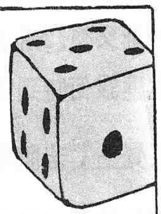

機率專題

淺論機率上的幾個極限定理

—洋 洋—

0.開場白 本文的主要目的是簡單地介紹機率上幾個常見極限定理的基本形式。由於篇幅及工具不足的限制，這些定理非但無法在此證明，就是定理的敍述恐怕也有許多不夠清楚的地方，更不用提它們的一般形式了。作者只希望本文的一些例子能夠引起讀者的興趣去進一步追究在許多看來完全隨意(random)的現象後面的規律性。

本文所涉及的一些定理在一般的機率教科書上都可以找到，譬如〔1〕—〔4〕。其中特別值得一提的是〔3〕，學習它只需要高中數學的程度，而且該書包含許多例子來幫助了解。事實上本文許多例子都來自〔3〕。

本文係在匆促間完成，一定有許多失誤之處。敬祈讀者們原諒及指正。

1.基本概念 機率空間具有三個要素：$(\Omega, \mathcal{F}, P)$，其中Ω是一個集合，稱做樣本空間；$\mathcal{F}$是一些事件的集合，它是一個非空的σ域，具有下列性質：

$F1.$ 對每一$A\in\mathcal{F}$，$A\subseteq\Omega$

$F2.$ $\phi\in\mathcal{F}$，$\Omega\in\mathcal{F}$

$F3.$ 若$A\in\mathcal{F}$則$A^c=\Omega-A\in\mathcal{F}$

$F4.$ 若$A_i\in\mathcal{F}$，$i=1, 2, \cdots$，

則 $\bigcup\limits_1^\infty A_i\in\mathcal{F}$

所以若A，B是兩事件，則$A\cap B$，$A\cup B$亦是事件。條件4或許不容易被立刻接受，高中課本上所討論的機率理論多半限於離散空間，在那裡我們只會遇到$\bigcup\limits_1^n A_i$，而不會遇到需無窮次運算的$\bigcup\limits_1^\infty A_i$。

機率空間的最後一要素P是Probability的簡寫，它是定義在$(\Omega, \mathcal{F})$上的一個非負函數，具有下列性質：

$P1.$ 對每一$A\in\mathcal{F}$，$P(A)\geqslant 0$，且$P(\phi)=0$，$P(\Omega)=1$。

$P2.$ 若$A_i\in\mathcal{F}$，$i=1, 2, \cdots$，彼此不相交，則

$$P\left(\bigcup_1^\infty A_i\right)=\sum_1^\infty P(A_i)。$$

由是我們可以推出

$$P(A^c)=1-P(A)$$
$$P(A\cup B)=P(A)+P(B)-P(A\cap B)$$

等公式。用白話來說$P(A)$就是事件A發生的機率，所以若$P(A)=1$我們可以說事件A一定發生（嚴格來說應是almost surely發生）。譬如說，若X代表投擲一公平骰子時的點數，則

$$P(X=1)=\frac{1}{6}$$

這表示投擲的結果是1，2，3，4，5或6的機會均等，換句話說，平均起來在6次投擲骰子中有一次會出現1點。注意這並不意味6次投擲骰子中就恰有一次是1點，或者600次投擲中就恰有100次出現一點。請參見第二節的大數法則。

在正式介紹大數法則的等定理之前，我們還需要一些專門術語。

我們稱X是Ω上的一個隨機變數(random variable)就是說X是$\Omega\to R$的一個好函數（一般來說X的函數值不需是實數）。所謂“好”係指在理論上我們能計算與X相關的一些機率，譬如

$$P\{w: X(w)\in[a,b]\},\quad a,b\in R \tag{1.1}$$

用術語來說X是$\mathfrak{F}$可測，即對任意實數a，b $\{w: X(w)\in[a,b]\}\in\mathfrak{F}$。特別來說

$$F(a)=P\{w: X(w)\leqslant a\} \tag{1.2}$$

稱為X的分佈函數（distribution function）。注意知道(1.1)與知道(1.2)是同一回事。

隨機變數X的期望值常用μ表示。依定義

$$\mu=EX=\int_{-\infty}^{\infty} x\,dF(x)$$

當X的值域只有可數多點a_1，a_2，…，上式可化簡成

$$\mu=EX=\sum_{i=1}^{\infty} a_i P(w: X(w)=a_i\}$$

譬如說X代表投擲一公正骰子所得的點數時，

$$\mu=\frac{1}{6}(1+2+3+4+5+6)$$
$$=3.5$$

我們稱隨機變數X_1，X_2，…，X_n是彼此獨立（mutually independent）如果對任意R上的區間A_1，A_2，…，A_n，

$$P(X_1\in A_1, X_2\in A_2,\cdots,X_n\in A_n)$$
$$=\prod_{1}^{n} P(X_i\in A_i)$$

注意上式中$P(X_i\in A_i)$是

$$P\{w: X_i(w)\in A_i\}$$

的簡寫。以後我們會經常採用此種簡寫。

我們稱隨機變數X_1，X_2，…，X_n是i.i.d.（independent and identically distrbuted）如果X_1，X_2，…，X_n除了彼此獨立外還有相同的分佈函數。統計上通稱X_1，X_2，…，X_n是依照F的隨機取樣（random sample），其中F是X_i們的共同分佈函數。

2.大數法則

假設隨機變數X_1，X_2，X_3，…，是i.i.d.，即彼此獨立而且有相同的分佈函數。若期望值$\mu\equiv EX_1$存在，且$S_n=\sum_{k=1}^{n} X_k$是X_k的部分和，則

a.（弱大數法則）$\frac{S_n}{n}$ converges in probability到μ，即對任意$\varepsilon>0$，

$$P\left(\left|\frac{S_n}{n}-\mu\right|>\varepsilon\right)\xrightarrow{n}0 \tag{2.1}$$

b.（強大數法則）$\frac{S_n}{n}$ converges almost surely 到μ，即

$$P\left(\frac{S_n}{n}\xrightarrow{n}\mu\right)=1 \tag{2.2}$$

由強弱的命名即知(1.1)是可由(1.2)導出。底下我們將用一些例子來說明大數法則。

例2.1　假設在一很大的賭場裡同時有許多賭局各自在進行。每一賭局的甲、乙兩方不停地投擲一公正的銅板，每次銅板出現正面時甲方即贏乙方 1 元，否則甲方付給乙方 1 元。換句話說，若 X_k 表示(固定一賭局)甲方在第 k 次投擲銅板的所得，則 X_k, $k=1, 2, \cdots$，是 i.i.d.，且 $P(X_k=\pm 1)=\frac{1}{2}$，$EX_k=0$。於是

$$\frac{S_n}{n}=\frac{\sum_{k=1}^{n} X_k}{n}$$

代表甲方在 n 次投擲後的平均所得。根據極限的定義，若取 $\varepsilon=0.01$，(2.1)式告訴我們存在一正數 N 使得 $n \geqslant N$ 時

$$P\left(\left|\frac{S_n}{n}-0\right|>0.01\right) \leqslant 0.01$$

換句話說，平均來說在 100 個賭局中，至多有一個賭局裡的甲方在 n 次投擲後的平均所得超出 0.01 元。但是這個可能存在的例外 1 局一般而言，會隨著 n 變動。

(2.2)式則表示(幾乎)在每一賭局裡，甲方的平均所得會漸趨於 0：

$$P\left\{w: \frac{S_n(w)}{n} \to 0\right\}=1$$

或者

$$P\left[w：對任意 \varepsilon>0，存在 N(w) 使得 n \geqslant N(w) 時，\left|\frac{S_n(w)}{n}\right| \leqslant \varepsilon 成立\right)=1$$

注意：N 是 w 的函數。說的更清楚一點，若取 $\varepsilon=0.01$，則在第一賭局時保證平均所得不超過 0.01 所需的投擲次數 N 與第二賭局裡相對應的投擲次數 N 不一定會相同。

例2.2　柏努尼試驗(Bernoulli trial)假設某甲以擲銅板的方式與某乙賭博，每次擲銅板前，某甲付賭資 a 元，若銅板出現正面則某甲可獲 1 元，否則某甲得不到錢。假設銅板出現正面的概率是 P，就一個賭徒來說，某甲最關心的事可能是這種賭博方法是否公平，並且在 n 次賭博後，他可能淨贏多少錢？

上述問題可用數學符號表示如下：令 X_k 代表某甲在第 k 次賭博中的所得，則

$$P(X_k=1)=p$$
$$P(X_k=0)=1-p$$
$$k=1, 2, \cdots\cdots$$

即 X_k，$k \geqslant 1$ 有相同的分佈函數。由於銅板的投擲與其前及其後的投擲無關，我們可以假設 X_k，$k \geqslant 1$，是彼此獨立的。令

$$S_n=\sum_{k=1}^{n} X_k$$

則 n 次賭博後某甲所淨贏的錢數可表成 S_n-na。若一個賭法是公平，那麼最低程度我們期望賭博的平均所得與每次所付賭資應相等，換句話說，我們期望 $\frac{S_n}{n} \approx a$。因爲

$$EX_1=1 \cdot P(X_1=1)+0 \cdot P(X_1=0)$$
$$=p$$

由(2.2)式

$$P\left(\lim_n \frac{S_n-na}{n}=p-a\right)=1$$

故唯有 $a=p$ 時才可能是公平賭博，此時根據極限之定義可知，幾乎在所有的賭局中，若賭博的次數 n 很大時，某甲所能淨贏的錢數 S_n-na 介於 $-n\varepsilon$ 與 $n\varepsilon$ 之間。

例2.3　正規數(normal number)在這個例子我們將確切表出 Ω，同時希望讀者能比較了解 almost surely 的意義。

設 x 是[0, 1]中的一實數，且令

$$x=0.a_1 a_2 a_3 \cdots\cdots$$

是它的十進位小數展開，其中每一 a_i 都是

0, 1, ……到 9 的一數位。如果我們採用無窮多個 9 來表有限小數（$\frac{a}{10^n}$ 的形式，a 是整數），則此種展開對每一 $0 \leqslant x < 1$ 是唯一的。譬如 $x = 0.12$ 的小數展開是 $x = 0.11999\cdots\cdots$。

令 $\Omega = \{x : 0 \leqslant x < 1\}$，$\boldsymbol{F}$ 是由所有區間 (a, b)，$0 \leqslant a < b < 1$，所生成的 σ 區（我們不在此處詳細說明），機率函數就是一般的長度函數

$$P\{x : x \in (a, b)\} = b - a \quad 0 \leqslant a < b < 1$$

定義：

$$X_k(x) = \begin{cases} 1, & \text{若 } x \text{ 的小數展開第 } k \text{ 位 } a_k \text{ 是 } 0 \\ 0, & \text{若 } x \text{ 的小數展開第 } k \text{ 位 } a_k \text{ 不是 } 0 \end{cases}$$

則不難證明 X_k，$k \geqslant 1$ 是 Ω 上的隨機變數，且他仍是 i.i.d.，

$$P(X_k = 1) = 0.1,$$

$$P(X_k = 0) = 0.9, \quad EX_1 = 0.1$$

於是 $\frac{S_n(x)}{n}$ 代表在 x 的小數展開的前 n 位裡數位 0 出現的頻率。根據強大數法則(2.2)

$$P\left\{x : \lim \frac{S_n(x)}{n} = 0.1\right\} = 1 \tag{2.3}$$

換句話說觀察自〔0, 1）任意排取的一實數 x 的小數展開時，我們“幾乎”都會發現到數位 0 出現的頻率是 1/10。

下面的事實不難證明：若

$$P(A_i) = 1, \quad i = 0, 1, \cdots, 9,$$

則 $$P\left(\bigcap_1^9 A_i\right) = 1。$$

若令 $A_i = \{x :$ 數位 i 在 x 的小數展開中出現的頻率是 $0.1\}$，而且定義實數 x 是**正規數**，如果在它的小數展開中每一數位出現的頻率是 $\frac{1}{10}$，則我們可結論：幾乎所有介於 0 到 1 的實數都是正規數。

同樣的方法加上一點小技巧可以證明

$$P\{x \in [0, 1) : \text{數字 } 12 \text{ 在 } x \text{ 的小數展開裡出現的頻率是 } \frac{1}{10^2}\} = 1 \tag{2.4}$$

$$P\{x \in [0, 1] : \text{數字} 4895 \text{在 } x \text{ 的小數展開裡出現的頻率是 } \frac{1}{10^4}\} = 1 \tag{2.5}$$

等等。譬如說，令

$$X_k(x) = \begin{cases} 1, & \text{如果 } x \text{ 的小數展開的 } a_k = 1,\ a_k = 2 \\ 0, & \text{其他情形} \end{cases}$$

則易見 X_k 是有相同的分佈函數，

$$P(X_k = 1) = \frac{1}{10^2}, \quad EX_k = \frac{1}{10^2}。$$

雖然 X_k，$k \geqslant 1$ 不再是彼此獨立，但 $\{X_{2k}, k \geqslant 1\}$ 是彼此獨立，$\{X_{2k-1}, k \geqslant 1\}$ 也是彼此獨立。根據強大數法則“幾乎”所有的 $x \in [0, 1]$ 都滿足

$$\lim_n \frac{\sum_{k=1}^n X_{2k}}{n} = \frac{1}{10^2},$$

$$\lim_n \frac{\sum_{k=1}^n X_{2k-1}}{n} = \frac{1}{10^2}$$

兩者相加即可得證(2.4)

$$\lim_n \frac{\sum_{k=1}^n X_k}{n} = \frac{1}{10^2}$$

上面的結果很明顯地不限於十進位小數展開。讓我們看看：另一個類似問題。假設打字機鍵盤上所有符號，包含 26 個字母，阿拉伯數字，空白鍵，標點符號，……等等共 50 個。假設許多猴子各自坐在打字機前不停地隨意敲打鍵盤，每次敲出任一符號的機會都是 $\frac{1}{50}$，莎士比亞的全集雖然很長，

畢竟是由有限個符號（包含空白符號在內），假設是M個所組成，於是由強大數法則

P {莎士比亞全集出現的頻率是 $(50)^{-M}$ } $=1$

因此"幾乎"每一隻猴子都一定可敲出莎士比亞全集（如果猴子有耐性而且活得長的話）！

在(2.3)－(2.5)的解釋裡我們一再用到"幾乎"兩個字，這表示(2.3)－(2.5)並非對所有 $x\in[0,1)$ 成立。譬如說，所有有限小數都不滿足(2.3)－(2.5)，因為它們的十進位展開差不多都是數位9。如果我們進一步要求正規數滿足所有類似(2.4)、(2.5)的公式，那麼任何一個有理數都不是正規數，由此可推論

$$P\{x\in[0,1]: x \text{是有理數}\}=0$$

這個事實也不難理解，如果我們事先知道

①$Q=$ {有理數} 是可數，即 Q 與正整數集合 $\{1, 2, \cdots\}$ 有一一對應的關係。

②若集合 A_i 都是可數，則 $\bigcup_1^\infty A_i$ 亦是可數

③$[x: 0\leqslant x<1]$ 是不可數。

根據可數之定義不難證明②。至於③，可用對角化的程序來證明，我仍不在此討論。

雖然幾乎所有的數都是正規數，到目前為止尚無人能夠證明一些常見的無理數如 $\sqrt{2}$，$\sqrt{3}$， π，e 等是否為正規數。

例2.4 科學測量：科學上經常要做一些測量，譬如測量長度、溫度等等。由於環境的微小變異，儀器的靈敏度、操作的程序總總因素，每次測量所得數值或多或少有一些誤差。消除這些誤差的一個常用方式是多做幾次測量：假設 $Y_1, Y_2, \cdots, Y_n$ 是 n 次測量（某物體溫度）所得值。然後取這些數值的平均值 $\left(\sum_{k=1}^n Y_k\right)/n$ 來做為該物體的溫度。這種作法的可行性其實是用到大數法則。每次測量數值 Y_k 可分成兩部分

$$Y_k=\theta+X_k$$

其中 θ 是物體的眞正溫度，X_k 則是誤差。由於前後測量之不相關及 X_k 為正為負的可能性均等，我們不妨假設 X_k，$k\geqslant 1$ 是 i.i.d.隨機變數且 $EX_k=0$。因此由(2.2)，almost surely

$$\lim\frac{\sum_1^n X_k}{n}=0$$

於是

$$\frac{\sum_1^n Y_k}{n}=\theta+\frac{\sum_1^n X_k}{n}\approx\theta$$

當 n 相當大時。

例2.5 Stone-Weirstrass 定理：這是分析上一個很有名的定理。它告訴我們連續函數可用多項式來逼近。說的更清楚一點，若 f 是定義在 $[0,1]$ 閉區間上的一個連續函數，則對任意 $\varepsilon>0$ 我們可以找到一多項式 $P(x)$（$P(x)$ 當然與 f 及 ε 有關）使得

$$\max_{x\in[0,1]}|P(x)-f(x)|<\varepsilon$$

大數法則的一個廣為人知的應用就是為上述定理提供了一個建設性的證明（Constructive proof）：定義Bernstein 多項式如下：

$$p_n(x)=\sum_{k=0}^n f\left(\frac{k}{n}\right)\binom{n}{k}x^k(1-x)^k$$

則

$$\lim_n\left(\max_{x\in[0,1]}|P_n(x)-f(x)|\right)=0 \qquad (2.6)$$

上式之證明大致如下：首先固定 $x \in [0, 1]$，令 X_k，$k \geqslant 1$，是 i. i. d. 隨機變數，且

$$P(X_1 = 1) = x$$
$$P(X_1 = 0) = 1 - x$$

於是

$$E\{f(\frac{S_n}{n})\}$$
$$= \sum_{k=0}^{n} f(\frac{k}{n}) \cdot P(S_n = k)$$
$$= \sum_{k=0}^{n} f(\frac{k}{n}) \binom{n}{k} x^k (1-x)^{n-k}$$
$$= p_n(x)$$

根據 (2.2)，$P(\frac{S_n}{n} \xrightarrow{n} EX_1 = x) = 1$，由於 f 是連續 $P(f(\frac{S_n}{n}) \to f(x)) = 1$，於是不難相信（可用測度論的有界收斂定理）逐點收斂成立

$$p_n(x) = E\{f(\frac{S_n}{n})\} \to f(x)$$

至於代表均勻收斂的 (2.6) 還需一點工夫才能證明。我們不再討論。有興趣的讀者可參考〔1〕〔2〕及〔4〕。

3. 中央極限定理（Central Limit Theorem）：

弱大數法則有些地方不盡完善，因爲對任意事先選定的正數 ε，不論它是多小，

$$P(|\frac{S_n - n\mu}{n}| > \varepsilon) \xrightarrow{n} 0。$$

這表示用 n 來除（$S_n - n\mu$）是稍微大了些。那麼是否可以找到一個比 n 小的尺度，用它來除（$S_n - n\mu$）時我們可以得到一些有意思的結果呢？中央極限定理是這個方向的重要結果。

設 $X_1, X_2, \cdots$ 是 i. i. d.，且 $\mu = EX_1$ 及變異數 $\sigma^2 = E(X_1 - \mu)^2$ 皆存在，則對任意實數 x，

$$F_n(x) \equiv P(\frac{S_n - n\mu}{\sqrt{n}\,\sigma} \leqslant x)$$
$$\to \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt。 \tag{3.1}$$

上式的 Φ 就是有名的標準正規分佈函數（Standard normal distribution），或稱 Gauss 分佈函數，因爲 Gauss 在研究天文測量的誤差問題時發現了它 (*)。Φ 的期望值 $\mu = 0$，變異數 $\sigma^2 = 1$：

$$\int_{-\infty}^{\infty} x\,\Phi'(x)\,dx = 0,$$
$$\int_{-\infty}^{\infty} x^2\,\Phi'(x)\,dx = 1。$$

由於 Φ 在統計上的重要性，Φ 的數值有表可查。譬如

$$\Phi(2.326) = 0.990,$$
$$\Phi(3) = 0.999。 \tag{3.2}$$

例3.1 雖然 (3.1) 式是一極限，在實際計算 $F_n(x)$ 時，我們常用 $\Phi(x)$ 來取代

$$F_n(x) \approx \Phi(x) \tag{3.3}$$

或者

$$P(y\sqrt{n}\,\sigma \leqslant S_n - n\mu \leqslant x\sqrt{n}\,\sigma)$$
$$\approx \Phi(x) - \Phi(y)$$

許多時候這種取代所導致的誤差相當小。設例 2.2 的 Bernoulli trial 的 $P = 1/2$，$n = 200$，我們想要計算 $P(95 \leqslant S_{100} \leqslant 105)$ 即在 200 次投擲一公正銅板當中，正面出現的次數介於 95 到 105 次的概率。顯然地，

$$P(95 \leqslant S_{200} \leqslant 105)$$
$$= \sum_{k=95}^{105} \binom{200}{k} 2^{-200}$$
$$(= 0.56325 \cdots\cdots)$$

但右式並不容易計算。利用(3.3)式，因

$$\mu = \frac{1}{2}, \quad \sigma = \sqrt{E\left(X_1 - \frac{1}{2}\right)^2} = \frac{1}{2},$$
$$P(95 \leqslant S_{200} \leqslant 105)$$
$$= P\left(\frac{-5}{\sqrt{50}} \leqslant \frac{S_{200} - 100}{\sqrt{200 \cdot 2^{-1}}} \leqslant \frac{5}{\sqrt{50}}\right)$$
$$\approx \Phi\left(\frac{1}{\sqrt{2}}\right) - \Phi\left(-\frac{1}{\sqrt{2}}\right)$$
$$\approx 2\,\Phi\left(\frac{1}{1.414}\right) - 1$$
$$= 0.56331$$

與正確值0.56325……的誤差不到0.02%。

讓我們再看另一個情形。當 $p = 0.1$，$n = 500$，

$$P(45 \leqslant S_{500} \leqslant 55)$$
$$\approx \Phi\left(\frac{5}{\sqrt{45}}\right) - \Phi\left(\frac{-5}{\sqrt{45}}\right)$$
$$\approx 0.3235 \cdots\cdots$$

與正確值0.3176…… 的誤差約2%。

注意：第一個情形誤差較小的理由是 $p = \frac{1}{2}$，故 S_n 的分佈是對稱的。

例3.2 保險問題：假如某甲（或某機構）只有一棟房子，那麼毫無疑問地每個人都會建議某甲投保火險，以確保他的財產安全。如果某甲有 1000 棟房子，那麼某甲是否仍該爲這 1000 棟房子全部投保呢？

讓我們假設房子是分散各處。因此，某一棟房子遭受火災與否與其他房子無關。令 $X_k = 1$ 或0表示第 k 棟房子在今年遭受或未遭受火災，於是

$$S_{1000} = \sum_{1}^{1000} X_k$$

就代表今年某甲房子遭到火災的總數。根據歷年來的統計資料我們可以估計每一棟房子在一年內遭到火災的機率，爲方便計我們假設每棟房子遭火災的機率都是 p

$$P(X_k = 1) = p$$
$$P(X_k = 0) = 1 - P$$

於是 X_k，$k \geqslant 1$，i. i. d. 的Bernoulli trial，根據中央極限定理我們可以估計 S_{1000} 的分佈情況。

以上說法也適用於保險公司。唯一的差別是向它投保火險的房子總數更大，也因此(3.3)更加適用。現在讓我們來看看保險費是如何計算的。大致上，保險費是保險公司利用 $\Phi(x)$ 估算它每年將賠償客戶的總數及保險公司所賺利潤的總和。

由於在一般情況中央極限定理已能應用於 S_{1000}，我們似乎可以建議某甲不投保火險，從而省下將被保險公司賺去的利潤。從這裡我們也可了解爲何當某甲只有一棟房子時他該投保。蓋此時數目太小，中央極限定理無法派上用場。

例3.3 假設之檢定（Testing Hypothesis）：在科學上我們經常會做一些假設（hypothesis）來解釋一些現象。由此可知用實驗數據（data）來檢定，假設正確與否是一件很重要的事。

假設某人在900次投擲一銅板中記錄到496次正面，究竟某人可否視此銅板公正？

假設是公正，應用中央極限定理於例2.2的 $p = \frac{1}{2}$，$n = 900$ 情形（$\mu = \frac{1}{2}$，$\sigma = \frac{1}{2}$）。

$$P(405 \leqslant S_{900} \leqslant 495)$$

$$= P\left(\frac{-45}{15} \leqslant \frac{S_{900}-450}{\sqrt{900}\,/\,2} \leqslant \frac{45}{15}\right)$$
$$\doteq \Phi(3)-\Phi(-3)$$
$$= 2\,\Phi(3)-1$$
$$= 0.998$$

換句話說在 1000 次的類似情形裡，正面出現次數高過 496 的次數約只有 2 次。因此，我們大致可認為此銅板不公正。

例3·4　**座位問題**：假設甲、乙公司每天都各自有一輛設備完全相同的火車同時離開台北前往高雄，假設 n 位旅客各自任意地選搭甲車或乙車，因此搭乘每一輛火車的人數是 $p=\frac{1}{2}$ 的 n 次 Bernoulli trial的結果。現假設甲車（或乙車）只有 $S<n$ 個座位，因超 S 位旅客搭乘甲車或乙車的機率 $f(s)$ 是正數此時就有旅客無位可坐。根據（3·3）式

$$f(s)=P(S_n>s)=P\left(\frac{S_n-\frac{n}{2}}{\sqrt{n}\cdot\frac{1}{2}}>\frac{2s-n}{\sqrt{n}}\right)$$
$$\approx\Phi\left(\frac{2s-n}{\sqrt{n}}\right)$$

如果 s 相當大而使得 $f(s)<0.01$ ，那麼 100 次中的 99 次座位都會夠。更一般情況，公司可以事先決定一個失誤率 α 然後算出 s 使得 $f(s)<\alpha$。譬如說，由（3.2,）當 $n=1000$ ，$\alpha=0.1$ 則　$S=537$　即足夠。如果兩家公司採用同樣的失誤率 0.01，則兩車共有 1074 個座位，其中 74 個將是空位。此種因競爭而導致的損失實在很小。同樣的道理，每車若有 549 個座位就是足以應付 1000 次中的 999 次。

4. 遞疊對數定理（Law of Iterated Logarithm）

4.遞疊對數定理（Law of Iterated Logarithm）：在第三節的開始我們提到了弱大數法則的缺點。其實強大數法則也有同樣的問題。譬如說在例 2.1 中我們只知道在賭博次數 n 相當大時，甲方所得 S_n 介於 $-n\varepsilon$ 與 $n\varepsilon$ 之間。由於 ε 可以任意小，$\pm n\varepsilon$ 並不能視為對 S_n 的一個很好估計。LIL 則確切地告訴我們 S_n 能好或壞到什麼程度：

在與中央極限定理相同的假設下，

$$P\left\{w:\ \overline{\lim_n}\ \frac{S_n(w)-n\mu}{\sqrt{2n\sigma^2\log\log n}}=1\right\}=1$$

由對稱性，

$$P\left\{w:\ \underline{\lim_n}\ \frac{S_n(w)-n\mu}{\sqrt{2n\sigma^2\log\log n}}=-1\right\}=1$$

此處和以下出現的對數函數 log 都是以 e 為底的自然對數。讓我們先解釋一個數列 $a_1a_2\cdots\cdots$ 的 $\overline{\lim}a_n$ 及 $\underline{\lim}a_n$ 的意義。如果我們把數列 $\{a_n\}$ 看成一個集合，且在實軸上逐一點出，那麼有些點的附近會被一再地點到，這種點我們稱做數列 $\{a_n\}$ 的叢聚點（accamulation point）。數列 $\{a_n\}$ 的叢聚點可能有很多個，其中最大的一個我們就把它記成 $\overline{\lim}\ a_n$，至於 $\underline{\lim}a_n$ 則是指叢聚點中最小的一個。譬如 $a_n=(-1)^n\frac{1}{n}$，則 $\overline{\lim}\,a_n=\underline{\lim}\,a_n=0$，這時 $0=\lim a_n$， 所以 $\lim a_n$ 與 $\{a_n\}$ 的最大點 $a_2=\frac{1}{2}$ 並不一樣。又譬如 $a_n=(-1)^n+\frac{1}{n}$　則　$\overline{\lim}\ a_n=1$，$\underline{\lim}\ a_n=-1$，此時因兩者不相等，$\lim a_n$ 不存在。

根據LIL，在例 2.1 中甲方賭博 n 次所得 S_n 滿足

$$-(1+\epsilon)\sqrt{2n\log\log n}\leqslant S_n\leqslant(1+\epsilon)\sqrt{2n\log\log n}$$

而且有許許多多次。（即有無窮多個 n），

$$S_n\geqslant(1-\varepsilon)\sqrt{2n\log\log^n}$$

這個結果比強大數法則的

$$-n\varepsilon\leqslant S_n\leqslant n\varepsilon$$

要準確多了，注意此處的 n 可能隨著賭局而

變動，也就是說 n 是 w 的函數。

同樣的道理在（2.3）式中的 S_n 滿足（$\mu = 0.1$，$\sigma = 0.3$）

$$P\{x\in(0.1)\ \overline{\lim_n}\frac{S_n(x)-\frac{n}{10}}{\sqrt{n\log\log n}} = 0.3\sqrt{2}\} = 1$$

換句話說我們不但知道數位 0 在 x 的小數展開裡出現的頻率是 0.2 我們也知道幾乎所有的 x 的小數展開式前面 n 個位數中 0 出現約達 $\frac{n}{10}+0.3\sqrt{2n\log\log n}$ 次的這種 n 有無窮多個。

5. 大離差（Large Deviations）與正規逼近（Normal Approximation）：弱大數法則告訴我們 $P\{|\frac{S_n}{n}-\mu|>\varepsilon\}$ 趨近於 0。在實際應用中我們常常希望知道這個量趨近於 0 的速度有多快。這方面的研究屬於大離差理論的範疇。它的一個基本結果是 Cramér 所證

假設 X_k，$k\geqslant 1$ 是 i.i.d，且 X_1 的 moment generating function

$$M(t)=E(e^{tX_1})=\int e^{tx}dF<\infty,$$

對任意實數 t 成立。定義

$$I(x)=\sup_{t\in R}(tx-\log M(t))$$

則對任意 R 上的一閉集 F，

$$\overline{\lim}\frac{1}{n}\log P(\frac{S_n}{n}\in F)\leqslant-\inf_{x\in F}I(x) \tag{5.1}$$

對 R 上的任一開集 G

$$\underline{\lim}\frac{1}{n}\log P(\frac{S_n}{n}\in G)\geqslant-\inf_{x\in G}I(x) \tag{5.2}$$

定理中的 I 稱做速率函數（rate function），$0\leqslant I(x)\leqslant\infty$，在 $I(x)<\infty$ 的部份它是一個凸函數，$I(\mu)=0$，且 $I(x)$ 在 $[\mu$，$\infty)$ 是漸增，在 $(-\infty$，$\mu]$ 是漸減。因此在許多情況下若 $x\neq\mu$ 則 $I(x)>0$。

譬如說在例 2.2 的 Bernonllitrial（$\mu=p$）此時

$$I(x)=\begin{cases} x\log\frac{x}{p}+(1-x)\log\frac{1-x}{1-p} & 若\ x\in(0.1) \\ -\log(1-p) & 若\ x=0 \\ -\log p & 若\ x=1 \\ \infty & 若\ x\notin[0,1] \end{cases}$$

於是若令 $F=(-\infty,p-\varepsilon]\cup[p+\varepsilon,\infty]$，則由（5.1）式

$$\overline{\lim}\frac{1}{n}\log P(|\frac{S_n}{n}-p|\geqslant\varepsilon)\leqq-\inf_{|x-p|\geqslant\varepsilon}I(x)$$
$$=-\min(I(p-\varepsilon),I(p+\varepsilon))$$

另一方面令 $G=(-\infty,p-\varepsilon)\cup(p+\varepsilon,\infty)$，則由（5.2）式

$$\underline{\lim}\frac{1}{n}\log P(|\frac{S_n}{n}-p|>\varepsilon)\geqq-\inf_{|x-p|>\varepsilon}I(x)$$
$$=-\min(I(p-\varepsilon),I(p+\varepsilon))$$

故而綜合可得

$$\lim\frac{1}{n}\log P(|\frac{S_n}{n}-p|\geqslant\varepsilon)=-\min(I(p-\varepsilon),I(p+\varepsilon))$$

令右式爲 a，則 $a<0$，且上式表示

$$P(|\frac{S_n}{n}-p|\geqslant\varepsilon)\approx e^{na}$$

換句話說 $P(|\frac{S_n}{n}-p|\geqslant\varepsilon)$ 是以指數的速度漸趨於 0。

上述所提是大離差理論的一些較古典的結果，最近十九年已有很多深入的發展。目前機率界有一些人從事這方面的研究。〔5〕是大離差理論的一本"比較"容易看懂的書。有興趣的讀者可以參考它。

與上面類似的一個問題是探討中央極限定

理（見（3.1）式）裡$F_n(x)$收斂到$\Phi(x)$的速度。除了中央極限定理的要求，若進一步要求$E|X_1|^3=r<\infty$，則 Berry-Essen 證明了存在一常數A_0使得

$$\sup_{x\in R}|F_n(x)-\Phi(x)|\leqslant\frac{A_0 r}{\sigma^3\sqrt{n}}$$

在更強的假設下，Cramer 及 Hsu（許寶騄得到下面的展式

$$F_n(x)=\Phi(x)+\frac{H_1(x)}{\sqrt{n}}+\frac{H_2(x)}{n}+\frac{H_3(x)}{n^{3/2}}+\cdots$$

其中$H_n(x)$是與Hermite多項式有關的函數。這些結果屬於Normal Approximation 的範圍。

(*) **編者註**：（倒霉的）De Moivre 比Gauss 早 76 年就導出Gauss 分佈了。有關的討論可參考

J.L.Leboui z & E.W. Montroll, None-quilibrium Phenomena II, From Stochastics to Hydrodynamics, Studies in Statistical Mechanics, Vol. XI. NH 1984.

P. 13 及其所別資料。

參考書

1. Y.S. Chow（周元燊）and H. Teicher, *Probability Theory*, Springer-Verlag, New York, 1978.
2. K.L. Chung 鍾開萊, *A Course in Probability Theory*, Academic Press, New York, 1974.
3. W. Feller, *An Introduction to Probability Theory and its Applications*, vol 1, 2nd ed., Wiley, New York, 1968.
4. W. Feller, *An Introduction to Probability Theory and its Applications*, vol 2, Inded., Wiley, New York, 1971.
5. D.W. Stroock, *An Introduction to the Theory of Large Deriations*, Springer-Verlag, New York, 1984.