銅板乾坤

黃文璋教授
國立高雄大學應用數學系

1 丟銅板

兩支球隊比賽前,常以丟銅板來決定誰先發球。打麻將前,也先丟骰子以決定由誰先開始。這其中我們不由自主地做了一些假設:第一,銅板(或骰子)是公正的,也就是兩(6)個面都同有1/2(1/6)的機會出現。第二,各次投擲爲相互獨立,也就是結果互不受影響。此二假設是否一定成立,當然只有天曉得。但小說裡不乏銅板並非公正(如在二月河(1994)pp.56-57,施琅準備至福建部署攻打台灣,臨離開北京前,康熙賜給他一百枚兩面字兒的銅錢),以及骰子被做過手腳(如金庸(1996)裡,韋小寶常用的骰子)之例子。

丟骰子可說是丟銅板的推廣:由兩個面變成6個面。當初製銅板者,一定沒想到銅板可被拿來當賭具用,更不會想到銅板在機率理論裡,扮演著極重要的角色。

在日常生活裡,有很多我們接觸的隨機現象(random phenomenon),往往恰有兩個可能的結果。譬如說生男或生女;考試及格或不及格,丟一骰子得到奇數或偶數點,自一袋中任取一球得到白球或非白球,某人是否得到某一特定的病,某股票是否上漲等。當然這一切,都可簡化爲成功(success)及失敗(failure)兩種結果。即在二結果中,指定一我們有興趣的,並將之稱爲成功,另一結果則稱爲失敗。

假設某隨機現象是可以重覆的,則每觀察一次這種隨機現象,便稱爲進行一次隨機實驗(random experiment)。而對一只有兩種可能結果的隨機實驗,便稱爲一伯努力試驗(Bernoulli trial)。有時進行一連串的伯努力試驗,然後數看看其中有幾次成功。

一伯努力試驗中,被稱爲成功的事件,不一定是我們喜歡的結果,而是二

結果中,我們較關心或有興趣者。例如,我們可能會記錄共有幾次交通事故,或共有幾人得病等。則每一次交通事故,或每有一人得病,便皆稱為一次成功。

在有些較複雜的隨機實驗中,若我們有興趣的是某特定事件的發生與否,則也會產生伯努力試驗。例如,假設觀察患有某病之病人的存活時間 T。雖然通常 T 取的值超過兩個,但若我們有興趣的為病人是否存活超過5年,則事件$E = \{T > 5\}$,便可稱為成功事件,至於其餘集$E^c = \{T \leq 5\}$,便稱為失敗事件。如此一來,此實驗就可視為一伯努力試驗了。又如,在某次選舉中,有候選人多位,投票情況當然很複雜。但若我們只關心選民是否會投給某特定候選人,則便化為只有二結果,伯努力試驗又出現了。

由上討論知,伯努力試驗的例子處處可見。而一只有二可能結果的實驗,可說是最簡單的實驗了(只有一個結果的實驗,其實並非隨機實驗,但可視為退化的隨機實驗)。又因銅板恰有兩個面,因此一伯努力試驗的隨機結構,與丟銅板會出現正面或反面的隨機結構相同,只要二者成功的機率相同。而丟銅板算是一最易操作的實驗,銅板遂在伯努力試驗中,擔任著隨傳隨到的工作。而由於此一簡單的伯努力試驗,竟也闖出一片天地,觸角既深又廣,水漲船高,銅板的身價遂也不同凡響。

此結果是令人發省的。計算機裡採用二進位,以0及1,便可表示出所有的數。在隨機世界裡,也可以只有二結果的伯努力試驗,而繁衍出種種的分佈。當初上帝先造了亞當,隨後便發現,要有個夏娃。沒有此一眾生之母,是無法生養眾多,在地上昌盛繁茂的(見舊約聖經創世記第九章)。從雌雄開始,諸如是非、恩怨、好壞、賢愚、美醜、高矮、胖瘦、快慢、陰陽、酸鹼、舊約新約、信與不信,及天堂與地獄,人們是如此習慣將事物二分。二分是一切分的開端,伯努力試驗的產生,就是源自這種二分的概念。太極生兩儀,兩儀生四象,四象生八卦,$\cdots$。此試驗成為許多分佈的源頭,其實是不足為奇的。

2　二項分佈

對一伯努力試驗,其結果以 ω 表之,便有一關於成功之指示函數(indicator function),設以 Y 表之。即

$$Y(\omega)=\begin{cases}1, & \text{若結果 } \omega \text{ 爲成功},\\ 0, & \text{若結果 } \omega \text{ 爲失敗。}\end{cases}$$

若成功之機率爲 p,則 Y 爲一離散型的(discrete)隨機變數(random variable),且機率密度函數(probability density function,簡寫爲p.d.f.)爲

$$(1)\qquad f(x)=\begin{cases}1-p & ,\text{若 } x=0,\\ p & ,\text{若 } x=1,\\ 0 & ,\text{其他}\end{cases}=\begin{cases}p^x(1-p)^{1-x} & ,\text{若 } x=0,1,\\ 0 & ,\text{其他。}\end{cases}$$

任一隨機變數,若以(1)爲p.d.f.,其中$0\le p\le 1$,便稱爲伯努力隨機變數,其分佈並稱爲伯努力分佈(Bernoulli distribution)。(1)中之p.d.f.與常數p有關,p 稱爲伯努力分佈的參數(parameter)。我們並以符號 $Ber(p)$ 表伯努力分佈。

有時我們觀測 n 個獨立的伯努力試驗,每次成功之機率皆設爲 p,又以 X 表總共成功之次數。若第 i 次之結果以 Y_i 表之, $i=1,2,\cdots,n$, $Y_i=1$ 表第i 次爲成功, =0表失敗,則易見

$$(2)\qquad X=Y_1+Y_2+\cdots+Y_n\text{。}$$

平常選舉開票時,往往以劃“正”來計票,每開出一張票,若支持某候選人,便在其名字(或號碼)下,依序劃一線。若開出 n 張票,則該候選人共得的票數,就是$Y_1+Y_2+\cdots+Y_n$,其中 Y_i 表第 i 張票的結果(=1或0)。當一組隨機變數獨立且有共同分佈(independent and identically distributed,簡寫爲i.i.d.),便稱爲一組隨機樣本(random sample)。

利用排列組合的技巧,可得

$$(3)\qquad P(X=i)=\binom{n}{i}p^i(1-p)^{n-i},\ i=0,1,\cdots,n,$$

其中

$$(4)\qquad \binom{n}{i}=\frac{n!}{i!(n-i)!},$$

而$i! = i(i-1)\cdots 3\cdot 2\cdot 1$, $i \geq 1$, $0! = 1$。$\binom{n}{i}$ 便是所謂二項式係數(binomial coefficient)。一以(3)爲p.d.f.之隨機變數,便稱爲有參數 n 及 p 之二項分佈(binomial distribution),以 $\mathcal{B}(n,p)$ 表之,其中 n 爲一正整數, $0 \leq p \leq 1$。當然 $Ber(p)$ 即爲 $\mathcal{B}(1,p)$。此二者皆爲離散型的分佈,因其取值的範圍爲一離散集合。由二項式定理(Binomial Theorem) 可驗證

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

故(3)式的確定義出一離散型的分佈。

有一些分佈由於常出現,因此稱爲常用分佈。當然所謂"常用"並無一標準。知道一隨機變數有某一特定的分佈,其好處是不用每次從頭推導其性質。而且有些書會列出某些常用分佈之機率值,確定該分佈之參數後,便可查表得知其機率值。

要注意的是,n,p 須固定,且 n 次伯努力試驗須爲i.i.d.,否則其成功次數便不是二項分佈了,見下例。

例1. 一袋中有100個球,其中有40個白球。隨機地取3球,每次取出後不放回,令 X 表其中之白球數,則可如下看出 X 並無二項分佈。

任取一球爲白球之機率爲0.4,所以有三個 $Ber(0.4)$ 試驗,但各次試驗並不獨立。此因若第1次取中非白球,則第2次有40/99之機率取中白球;若第1次取中白球,則第2次有39/99之機率取中白球。第2次取中的結果 Y_2 受到第1次的結果 Y_1 之影響, Y_1 與 Y_2 不獨立。因此 X 並無二項分佈。

不難看出,若每次取出球後放回,則 X 有 $\mathcal{B}(3,0.4)$ 分佈。

二項分佈當 n 較大時,計算上便有些困難。例如,若獨立地丟一銅板100次,每次成功的機率設爲1/2,總共之正面數 X,便有 $\mathcal{B}(100,1/2)$ 分佈。則 $X=30$ 之機率爲

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

$\binom{100}{30}$ 自然是不太好算的。儘管如此,一方面是常遇到,一方面是由簡單的伯努力試驗出發,但又不至於過分簡單(伯努力試驗的變化太少),二項分佈

遂成爲一重要且基本的分佈。

另外,丟銅板會得到一數列的正面或反面,如果將正面當做1,反面當做0,就得到一數列的1或0。總共幾個正面,也就成爲數列的和,(2)式便出現了。畢達哥拉斯(Pythagoras,約西元前580-500年)曾說萬有皆數(All is number)。我們往往喜歡將一切數量化。譬如說,每人有一身分證字號,進學校有學號,郵局也有局號等。我們可經由丟銅板,得到一串0,1的數字,再以之來描述如生男(以1代表),生女(以0代表)等兩個結果的隨機現象。如若生三個小孩,則會有000,001,010,100,011,101,110,111等8個可能的結果,因此會有兩個女孩(即001,010,100)之機率爲3/8,與用二項分佈$\binom{3}{2}(\frac{1}{2})^1(\frac{1}{2})^2$ 算出的吻合。

3 幾何分佈

持續地丟一出現正面的機率爲 $p>0$ 之銅板,每次的結果設爲獨立,直至出現一正面才停止,總共須丟幾次?與二項分佈裡不同的是,此次丟的次數並非固定,運氣好只要丟一次,運氣不好就可能要丟很多次了。如果是丟 x 次停止,則前 $x-1$ 次須皆爲反面,最後一次須爲正面。而每次出現反面之機率爲 $1-p$,出現正面之機率爲 p。因此若令 X 表總共須丟的次數,則

$$P(X=x)=(1-p)^{x-1}p,\ x=1,2,\cdots \text{。} \tag{5}$$

具有以(5)爲p.d.f.之隨機變數,便稱爲有參數 p 之幾何分佈(geometric distribution),以 $\mathcal{G}e(p)$ 表之。(5)式定義出一幾何級數(即等比級數),這是此分佈名稱之由來。也易驗證

$$\sum_{x=1}^{\infty}P(X=x)=\sum_{x=1}^{\infty}(1-p)^{x-1}p=\frac{p}{1-(1-p)}=1\text{。}$$

故此爲一離散型的分佈,取值在 $\{1,2,\cdots\}$。當然如果 $p=1$,則X 爲一退化的隨機變數,且 $P(X=1)=1$。一般假設 $0<p<1$。

附帶一提,有時會對反面數 Y 有興趣,即在得到一正面前,究竟白丟幾次?由於 $Y=X-1$,因此易見

$$P(Y=y)=(1-p)^y p,\ y=0,1,\cdots \quad (6)$$

有時幾何分佈是指其p.d.f.之型式如(6)。爲了區分,不妨對p.d.f.如(5)者,稱爲自1開始之幾何分佈,如(6)者,稱爲自0開始之幾何分佈。

幾何分佈就是進行一數列之獨立的伯努力試驗,但次數並未事先固定,而是直至成功一次才停止。這種情況是常可見的,如相親、試驗一新技術及革命等。

例2. 丟一骰子,直至出現一6點才停止,則總共丟的次數 X 便有 $\mathcal{G}e(1/6)$ 分佈。

由(5)又得

$$P(X\geq x)=\sum_{i=x}^{\infty}(1-p)^{i-1}p=(1-p)^{x-1},\ x=1,2,\cdots, \quad (7)$$

有一簡單的型式。由此又得

$$P(X\leq x)=1-P(X\geq x+1)=1-(1-p)^x,\ x=1,2,\cdots$$

在此對一隨機變數 Z,

$$F(x)=P(Z\leq x),\ -\infty<x<\infty,$$

稱爲 Z 之分佈函數(distribution function)。離散型隨機變數之分佈函數爲一階梯函數(step function)。

4 負二項分佈

執行一數列之獨立的伯努力試驗,每次成功的機率爲 p,直至得到一成功才停止,則所需試驗次數便有幾何分佈。一個很自然的推廣是,給定一

正整數 r,持續進行伯努力試驗,直至r 次成功才停止。則總共之試驗次數 X,便稱爲有參數 r 及 p 之負二項分佈(negative binomial distribution),以 $\mathcal{NB}(r,p)$ 表之。一般假設 $0<p<1$。若 $p=1$,則 $P(X=r)=1$。

若 $X=x$,由於最後一次爲成功,且第一次至第 $x-1$ 次中,有 $r-1$ 次成功,$x-r$ 次失敗,故

$$P(X=x)=\binom{x-1}{r-1}(1-p)^{x-r}p^r,\ x=r,r+1,\cdots \text{。} \tag{8}$$

(8)式的確定義出一機率分佈,其證明要用到下述推廣的二項式定理。

定理1. 對 $\forall c\in R$, $|x|<1$,

$$(1+x)^c=\sum_{n=0}^{\infty}\binom{c}{n}x^n \text{。} \tag{9}$$

此處對任一實數 c,

$$\binom{c}{n}=\frac{c(c-1)\cdots(c-n+1)}{n!},\ n\ge 1,\ \binom{c}{0}=1 \text{。} \tag{10}$$

又當 c 爲一非負整數時,由(10)得

$$\binom{c}{n}=\frac{c!}{n!(c-n)!},\ 0\le n\le c \text{。}$$

(10)即推廣 c 不爲非負整數時 $\binom{c}{n}$ 的意義。由(10)可看出,當c 爲一非負整數時,對 $\forall n>c$, $\binom{c}{n}=0$,此時(9)式之右側成爲一有限的級數。除此情況外(即 c 不爲非負整數),(9)式之右側皆爲一無限級數。

定理1即將二項式定理推廣到適用任意次方,此爲牛頓(Newton, 1642-1727)在數學的諸多發現中,相當值得激賞的一件,爲他早期的研究成果之一。

利用定理1,對 $\forall 0<p<1$,$r>0$,

$$\begin{aligned}p^{-r}&=(1-(1-p))^{-r}\\&=\sum_{i=0}^{\infty}\binom{-r}{i}(-(1-p))^i\end{aligned}$$

$$
\begin{aligned}
&= \sum_{i=0}^{\infty} \frac{(-r)(-r-1)\cdots(-r-i+1)}{i!}(-1)^i(1-p)^i \\
&= \sum_{i=0}^{\infty} \frac{(r+i-1)(r+i-2)\cdots r}{i!}(-1)^i(-1)^i(1-p)^i \\
&= \sum_{i=0}^{\infty} \binom{r+i-1}{i}(1-p)^i \\
&= \sum_{i=0}^{\infty} \binom{r+i-1}{r-1}(1-p)^i \\
&= \sum_{x=r}^{\infty} \binom{x-1}{r-1}(1-p)^{x-r} \text{ 。}
\end{aligned}
$$

故

$$
\sum_{x=r}^{\infty} P(X=x) = \sum_{x=r}^{\infty} \binom{x-1}{r-1}(1-p)^{x-r}p^r = p^{-r}p^r = 1 \text{ 。}
$$

即得證(8)式的確定義出一機率分佈。

在前述推導中,用到次方爲負之二項式定理,這是負二項分佈命名的由來。

如果以丟銅板爲例,二項分佈就是丟的次數固定,看其中出現幾次正面,而負二項分佈則是要求出現的正面數固定,看需要丟幾次。

若 X 有 $\mathcal{NB}(r,p)$ 分佈,則對 $\forall x = r, r+1, \cdots,$

$$
P(X > x) = P(U < r), \tag{11}
$$

其中 U 有 $\mathcal{B}(x,p)$ 分佈,此式成立是因 $X > x$,若且唯若前 x 次伯努力試驗中,成功次數小於 r。由(11)式得對 $\forall x = r, r+1, \cdots,$

$$
\begin{aligned}
(12)\ P(X \le x) &= 1 - P(X > x) \\
&= 1 - P(U < r) = 1 - P(U \le r-1) = P(U \ge r) \text{ 。}
\end{aligned}
$$

因此負二項分佈的分佈函數可經由某一二項分佈的分佈函數表示出。

前面指出隨機變數若有二項分佈,則可表示成一些i.i.d.伯努力隨機變數之和(見(2)式)。類似地,負二項分佈,也可表示成一些i.i.d.幾何隨機變數

之和。即若 X 有 $\mathcal{NB}(r,p)$ 分佈,則 X 可表成

$$X=\sum_{i=1}^{r}W_i, \tag{13}$$

其中 $W_1,W_2,\cdots,W_r$ 爲i.i.d.的 $\mathcal{G}e(p)$ 隨機變數。

(13)式之成立,由負二項分佈之產生方式立即可得知。因要得到 r 個正面,便要先得到第1個正面(丟了 W_1 次),再得第2個正面(丟了 W_2 次),餘類推。則X 便爲 $W_1,W_2,\cdots,W_r$ 之和,易見這些隨機變數皆獨立,且均有 $\mathcal{G}e(p)$ 分佈。

5 波松分佈

在二項分佈 $\mathcal{B}(n,p)$ 中,若 n 很大時會如何?顯然只要 $p>0$ 但固定,則成功次數很可能也會很大(注意,不論 n 多大,成功次數均可能會很小)。但若 n 很大且 p 很小,而 np 卻適當地大,此時成功次數會有什麼近似的分佈?我們有底下定理。

定理2.設隨機變數 X_n 有 $\mathcal{B}(n,p_n)$ 分佈,$n\geq 1$,且滿足$\lim_{n\to\infty}np_n=\lambda$,$0<\lambda<\infty$,則

$$P(X_n=k)\longrightarrow \frac{e^{-\lambda}\lambda^k}{k!},\ k=0,1,\cdots 。 \tag{14}$$

證明.首先有

$$\begin{aligned}P(X_n=k)&=\binom{n}{k}p_n^k(1-p_n)^{n-k}\\&=\frac{1}{k!}\frac{n(n-1)\cdots(n-k+1)}{n^k}(np_n)^k(1-p_n)^{n-k}。\end{aligned}$$

再由 $n\to\infty$ 時,

$$\frac{n(n-1)\cdots(n-k+1)}{n^k}=(1-\frac{1}{n})(1-\frac{2}{n})\cdots(1-\frac{k-1}{n})\longrightarrow 1,$$

$$(np_n)^k\longrightarrow\lambda^k,$$

且

$$(1-p_n)^{n-k}\longrightarrow e^{-\lambda},$$

使得證(14)成立。

註1. 此處用到若 $n \to \infty$ 時,$a_n \to 0$,且 $a_n b_n \to c$, 其中 $|c| < \infty$,則 $n \to \infty$ 時,$(1+a_n)^{b_n} \to e^c$。又 e 爲自然對數的底,其值約爲2.71828。

一隨機變數 X 若滿足

(15) $$P(X=k) = \frac{e^{-\lambda}\lambda^k}{k!},\ k=0,1,\cdots,$$

其中 $\lambda > 0$,便稱有參數 λ 之波松分佈(Poisson distribution),以$\mathcal{P}(\lambda)$ 表之。由於

$$\sum_{k=0}^{\infty}\frac{e^{-\lambda}\lambda^k}{k!} = e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} = e^{-\lambda}e^{\lambda} = 1,$$

故(15)的確定義出一p.d.f.。此處用到對 $\forall \lambda \in R$,

(16) $$e^{\lambda} = \sum_{k=0}^{\infty}\frac{\lambda^k}{k!},$$

爲微積分中的一重要結果。

例3. 設 X 有 $\mathcal{P}(2)$ 分佈,則

$$\begin{aligned} P(X=0) &= e^{-2} \doteq 0.1353, \\ P(X=1) &= 2e^{-2} \doteq 0.2707, \\ P(X=2) &= 4e^{-2}/2 \doteq 0.2707, \\ P(X=3) &= 8e^{-2}/6 \doteq 0.1804, \end{aligned}$$

而

$$P(X \geq 2) = 1 - P(X=0) - P(X=1) \doteq 0.594\text{。}$$

前面提過,對二項分佈 $\mathcal{B}(n,p)$,當 n 較大時,一般而言,其機率值並不好算。即使是數值表,由於有兩個參數,針對不同的 n 及 p,要列出很多表才行。但波松分佈只有一參數,且其p.d.f.之極大值發生於 $[\lambda]$(若 λ 不爲一正整數),或 $\lambda-1$ 及 λ (若 λ 爲一正整數),其中 $[\lambda]$ 表小於或等於 λ 之最大整

數。因此若 X 有 $\mathcal{P}(\lambda)$ 分佈,我們往往只需求 k 不太大時之 $P(X=k)$(因對較大的 k,此機率值很小),此時$k!$ 不難算,而 $e^{-\lambda}$ 及 λ^k 用計算機立刻可算出。要知次方的計算比二項式係數 $\binom{n}{k}$ 容易求,尤其是 n 較大時(一般桌上型計算機只能算到69!)。所以 n 很大時,以波松分佈之機率值來做爲二項分佈之機率值的近似值(即Poisson approximation to the binomial distribution,常簡稱爲波松近似),是有實際的便利。但所謂 np 要適當地大,多大才算適當呢?通常 $np \leq 7$時,波松分佈便對二項分佈提供一很好的近似值。當然所謂"很好"的標準因人而異,如Olkin et al.(1994)p333,便認爲 $np \leq 20$ 就可以了。

在給下例之前,令

$$b(k;n,p)=\binom{n}{k}p^k(1-p)^{n-k},\ k=0,1,\cdots,n,$$
$$p(k;\lambda)=e^{-\lambda}\lambda^k/k!,\ k=0,1,\cdots,$$

分別表 $\mathcal{B}(n,p)$ 及 $\mathcal{P}(\lambda)$ 之p.d.f.。

例4.對一有500員工之公司,令 X 表生日爲元旦之員工數,則便可以 $\mathcal{B}(500, 1/365)$ 當做 X 之分佈。表1顯示以參數 $\lambda=500/365$ 之波松分佈來求 X 之機率值還蠻精確。

k	0	1	2	3	4	5	6
$b(k;500,1/365)$	.2537	.3484	.2388	.1089	.0372	.0101	.0023
$p(k;500/365)$	.2541	.3481	.2385	.1089	.0373	.0102	.0023

表1.以波松分佈做爲二項分佈之近似

6 常態分佈

伯努力試驗是否只能引出離散型的分佈呢?持續地丟一出現正面機率爲 p 之銅板,則不論 p 多小,只要次數 n 夠大,會得到不少正面的機會都會很大。因此在上一節,當 np 不算小時,我們並不建議以$\mathcal{P}(np)$ 來扮演

$\mathcal{B}(n,p)$ 分佈之近似的角色。幸好萬能的上帝不會留下此缺憾的,常態分佈(normal distribution)因應而生。

隨機變數 X 若以下述 f 爲其p.d.f.,便稱爲有參數 μ 及 σ^2 之常態分佈,以 $\mathcal{N}(\mu,\sigma^2)$ 表之:

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-(x-\mu)^2/2\sigma^2},\ -\infty<x<\infty, \tag{17}$$

其中 $\mu\in R,\sigma>0$ 爲二常數。

欲證明(17)式所定義出之函數確爲一p.d.f.,即

$$\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sigma}e^{-(x-\mu)^2/2\sigma^2}dx=1, \tag{18}$$

要用到重積分的技巧,可查一般微積分的書。

我們有下述定理。

定理3. 設 X_n 有 $\mathcal{B}(n,p)$ 分佈,則對任二整數 $k\geq j\geq 0$,

$$\begin{aligned}&P(j\leq X_n\leq k)\\&=P(j-0.5\leq X_n\leq k+0.5)\\&=P(\frac{j-np-0.5}{\sqrt{np(1-p)}}\leq\frac{X_n-np}{\sqrt{np(1-p)}}\leq\frac{k-np+0.5}{\sqrt{np(1-p)}})\\&\doteq\Phi(\frac{k-np+0.5}{\sqrt{np(1-p)}})-\Phi(\frac{j-np-0.5}{\sqrt{np(1-p)}})。\end{aligned} \tag{19}$$

在此

$$\Phi(x)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-u^2/2}du,\ -\infty<x<\infty, \tag{20}$$

表 $\mathcal{N}(0,1)$ 分佈(即標準常態分佈, standard normal distribution)之分佈函數。至於$\mathcal{N}(0,1)$ 分佈之p.d.f.,有時以 ϕ 表之,即

$$\phi(x)=\frac{1}{\sqrt{2\pi}}e^{-x^2/2},\ -\infty<x<\infty。 \tag{21}$$

設隨機變數 X 有 $\mathcal{N}(\mu,\sigma^2)$ 分佈,則對任二實常數 a,b, $aX+b$ 有 $\mathcal{N}(a\mu+b,a^2\sigma^2)$ 分佈,此利用變數代換(change of variable)的技巧即可得

到。因此 $Y=(X-\mu)/\sigma$ 有 $\mathcal{N}(0,1)$ 分佈。故

$$(22)\qquad \begin{aligned} P(X\le x) &= P((X-\mu)/\sigma \le (x-\mu)/\sigma) \\ &= P(Y\le (x-\mu)/\sigma)=\Phi((x-\mu)/\sigma)\text{。} \end{aligned}$$

所以常態分佈有一其他分佈所沒有的特點,那就是只需 $\mathcal{N}(0,1)$ 分佈的數值表,而對任一 $\mathcal{N}(\mu,\sigma^2)$ 分佈,經由(22)式,便可得到其分佈函數。

定理3給出一種常態趨近(normal approximation),爲中央極限定裡(Central Limit Theorem)之一特例。此定理指出若 X_n 有 $\mathcal{B}(n,p)$ 分佈,則

$$Z_n=\frac{X_n-np}{\sqrt{np(1-p)}}$$

之分佈近似於 $\mathcal{N}(0,1)$ 分佈。而(19)中的 $k+0.5$ 及 $j-0.5$ 稱爲連續性的更正(continuity correction)。當 n 很大時,因0.5與 $\sqrt{np(1-p)}$ 相比很小,此項當然可以忽略。不過若 n 不是特別地大,有這一項會較精確。根據經驗法則(a rule of thumb),只要 np 及 $n(1-p)$ 皆大於5,則通常(19)便給出一很好的近似值。又在Olkin et al.(1994)p.405,認爲根據他們多年的經驗,當 $np(1-p)\ge 3$,此近似值便很好。例如,$p=1/2$ 時,$n\ge 12$;$p=0.2$ 時,$n\ge 19$。而前法之 n 分別爲10及25。只要 p 不要太小或太大,則 n 不需太大,便可以常態分佈來做爲二項分佈之近似分佈。值得注意的是,在二項分佈趨近至波松分佈裡,np 要不太大(≤ 7)才適用,但np 及 $n(1-p)$ 要不太小(≥ 5),常態趨近才適用。

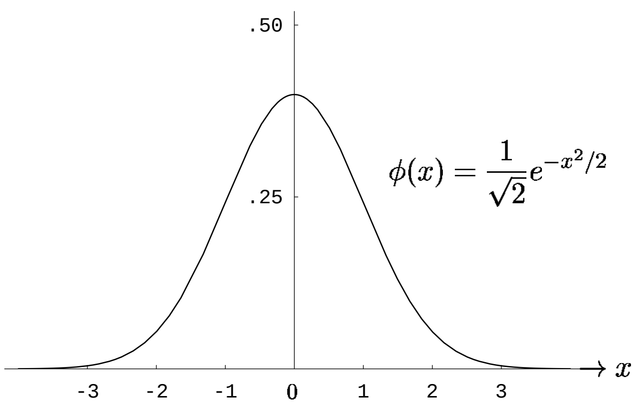

圖1. 標準常態分佈之機率密度函數

由圖1可看出標準常態分佈之p.d.f.爲一鐘形(bell-shaped curve), 對稱於直線$x = 0$, 即$\phi(-x) = \phi(x)$, 極大值發生在$x = 0$, 但極大值並不太大, 等於$1/\sqrt{2\pi}$, 約爲0.3989 。又也可看出對$\forall x \in R$, $P(X \leq x) = 1 - P(X \leq -x)$, 即

$$\Phi(-x) = 1 - \Phi(x), -\infty < x < \infty \text{ 。} \tag{23}$$

由圖2可看出, 即使是$\mathcal{B}(8, 0.5)$, 其p.d.f.與$\mathcal{N}(4, 2)$之p.d.f.已算吻合了 。雖n只是小小的8, 而$np = n(1-p) = 4$ 。

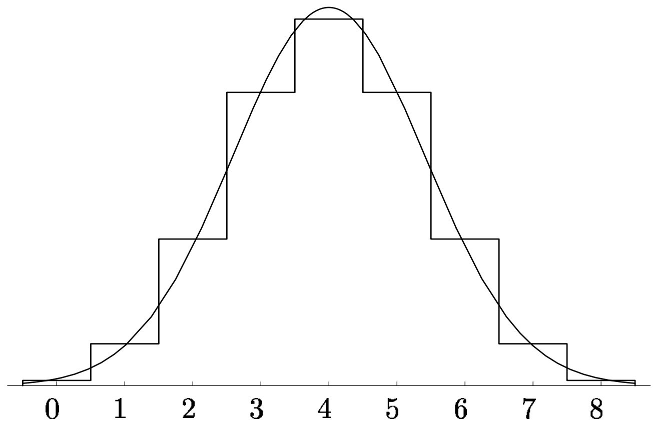

圖2. $\mathcal{B}(8, 0.5)$及$\mathcal{N}(4, 2)$之機率密度函數圖形

例5.某飲料公司確信其產品之市場佔有率爲10% 。某調查公司訪問2,500位消費者, 獲知其中有211位採用此公司之產品 。若10%之佔有率爲可信, 求調查結果會有不超過211位消費者採用該公司產品之機率 。

解. 令X表2,500位消費者中採用該公司產品之人數 。則X有$\mathcal{B}(2, 500, 0.1)$分佈, 且所求即

$$P(X \leq 211) = \sum_{i=0}^{211} \binom{2,500}{i} 0.1^i 0.9^{2,500-i} \text{ 。}$$

這當然是一不易知其大小的數字 。以常態趨近法, 因np=2,500·0.1=250,及$n(1-p)$=2,500·0.9=2,250均夠大,且$(np(1-p))^{1/2} = (2{,}500 \cdot 0.1 \cdot 0.9)^{1/2} =$

15,故

$$P(X \le 211) \doteq \Phi\left(\frac{211-250+0.5}{15}\right) \doteq \Phi(-2.567) \doteq 0.0051 \text{。}$$

由於此機率值是如此地小,所以要嘛是此調查得到一組不太尋常的樣本(假設 $p=0.1$ 還是正確的),要嘛是該公司的資訊可能有誤,即該公司之市場佔有率小於0.1,且可能較接近211/2,500$\doteq$ 0.0844。

7 其他相關分佈

如例1顯示,在伯努力試驗中, 若每次成功的機率不一樣, 則 n 次試驗後, 所得成功次數就不是二項分佈了。假設袋中有 N 個球,其中有 D 個白球, $N-D$個非白球,自其中隨機地取出 n 個球,每次取出後不放回。令 X 表總共取得之白球數,則

$$(24) P(X=k) = \frac{\binom{D}{k}\binom{N-D}{n-k}}{\binom{N}{n}},\ \max\{0, n-N+D\} \le k \le \min\{n, D\} \text{。}$$

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 $n-k$ 當然要不超過全部之非白球數 $N-D$,而 $k \ge 0$ 又要成立,因此 $k \ge \max\{0, n-N+D\}$。

(24)式便定義出一超幾何分佈(hypergeometric distribution),以 $\mathcal{H}(N, D, n)$ 表之,有三個參數。此分佈在品質管制(quality control)的討論裡常出現。

例6.設有 $N=200$ 個零件,其中有 $D=4$ 個不良品(不良率爲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

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

同理可求出

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

此兩機率值可說非常地小。

例6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。在漢朝劉向撰,東漢高誘註的戰國策第四篇(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3個使用。若其中壞了一個尚可忍受,因機率約爲0.0582(見例6)。若壞了兩個,可能便要找公司退貨了(因機率小於千分之一,約爲0.000895)。若3個皆壞,大概便不相信不良率才2%而已。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2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的機率很低之情況中,顯現其影響力。戰國策第二十三篇(魏策二),另有一則“三人市虎”的典故,也是相似的道理:夫市之無虎明矣,然而三人言而成虎。要知有時是可以偏概全的。

銅板之兩個面,若推廣到 k 個面會如何,骰子是 $k=6$。假設有一 k 多面體,每個面出現的機率分別爲 $p_1, p_2, \cdots, p_k$,其中 $p_1+p_2+\cdots+p_k=1$,且 $0 \le p_i \le 1$, $\forall i=1,2,\cdots,k$。隨機地丟 n 次後,設各面分別出現

$X_1, X_2, \cdots, X_k$ 次,則可得

$$P(X_1 = n_1, \cdots, X_k = n_k) = \binom{n}{n_1, \cdots, n_k} p_1^{n_1} \cdots p_k^{n_k}, \quad 0 \le n_i \le n, n_1 + n_2 + \cdots + n_k = n \text{。} \tag{25}$$

其中多項式係數(multinomial coefficient)

$$\binom{n}{n_1, \cdots, n_k} = \frac{n!}{n_1! \cdots n_k!} \text{。} \tag{26}$$

滿足(25)之 k 維的隨機變數,便稱有多項分佈(multinomial distribution),參數爲 $n, p_1, \cdots, p_k$,以 $\mathcal{M}(n, p_1, \cdots, p_k)$ 表之。

例7.丟一均勻的骰子12次,則每個面均出現兩次的機率爲

$$\binom{12}{2, \cdots, 2} (\frac{1}{6})^{12} = \frac{12!}{(2!)^6} (\frac{1}{6})^{12} \doteq 0.00344 \text{。}$$

可看出丟12次,每面出現的次數均相等並不容易。換句話說,各面出現的次數參差不等,反而是常態(機率約爲0.99656)。這是違反一般人的直觀。

8 結語

由丟銅板出發,我們引出了一連串重要的分佈。尤其連在分佈中居於獨尊地位的常態分佈都出現了,銅板角色之重要可見。難怪清軍進攻台灣前要丟銅板,難怪在桃色交易(Indecent Proposal,由勞勃瑞福(Robert Redford)及黛咪摩兒(Demi Moore)主演)那部電影裡,銅板的出現那一面,可抵百萬美元的身價。

波松及常態分佈,分別是離散型及連續型中極重要的分佈,此二分佈彷如交通樞紐,可延伸至不少分佈。一個簡單的伯努力試驗,在整個機率學的發展中,影響卻是深遠無比。只由一個銅板便可引出一機率世界。

重要的分佈中,以人名命名的並不多。波松(Poisson,1781-1840)爲法國數學家及物理學家,西元1837年出版"司法機率的研究"(Recherches sur la probabilité des jugements)一書。在該書中,波松由二項分佈的極限,得到

此一日後以他命名的機率分佈。波松雖得到此一分佈,但他並未繼續探討此分佈的其他性質。西元1898年,波蘭數學家Bortkewitz在他寫的Das Gesetz der kleinen Zahlen(小數法則,Law of Small Numbers)一書中,描述了波松分佈的一些應用。他給了幾個自殺及意外死亡的數據,均吻合波松分佈。波松本人雖然推導出波松分佈。但將此分佈發揚光大的卻是Bortkewitz。

二項分佈趨近至常態分佈,是機率論中第一個關於極限的定理。由於早期沒有計算的機器,爲了處理諸如銅板丟很多次後,共得之正面數的分佈情況,幾位著名的機率學家,如De Moivre(1667-1754)及Laplace(1749-1827),相繼投入這方面的探討。由今日的眼光來看,他們的結果不過是中央極限定理的特例。利用特徵函數(characteristic function),中央極限定理可簡單地被證出,並且是對一般的i.i.d.的隨機變數,見黃文璋(1994)p.207。而De Moivre 及Laplace之證法,可說是極冗長的,可參考黃文璋(1994)pp.242-248,或蔡聰明(1995)。讀者不妨去翻閱他們的證法(目前見到的證法當然已被修飾過),當可體會在缺乏近代有效的工具下,他們洞察先機的能力,計算細膩的功夫,是令人佩服的。雖然他們開創性的貢獻很大,但常態分佈並未以他們命名,即使更一般的中央極限定理(這是西元1920年, Pólya (1887-1985)所取的),也沒有以人命名。波松可說是一位幸運的學者。

最後貫穿本文的伯努力,指的是瑞士數學家James Bernoulli(1645-1705)。他是Bernoulli數學大家族(在十七、十八世紀三代間出了八位數學家,其中五位對機率有貢獻)中,最著名的三位之一,另二位爲他弟弟John Bernoulli (1667-1748) 及John 的次子Daniel Bernoulli (1700-1782)。James 及John 爲英文名,在德文中分別爲Jakob 及Joham, 而在法文中分別爲Jacques 及Jean 。不同的作者有時會採不同的型式的名,其實爲同一人。此家族之事蹟可參考蘇淳、劉鈍譯(1991)。Keynes (1921) (凱因思,1883-1946,著名的英國經濟學家),稱(James) Bernoulli 爲數理機率的眞正創立者(The real founder of mathematical probability) 。Bernoulli 死後,西元1713年,他在機率論方面的開創性結果: Ars Conjectandi (The Art of Conjectaring ,猜測的藝術)才告出版。伯努力大數法則(Bernoulli Law of Large Numbers) 亦包含於此書中。在數理統計及機率論方面,有一活躍的國際性組織,就叫Bernoulli Society for Mathematical Statistics and Probability ,簡稱Bernoulli Society。此家族之爲機率統計界所肯定由此可見。

習 題

1. 試說明離散型隨機變數之分佈函數爲一階梯函數。

2. 設A, B 二人下9盤棋,每盤 A,B 贏之機率各爲0.6及0.4。試求 A 贏較多盤之機率。

3. 一婦人決定生到第2個男孩便停止,又假設生男生女之機率均爲0.5,且每次相互獨立。求生孩子的總數大於5之機率。

4. 某盒20個裝之零件中有4件不良品,隨機取5件(每次取出後不放回),若不良品數不超過1件,則該盒零件視爲檢驗合格。試求檢驗合格之機率。

5. 將 $2r$ 個球隨機地放進 r 個盒中,令 X_i 表第 i 個盒子中之球數。試求每個盒子中恰有二球之機率。

6. 設 A, B 二支球隊進行7戰4勝的比賽,每場球 A 勝之機率設爲0.6。
(i)分別求比賽結果爲 A 4比0,4比1,4比2,及4比3勝之機率;
(ii)分別求 A 及 B 贏得比賽之機率;
(iii)分別求比賽在第4場,第5場,第6場,及第7場結束之機率。

7. 設隨機變數 X 有 $\mathcal{P}(\lambda)$ 分佈,其p.d.f.以 $f(x)$ 表之。試證 $f(x) = (\lambda/x)f(x-1)$,$\forall x \geq 1$,並討論 f 之極大值發生處。

8. 對固定的 n 及 p ,試證若 $0 \leq k < (n+1)p$, $b(k;n,p)$ 爲增函數,若 $(n+1)p < k \leq n$, $b(k;n,p)$ 爲減函數,並求 $b(k;n,p)$ 之極大值發生處。

9. 試繪 $\mathcal{N}(0,1)$ 分佈之p.d.f.圖形,並指出其漸近線、極值及反曲點(reflection points)。

10. 假設每張彩券中獎之機率爲0.01。試估計200張彩券中,

(i)皆未中獎之機率;

(ii)至少有4張中獎之機率。

11. 某校有學生730人。試估計其中

(i)恰有3位學生生日爲教師節之機率;

(ii)至少有3位學生生日爲教師節之機率。

12. 設每頁書會有打字錯誤之機率約爲0.02,且各頁書是否有錯誤爲獨立。對一本500頁的書,求有錯誤之頁數少於2頁之機率。又以波松近似來估計此機率。

13. 投擲一公正的骰子12,000次。

(i)試估計點數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,你是否會懷疑此非一公正的骰子?試說明之。

參考文獻

1. 二月河(1994).康熙大帝一玉宇呈祥<下>。巴比倫出版社,台北。

2. 金庸(1996).鹿鼎記,第三版。遠流出版社,台北。

3. 黃文璋(1994).機率論講義。全華科技圖書股份有限公司,台北。

4. 蔡聰明(1995).什麼是機率與機率法則?數學傳播季刊,第19卷第1期,50-63。

5. 蘇淳、劉鈍譯(1991). Bernoulli們:一個學者家族。數學傳播季刊,第15卷第1期,47-56。

6. Keynes, J.M.(1921). *A Treatise on Probability.* Macmillan Company, London.

7. Olkin, I., Gleser, L.J. and Derman, C.(1994). *Probability Models and Applications*, 2nd. ed. Macmillan College Publishing Company, New York.