樣本空間中的元素就已是數字, 因種種目的, 我們仍可能會考慮它的某一數值函數 。眾所周知, 在數學中我們一向常考慮各式各樣的函數 。假設隨機選取一蘋果, 且秤其重量, 則可取樣本空間為區間$(0, \infty)$ 。若蘋果每單位重量的價格為a元, 則將蘋果的重量u, 對應至$X = au$, 而得其價格 。

由於以上這些原因, **隨機變數**(random variable)的概念便自然地產生了 。如前所述, 在初等的機率論裡, 由於所有事件的集合, 通常取成最大, 即包含樣本空間的所有子集合, 因此所謂隨機變數, 不過就是一個定義在樣本空間上的實數值函數 。若以X表一隨機變數, 則此函數的定義域為樣本空間, 對應域為實數集合R 。在較高等的機率論裡, 事件的集合不見得是取最大者, 則隨機變數便需額外的條件, 詳情可參考黃文璋(2011b)一文 。底下除非特別聲明, 否則我們都僅考慮初等的情況 。在同一樣本空間, 可定義出無限多個隨機變數 。引進隨機變數, 可能是基於數值化, 也可能只是純粹數學的目的 。

簡單講, 隨機變數就是一個函數 。既然如此卑之無甚高論, 那為何會讓不少人得經過多年後, 才終於豁然貫通, 了解其意義?

首先, 以往我們對函數, 常以f, g, h, 或大寫的F, G, H等命名; 函數的變數, 則以x, y, z等表示 。例如, 我們常說給定一函數$f(x)$ 。即使有時因討論的函數較多, f, g, h等符號不夠用了, 取名為$a(x)$, $b(x)$等, 或藉助希臘字母, 以$\xi(x)$, $\eta(x)$等來表示亦可 。但對隨機變數, 我們往往名之為X, Y, Z等 。又以往的函數, 變數常是附在一起的, 如$f(x)$, $g(y)$等 。但隨機變數, 雖明明是函數, 其變數卻不常與函數一起出現 。例如, 我們會

寫$f(x)=3$, 或$g(y)>5$等。而對隨機變數, 卻寫成$X=3$, 或$Y>5$等。尤有進者, 除了在一開始定義時, 說隨機變數為一實數值函數, 之後我們幾乎不稱它為函數, 連名稱都含有“變數”。因此較不易讓人與以往所學的函數, 連結在一起。

再看一個原因。樣本空間常以希臘字母Ω表示。隨機變數定義於Ω, 至於Ω中的元素如何出現, 則是隨機的, 依一機率函數來描述。與數學裡的情況相比, 我們說解$f(x)=3$, 其中x為實數, 就是找出所有滿足$f(x)=3$的實數x, 沒有那一實數會較易或較不易出現。但對隨機變數X, 我們較少說解$X=3$。X如果是由Ω映至R, 是可以有一滿足$X(\omega)=3$之解集合$\{\omega|X(\omega)=3,\ 其中\omega\in\Omega\}$(此集合常簡寫為$\{X=3\}$)。但我們通常不會就此打住, 而會去求諸如此集合之機率等。甚至後來經過一番改造, 我們常略去解$X=3$的過程, 而直接給出事件$\{X=3\}$的機率。換句話說, 雖然隨機變數本身是一函數, 但常另有跟機率有關的函數伴隨它, 指揮該隨機變數的行為。那些伴隨函數的性質, 就比較接近我們以往所遇到的函數。因此那些伴隨的函數, 便採用一般我們熟悉的函數之符號, 如f, g, h, 及F, G, H等。這也可以解釋為什麼在隨機變數這一階段, 我們不採慣用的函數符號。

第三個原因是, 如前所述, 有時事件的集合取得較小, 使得並非每一由Ω映至R的函數, 皆可當做隨機變數。這點便造成隨機變數的意義不是那麼容易理解了。

此外, 隨機變數的概念, 有時可能未精準地講授。我們來看現行高中數學課程綱要中, 是如何介紹隨機變數。在機率統計II中寫著:

> 生活中所接觸的變量(variables)常常具有隨機現象, 比如甲乙兩人猜拳n次, 甲贏乙的次數; 投擲銅板n次, 出現正面的次數; $\cdots$等等。這些具有隨機性(不確定性)的變量就稱作隨機變量(也叫隨機變數)。$\cdots$但隨機變數不需用機率空間上的函數來嚴格定義。

首先, 強調不需用機率空間上的函數, 來嚴格定義隨機變數, 此講法便有問題。在上一節我們已指出, 機率空間中有三基本要件, 但隨機變數僅定

義於其中的樣本空間, 因此不宜視為機率空間上的函數。且只說不需嚴格定義, 卻未說該如何定義。而所舉的例子, 皆為隨機現象觀測的結果原本便已是數值, 未能突顯數值化是隨機變數概念產生的主因。另外, 有教科書曾出現下述講法:

> 我們常將試驗的結果賦予數值, 稱為隨機變量。$\cdots$試驗的結果可以用一個變量來表示時, 我們稱這個變量為隨機變量。

前面一句還好, 後面一句就沒那麼妥當, 恐讓學生以為隨機變數僅針對試驗的結果為數值, 此句實有蛇足之嫌。底下為另一教科書裡的講法:

> 隨機表示結果的不可預知, 而變數表示每次結果會有不同的變化。

"結果的不可預知", 與"每次結果會有不同的變化", 不知其間有何差異? 為何一名之為隨機, 一名之為變數? 學生心中原有過去熟知的變數概念, 如今加上這樣的說明, 恐讓學生對"變數"一詞的意義更感迷惑。

其實在很多隨機試驗裡, 我們早就隱含地使用隨機變數的概念了。例如, 投擲一公正骰子兩次, 令X表點數和; 投擲一公正銅板20次, 令Y表所得正面數等。對於前述X, 可以想成原先觀測到(u, v), 其中u, v各為$1, \cdots, 6$中的某數, 再轉換至$X = u + v$。但也可想成一開始所觀測到的結果就是X。又雖隨機變數常以大寫的英文字母表示, 但其觀測值則以對應的小寫字母表示。有時會說"隨機變數X取值x", 或說"求$X = x$之機率"。

仍以Ω表樣本空間。如果Ω本來就是實數的一個子集合, 則對每一Ω中的元素ω, 令$X(\omega) = \omega$, 為一很自然的隨機變數。例如, 一袋中有10張紙牌, 分別寫數字1至10。隨機地取一張, 並觀測所得點數。取$\Omega = \{1, 2, \cdots, 10\}$, 令$X(\omega) = \omega$, 便得一隨機變數。另一種講法是, 自袋中隨機地取一張紙牌, 並令X表所得點數。這種講法就是直接引進隨機變數。其他如令Y表某袋中的紅球數, 令Z表某校戴眼鏡的學生人數等, 都是直接引進隨機變數之例。紅球、戴眼鏡者, 本來皆非數字, 但因我們將觀測的結果數值化, 或者說考慮數值上的特性, 便產生了隨機變數。

在同一樣本空間Ω中, 可定義無限多個隨機變數。即使$\Omega = \{0, 1\}$只有兩個元素, 對任二實數a, b, 令$X(0) = a$, $X(1) = b$, 便得一隨機變數。常數也是一隨機變數。即對一實數c, 若對每一Ω中的ω, 令$X(\omega) = c$, 則X為一隨機變數。這種只取一個值之隨機變數, 由於並不會有變化, 有時稱之為退化的隨機變數。

隨機變數已經是一函數, 對一隨機變數X, 我們又引出另一重要的函數, 即**分佈函數**(distribution function), 又稱分布函數, 或分配函數:

$$F(x) = P(X \leq x),\ x \in R\ 。$$

統計裡有時稱為**累積分佈函數**(cumulative distribution function), 但"累積"一詞, 其實多餘。對隨機變數X, 給定一實數x, $F(x)$乃表一機率值。由於我們常對形如$\{X \leq x\}$的事件之機率感興趣, 所以特別以函數$F(x)$表此機率值, 而且簡單地寫為$P(X \leq x)$, 而非$P(\{X \leq x\})$。一些常會用到的與X有關的機率值, 都可藉由F表示出來。分佈函數與隨機變數之關係, 可說極為密切。

分佈函數是一定義域為實數集合, 對應域為區間$[0, 1]$的函數。由機率空間, 隨機變數, 至分佈函數, 當只給分佈函數時, 源頭之機率空間為何, 可能毫不清楚。自完全不同的機率空間, 有可能會得到相同的分佈函數。這就像紙雖然是由樹木製造出來, 但當我們振筆疾書時, 往往不會知道這張紙當初是在那一棵樹上。而不同的樹木, 也可能造出品質無分軒輊的紙來。不過有時討論較細膩的問題, 就仍須把機率空間找出來。

如果我們只討論機率, 有時只須知隨機變數之分佈, 這時機率空間並無關緊要。隨機變數的"名字"亦不重要, 稱為X, 或Y等皆可。

對一隨機變數X, 除了伴隨分佈函數F, 亦有一與其關係密切的函數, 即**機率密度函數**(probability density function), 或是**機率質量函數**(probability mass function), 前者是針對**連續型**, 後者是針對**離散型**。不過為了簡便, 常不分連續或離散, 皆以機率密度函數稱之。底下我們給其定義。

設X為一隨機變數, 以F為分佈函數。若X為離散型, 則下述函數f, 便稱為X之機率密度函數:

$$f(x) = P(X = x),$$

其中x屬於某一離散的集合。此集合若以B表之, 則f滿足$\sum_{x \in B} f(x) = 1$。若X為連續型, 則其機率密度函數, 乃定義為滿足下式之非負函數f:

$$F(x) = \int_{-\infty}^{x} f(t)dt,\ x \in R\ 。$$

對一隨機變數X, 我們常想粗略地知道其值究竟多大, **期望值**(expectation, 或稱expected value, mean)$E(X)$, 就是常被拿來扮演這種以一單一的值, 來代表一隨機變數大小的角色。設X之機率密度函數為f, 則

$$E(X) = \begin{cases} \int_{-\infty}^{\infty} xf(x)dx, & \text{連續型}, \\ \sum_x xf(x), & \text{離散型}\ 。 \end{cases}$$

期望值這個名詞, 也許會令人感到有些困擾。投擲一公正的骰子一次, 令X表所得之點數。則X之機率密度函數為

$$f(x) = \frac{1}{6},\ x = 1, 2, \cdots, 6\ 。$$

因此投擲骰子之期望值為3.5。點數為整數, 無論怎麼投擲都得不到3.5, 怎會期望得到3.5呢? 所以暫時不要去想期望值字面上的含意, 那是不易想通的。不過期望值是一隨機變數之"最佳"常數的代表值。所謂最佳, 指的是誤差平方的期望值最小。

果園裡生產的蘋果, 大小頗有差異。有人問你, 蘋果每個重量大約是多少, 如何回答呢? 人們往往想以一常數a, 來代表一隨機變數X。很難說那一個a最好, 誤差$X - a$仍是隨機的。但在誤差平方$(X - a)^2$之期望值$E((X - a)^2)$最小下, 期望值$E(X)$便是最好的選擇。

在機率論裡, 我們引進了隨機變數的概念。例如, 學生不再如小學的練習裡, 每位體重30公斤, 求25位學生的總體重, 而是讓每位學生的體重有一機率分佈。雖然這樣的模式較合理, 但因無法掌握隨機變數之大小, 我們才想要有一代表值。由於具備很多好性質, 期望值遂常被拿來當做隨機變數之代表值。期望值像是隨機變數分佈之一個核心, 隨機變數可能取

的值, 散佈在期望值的左右。其他亦常被拿來當做隨機變數之代表值者, 尚有**中位數**(median)及**眾數**(mode)等。

我們常要藉助統計做各種決策。投資某項事業, 可能成功而獲利若干, 可能失敗而損失若干。那到底要不要投資呢? 不能只看到成功機率很大(說不定獲利很小), 也不能只看到獲利很大(說不定成功機率很低), 算一算淨所得之期望值, 是一簡單的決策依據。判斷出門要不要帶雨傘, 也是先盤算一下期望損失(先估計下雨的機率, 及下雨不帶傘的損失等)。期望值在做決策時, 便常扮演重要的角色。

不論期望值是多好的一個隨機變數之代表值, 難免會偏離該隨機變數, 如何量測偏差究竟有多大? 這就是**變異數**(variance)與**標準差**(standard deviation)的功能。

設有一隨機變數X, 且設期望值$E(X)$存在(期望值並不一定會存在)。則X與$E(X)$之離差為$X - E(X)$。此值有正有負, 但期望值$E(X - E(X)) = E(X) - E(X) = 0$。離差之期望值為0, 亦為期望值之一性質。但我們往往是對離差之絕對值較感興趣。此正如射飛鏢時, 有時偏右有時偏左, 總不能得意地說平均命中紅心。

離差之絕對值$|X - E(X)|$仍為一隨機變數, 如何衡量其大小呢? 取期望值為一辦法。也就是求離差絕對值的"平均"。但有關絕對值函數之求和或積分, 往往有些麻煩。對於$E(|X - E(X)|)$, 須討論何時$X \geq E(X)$, 何時$X < E(X)$, 這些不等式不見得好解。因此通常的作法, 乃考慮離差平方之期望值。我們便將變異數定義為

$$\mathrm{Var}(X) = E((X - E(X))^2)\text{。}$$

由於是平方的期望值, 像是距離的平方。因此若將變異數開平方, 就像使其回到距離, **標準差**便產生了。

如同期望值, 標準差一詞可能亦會令人感到迷惑。取名"標準", 並不表示隨機變數有較大的機率(如大於1/2), 與期望值之差距不超過1個標準差。曾有教科書上出現這樣的誤解。我們給一反例如下:

設有一隨機變數X, 滿足

$$P(X=0)=0.1,\ P(X=2)=P(X=-2)=0.45\text{ 。}$$

即只取$-2, 0, 2$等3個值 。則

$$E(X)=0,$$
$$\mathrm{Var}(X)=E(X^2)=0.1\cdot 0^2+0.45\cdot 2^2+0.45\cdot(-2)^2=3.6\text{ 。}$$

因此X之標準差$=\sqrt{3.6}\doteq 1.8972<2$ 。但有$0.9(=P(X=2)+P(X=-2)=0.45+0.45)$之機率, X取值與期望值0之差距超過標準差 。

與數學不同, 在統計裡允許誤差的存在, 我們並不要求百分之百的精準 。以估計一銅板出現正面之機率p為例 。除非此銅板兩面皆為正面或皆為反面(即$p=0$或1), 否則不論投擲再多次, 皆難以估計出正確的p值 。因此對於估計, 我們只能要求誤差在可接受的範圍 。而這誤差其實也是隨機的, 故我們才考慮“平均式”(即取期望值)的誤差 。大家常說, 雖不中亦不遠矣 。但不遠到底是多遠? 標準差的功能, 就是對誤差之大小, 提供一參考值 。如果說期望值是描述一隨機變數分佈的“集中處”, 標準差便是描述分佈對該集中處之偏離程度 。標準差愈小(或等價地說變異數愈小), 表隨機變數之分佈較集中在期望值附近; 標準差愈大, 表隨機變數之分佈, 分散的較開 。人們常說命運, 命就有點像是期望值, 運則像是標準差 。命也許不太好(期望值不大), 但由於有變異的關係, 使得有時好運會發生 。

假設兩組學生參加測驗, 每組各有3人 。平均成績皆為60分, 但第一組之成績為61, 60, 59; 第二組之成績為90, 60, 30 。第一組學生的程度很接近, 而第二組學生的程度則有很大差異 。可看出光由平均成績, 並無法了解兩組學生程度差異之情況 。對於數據, 光知道平均(或期望值), 常是不夠的 。例如, 射箭時, 不能只說平均命中紅心; 要泡溫泉, 不能只看到平均水溫攝氏42度, 符合標準, 因說不定有時80度太燙, 有時10度太冷, 讓人根本不敢下去 。因此對於隨機變數, 除了期望值, 我們尚須給出標準差 。有了期望值及標準差, 對一隨機變數的中心, 及散佈範圍, 才有一較清晰的概念 。

如同函數可以有多個變數, 我們也可以有**多維**(或說多變數, 多變量)**隨機變數** 。**聯合分佈函數**,及**聯合機率密度函數**, 也就對應產生 。例如, 二

隨機變數X, Y之聯合分佈函數的定義為

$$F(x,y) = P(X \le x, Y \le y),\ x, y \in R\ 。$$

統計分析要先收集資料, 而很多資料有不只一個變數, 多變數的討論, 為統計學中重要的題材。此處只是對隨機變數給一初步的介紹, 因此不多著墨。我們只給幾個概念。

設二隨機變數X, Y, 以$f(x,y)$為聯合機率密度函數, X, Y之機率密度函數則分別為$f_X(x)$及$f_Y(y)$表之。若

$$f(x,y) = f_X(x) f_Y(y),\ x, y \in R,$$

則稱X與Y**相互獨立**(mutually independent), 或只簡單地說**獨立**。

二隨機變數獨立, 表知道其中之一的值, 對另一變數的分佈, 沒有影響。在某些情況, 有可能一隨機變數為另一隨機變數之函數, 兩者可說關係極度密切, 但卻相互獨立。隨機變數的獨立, 有其特殊的意義, 與平常所說的獨立自主, 含意不同。

同理我們也可以定義n個隨機變數之獨立。另外, 二隨機變數若不獨立, 如何來表示兩者之關係, 或說互相變化之情況, 也是統計裡常討論的題材。

設以F, G分別表二隨機變數X及Y之分佈函數。X與Y**分佈相同**, 表對每一實數x, $F(x) = G(x)$。也可說X與Y有共同的分佈。但分佈相同, 與隨機變數相同是完全不一樣的, 即使二截然不同的隨機變數, 都可有相同的分佈函數。舉例如下。投擲一公正的骰子, 並觀測所得點數, 則樣本空間$\Omega_1 = \{1,2,3,4,5,6\}$。對$\omega \in \Omega_1$, 令$X(\omega) = 1$, 若$\omega$為偶數; $X(\omega) = 0$, 若ω為奇數。其次投擲一公正的銅板, 並觀測會得那一面, 則樣本空間$\Omega_2 = \{$正面, 反面$\}$。令$Y($正面$) = 1$, $Y($反面$) = 0$。X, Y為二相異的隨機變數(定義域便不同), 但X, Y分佈相同。

n個隨機變數$X_1, \cdots, X_n$, 若相互獨立, 且有共同的分佈(independent and identically distributed, 簡稱iid), 便稱此為一組**隨機樣本**(random sample)。在統計分析裡常會接觸隨機樣本, 很多統計的理論都基於隨機樣本。

4 伯努力試驗

在日常生活裡, 有很多我們接觸的隨機現象, 往往恰有兩種可能的結果。譬如說生男或生女, 考試及格或不及格, 投擲一骰子是否得到點數3, 選民是否會投給某特定候選人等。上述這一切, 都可簡化為成功及失敗兩種結果。即在二結果中, 指定一為成功, 另一結果則稱為失敗。只有兩種可能的結果, 可說是會令人感興趣之最簡單的隨機現象了。比此更簡單的, 便是退化的情況, 只有一種結果。

一試驗若只有兩種可能的結果, 便稱為**伯努力試驗**(Bernoulli trial)。重複且獨立地觀測此試驗, 即得一數列之伯努力試驗。

對一伯努力試驗, 令樣本空間$\Omega=\{$成功,失敗$\}$, 又設$P(\{$成功$\})=p$, $P(\{$失敗$\})=1-p$, 其中$0\le p\le 1$。在Ω上定義隨機變數X, 令$X($成功$)=1$, $X($失敗$)=0$。此為一僅取0, 1二值之離散型的隨機變數, $P(X=1)=p$, $P(X=0)=1-p$。我們稱X有參數p之**伯努力分佈**(Bernoulli distribution), 以$\mathcal{B}er(p)$表之。

對一數列n個獨立的伯努力試驗, 我們常對總成功數有興趣, 而不關心究竟那幾次成功。假設每次成功機率皆為p, 又以S_n表總成功數。則

$$S_n=X_1+X_2+\cdots+X_n,$$

其中如前, 若第i次成功, 則$X_i=1$, 否則$X_i=0$, $i=1,\cdots,n$。$X_1,\cdots,X_n$即為iid的隨機變數, 以$\mathcal{B}er(p)$為共同分佈。

利用排列組合的技巧, 可得

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

其中

$$\binom{n}{i}=\frac{n!}{i!(n-i)!}, \tag{2}$$

而$i!=i(i-1)\cdots 3\cdot 2\cdot 1$, $i\ge 1$, 且令$0!=1$。$\binom{n}{i}$便是所謂**二項式係數**(binomial coefficient)。S_n稱為有參數n及p之**二項分佈**(binomial distribution), 以$\mathcal{B}(n,p)$表之, 其中n為一正整數, $0\le p\le 1$。當然$\mathcal{B}er(p)$即

為$\mathcal{B}(1,p)$。$\mathcal{B}(n,p)$亦為離散型的分佈。由**二項式定理**(binomial theorem)可驗證

$$\sum_{i=0}^{n} P(S_n = i) = \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} = (p+1-p)^n = 1 \text{ 。}$$

故(1)式的確定義出一離散型的分佈。又可求出期望值$E(S_n) = np$, 及變異數$\mathrm{Var}(S_n) = np(1-p)$。

要注意的是, n, p須固定, 且n次伯努力試驗須為iid, 否則其總成功數, 便不見得是二項分佈了。底下給一例。

例1.假設一袋中有100個球, 其中有40個白球, 60個其他顏色的球。依序隨機地取(即每次取球時, 袋中每一球皆有相同的機率被取中)二球, 每次取出後不放回, 令X表共取中的白球數。底下來看X並無二項分佈。

令$Y_i = 1$, 若第i次取中白球, 否則$Y_i = 0$, $i = 1, 2$。顯然Y_1有$\mathcal{B}er(0.4)$分佈。其次若$Y_1 = 0$, 即第1次未取中白球, 則第2次有40/99之機率取中白球; 若$Y_1 = 1$, 即第1次取中白球, 則第2次有39/99之機率取中白球。因$Y_2 = 1$之機率與Y_1之值有關, 故Y_1與Y_2不獨立。由上討論並得

$$P(Y_2 = 1) = \frac{60}{100} \cdot \frac{40}{99} + \frac{40}{100} \cdot \frac{39}{99} = 0.4 \text{ 。}$$

即Y_2亦有$\mathcal{B}er(0.4)$分佈。因

$$P(X = 0) = \frac{60}{100} \cdot \frac{59}{99} \neq 0.36 = \binom{2}{0} 0.4^0 \cdot 0.6^2,$$

故(1)式不成立, 因此X並無二項分佈。

附帶一提, 若重複前述試驗, 即每次隨機地取一球, 取出後皆不放回, 且類似地定義出隨機變數, $Y_1, \cdots, Y_{100}$。則可證明雖$Y_1, \cdots, Y_{100}$並不獨立, 但分佈皆為$\mathcal{B}er(0.4)$。此可說明在摸彩活動中, 先後抽獎順序, 並不影響中獎機率。當然若情況改為每次取出後放回, 則對每一$n \geq 1$, $Y_1, \cdots, Y_n$(此時n可大於100)便為iid, 以$\mathcal{B}er(0.4)$為共同分佈。

上例為一伯努力試驗中, 各次試驗不獨立之例。生活裡屢見這種取出後不放回的情境。底下來看, 總成功數若不具二項分佈, 會有什麼分佈?

考慮一般的情況。假設袋中有N個球, 其中有D個白球, $N-D$個非白球, 自袋中依序隨機地取出n個球, 每次取出後不放回, 令X表總共取得之白球數。則

$$(3)\quad P(X=k)=\frac{\binom{D}{k}\binom{N-D}{n-k}}{\binom{N}{n}},\ \max\{0,n-N+D\}\le k\le \min\{n,D\},$$

其中min, max分別表較小及較大者。我們說明k之範圍的由來。因只取n個球, 且白球數共只有D個, 所以k不能超過n及D, 即$k\le\min\{n,D\}$。又取中之非白球的個數$n-k$, 當然不可超過全部之非白球數$N-D$, 而$k\ge 0$又要成立, 因此$k\ge\max\{0,n-N+D\}$。

(3)式便定義出一常見的**超幾何分佈**(hypergeometric distribution), 以$\mathcal{H}(N,D,n)$表之。此分佈在有關**品質管制**(quality control)的探討裡常出現。對X有$\mathcal{H}(N,D,n)$分佈, 可證明(見黃文璋(2003b)pp.251-252)

$$E(X)=\frac{nD}{N},\ \mathrm{Var}(X)=\frac{nD(N-D)(N-n)}{N^2(N-1)}\ 。$$

我們給一超幾何分佈之例如下。

例2.設有$N=200$個零件, 其中有$D=4$個不良品(不良率為2%)。隨機地取3個樣本, 每次取出後不放回, 則樣本中之不良品數X有$\mathcal{H}(200,4,3)$分佈。因此

$$\begin{aligned}P(X=0)&=\frac{\binom{4}{0}\binom{200-4}{3-0}}{\binom{200}{3}}=\frac{1{,}235{,}780}{1{,}313{,}400}\doteq 0.9409,\\ P(X=1)&=\frac{\binom{4}{1}\binom{200-4}{3-1}}{\binom{200}{3}}=\frac{4\cdot 19{,}110}{1{,}313{,}400}\doteq 0.0582\ 。\end{aligned}$$

同理可求出

$$\begin{aligned}P(X=2)&\doteq 0.000895,\\ P(X=3)&\doteq 3.046\cdot 10^{-6},\end{aligned}$$

此二機率值可說非常地小。

例2給出一種**小樣本事件**。通常我們做決策要仰賴大樣本, 小樣本怎會值得重視? 大家都聽過“曾參殺人”的典故吧。在戰國策的秦策二: 昔者曾子處費, 費人有與曾子同名族者, 而殺人, 人告曾子母曰“曾參殺人”。曾子之母曰“吾子不殺人”, 織自若。有頃焉, 人又曰“曾參殺人”, 其母尚織自若也。頃之, 一人又告之曰“曾參殺人”。其母懼, 投杼踰牆而走。夫以曾參之賢, 與母之信也, 而三人疑之, 則慈母不能信也。

曾子是孔子的弟子, 戰國策記載, 以他的賢能及其母對他的信任, 但接連三個人告訴其母他殺人, 其母對他的信心便動搖了。

如果某公司宣稱其生產之某零件不良率僅2%。有一回你買了一盒有200個, 並取出三個使用。若其中壞了一個尚可忍受, 因機率約為0.0582。若壞了兩個, 可能便要找公司退貨了, 因機率才約為0.000895。若三個皆壞, 大概便不相信不良率才2%而已, 因機率僅約百萬分之3。有些事件我們原先的認知是不太會發生, 偶而碰到一次只是覺得運氣不好。碰到第二次時, 心裡便可能覺得怪怪的。若再多碰到一、二次, 便很可能覺得要嘛有人搞鬼, 要嘛這件事發生的可能性其實不是那麼低。小樣本就是在這類發生機率很低之情況中, 顯現其影響力, 要知有時是可以偏概全的。

超幾何分佈在做民調時亦常會出現。例如, 想知道某特定候選人A所獲得之支持率。設有N個選民, 其中有D個支持A, $N-D$個不支持, 隨機地取n個選民回答問卷。令X表總共支持A之人數。若每次取出後放回, 則X有二項分佈。只是總不能重複問同一個人, 因此通常是取出後不放回的情況, 則X便有超幾何分佈。

直觀上, 若取出的樣本數n, 與總人數N相比很小, 則雖取出後不放回, X之分佈應近似於二項分佈。在丁村成(2010)一文, 遂“證明”當$N \to \infty$時,

$$\frac{\binom{D}{k}\binom{N-D}{n-k}}{\binom{N}{n}} \to \binom{n}{k}\left(\frac{D}{N}\right)^k\left(1-\frac{D}{N}\right)^{n-k}, \tag{4}$$

並說此為“超幾何分配之極限是二項分配的事實”。(4)式右側為$\mathcal{B}(n, D/N)$分佈之機率密度函數, 此二項分佈即為該文作者想表達的超幾何分佈之極

限。可惜只要具正確極限概念的讀者, 應已看出(4)式有誤。

既然讓N趨近∞, 極限怎能還有N? 此犯了極限理論裡, 可說是最基本的錯誤。在丁村成(2010)一文, 一開始便舉了幾個取樣的例子, 以說明超幾何分佈如何產生。既然源自於取樣, N自然是有限, 不會趨近∞。就算不理會取樣, 只純就數學討論, $N \to \infty$時, D若維持有限, 則$D/N \to 0$, 且$1 - D/N \to 1$。如此除非$k = 0$, 否則(4)式右側將趨近0。這是對的, D若固定, 當$N \to \infty$時, 所取的n個樣本裡, 會至少有1個出自那D個的機率, 將趨近0。也就是會有0個出自那D個的機率趨近1。

如何修正(4)式? 仍純就數學而言, 設$N \to \infty$時, D也隨之增加, 且$D/N \to p$。則可證明

$$\frac{\binom{D}{k}\binom{N-D}{n-k}}{\binom{N}{n}} \to \binom{n}{k} p^k (1-p)^{n-k},\ k = 0, 1, \cdots, n\ 。 \tag{5}$$

就實務上來說, 不會有$N \to \infty$之情況。但如果n/N很小, 即取的樣本數, 與母體數相比很小(做民調或檢驗時, 此條件常會滿足), 則(4)式左、右兩側二機率值之差異便不大。這是以二項分佈做為超幾何分佈之近似的含意, 而非如丁村成(2010)所述之極限結果。機率的概念本已不易掌握, 若又涉及極限, 更不可輕率以對。

超幾何分佈源自於抽樣, 且是一種**簡單隨機抽樣**(simple random sampling)。曾有教科書說:

> 當母群體中每一元素被抽中的機會均等時, 從母群體中隨機抽取所需樣本的方法, 稱為簡單隨機抽樣法。

底下為另一說法:

> 設母體個數為N, 從中抽取n個樣本, 則每一個體被抽到的機率都是n/N, 此種抽樣法稱為簡單隨機抽樣法。

此二講法乃類似, 只是皆不正確。

正確的定義是, 假設欲自有N個元素的母體中, 抽取n個樣本, 若母體那$\binom{N}{n}$個有n個元素的子集合, 皆有相同的機率被取中, 則此法稱為**簡單隨**

機抽樣法, 得到的樣本即為**簡單隨機樣本**。此定義導致母體中之每一元素, 被取中的機率皆相同, 即n/N。反之, 即使滿足前述二課本中的條件, 也不一定滿足簡單隨機抽樣法的定義。我們給一反例如下。

假設母體中有a, b, c, d四個元素, 要抽取兩個樣本。即$N = 4$, $n = 2$。現各以$1/2$的機率取$\{a, b\}$, 及$\{c, d\}$。顯然母體中的四個元素, 每一被取中的機率皆為$2/4 = 1/2$, 所以這種取法, 符合前述二課本上的要求。但$\{a, c\}$, $\{a, d\}$, $\{b, c\}$, $\{b, d\}$等, 皆不會被取中, 故此非簡單隨機抽樣。可看出此抽樣法是偏頗的, 有四個組合必不會產生。執行抽樣時, 當然不可有這種偏差, 否則將無法達到一葉知秋的目的。

讀者易見, 我們所給的定義是較強的, 必須母體全部有n個元素之子集合, 每一皆有相同的機率$1/\binom{N}{n}$被取中。當此條件成立時, 極易推導出, 母體中的任一元素, 甚至任意有k個元素的子集合, $k = 1, 2, \cdots, n$, 也皆有相同的機率被取中。又若慢慢取, 每次只隨機取一樣本, 取出後不放回, 持續取n個樣本, 則所得即為一組簡單隨機樣本。

通常簡單隨機抽樣法是指取出後不放回。若是取出後放回, 反而得特別強調。

二項分佈當n較大時, 計算上便有些困難。例如, 若獨立地投擲一銅板100次, 每次成功的機率設為$1/2$, 則總共之正面數X, 便有$\mathcal{B}(100, 1/2)$分佈。由此得$X = 30$之機率為

$$P(X = 30) = \binom{100}{30}\left(\frac{1}{2}\right)^{100},$$

$\binom{100}{30}$自然是不太好計算的。為解決此困難, 遂發展出中央極限定理, 我們留至第6節再談。

另外, 由伯努力試驗, 在不同的情境下, 可得好幾個機率裡重要的分佈, 可參考黃文璋(2003a)一書“銅板乾坤”, 此處不多討論。

5 大數法則

只要懂點機率統計的人, 都聽過**大數法則**(law of large numbers)。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。弄懂極限, 乃進入較高深數學的第一步, 這得看一般微積分的書。底下我們只略做介紹。

任給一實數列$\{a_n,\ n \geq 1\}$, 當$n \to \infty$時, a_n不見得會趨近至某一定值。例如, 若$b_n = (-1)^n$, $n \geq 1$, 則此數列為-1, 1, -1, 1, $\cdots$, 不論n多大, 數列都是-1, 1交錯著進行, 不會接近任一定值; 若$c_n = n$, $n \geq 1$, 則當n愈來愈大, c_n將無止盡地增大。此二數列皆稱為**發散**。但若取$a_n = (1/2)^n$, $n \geq 1$, 則隨著n之增大, a_n將任意接近0。我們以

$$n \to \infty\text{時},\ a_n \to 0$$

表之, 也可寫成

$$\lim_{n\to\infty} a_n = 0\ 。$$

對於$n \to \infty$時, $a_n \to a$, 其中a為一實數, 是說當n不斷地增大, a_n可任意接近a。此表a_n與a之差距$|a_n - a|$可任意小。而怎樣是任意小? 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$|a_n - a|$須小於被指定的ε。但何時$|a_n - a| < \varepsilon$? 可沒要求對每一$n \geq 1$, $|a_n - a| < \varepsilon$都要成立, 而是說n很大時, $|a_n - a| < \varepsilon$才須成立。但n怎樣才算很大? 如果從某一項n_0開始, $|a_n - a| < \varepsilon$皆成立, 大約便只好服氣了。於是產生了下述極限的定義: 假設有實數列$\{a_n,\ n \geq 1\}$, 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$\lim\limits_{n\to\infty} a_n = a$表之。

上述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。

講大數法則前, 可能得先正名。**巨數法則**, 雖然英文名為law of truly large numbers, 但其實與大數法則並不太相干。在一般正規的機率論書

中, 並不會提到此法則。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)。

雖明明與大數法則大異其趣, 但由於都涉及大樣本, 有些人遂將二者混在一起。如Shermer(2004, 姚若潔譯(2004)為其翻譯)一文中有底下一段:

> 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.

不但提到"Law of Large Numbers", 且說發生機率百萬分之一的怪異事件, 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8月底, 美國人口超過3.12億。

Shermer為著名通俗科學雜誌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, 恐易引起誤解。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"(見姚若潔譯(2004)), 也該修正為"發生機率很小的事件, 若樣本數較少時, 會有這種事件發生的機率不高; 若樣本數較大時, 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, 並不會隨試驗數之多寡而改變。另外, 最後一句話, 宜改為"在一個人身上每天發生機率為百萬分之一的怪異事件, 在美國平均一天可發生295次"才較恰當。

一事件發生的機率p雖然很小, 若重複觀測n次, 且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, 將愈來愈接近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, 任一聳人聽聞的事件, 其發生便都不必驚訝。

巨數法則可用來解釋何以生活上處處有巧合。關於巧合事件之無所不在, 可參考黃文璋(2003a)"純屬巧合"一章。

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)。大數法則是說, 若一試驗(或觀測), 能重複且獨立地進行, 則觀測值之平均, 將任意接近期望的結果。比較正式一點地說, 就是隨機產生樣本之平均, 當樣本數很大, 將有很大的機率, 接近母體之期望值。

機率論早期的發展, 常對某事件是否發生有興趣。如上節所說, 這種發生或不發生, 兩個結果的觀測, 便產生伯努力試驗。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∞, 事件發生之相對頻率接近p之機率, 將趨近1。

伯努力所指的"機率趨近1"是什麼意思?

令$n(A)$表獨立且重複地觀測n次, 事件A所發生之次數。則$n(A)$有$\mathcal{B}(n, p)$分佈。即

$$P(n(A)=k)=\binom{n}{k}p^k(1-p)^{n-k},\ k=0,1,\cdots,n\text{。} \tag{6}$$

伯努力指出, 當n很大時, 事件A發生之相對頻率$n(A)/n$與p之差距, 應極可能很小。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$n(A)=0$。投擲一公正銅板100次, 100次全出現正面的機率當然很低, 僅$1/2^{100}$, 微乎其微。但若一直重複執行此試驗(每回投擲100個銅板), 譬如說做了2^{100}回, 則如果其中出現一回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$2^{110}/2^{100} = 2^{10} = 1,024$(回)。此即巨數法則。

註1.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。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。即要四百多萬年才能模擬完。野史裡偶有投擲出100個正面的記載, 見黃文璋(2003a)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, 在不作假之下, 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。伯努力所說n很大時, $n(A)/n$與p之差應極可能很小, 表只要n夠大, $|n(A)/n - p| \le \varepsilon$之機率應很大, 其中$\varepsilon$為任一正數。而前述機率即

$$\sum_{|k/n-p|\le\varepsilon} P(n(A) = k) = \sum_{|k/n-p|\le\varepsilon} \binom{n}{k} p^k (1-p)^{n-k} \text{。} \tag{7}$$

伯努力辛苦地證明$n \to \infty$時, 上式右側和趨近1。事實上, 後人利用**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(2010a)p.124。但伯努力的原創性, 至今仍被推崇。

伯努力所指的事件發生之相對頻率, 可表示為

$$\frac{n(A)}{n} = \frac{X_1 + \cdots + X_n}{n} = \frac{S_n}{n} = \overline{X}_n,\ n \ge 1, \tag{8}$$

其中$X_1, \cdots, X_n$為iid之隨機變數, 以$\mathcal{B}er(p)$為共同分佈, 而$S_n = X_1 + \cdots + X_n$, $n \ge 1$, 有$\mathcal{B}(n,p)$分佈。伯努力即指出, iid以$\mathcal{B}er(p)$為共同分佈的隨機變數X_i, $1 \le i \le n$, 當$n \to \infty$時, **樣本平均**(sample mean)$\overline{X}_n$, 會很接近p的機率趨近1, 即會稍偏離p的機率趨近0。

讀者應可看出大數法則與巨數法則, 乃不同的兩個法則。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, 大數法則卻是說事件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, 一特定事件將"大約"發生多少次。因n很大時, 事件A發生"約"np次, 故即使p很小, 只要$np > 1$, 則看到事件A發生, 就不足為奇。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。伯努力之後的機率學家們, 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$-\infty < E(X_1) < \infty$), 則當$n \to \infty$時, $\overline{X}_n$會**機率收斂**(converges in probability)至$E(X_1)$, 以$\overline{X}_n \xrightarrow[n \to \infty]{P} E(X_1)$表之。在此機率收斂的定義如下:

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

$$\lim_{n \to \infty} P(|Y_n - Y| > \varepsilon) = 0, \forall \varepsilon > 0, \tag{9}$$

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。

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

辛欽所證出的結果, 後來被稱為**弱大數法則**。你可預期亦有**強大數法則**, 此處略過不表, 可參考黃文璋(2010a)5.7節。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, 樣本平均與隨機變數之期望值$E(X_1)$"必會"很接近, 這點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, $\overline{X}_n$與期望值$E(X_1)$很接近的機率很大。即$\overline{X}_n$不論與$E(X_1)$差距多小之機率要多大, 皆能辦到, 只要樣本數n夠大。

弱大數法則並沒有說, n很大時, $\overline{X}_n$會等於$E(X_1)$。有些初學者誤以為如此。教科書中也有下述離譜的講法:

機率裡的期望值就是統計試驗中大量數據的平均值。

即使說n很大時, $|\overline{X}_n - E(X_1)|$必很小, 此講法也是錯的。這些皆屬於缺乏隨機的概念。正確的說法是, n很大時, $|\overline{X}_n - E(X_1)|$很小的機率很大。

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, 你覺得那一種講法較精準?

弱大數法則不只針對iid的隨機變數才成立, 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。

對iid的隨機變數, 期望值若不存在, 則弱大數法則就不適用了。設X_n, $n \geq 1$, 為iid的隨機變數。若以$\mathcal{B}er(1/2)$為共同分佈, 此時圖1給出$\overline{X}_n$之一模擬圖形, $1 \leq n \leq 1,000$。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。其次若共同分佈為參數2.5及1之**柯西分佈**(Cauchy distribution), 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2給出此時$\overline{X}_n$之模擬圖形, $1 \leq n \leq 1,000$。$\mathcal{C}(2.5, 1)$分佈之期望值不存在, 圖形顯示$\overline{X}_n$之振盪頗大。

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

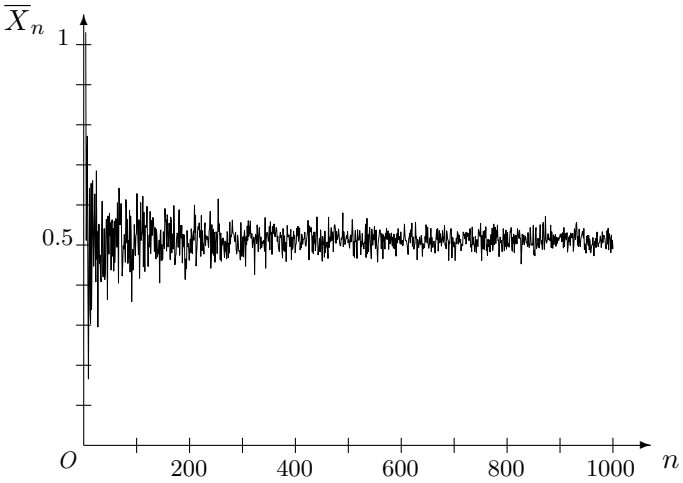


圖1. $\mathcal{B}er(1/2)$分佈$\overline{X}_n$, $1 \le n \le 1,000$, 之模擬圖

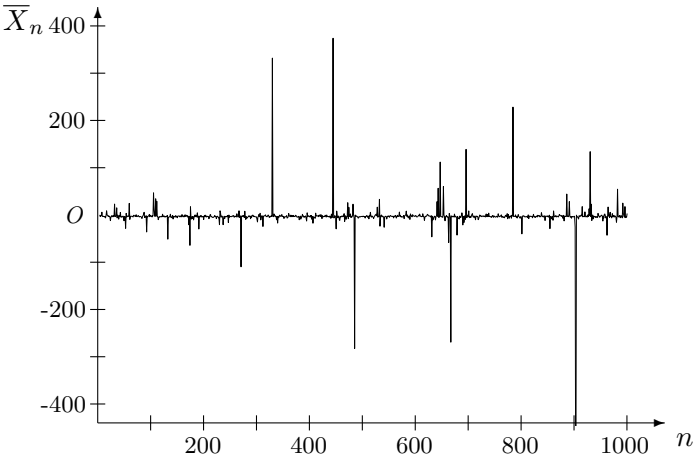


圖2. $\mathcal{C}(2.5,1)$分佈$\overline{X}_n$, $1 \le n \le 1,000$, 之模擬圖

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。正確的說法分別是: 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$\overline{X}_n$。雖然$X_1 = 1$, 但自X_2起, 應沒有人覺得會是1了。隨著n之增大, 你愈來愈相信$\overline{X}_n$將很接近0。何以我們那麼“確定”? 因弱大數法則告訴我們, $n \to \infty$時, $\overline{X}_n \xrightarrow[n \to \infty]{P} E(X_1) = 0$。

由於有上述這類困擾, 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)。

在統計裡, 很多不錯的統計估計方法, 都基於弱大數法則。著名的**動差估計法**(method of moment estimator)即為一例。大數法則除支持頻率對機率的解釋, 在統計學裡用途亦不小。

6 中央極限定理

如前所述, 生活裡伯努力試驗可說處處可見。而若有n個獨立的伯努力試驗, 每次成功機率皆為p, 則總成功數S_n便有$\mathcal{B}(n, p)$分佈。二項分佈也就是如此容易產生。在機率論發展的早期, 由於沒有快捷的計算工具, 因此尋找易於處裡的方法, 來計算二項分佈的機率值, 便很迫切。

伯努力的大數法則指出, n很大時, S_n/n差不多就是p。即只要n夠大, S_n/n應有很大的機率, 只在p附近變動。有些人因此以為, 當n很大時, S_n便應很接近期望值np。此想法導致獨立地投擲一公正銅板($p = 1/2$)多次後, 會以為正、反兩面數將差不多。甚至加以引申, 以為樂透彩若公正, 則開獎多期後, 各號碼累積出現的次數, 應大致相同。這並不正確。事實上, 開獎期數愈多, 各號碼累積出現次數之差異, 反該預期會愈大。可參考黃文璋(2005)圖2。不妨以一常數數列來說明。取$a_n = n + \sqrt{n}$,

$n \geq 1$。雖n很大時, $a_n/n = 1 + 1/\sqrt{n}$很接近1, 但a_n/n與1各放大n倍後, a_n與n之差$\sqrt{n}$, 卻隨著n之增大而無止盡地增大。由此可理解, n很大時, 雖S_n/n大致在p附近, 但S_n卻很可能會愈偏離np。

直觀上, 對一固定的p, 不論p多小, 有如聚沙成塔, 只要n夠大, 則n次試驗後的總成功數S_n很可能會很大。但S_n究竟大到什麼地步? 也就是說, 能否了解S_n隨n成長的方式? 如在天文學裡, 以光年來量度很遠的距離。有沒有什麼辦法, 將S_n轉換成不是那麼大, 使求其機率值, 變得較有意義?

對S_n有$\mathcal{B}(n,p)$分佈, 則

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

由於$\binom{n}{k}$中有階乘的關係, 使得當n較大時, $P(S_n = k)$有多大, 除非k很靠近0或n, 否則便不易掌握。我們來看比較簡單$p = 1/2$, 且n為偶數的情況。利用可近似階乘的**史德林公式**(Stirling formula, 可參考黃文璋(2010a)p.128), 得$n \to \infty$時, S_n等於期望值$n/2$之機率

$$P(S_n = \frac{n}{2}) = \binom{n}{n/2} \left(\frac{1}{2}\right)^n \sim \frac{\sqrt{2}}{\sqrt{\pi n}}, \tag{10}$$

其中符號"$\sim$"表$n \to \infty$時, 兩側比值趨近1。因n很大時, $\sqrt{2}/\sqrt{\pi n}$很小, 因此(10)式左側也必很小。印證我們之前所說的, 投擲一公正銅板n次(n為偶數), 若n很大時, 乃極不容易出現$n/2$(半數)次正面。另一方面, 對$\mathcal{B}(n, 1/2)$分佈, 其中n為偶數, 可證明其機率密度函數之極大值, 發生在$S_n = n/2$處。圖3給出$\mathcal{B}(20, 1/2)$之(機率)**直方圖**(histogram)。所謂直方圖乃描述離散型隨機變數取各可能值之機率, 其中每一長方條之底部中點, 為所取的值, 面積則為會取該值之機率。所以長方條之面積和為1。圖3顯示, 此直方圖在$n = 10$時最高, 並往兩側下降。

對二固定的整數r, s, 其中$0 \leq r \leq s$, 可證明當n很大時,

$$P(r \leq S_n \leq s) = \sum_{k=r}^{s} \binom{n}{k} \left(\frac{1}{2}\right)^n,$$

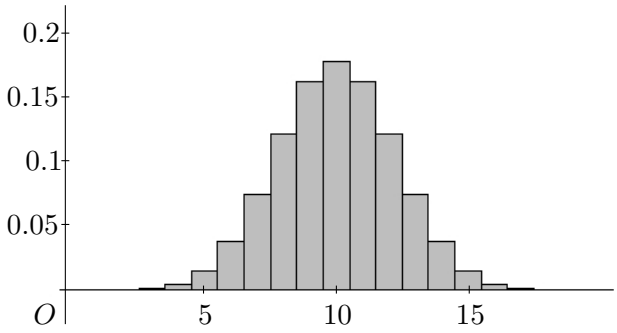


圖3. $\mathcal{B}(20, 0.5)$分佈直方圖

也將很小。事實上, 仍考慮n為偶數, 利用史德林公式, 對每一固定整數k, 可得n很大時,

$$P(S_n = \frac{n}{2} + k) \sim \frac{\sqrt{2}}{\sqrt{\pi n}} \text{ 。} \tag{11}$$

由上討論知, 除非在$n/2$附近有"夠多"的項, 否則S_n落在$n/2$附近一區間$[n/2 + a, n/2 + b]$之機率將很小, 其中整數a, b滿足$-n/2 \leq a \leq b \leq n/2$。即$n$很大時,

$$P(\frac{n}{2} + a \leq S_n \leq \frac{n}{2} + b) = \sum_{k=a}^{b} P(S_n = \frac{n}{2} + k) \doteq 0 \text{ 。} \tag{12}$$

而且所謂夠多的項, 由(11)式看出, 隨著n之增大, 所需項數, 應須能與$\sqrt{n}$相匹比。

明乎此, **隸美弗**(Abraham de Moivre, 1667-1754)以適當的a_n, b_n取代(12)式中的a, b。仍利用史德林公式, 經過一番取極限的過程, 他得到那些機率值的和, 可表示為一個積分, 並在1733年發表此結果。這個劃時代的發現, 也就是後來**中央極限定理**的雛型, 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, **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裡, 推廣棣美弗$p = 1/2$的結果至一般的p。但如同棣美弗, 拉普拉斯的曠世巨著, 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, **里阿普那夫**(Aleksandr M. Lyapunov, 1857-1918), 給出此定理較一般的敘述, 及嚴密的證明。由於有前述這一段歷史, 中央極限定理又稱**棣美弗-拉普拉斯定理**。有關中央極限定理的歷史, 可參考Tijms(2004)p.169。

對S_n有$\mathcal{B}(n,p)$分佈, 拉普拉斯即證出對實數$\alpha \le \beta$, $n \to \infty$時,

$$
\begin{aligned}
(13) \qquad & P(np + \alpha\sqrt{np(1-p)} \le S_n \le np + \beta\sqrt{np(1-p)}) \\
& \to \Phi(\beta) - \Phi(\alpha) = \int_\alpha^\beta \phi(u)du,
\end{aligned}
$$

其中

$$
(14) \qquad \Phi(x) = \int_{-\infty}^{x} \phi(u)du, \ x \in R,
$$

$\Phi(\infty) = \int_{-\infty}^{\infty} \phi(x)dx = 1$, 而

$$
(15) \qquad \phi(x) = \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}, \ x \in R \text{。}
$$

(15)式定義出一**標準常態**(以$\mathcal{N}(0,1)$表之)**分佈**之機率密度函數。對$x \in R$, $y = \phi(x)$之圖形對稱於y軸, 是所謂**鐘形曲線**(bell-shaped curve), 圖形下與x軸間之面積為1。圖4給出其圖形, 注意橫軸及縱軸之單位長不同。圖形最高點發生在$x = 0$, 其值為$\phi(0) = 1/\sqrt{2\pi}$, 約為0.39894。故若兩軸單位長取的相同, 則圖形將更呈現平坦, 便不易看似鐘形了。

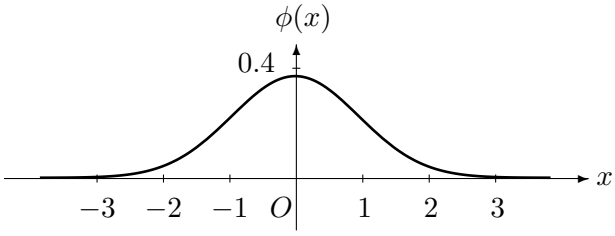


圖4. $\mathcal{N}(0,1)$分佈之機率密度函數圖形

(13)式之另一表示法為, 對$\alpha \le \beta$, $n \to \infty$時,

$$
(16) \qquad P(\alpha \le \frac{S_n - np}{\sqrt{np(1-p)}} \le \beta) \to \Phi(\beta) - \Phi(\alpha) \text{。}
$$

上式給出了所謂**二項分佈趨近至常態分佈**。底下我們略為說明。

首先, 對S_n有$\mathcal{B}(n,p)$分佈, np, 及$\sqrt{np(1-p)}$分別為S_n之期望值及標準差。我們說過, n很大時, S_n也幾乎是很大。S_n的"核心"是np, 隨著n之增大, 此核心已移至很遠處。將S_n減去期望值np, 則$S_n - np$的"核心"便成為0了。此步驟有如經過座標平移, 將很遠的點拉回來。再經改變尺度, 單位長由1成為$\sqrt{np(1-p)}$, 則S_n便不再是遙不可及的龐然大物。而令人感到驚訝的是, (13)式指出, S_n的直方圖, 從np量起, α個標準差至β個標準差間, 那些長方條面積和, 當$n \to \infty$時, 可表為$y = \phi(x)$之圖形下與x軸間, 介於$x = \alpha$至$x = \beta$間的面積。

如(16)式左側, 對一隨機變數減去期望值, 再除以標準差, 此過程便稱為將其**標準化**。

隨機變數若有二項分佈, 則可表為iid伯努力隨機變數之和。將其標準化後, 會趨近至$\mathcal{N}(0,1)$分佈。只限伯努力隨機變數之和嗎? 其他分佈呢?

中央極限定理, 差不多可以說是機率論裡, 最重要的一個定理。此定理是說, 若有夠多的獨立且同分佈隨機變數之和, 則經標準化後, 當n很大時, 將可以$\mathcal{N}(0,1)$分佈做為近似。設對每一$n \geq 1$, $X_1, \cdots, X_n$為iid的隨機變數, 且設$\mu = E(X_1)$, 及$\sigma^2 = \mathrm{Var}(X_1)$皆存在。中央極限定理指出, 對任二實數$\alpha < \beta$,

$$\lim_{n\to\infty} P\left(\alpha \leq \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \leq \beta\right) = \Phi(\beta) - \Phi(\alpha)\ 。 \tag{17}$$

令$\overline{X}_n = (X_1 + \cdots + X_n)/n$, $n \geq 1$。則(17)式導致

$$\lim_{n\to\infty} P\left(\alpha \leq \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \leq \beta\right) = \Phi(\beta) - \Phi(\alpha),\ \alpha < \beta\ 。 \tag{18}$$

這是以樣本平均來表示的中央極限定理。又$\Phi(x)$滿足

$$\Phi(-x) = 1 - \Phi(x),\ x \in R\ 。 \tag{19}$$

利用(17)及(19)式, 得(取$\beta = c$, $\alpha = -c$)n很大時,

$$P\left(\left|\frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}\right| \leq c\right) \doteq 2\Phi(c) - 1,\ c \geq 0, \tag{20}$$

由(20)式即得

$$P\left(|\overline{X}_n-\mu|\le\frac{c\,\sigma}{\sqrt{n}}\right)\doteq 2\Phi(c)-1,\ c\ge 0\ 。\tag{21}$$

上式又導致

$$P\left(|\overline{X}_n-\mu|>\frac{c\,\sigma}{\sqrt{n}}\right)\doteq 2(1-\Phi(c)),\ c\ge 0\ 。\tag{22}$$

令$c\,\sigma/\sqrt{n}=\varepsilon$, 由上式得

$$P(|\overline{X}_n-\mu|>\varepsilon)\doteq 2(1-\Phi(\varepsilon\sqrt{n}/\sigma)),\ \varepsilon\ge 0\ 。\tag{23}$$

弱大數法則指出, 對每一$\varepsilon>0$, n很大時, $P(|\overline{X}_n-\mu|>\varepsilon)$將很小。但究竟多小, 便沒說了。中央極限定理便給出n很大時, $P(|\overline{X}_n-\mu|>\varepsilon)$約為$2(1-\Phi(\varepsilon\sqrt{n}/\sigma))$。且不論$\{X_n,n\ge 1\}$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。因此雖然無法保證n很大時, $|\overline{X}_n-\mu|$不會超過ε, 但只要變異數σ^2存在, 則此機率$P(|\overline{X}_n-\mu|>\varepsilon)$的大小, 便可掌握, 即約$2(1-\Phi(\varepsilon\sqrt{n}/\sigma))$。此值隨著$n$之趨近$\infty$而趨近0(因$n\to\infty$時, $\Phi(\varepsilon\sqrt{n}/\sigma)\to 1$)。另外, 若(23)式成立, 即得$n\to\infty$時, $P(|\overline{X}_n-\mu|>\varepsilon)\to 0$, 弱大數法則便成立了。因此乍看之下, 中央極限定理是一比弱大數法則還強的定理。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, 可參考黃文璋(2010a)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, 皆可以常態分佈來描述。諸如人的智商、身高, 及膽固醇含量等, 這些隨機的量, 常可視為很多微小的隨機效應累積所造成, 因此往往能用常態分佈來描述。19世紀時, **奎特雷**(Adolphe Quetelet, 1796-1874)及**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。這是此分佈被稱為**常態**的一個主因, 也隱含其他分佈不算常態。常態分佈就是如此被推崇。

常態分佈又稱**高斯分佈**(Gaussian distribution)。**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。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**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**波里亞**(George Pólya, 1887-1985)所取的。時至今日, 中央極限定理在機率論裡的地位, 仍是屹立不搖。

由於當$x \to \infty$, $\Phi(x)$趨近1的速度很快, 因此$1-\Phi(x)$及$2(1-\Phi(x))$趨近0的速度亦很快。例如, $1-\Phi(4)$就已約是$3.17 \cdot 10^{-5}$了。表1給出一些與$\Phi(x)$有關之近似值。

表1. 一些關於$\Phi(x)$之值

x	$\Phi(x)$	$1-\Phi(x)$	$2\Phi(x)-1$	$2(1-\Phi(x))$
1	0.8413447	0.1586553	0.6826895	0.3173105
2	0.9772499	0.0227501	0.9544997	0.0455003
3	0.9986501	0.0013499	0.9973002	0.0026998
4	0.9999683	0.0000317	0.9999367	0.0000633
5	0.9999997	0.0000003	0.9999994	0.0000006

圖5給出本文一開始所說德國10馬克上, 所放的一般常態分佈$\mathcal{N}(\mu, \sigma^2)$之機率密度函數及其圖形, 其中參數$\mu$為實數, $\sigma > 0$。

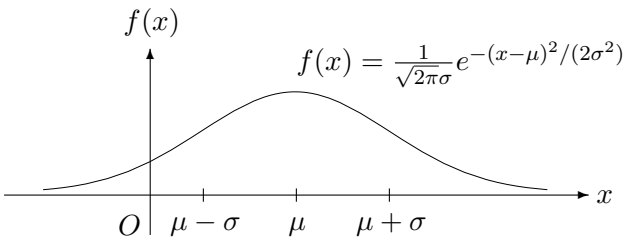


圖5. $\mathcal{N}(\mu, \sigma^2)$分佈之機率密度函數及其圖形

設X有$\mathcal{N}(\mu, \sigma^2)$分佈, 則$\mu$, σ分別為X之期望值及標準差。常態分佈有個特性, 即$(X-\mu)/\sigma$有$\mathcal{N}(0,1)$分佈。也就是經標準化後, $\mathcal{N}(\mu, \sigma^2)$成為標

準常態分佈。再度利用(19)式, 得

$$P(|X-\mu| \leq \sigma) = P(|\frac{X-\mu}{\sigma}| \leq 1) = 2\Phi(1) - 1 \doteq 0.6827,$$
$$P(|X-\mu| \leq 2\sigma) = P(|\frac{X-\mu}{\sigma}| \leq 2) = 2\Phi(2) - 1 \doteq 0.9545,$$
$$P(|X-\mu| \leq 3\sigma) = P(|\frac{X-\mu}{\sigma}| \leq 3) = 2\Phi(3) - 1 \doteq 0.9973 \text{。}$$

即X若有$\mathcal{N}(\mu, \sigma^2)$分佈, 則標準化後的$X$, 落在距期望值1,2,3個標準差內之機率, 分別約為0.68, 0.95及0.997, 與μ, σ^2之值皆無關。這就是中學教科書裡, 所謂68-95-99.7法則(規則)。有些還稱此為"68-95-99.7經驗法則", 或就只稱"經驗法則"。尚有作者則稱此為"常態分配的經驗法則"(如丁村成(2010))。事實上, 是有人說68-95-99.7 rule, 或three-sigma rule, 或empirical rule, 但並無68-95-99.7 empirical rule。並且前述三種稱呼, 均未見於一般機率統計的書中。各種中西文統計學名詞, 及統計大詞典中, 也少有選錄者。對隨機變數X具$\mathcal{N}(\mu, \sigma^2)$分佈, 會有所謂68-95-99.7的性質, 是先"證明"出$(X-\mu)/\sigma$有$\mathcal{N}(0,1)$分佈; 然後"計算"出表1, 絕不是憑什麼"經驗"而得到的。

例3. 如果有人告訴你, 他獨立地投擲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, 你知道大約會得到5,000個正面。但由(10)式(n =10,000), 會恰好得到5,000個正面的機率, 僅約0.008。因此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"。即如果離5,000過遠, 會令人懷疑銅板不是公正。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。

若銅板為公正, 則要得到至少5,250個正面之機率為

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

前面說過了, 因有$\binom{10,000}{k}$這項, 除非k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

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

利用中央極限定理, 並由表1, 得$P(X_1 + \cdots + X_n - n\mu > 5\sigma\sqrt{n})$之機率約為$3 \cdot 10^{-7}$(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, 我們推測銅板正面出現的機率很可能大於1/2。

將μ及σ皆以1/2代入(20)式, 得

$$(24)\quad P(X_1 + \cdots + X_n \in [\frac{n}{2} - \frac{c\sqrt{n}}{2}, \frac{n}{2} + \frac{c\sqrt{n}}{2}]) \doteq 2\Phi(c) - 1,\ c \geq 0\ 。$$

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(離期望值5,000不超過3個標準差)之機率, 約為0.9973。此機率通常便已被認為夠大了。即認為所得正面數, 不太會落於此區間外。此概念我們下一節會再討論。

(24)式顯示, 當n夠大時, 獨立地投擲一公正銅板n次, 所得正面數落在一以$n/2$為中心之對稱區間的機率若固定, 譬如說皆為$2\Phi(c) - 1$, 則該區間之半徑$c\sqrt{n}/2$, 將隨著n之增大而增大。但此半徑若除以投擲數n, 得相對半徑$c/(2\sqrt{n})$, 將隨n之增大而漸減。對同一機率, 隨著投擲數n之增大, 所得正面數落在的範圍會愈來愈大。但相對投擲數而言, 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次, 佔投擲數之5%; 投擲1,000,000次, 區間半徑成為500次, 區間雖然變長, 但半徑卻只佔投擲數之0.05%。相對於投擲數, 後者乃是一較無關痛癢的小區間。

再看一個例子。

例4. 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, 他答對21題。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。若答對的題目過多, 顯示該生可能並非完全不會。如今該生答對21題, 21題夠不夠多呢? 我們來分析看看。

假設隨機地答。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

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

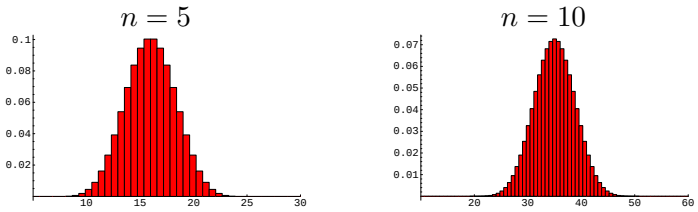
經查表得$1 - \Phi(2.776) \doteq 0.0027$。也就是若隨機地猜, 則會猜中至少21題之機率約為0.0027。此機率夠小, 足以讓我們得到該生很可能並非完全用猜的之推論。

例3及例4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。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3中的投擲銅板, 或例4中的考試答題, 即使銅板為公正, 是可能投擲出5,250個正面; 答題全用猜的, 也有可能答對21題, 甚至50題全猜對也非不可能。因為此二事件發生之機率皆為正。由於眞相為何, 常就是無法判定, 所以採用"假設"一詞。而得到的推論乃是**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。當機率計算不易, 若樣本數夠大, 往往中央極限定理便可派上用場, 以簡化計算。

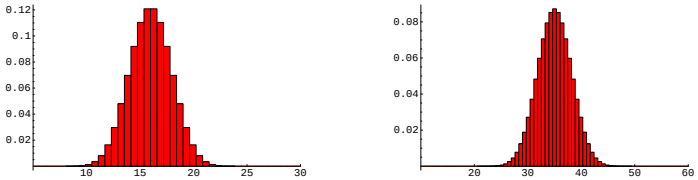
很多機率統計書上說當樣本數$n \geq 30$, 中央極限定理便適用。即若有iid的隨機變數$X_1, \cdots, X_n$, 且$\mu = E(X_1)$, $\sigma^2 = \mathrm{Var}(X_1)$皆存在, 則當$n \geq 30$, $(X_1 + \cdots + X_n - n\mu)/(\sigma\sqrt{n})$, 或$(\overline{X}_n - \mu)/(\sigma/\sqrt{n})$, 便皆可以$\mathcal{N}(0,1)$分佈來近似。這點比較微妙, 留待本節最後再說明。若X_i有

$\mathcal{B}er(p)$分佈,書上更說, 根據**經驗法則**(a rule of thumb, 這才眞是經驗法則), 通常只要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?

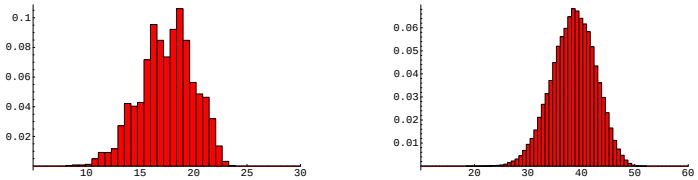
事實上, 若X_i's共同分佈之機率密度函數, 對稱於其期望值, n往往不用太大, 中央極限定理便已適用。否則n就得較大些。以p_i表骰子出現i點之機率, $i=1,\cdots,6$。於三組不同的p_i's下, 圖6給出$n=5$, 及$n=10$, $X_1+\cdots+X_n$直方圖。在(a)圖中, p_i's皆為1/6, 則即使$n=5$, 直方圖便已讓人產生常態分佈的"感覺", 但仍呈鋸齒狀。在(b)圖中, p_i's為對稱($p_1=p_6$, $p_2=p_5$, $p_3=p_4$), $n=10$時, 直方圖已有些平滑了。而可看出在(c)圖中, 即使p_i's較偏斜, $n=5$時, 直方圖雖不會讓人聯想到常態分佈, 但$n=10$時, 直方圖即使仍有些鋸齒狀, 但已大致已會讓人想到常態分佈了。底下我們對這種以圖形來說明中央極限定理, 略做說明。



(a) $p_1=p_2=p_3=p_4=p_5=p_6=1/6$



(b) $p_1=0.1,\ p_2=0.15,\ p_3=0.25,\ p_4=0.25,\ p_5=0.15,\ p_6=0.1$



(c) $p_1=0.2,\ p_2=0.1,\ p_3=0,\ p_4=0.1,\ p_5=0.4,\ p_6=0.2$

圖6. 各面出現機率不同的3個骰子, 分別投擲5次及10次, 所得點數和之直方圖

我們先來看一下中學教科書怎麼說。在中學裡介紹中央極限定理, 大抵僅針對二項分佈, 只是有些詮釋並不太正確。底下為一些教科書中的講法:

> 在參數是(n,p)的二項分布中, 當試驗的次數n足夠大時, 成功次數X的機率分布會近似於平均數μ為np, 標準差σ為$\sqrt{np(1-p)}$的常態分布。
>
> 當n足夠大時, 二項分配的圖形近似一平滑的鐘形曲線, 此時二項分配近似於常態分配。
>
> 可看出當$n=32$而$p=0.5$時, 二項分配機率圖是對稱且呈鐘形, 接近常態分配。$\cdots$事實上, 對任意$0<p<1$, 由中央極限定理可知: n夠大時, X會接近常態分配。
>
> $\overline{X}_n$的相對次數直方圖長相非常接近鐘形。
>
> n夠大, 隨機變數$\hat{p}=X/n$近似常態分布, 其平均數$\mu=p$, 標準差$\sigma=\sqrt{p(1-p)/n}$, $\cdots$。
>
> 當樣本數n夠大時, $\hat{p}$的分佈會趨近於平均數為p, 標準差為$\sqrt{p(1-p)/n}$的常態分佈。

我們仍是一字不改, 而如前所述, 分佈、分布、分配乃通用。

所謂n足夠大, 當然包含n無止盡地增大。對S_n有$\mathcal{B}(n,p)$分佈, 我們已說過, 不論p為何, n愈大時, S_n將很可能也是很大。以機率的語言來說, 就是$n\to\infty$時, S_n機率收斂至∞。同理, 在上一節討論大數法則時已說了, $n\to\infty$時, S_n/n機率收斂至p。因此說, n夠大時, S_n及S_n/n會分別近似那一常態分佈, 都非正確。其次來說明圖形。隨著n之增大, $\mathcal{B}(n,p)$分佈之直方圖, 將愈趨貼近水平座標軸(想想最大高度趨近0), 怎會近似鐘形曲線(見圖7)? 至於n很大時, S_n/n的直方圖, 將如底部中點為p之一座尖塔, 高聳入雲(見圖8), 也絕不會讓人聯想到鐘形曲線或常態分佈。又"趨近"為極限下的行為。前面所引最後一則中, 樣本數n僅為"夠大", 便接到"會趨近", 此並非極限裡的說法。

當n持續增大, 想圖示中央極限定理, 都要很謹慎, 絕不可想當然

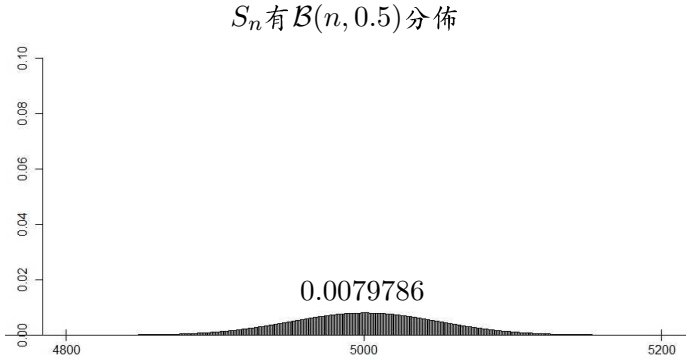


圖7. S_n之直方圖, $n = 10,000$

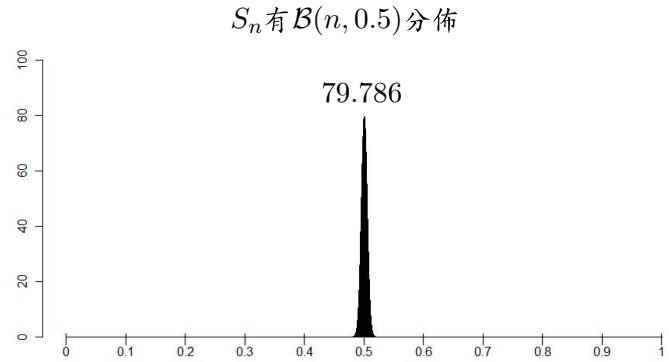


圖8. S_n/n之直方圖, $n = 10,000$

耳。記住, 對iid的隨機變數, 中央極限定理是說, 樣本和$X_1 + \cdots + X_n$, 及樣本平均$\overline{X}_n$, 都要經標準化後, 才會近似$\mathcal{N}(0,1)$分佈。並不是說$X_1 + \cdots + X_n$, 及$\overline{X}_n$有近似的$\mathcal{N}(0,1)$分佈, 這點須特別強調。關於圖示中央極限定理的進一步說明, 可參考黃文璋(2011c)一文。

事實上, 在Feller(1968)p.185建議, 對S_n有$\mathcal{B}(n,p)$分佈, 及$0 \le a \le b$, 若以下式來近似S_n介於a, b二整數間之機率

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

則n宜"適當地大"(注意並非要求"n夠大")。原因乃若不是採(13)或(16)式, 則給定a, b, $0 \le a \le b$, 當n很大時, $P(a \le S_n \le b)$根本就很小, 這時上式並無功能(此時右側當然也很小)。

又必須一提的是, 由於獨尊的名稱“中央”, 與涉及的分佈名之為“常態”, 有些人遂誤以為機率統計裡只有常態分佈。要知常態分佈雖無所不在, 所謂萬物有常, 但在實務上, 仍要謹慎判斷是否適合用來當模型。尤其有時明明一看就非具常態分佈, 更不可輕率採用。如2011年7月號科學月刊“Hello Kitty的數學”一文中, 對蒐集到一套時所要買的次數, 寫著“根據常態分佈的50-68-95法則$\cdots$”, 實不知彼處常態分佈從何而生? 似亦為想當然耳。切記, 並非任意獨立隨機變數和, 當樣本數夠多, 只要經標準化後, 便可以常態分佈來近似, 而是有須滿足的條件。另外, 有教科書說:

> 這類資料的次數分布圖呈現中間較高, 且左右對稱的鐘形時, 我們就稱這組資料呈現常態分佈。

亦有作者說(見丁村成(2010)):

> 分佈曲線都是呈現單一高峰的左右對稱曲線, 這種曲線稱為常態曲線。

此類講法都是誤導。到底何謂鐘形? 即使常見分佈裡, 除常態分佈外, 便有好幾個也具“單一高峰左右對稱”的機率密度函數圖形。如柯西分佈, t分佈, **拉普拉斯分佈**(Laplace distribution), 及**羅吉斯分佈**(Logistic distribution)等。圖9給出$\mathcal{C}(0,1)$分佈之機率密度函數圖形, 僅就形狀而言, 你能分辨圖9不像鐘形嗎?

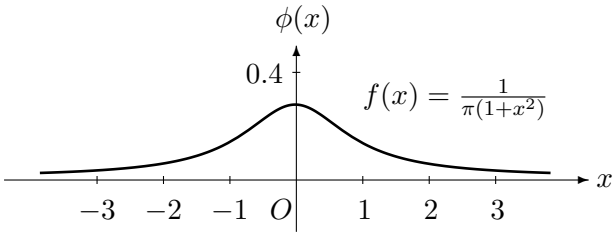


圖9. $\mathcal{C}(0,1)$分佈之機率密度函數圖形

另有教科書說:

> 對一組數據畫其直方圖後, 若將每個長條頂端中點以折線連接會有很多種不同的式樣, 但最後看到的是中間高而往左右兩邊下降, 近似鐘形。

這樣講對常態分佈的鐘形乃過度推崇。即使左右對稱都不見得是常態分佈, 何況很多數據連左右對稱都不是, 當然更不會有什麼近似鐘形。

最後, 前面提到很多書上說, 當樣本數$n \geq 30$, 中央極限定理便適用。但中央極限定理畢竟是極限下的結果, 而實際應用時, 樣本數當然是有限。就算樣本數超過30, 利用中央極限定理, 以常態分佈來近似標準化後的$X_1 + \cdots + X_n$, 或$\overline{X}_n$, 須在意此近似究竟好不好? 換句話說, 誤差是否夠小? 要知你可以任意近似, 但誤差大小可不能置之不理。而誤差大小, 乃與$X_1, \cdots, X_n$之共同分佈為何有關。舉例來看, 假設X_i's之共同分佈為$\mathcal{B}er(0.02)$, 且$n = 100$。則$X_1 + \cdots + X_{100}$有$\mathcal{B}(100, 0.02)$分佈。而$\mathcal{B}(100, 0.02)$之分佈可以$\mathcal{P}(2)$分佈做很好的近似。此處$\mathcal{P}(\lambda)$表參數λ之**波松分佈**(Poisson distribution), 其機率密度函數為

$$f(k) = \frac{e^{-\lambda}\lambda^k}{k!},\ k = 0, 1, 2, \cdots,$$

乃一離散型的分佈。圖10給出$\mathcal{P}(2)$分佈之直方圖, 這當然絕不像常態分佈, 再怎麼標準化都沒用。對$\mathcal{B}er(p)$分佈, 只要降低p, 則對更大的n, 仍有這種波松近似的結果。這是所謂**二項分佈趨近波松分佈**, 可參考黃文璋(2010a)5.1節。這類例子不少。如1,000個iid以$\mathcal{P}(0.001)$為共同分佈之隨機變數, 其和有$\mathcal{P}(1)$分佈(可參考黃文璋(2010a)第二章定理6.8)。雖樣本數高達1,000, 常態近似對此卻不適用。

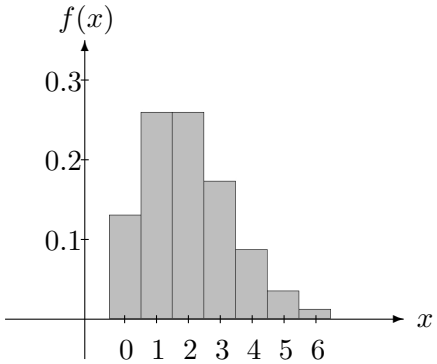


圖10. $\mathcal{P}(2)$分佈之機率密度函數圖形

7 信賴區間

人們常在做估計。如氣象局估計即將登陸颱風的降雨量, 政府估計明年經濟成長率, 學生拿到入學考成績單後, 估計能進那一校系等。有些情況很複雜, 沒那麼容易估計, 來看兩個較簡單的例子。

欲估計一銅板出現正面的機率p。一個簡單的方法是投擲若干次, 譬如說n次, 且得到X個正面, 則以正面數出現的平均, 或說相對頻率X/n, 來估計p。

想知道某地區最高的人之身高, 有何方法? 若能取得該地區所有人身高的資料, 立刻便知道了。只是可能沒有這種資料, 也很難將所有人都找來量測。仍有一個簡單的方法, 但這回不再是求平均了, 而是以所見過該地區最高者之身高來估計。只要你見過的人夠多, 且大致是隨機出現, 則此估計法將夠準確。

是否一定要這樣估計呢? 當然不見得。對於銅板, 你可以認為由於是政府發行的, 一定公正, 即$p = 0.5$。對於身高, 你也可根據過去經驗, 不做量測, 就估計最高者, 其身高必至少是188公分。所以若見到的最高者未達此值, 便估計188公分, 否則以所見到的最高者來估計。

上述二例皆屬點估計, 以一量來估計某未知參數。點估計要剛好命中的機率, 通常微乎其微, 於是便發展出**區間估計**(interval estimation), 以一區間來估計某參數。而這種區間, 便稱為**信賴區間**(confidence interval)。

中央極限定理在建立信賴區間時, 常也能發揮功能。底下給一情況來說明。

假設$X_1, \cdots, X_n$為iid之隨機變數, 且設$\mu = E(X_1)$, $\sigma^2 = \mathrm{Var}(X_1)$皆存在。對於$\mu$, 我們知道常以樣本平均$\overline{X}_n$來估計。底下我們來看, 如何以一區間來估計$\mu$。即使不知$X_i$'s之共同分佈為何, 當$n$夠大, 且$\sigma$已知, 利用(22)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

$$(25)\quad P\left(-z_{1-\alpha/2} \cdot \sigma/\sqrt{n} \le \overline{X}_n - \mu \le z_{1-\alpha/2} \cdot \sigma/\sqrt{n}\right) \doteq 1 - \alpha\text{。}$$

在此設Z有$\mathcal{N}(0,1)$分佈, 則對$\forall 0 < \alpha < 1$, z_α滿足$P(Z \le z_\alpha) = \alpha$。即

圖4中, 在$x = z_\alpha$左側面積為α, 因此$\Phi(z_\alpha) = \alpha$ 。(25)式等價於

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

故當σ已知, 區間

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

便稱為參數μ之一近似的(“近似的”三字常被省略)$1-\alpha$, 或者說$100(1-\alpha)\%$, **信賴區間** 。通常α取為一較小的正數, 如0.01, 0.05, 0.10等 。μ的$1-\alpha$信賴區間自然不唯一, 只要μ落在該區間之機率為$1-\alpha$便可 。譬如說

$$[\overline{X}_n - z_{1-\alpha/3} \cdot \sigma/\sqrt{n},\ \overline{X}_n + z_{1-2\alpha/3} \cdot \sigma/\sqrt{n}]$$

亦是 。可看出有無限多個μ之$1-\alpha$信賴區間 。一般在同樣的α下, 會取信賴區間的長度愈短愈好 。

表2給出一些常用的z_α之近似值 。至於一般的α, 就要藉助查表或使用計算機而得 。

表2. 常用之z_α值

α	0.90	0.95	0.975	0.99	0.995	0.999
z_α	1.282	1.645	1.960	2.326	2.576	3.090

如果σ未知, 因以一區間來估計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, 來取代σ, 其中

(28) $$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2,\ n \ge 2 \text{ 。}$$

$1-\alpha$, 或$100(1-\alpha)\%$, 稱為上述信賴區間之**信心水準**(confidence level), 或**信賴係數**(confidence coefficient) 。我們來看信賴區間的概念, 如何應用至做民調的情況 。

例5. 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, 總不能同一個人問他兩次同樣的問題。不過通常選民數相對於樣本數n很大, 所以取出後不放回, 由於樣本不獨立所造成的誤差, 就忽略好了。即考慮簡化後的情況, 假設$X_1, \cdots, X_n$獨立, 且$P(X_i = 1) = p,\ P(X_i = 0) = 1 - p$。因$\mu = E(X_i) = p,\ \sigma^2 = \mathrm{Var}(X_i) = p(1-p)$, 由(23)式得

$$P(|\overline{X}_n - p| > \varepsilon) \doteq 2(1 - \Phi(\varepsilon\sqrt{n}/(p(1-p)))),\ \varepsilon \geq 0\ 。\tag{29}$$

只要n夠大, 上式給出$\overline{X}_n$與p之差異大於ε之機率的近似值。

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

$$P(|\overline{X}_n - p| > \varepsilon) < \alpha, \tag{30}$$

或等價地,

$$P(|\overline{X}_n - p| \leq \varepsilon) \geq 1 - \alpha, \tag{31}$$

則n要多大呢? 由(29)式,

$$P(|\overline{X}_n - p| \leq \varepsilon) \doteq 2\Phi(\varepsilon\sqrt{n}/(p(1-p))) - 1 \geq 1 - \alpha\ 。\tag{32}$$

上式導致

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

即

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

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

$$n \geq \frac{z_{1-\alpha/2}^2}{4\varepsilon^2} \tag{33}$$

即可。當然因(29)式為一近似的式子, 所以此n亦為一近似值。

例如, 若$\varepsilon = 0.03$, $\alpha = 0.05$(這是一般做民調時所設定的ε及α), 則由表2, $z_{1-\alpha/2} = z_{0.975} \doteq 1.960$, 而$1.960^2/(4 \cdot 0.03^2) \doteq 1{,}067.1$。故取樣本數$n \geq 1{,}068$。實際抽樣後, 成功的樣本數往往不會恰好是1,068, 代入(33)式(令等式成立), 而解出誤差ε之值。$n > 1,068$時, ε便小於0.03; 反之ε則大於0.03。

附帶一提, 有些民調結果宣稱"在有效樣本數1,068下, 百分之95的信賴區間為$\cdots$, 且誤差不超過3%", 這種數據一看便有些可疑。因民調通常是多位訪員同時進行, 因此停止時, 成功的樣本數, 可說很難恰好是當初所設定的目標1,068。因此教科書上舉的例子, 如果盡是"成功訪問1,068位選民", 也都非恰當。

另外, 你可能覺得以1/4取代$p(1-p)$太粗糙。此點倒還好, 因這樣會造成須取較大的n, 並不會減小精確性。又(29)式也只是一近似公式而已。由於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(33)式, 在$\alpha = 0.05$下, 當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, 因常得在短時間內完成取樣, 困難度將增加很大, 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。

要特別提醒的是, 如第4節中所述, 簡單隨機抽樣所得樣本並不獨立, 須樣本數相較於母體數很小, 此時誤差才較小。有些教科書未留意此點, 如有書上說:

> 簡單隨機抽樣調查一千個合格選民, 就相當於丟擲一枚正面機率為p的銅板一千次的試驗, 而推論甲候選人支持率的問題, 就相當於推論銅板正面之機率的問題。

此說法當然為錯誤。有些作者如江前佑(2006)及丁村成(2010), 忽視X_1,

$\cdots$, X_n不獨立, 而去證明樣本獨立下, $\overline{X}_n$變異數的公式, 所得當然也是錯的。

統計實務裡, 為了簡化常採近似。使用這些方法時要清楚知道能近似的原因。否則若遇不適用近似的情況, 而貿然採用, 所得結果便不可靠了。例如, 假設選民數有1,000人, 其中有300人支持某候選人, 即$p = 0.3$。今採簡單隨機抽樣, 且$n = 1,000$, 即採普查。則樣本數等於母體數, 二者比值為1, 當然不是很小。這時若仍宣稱樣本$X_1, \cdots, X_{1,000}$為近似獨立, 便很荒謬了。事實上, 此時$\overline{X}_{1,000} = p = 0.3$, 為一常數, 並不隨機。中央極限定理在此完全不適用。

在例5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$\overline{X}_n = \overline{x}_n$後, 民調裡的說法是"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$[\overline{x}_n - 0.03, \overline{x}_n + 0.03]$。

例6. 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5。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(27)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

$$(34)\left[\overline{X}_n - z_{1-\alpha/2}\sqrt{\overline{X}_n(1-\overline{X}_n)/n}\,,\ \overline{X}_n + z_{1-\alpha/2}\sqrt{\overline{X}_n(1-\overline{X}_n)/n}\right]\text{。}$$

此處以$\sqrt{\overline{X}_n(1-\overline{X}_n)}$取代$\sigma$, 較以$S_n$取代$\sigma$, 形式簡單些。

對42取6的樂透彩, 若開獎公正, 則開出的頭獎號碼, 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 。

我們提過, 在高中引入信賴區間, 可說是問題重重。底下略做說明 。在高中數學裡, 為了近似, 先介紹中央極限定理, 且以最簡單的情況, 即伯努力分佈為例 。講完中央極限定理後, 便以民調當應用, 因引入信賴區間, 才是大費周章介紹中央極限定理的主要目的 。只是民調的取樣, 為簡單隨機抽樣, 也就是取出後不放回 。既然是取出後不放回, 則取出的樣本便不獨立 。在第4節我們已說了, 此時涉及的是超幾何分佈, 而非二項分佈 。只是要講清楚為什麼在此情況下, 中央極限定理仍適用, 可不是那麼容易 。高中數學課本, 寫到這裡, 彆扭不產生也難 。於是或不講原因, 讓細心的學生充滿疑慮; 或講得一團混亂, 製造出更多困擾 。

另外, 信賴區間的含意, 也讓很多人產生困惑 。我們必須略做闡釋, 也可參考黃文璋(2006)一文 。

在取樣前, (27)式為一隨機區間 。取樣後, 得到$\overline{X}_n=\overline{x}_n$, 代入(27)式, 則得一常數區間, 即

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

此固定區間或包含μ, 或不包含, 說有$1-\alpha$的機率會包含μ, 難免令人難以置信 。以投擲來估計銅板出現正面的機率p 。假設經投擲100次後, 得到p之一95%信賴區間(0.33,0.45) 。如何解釋95%呢? 考慮下述三選項:

1. 有95%的信心, 區間(0.33,0.45)包含p 。
2. p落在區間(0.33,0.45)之機率為0.95 。
3. 若持續重複前述投擲的實驗, 每次得一信賴區間, 則那些區間中, 約有95%會包含p 。

有些人可能知道了, 一般認為選項3乃正確解釋, 此即對機率的頻率觀點。不過要注意, 是“約有”95%會包含p。有些作者如江前佑(2006)及丁村成(2010), 分別解釋為“有95%包含眞正的p”, 及“有足夠的信心使其中的95%$\cdots$”, 少掉了“約”, 一字之差, 便未能顯現隨機的概念。選項1說了等於沒說, 並且也非十分恰當的講法。因所謂“信心”(或“信賴”), 是對整個建立該信賴區間過程的信心, 而非只是對所得到的那一固定區間的信心。至於選項2, 雖p並非隨機變數, 但倒也不要覺得有此看法很荒謬, 這其實是對機率主觀的解釋。原因是由一本為隨機的區間, 經取樣而得一固定, 即不再隨機的區間, 若根本不知實際p值為何, 只好由當初此區間產生的方式, 主觀地認為此區間包含p的機率為0.95。此正如在**汽車與山羊問題**(Car-Goat Problem)中(可參考黃文璋(2010b), 及(2011d)), 有三扇門, 其中一扇後面有汽車, 另兩扇後各只有一頭山羊。你選定一扇門後, 主持人打開另兩扇門中的一扇, 門後是山羊, 問你是否要更改選擇。汽車究竟在那扇門之後, 在節目開始前早已確定, 並非隨機。但由於你不知, 所以仍可依“自認合理”的假設下, 大算機率。並根據選擇更改或不更改, 何者得到汽車之機率較大, 而做決定。當然, 如果你知道p之值, 就不一樣了。譬如說取一介於0,1之間的p, 然後經由模擬, 產生一信賴區間, 這時你完全知道此區間是否包含p, 就不能說有0.95之機率p落在此區間。

底下為一類似的情況, 但也困擾不少人。假設有一公正銅板, 投擲後遮蓋住, 不讓人看見。則可否說會是正面的機率為1/2? 答案是可以, 仍可視為主觀機率。雖是否正面朝上已經確定了(有人因此說正面的機率不是1便是0), 但因你並不曉得, 所以仍可談機率。

試驗若僅做1次, 則不是發生就是不發生, 這時看不出機率的作用。例如, 假設有一正面出現機率為0.9之銅板, 要人猜投擲之後出現正面或反面? 答案當然很明顯該猜正面。但若有人不信邪, 選擇反面, 他可能會猜對, 因發生機率0.1為正。這時他是否可很得意地說, 打敗機率了嗎? 可以預期, 若持續投擲, 則堅持選反面者, 將大部分是猜錯的。機率的功能, 便是在這種情況下顯現。同理, 試驗若能一再重複, 則即使主觀的機率, 也能因此驗證其主觀是否夠客觀, 以及該主觀機率是否須做調整。以前述認

為經取樣後, 所得特定信賴區間包含p的機率乃0.95為例, 弱大數法則告訴我們, 只要重複多次, 則將有約95%的區間會包含p, 不就印證其主觀機率有道理! 要知主觀機率也並非能假主觀之名, 而太隨意。只要能重複的試驗, 其主觀機率是否夠客觀, 便不難檢驗了。

$1-\alpha$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$2z_{1-\alpha/2}\cdot S_n/\sqrt{n}$), 但各區間之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p。例如, 若$\alpha=0.05$, 且得到100個信賴區間, 則其中將約有95個左右會涵蓋p。實際上有幾個涵蓋p當然不一定。但在取樣前, 我們知道(不妨忽略這些95%信賴區間只是"近似的"), 會涵蓋p之區間數, 此為一隨機變數, 有$\mathcal{B}(100,0.95)$分佈, 可能有95個, 也可能有85個(此機率很小, 利用中央極限定理, 得此機率小於$\Phi(-4.588)<\Phi(-4)\doteq 3.164\cdot 10^{-5}$, 精確值為$\Phi(-4.588)\doteq 2.238\cdot 10^{-6}$), 或其他值, 但大致是95個左右。

本文在前言中指出, 課程綱要裡, 要學生以模擬或實驗, 來解讀"何以大多數的學生所得的信賴區間都會涵蓋p"之講法不妥。各位現在應知道了, 只要班級數夠多, 或實驗次數夠多, 則若有某一班, 全班學生所得的信賴區間, 有好幾個未涵蓋p, 乃不足為奇。詳細討論可參考黃文璋(2006)一文。

另外, 求信賴區間並不一定要用到中央極限定理。在估計期望值時, 如果共同分佈就已是常態分佈了, 那(27)式給出μ一恰好是(而非近似)$1-\alpha$之信賴區間。估計量若非樣本平均(如本節一開始估計身高之例), 很可能也用不到中央極限定理。在黃文璋(2003b)9.2節, 更有不少不需中央極限定理, 便能給出信賴區間之例。所以課程綱要中說, 估計背後的理論是中央極限定理, 此講法實有待商榷。

我們再以保險為例。保險對保户而言, 當然不是一**公正賽局**(fair game)。對於意外險, 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。這一年若平安無事, 則幾千元的保費便

有如蒸發掉了。此時應感慶幸, 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, 且保意外險, 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。保險每人保的金額、項目不盡相同; 因工作性質, 及年齡等原因, 每人保險費率也可能不同。底下我們以一簡單的賭博之例, 來顯示"保户愈多, 保險公司破產的機率便愈小"之現象。

例7. 假設每次投注1元, 贏則得2元(淨得1元), 輸則1元為賭場收走。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,

$$(35)\quad P(\sum_{i=1}^n X_i \geq 1) = P\left(\frac{\sum_{i=1}^n X_i - n(2p-1)}{2\sqrt{p(1-p)}\sqrt{n}} \geq \frac{1 - n(2p-1)}{2\sqrt{p(1-p)}\sqrt{n}}\right)$$
$$\doteq 1 - \Phi\left(\frac{(n+1)/2 - np}{\sqrt{np(1-p)}}\right)。$$

舉個例來看上式。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的近似值。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户愈多, 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。因並非所有保户同時加入保險, 且保費為一次收齊, 若才收了幾份保費, 就有人要理賠, 則便付不出理賠金了。

例8. 承上例。賭場想知道有0.99的機率, 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$P(\sum_{i=1}^n X_i \leq -a)$。令

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

因$z_{0.99} \doteq 2.326$, 由上式得

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

故

$$a \doteq n(1-2p) - 2.326\sqrt{np(1-p)}\text{。} \tag{36}$$

即有0.99的機率, 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0.99的機率約至少可淨贏的金額

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

若對機率與信心二詞之區隔仍有疑惑者, 可參考黃文璋(2011e)一文。

8 結語

一般認為, 機率統計乃比微積分不易教。後者有物理意義, 前者卻缺乏, 因此較難用直觀的方式來教機率統計。近年來, 統計的題材陸續被引入高中。只是要在中學數學課本少少的篇幅裡, 將一些有點深奧的機率統計概念講清楚, 乃屬高難度。對教科書的編輯, 為一大挑戰。導致現行教科書中於此部分, 往往語焉不詳, 讓教師與學生, 皆常產生困惑。而熱心教師所發表的心得, 也無可避免充斥不少錯誤。對基本概念的詮釋未能很精準, 這點國外較初等的大學機率統計教科書, 也不例外。看來機率統計要普及化, 的確是困難, 不分中外。

本文既名之為"庶民中央極限定理", 顯然初衷是設定普羅大眾(proletariat, 平民百姓)皆能讀得懂。這當然不容易, 退而求其次, 若能讓至少具備高中數學程度者, 便可以吸收, 也就不枉寫此文了。因希望對在高中任教的數學老師, 及一般系所的大學生及研究生, 於理解機率統計的內涵, 能有些幫助, 所以下筆時雖力求淺白, 但仍儘量嚴謹, 務求正確。為導正概念, 也將個人所知, 教科書及文章上出現的一些問題, 拿出來討論, 用意是避免類似錯誤一再發生。而若本文有任何錯誤, 也歡迎讀者指出。對機率統計概念之解釋, 是很容易犯錯, 有時還有不同的觀點。須反覆切磋琢磨, 以求精進。由於是以目前中學裡, 與中央極限定理相關的題材為骨幹,

所以雖基礎的機率統計中, 尚有些也常讓人迷惑的概念, 如獨立性, 及條件機率等, 本文皆未多著墨。可參考黃文璋(2010b), 及(2011d)二文。

機率統計既然要深耕, 就得嚴肅以對, 釐清其中隨機的本質, 而不是僅視其為較簡單的數學。如此才能生根, 提高國民統計素養, 達到在高中引進機率統計題材之目的。

參考文獻

1. 丁村成(2010). 談高中新教材中機率統計的缺失與改進。數學傳播季刊, 34(4): 9-23。
2. 江前佑(2006). 關於高中統計課程的一些說明。高中數學電子報第10期(2006年5月31日)。
3. 姚若潔譯(2004). 奇蹟? 機率!?。科學人, 31期(2004年9月號): 18。
4. 黃文璋(2003a). 隨機思考論。華泰文化事業股份有限公司, 台北。
5. 黃文璋(2003b). 數理統計。華泰文化事業股份有限公司, 台北。
6. 黃文璋(2005). 應隨機以恆周。科學發展月刊, 394期(2005年10月號): 68-73。
7. 黃文璋(2006). 統計裡的信賴。數學傳播季刊, 30(4): 48-61。
8. 黃文璋(2007). 統計裡的估計。數學傳播季刊, 31(1): 3-20。
9. 黃文璋(2008). 誰與達爾文爭鋒。心在南方(http://www.stat.nuk.edu.tw/South.asp), 2008年4月7日。
10. 黃文璋(2009). 統計思維。數學傳播季刊, 33(4): 30-46。
11. 黃文璋(2010a). 機率論, 第二版。華泰文化事業股份有限公司, 台北。
12. 黃文璋(2010b). 機率應用不易。數學傳播季刊, 34(1): 14-28。
13. 黃文璋(2011a). 機率空間。黃家小館(http://huang.nuk.edu.tw/cindex.htm)。

14. 黃文璋(2011b). 隨機變數。黃家小館(http://huang.nuk.edu.tw/cindex.htm)。

15. 黃文璋(2011c). 關於中央極限定理之圖示。黃家小館(http://huang.nuk.edu.tw/cindex.htm)。

16. 黃文璋(2011d). 認識機率。數學傳播季刊, 35(2): 32-44。

17. 黃文璋(2011e). 對機率要有信心。黃家小館(http://huang.nuk.edu.tw/cindex.htm)。

18. Diaconis, P. and Mosteller, F.(1989). Methods for studying coincidences. *Journal of the American Statistical Association*, 84, 853-861.

19. Feller, W.(1968). *An Introduction to Probability Theory and Its Applications*, Vol. 1, 3rd ed. John Wiley & Sons, New York.

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

21. Tijms, H.(2004). *Understanding Probability*: *Chance Rules in Everyday Life*. Cambridge University Press, Cambridge.

附錄一　對機率要有信心

黃文璋
國立高雄大學應用數學系

1　前言

在高中數學學科中心網站的討論區看到一篇“機率或信心Q&A”的文章(以下稱為Q&A), 由題目即知該文試圖解釋機率與信心二名詞。

處在一個生活中已缺不了統計的時代, 具備足夠的統計知識, 似乎是必須的。這可能是近年來, 高中數學增加不少統計題材的主因。但我一直不贊成高中數學裡教太多統計。因要在一向講究精準及不變性的高中數學裡, 讓學生理解涉及隨機概念的題材, 先天便有其困難。尤其信賴區間, 在一般大學統計教科書中, 是放在較後面的章節。學生接觸信賴區間時, 已具備夠多的機率統計基礎。如今卻跳過一切預備知識, 堂而皇之地進入高中數學裡, 引發的困惑, 多過吸收到的統計知識, 也就不足為奇。有些認眞的高中教師, 想釐清一些概念, 但由於高中數學裡, 能用的工具極有限, 因此很多釐清的文章, 往往淪為郢書燕說。既未澄清任何概念, 反讓原本清晰的部分, 也模糊起來。本文便擬對Q&A, 略做修正及補充。

2　機率的意義

在Q&A中, 以下述問題開場:

> 在某一次教師研習的綜合座談中, 有老師提到以下的問題:
>
> 投擲1只骰子(傳統6面骰字, 點數1,2,3,4,5,6,點數1,4為紅色, 其他點數為黑色), 擲出1 點的機率為何? 若已知擲出的顏色為紅色, 請問擲出1點的機率為何?
>
> 這個問題在學生沒有學過"信賴區間與信心水準"之前是沒有疑義的, 還沒有擲出骰子前, 擲出1點的機率是1/6; 在擲出的顏色為紅色的條件下, 擲出1點的機率為1/2。但學生學過信賴區間之後會說, 不, 老師, 既然骰子已擲出, 則骰子是1點的機率不是1就是0, 不會是1/2或其他任何數字。
>
> 結果, 學得越好的學生心中的疑惑卻越深。我們該如何回應這個學生? 在條件機率的情境下, 應該使用「機率」或「信心」?

Q&A作者如下回答:

> 擲一個公正骰子一次, 出現一點的機率$=1/6$。這是由大數法則得到的, 故以機率稱之。擲一個公正骰子一次, 在已知出現紅色點數的條件下, 出現一點的機率$=1/2$。這是由大數法則得到的, 故仍以機率稱之。

其中我們除了加上幾個標點外, 其餘一字未改。

我們說投擲一個公正骰子, 此處"公正"一詞乃口語, 其含意為骰子各面出現的機率相同, 即皆為1/6。這是骰子公正下的結果, 與"大數法則"並無關係。就如若假設生男、生女的機會相同, 那便表生男、生女的機率各為1/2。僅是一件事不同的講法而已。說是由大數法則得到, 不但不是事實, 而且易讓人對機率生畏, 後果是令師生連對中文字的運用, 都將失去

信心。Q&A作者說“學得越好的學生心中的疑惑卻越深”, 豈不就是因常看到這些夾纏不清的講法所造成?

前述Q&A所提出的問題, 最後有句“在條件機率的情境下, 應該使用機率或信心?” 依題意, 在擲出骰子的顏色為紅色之條件下, 擲出1點的機率就是1/2, 毫無疑義。提到什麼在條件機率的“情境”下, 提到什麼“信心”, 只有令人更為困惑而已。本來無一物, 就不要惹塵埃。

現在來看“骰子已擲出, 則骰子是1點的機率不是1就是0, 不會是1/2或其他任何數字”, 此Q&A中所提學生的困擾。欲解此疑惑, 歸根究柢, 仍是要回到機率的意義。

事實上, 只要是隨機現象, 便能談機率。有人敲門, 是男是女? 你可說男女機率各1/2。門外那人的性別當然是確定的, 並不會忽男忽女。但由於你不知, 對你而言, 可能男可能女, 這樣的“未知”, 造成“你”可視門外者之性別為一隨機現象。而基於社會上男女約各半的認知, 在無其他資訊下, 1/2的機率便是這樣產生。此與Q&A中, 已知擲出的顏色為紅色(點數有1,4二可能), 則擲出1點的機率為1/2, 乃一樣的原理。但若你聽到門外傳來高跟鞋的走路聲, 說不定會說8成是女的。這裡便用到條件機率, 給定的條件是“敲門者穿高跟鞋”。當然, 若你以往的經驗, 是男生也有不少穿高跟鞋者, 因此認為敲門者乃女生的機率為0.6, 亦無不可。對同一隨機現象, 每個人認定的機率, 可以很主觀的。但要注意一點, 一旦門打開, 則門外那人, 不是男就是女。即這時敲門者為女生的機率, 將不是1就是0, 不能再說0.8或0.6了。機率裡自有一套邏輯, 而既然是邏輯, 就要合理。若明明已看到是女生(或男生), 還說女生的機率為0.8; 或若已假設骰子為公正, 卻認為投擲一次後, 會得到偶數的機率為0.7, 那就絕非在談機率了。

上述那類例子處處可見。投擲一銅板, 待落地後蓋起來, 要人猜正面或反面? 拿個蘋果放背後, 要人猜蘋果或橘子? 這些都是對有些人並非隨機現象, 但對某些人卻為隨機現象之例。又如考試前一天仍可猜題, 題目不是早已出好了嗎? 怎還有人在說“這題很可能會考”? 也是類似的道理。再給一例。大家聽過“汽車與山羊”的問題嗎? 此問題亦曾出現在電影“決勝21點” 中。有三扇門, 其中一扇門後有汽車, 另兩扇門後為

山羊。你選定一扇門後, 主持人打開另兩扇門之一, 問你是否要更改選擇。選擇要不要更改, 當然是依更改後得到汽車的機率是否會變大? 你看, 汽車在那一扇門之後, 也是早已確定, 但仍可談機率。汽車與山羊問題之詳細討論, 可參考黃文璋(2010)一文。

至於為什麼可以說骰子為公正, 即何以知道骰子各點數出現的機率皆為1/6? 這有幾種可能的想法, 牽涉到"機率"一詞幾種不同的意義, 可參考黃文璋(2011)一文。我們列出幾種常見的想法如下。第一種是因骰子有6個面, 基於相同的可能性(古典機率), 導致每個面出現的機率皆為1/6。第二種想法可稱為主觀的, 說不定就是覺得沒有道理那一面較易出現。第三種可能, 乃由過去多次投擲的經驗, 觀察到骰子各面出現的"相對頻率"(即某面出現的次數除以總投擲數)難分軒輊。還有一種可能(這是不少數學家採用的), 就是此為一假設。若骰子不是虛擬, 而是真有一骰子, 則不論基於上述那一種想法, 究竟是否可採信骰子各面出現的機率皆為1/6, 可以做一統計檢定。因此, 在上述第三種對機率的想法裡, 骰子各面出現的相對頻率, 是否夠接近到足以視為相等, 便可藉助統計檢定來判定。而也是在第三種想法裡, 某面出現的機率為1/6, 導致多次投擲後, 該面出現的相對頻率將接近1/6, 才用到大數法則。但要有隨機的概念, 即使是公正的骰子, 不論投擲再多次, 各面出現的相對頻率, 都很難相等。有些人以為投擲數夠多後, 會使各面出現的相對頻率"相等", 這完全是錯的。

另外, 必須一提的是, 在隨機世界中, 一切都是假設, 只看你接受那一個。骰子是否公正, 不論投擲再多次, 仍是天曉得。統計檢定裡, 依無罪推定的原則, 及在給定所能容忍之犯錯機率下, 有一套程序, 以判定該接受, 或拒絕(更保守的講法是"不能接受")那一假設。除非不是隨機現象(如銅板兩面皆為正), 否則不論證據再顯著(如投擲銅板100次都得正面), 其推論均可能犯錯。我們僅能以較好的統計方法, 減小犯錯機率。

3 信心水準

Q&A的第二部分如下:

中央極限定理

- 從平均數為μ, 標準差為σ的母群體中, 以簡單隨機抽樣法抽取n個樣本, 以$\overline{X}$表示樣本平均值, 當樣本數n足夠大時, 則$\overline{X}$的期望值為μ, 變異數為σ^2/n 。
- CLT: $\overline{X}$分配近似常態分配$\mathcal{N}(\mu, \sigma^2/n)$ 。
- 平均數標準化後, 分佈近似$\mathcal{N}(0,1)$的分佈, 即

$$\frac{\overline{X}-E(\overline{X})}{\sqrt{Var(\overline{X})}} = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \to \mathcal{N}(0,1),$$

$$P\left(-1.96 < \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} < 1.96\right) \approx 0.95 \text{ 。} \qquad (*)$$

由$(*)$式子得到

$$\begin{aligned} & P\left(\mu - 1.96\sigma/\sqrt{n} < \overline{X} < \mu + 1.96\sigma/\sqrt{n}\right) \qquad (**) \\ & = P\left(\overline{X} - 1.96\sigma/\sqrt{n} < \mu < \overline{X} + 1.96\sigma/\sqrt{n}\right) \approx 0.95 \text{ 。} \end{aligned}$$

但是, 我們僅作一次抽樣, 得到樣本平均數$\overline{x}$, 標準差S, 95%信賴區間$[\overline{x}-1.96S, \overline{x}+1.96S]$, 此區間有包含到眞正的$\mu$值, 要不就是沒有包含到眞正的$\mu$值 。$(**)$式中, $\overline{X}$為隨機變數, 所以是重複做相同的抽樣很多次以後所得到機率, 故$(**)$的結果, 仍以機率表示 。

但是受限於人力、物力、財力, 我們僅作一次抽樣, 故稱$[\overline{x}-1.96S, \overline{x}+1.96S]$為$\mu$的95%信賴區間, 我們有95%信心說$\mu$會落在此區間內 。

此處, 我特別以μ、$\overline{X}$表示, 希望不要被p、$\widehat{p}$搞混, 但其意思是相同的 。

Q&A中缺乏足夠文字的鋪陳, 文字也屢有跳躍或不明確處。例如, 根本沒有p、$\widehat{p}$, 何以說不要搞混? 因此有時不太知道作者的意思。但再度,第一個黑點後那段敘述, 與中央極限定理沒有關係。此句並不嚴謹, 本擬設法修正此句, 但如前所述, 因Q&A的作者並未加以說明, 無法了解此句與全文之關聯。為避免誤導, 此句宜刪除, 所以就不費神去修改。

簡單隨機抽樣, 分取出後放回及取出後不放回兩種。若是前者, 則對所有的樣本數n, $\overline{X}$的期望值必為μ, 變異數必為σ^2/n, 不需n很大。若是後者, 如做民調時, 從母體中(設有N人)隨機抽取n個樣本(名字, 或電話)後不放回, 欲估計母體中對某一議題支持之比率, 則知道N, n及μ, 支持的人數X有超幾何分佈(hypergeometric distribution), 期望值為$n\mu$。因此支持的比率$\overline{X}(=X/n)$之期望值, 對每一樣本數n皆為μ。至於$\overline{X}$的變異數給在下式:

$$Var(\overline{X}) = \frac{N-n}{N-1}\frac{\sigma^2}{n},$$

稍複雜些。可看出只有當n與N相比很小時, $(N-n)/(N-1)$才會接近1。所以無論那一種簡單隨機抽樣, $\overline{X}$的期望值皆為μ。至於$\overline{X}$的變異數, 對取出後放回, 等於σ^2/n; 對取出後不放回, 只要n/N很小(而非n足夠大), 便接近σ^2/n。也就是都不需要用到中央極限定理。

必須一提的是, 中央極限定理通常是針對取出後放回的簡單隨機抽樣, 或更一般地, 針對獨立且有共同分佈之樣本。另外, 對於取出後不放回的簡單隨機抽樣, 在n與N相比很小下, 說n足夠大時(當然仍要比N小), $\overline{X}$的變異數"為"σ^2/n固然不對, 但"接近"或"近似"也都不是貼切的講法。因n夠大時, $\overline{X}$之變異數及σ^2/n, 二者皆近似於0。而二近似於0的數, 自然彼此近似。

第二個黑點後那句也宜刪除。中央極限定理, 並非在講$\overline{X}$的分佈, 乃是針對標準化後$\overline{X}$之分佈。對獨立且有共同分配之樣本, 當n很大時, $\overline{X}$之變異數σ^2/n近似於0。而一隨機變數之變異數很小, 表示它有很大的機率近似於常數。即此時$\overline{X}$近似於μ之機率很大。也就是對$\overline{X}$, 與它較相關的是大數法則, 並非中央極限定理。因此這句就不理它了, 直接刪掉。

在第三個黑點裡, 幾個式子大抵可接受。文字敘述中以樣本標準差S取代(**)中的σ也無妨, 仍為一近似的式子。只是第一個式子中有趨近的符號"$\rightarrow$", 因此宜在"即"之後, 加上"n趨近∞時", 或"$n \rightarrow \infty$時"。不過"所以是重複做相同的抽樣很多次以後所得到機率"這句一定得刪除, 完全畫蛇添足, 徒增困擾。$\overline{X}$為隨機變數, 是使μ落在區間有"機率"可言之原因, 與做多少次抽樣無關。再來"受限於人力、物力、財力, 我們僅作一次抽樣"也宜刪除。每次依同樣步驟抽樣, 不論得到再多的區間, 每一個皆是μ的95%信賴區間, 且95%永遠稱為"信心水準", 不會變成"機率"0.95。在估計時, 即使"人力、物力, 及財力" 皆不設限, 每次也只會(或說只想)得到一個信賴區間, 否則究竟要"信賴"那一個? 但在不受限制下, 取樣愈多(即n愈大), 所得的信賴區間長度將愈短。亦即對μ的估計將更準確, 這是花更多精力後的代價。

從估計的觀點, 區間較一個點更可"信賴"。想想醫生宣佈病人可再活1年, 跟說大約可再活6個月至3年間, 何者讓人覺得此醫生較科學, 更值得信賴? 對於估計, 譬如估計一銅板出現正面的機率μ, 在實際取樣前, μ有一機率值會落在信賴區間。如前述(**)式, 其中有一"機率"0.95。但我們一開始就不稱0.95"機率區間"。因就是為了估計才得到該區間, 接續便要取樣。而一旦得到一固定的信賴區間, 說此一特定的區間, 有0.95的機率包含μ, 難免讓人覺得不太妥當, 而又不易解釋清楚, 所以當初引進信賴區間概念之統計學家, 才自始便不稱0.95為機率。但μ的眞實值畢竟乃屬未知, 而數據會說話, 你還是覺得你的投擲結果, 總有值得"信賴"處。即除非已經知道μ值了, 否則有一定的"信心", μ會落在此區間。95%, 便用來描述該信心之大小, 稱為信心水準。至於Q&A的第一部分, 即原先大家所熟知的機率及條件機率的概念, 與信賴區間根本無關, 不會因引進信賴區間的概念後而改變。

只要是為隨機現象(即使對有些人並非隨機), 就能談機率。引進信賴區間及附隨的信心水準後, 原先的機率不會變成"信心"。對機率要有信心。

參考文獻

1. 黃文璋(2010). 機率應用不易 。數學傳播季刊, 34(1): 14-28 。

2. 黃文璋(2011). 認識機率 。數學傳播季刊, 35(2): 32-44 。