

第四章

隨機樣本及其他

4.2 收斂的概念

在數學裡極限為一重要的概念。在機率論裡也是一樣，當樣本數無止盡的增大，一些關於樣本的量會有什麼極限的結果？雖然無限的樣本數，只是一理論上的假設，但極限之下的結果(往往變得較簡單)，常可做為有限樣本數之下很好的近似值。

本節只簡單地介紹兩種收斂的型式。詳細有關收斂的討論，可見一般較高等的機率論的書，如Chung(1974)。

我們先看上一章柴比雪夫不等式之一些應用。

例2.1 某市長候選人欲估計選民對他的支持度 p ，也就是隨機地取一位選民會支持他之機率。現希望誤差不超過 $\varepsilon = 0.05$ 之機率至少是0.9。若利用柴比雪夫不等式，則至少要訪問多少選民？

解.欲建立一機率模式。令 $X_1, \dots, X_n$ 為i.i.d., 以 $Ber(p)$ 為共同分佈之隨機變數。 $X_i = 1$ 或0, 依第 i 個人是否支持該候選人而定。則樣本平均 $\bar{X}_n = \sum_{i=1}^n X_i/n$ 為此 n 個人中支持該候選人之比率。我們想求 n , 使得

$$(2.1) \quad P(|\bar{X}_n - p| < \varepsilon) \geq \delta.$$

2 第四章 隨機樣本及其他

因 $E(X) = p$, $\text{Var}(X) = p(1-p)$, 由定理1.3,

$$E(\overline{X}_n) = p, \quad \text{Var}(\overline{X}_n) = \frac{p(1-p)}{n}。$$

故第三章(8.22)式導致

$$(2.2) \quad P(|\overline{X}_n - p| < \varepsilon) \geq 1 - \frac{p(1-p)}{n\varepsilon^2}。$$

又因對任意 $0 \leq p \leq 1$, $p(1-p) \leq 1/4$, 故

$$(2.3) \quad P(|\overline{X}_n - p| < \varepsilon) \geq 1 - \frac{1}{4n\varepsilon^2}, \forall \varepsilon > 0。$$

因此只要

$$1 - \frac{1}{4n\varepsilon^2} \geq \delta,$$

則(2.1)式成立。

現取 $\varepsilon = 0.05$, $\delta = 0.9$, 則由

$$1 - \frac{1}{4n \cdot 0.05^2} \geq 0.9,$$

解出

$$n \geq \frac{1}{4 \cdot 0.1 \cdot 0.05^2} = 1,000。$$

即 n 至少要 1,000。要留意的是, 此 n 值有可能過大。一方面是因由柴比雪夫不等式所得到的第三章(8.24)式並非很精確, 一方面以 $1/4$ 取代 $p(1-p)$ 也會造成一些誤差。稍後我們將利用中央極限定理, 得到一較佳的 n 之下界。

在給柴比雪夫不等式之另一應用前, 我們先給一定義。

首先對 $X_1, \dots, X_n$ 為一組隨機樣本, 以 F 為共同分佈函數。令

$$(2.4) \quad F_n(x) = \frac{X_i \leq x \text{ 之個數}}{n}, x \in R。$$

$F_n(x)$ 稱為此組樣本之經驗分佈函數(empirical distribution function), 為一階梯函數。這是一種用來估計分佈函數 F 的方法。

例2.2 設 $X_1, \dots, X_5$ 為一組隨機樣本。今得到觀測值 3, 6, 5.5, 3, 4。則

$$F_5(x) = \begin{cases} 0 & , x < 3, \\ 0.4 & , 3 \leq x < 4, \\ 0.6 & , 4 \leq x < 5, \\ 0.8 & , 5.5 \leq x < 6, \\ 1 & , x \geq 6. \end{cases}$$

例2.3 設 $F(x)$, $F_n(x)$ 分別為隨機樣本 $X_1, \dots, X_n$ 之分佈函數及經驗分佈函數。試證對 $\forall x \in R$,

$$(2.5) \quad P(|F_n(x) - F(x)| < \varepsilon) \geq 1 - \frac{1}{4n\varepsilon^2}, \forall \varepsilon > 0.$$

證明. 首先, 可看出對一固定的 $x \in R$, $F_n(x)$ 為 n 個 i.i.d. 伯努力隨機變數之樣本平均,

$$F_n(x) = \bar{I}_n = \frac{1}{n} \sum_{i=1}^n I_i,$$

其中指示函數

$$I_i = \begin{cases} 1, & X_i \leq x, \\ 0, & X_i > x. \end{cases}$$

又

$$p = E(I_i) = P(X_i \leq x) = F(x)。$$

故由(2.3)式, 對 $\forall x \in R$,

$$P(|F_n(x) - F(x)| < \varepsilon) = P(|\bar{I}_n - p| < \varepsilon) \geq 1 - \frac{1}{4n\varepsilon^2}, \varepsilon > 0。$$

證畢。

對同一分佈函數 F , 雖每次會有不同的觀測值, 因此會得到不同的 $F_n(x)$ 。但上例顯示只要 n 夠大, $F_n(x)$ 為 $F(x)$ 之一很好的估計。又對 $\forall \varepsilon > 0$, 當 $n \rightarrow \infty$ 時, 因 $4n\varepsilon^2 \rightarrow \infty$, 故得

$$(2.6) \quad \lim_{n \rightarrow \infty} P(|F_n(x) - F(x)| < \varepsilon) = 1。$$

根據**墨菲定律**(Murphy's law), 該錯的總是會錯(If something can go wrong it will)。有人認為這是於西元1917年出生在美國, 而服務於海軍的Edward A. Murphy最先說的。時至今日已有成千上萬則墨菲定律了, 有些甚至還有系理(corollary)呢。多半是一些人們的經驗法則, 或平常自我解嘲、諷刺或無可奈何時脫口說出的。如排隊時可能會覺得“另一條線總是較快”(The other line always moves faster); 欲借款時可能會認為“要獲得貸款須先證明不需要它”(In order to get a loan, you must prove you don't need it)。其他有趣的尚有“笑一下吧, 明天會更糟”(Smile, tomorrow will be worse)。讀者也可有自己的墨菲定律。如勸人照相時可說“現在照永遠比以後照好看”。就“該錯的總是會錯”此則而言, 從機率的角度的看是對的。若實驗中某事件發生的機率為正, 則持續重複此實驗, 該事件遲早會發生。事實上不僅於此, 設某事件發生之機率為 p , 若持續重複實驗, 則該事件發生之**相對頻率**(relative frequency)將接近 p 。此性質為機率

中之一極重要的定理—**大數法則**(Law of large numbers)之一特例。

註2.1 網路上可查到Murphy's law有許多則。但並未有太多字典列有此名詞。在Random House 書局出版的Webster's Unabridged Dictionary, Second Edition(2001)裡, 對Murphy's Law 之解釋為: The facetious proposition that if something can go wrong, it will. Also called Murphy's First Law; after a fictions Murphy, allegedly the name of a bungling mechanic in U.S. Navy educational Cartoons of the 1950's。Webster(韋伯)字典認為Murphy為虛構的人物。在The Oxford English Dictionary, Second Edition(1989) 有較多對於Murphy's law 的說明。

大數法則是說當樣本數夠多, 樣本平均會收斂至母體期望值。這個法則可讓大家心理舒服一點。也許人生運氣有好有壞, 但長期下來, 平均的表現, 與真正該有之表現(指期望值), 差異便不大。在伯努力試驗的情況下, 樣本平均即為事件發生之相對頻率(見例2.1), 而母體期望值即事件發生之機率。印證之前所述大數法則可應用至事件發生之相對頻率。

在例2.1中, 我們有一數列之i.i.d.的伯努力隨機變數, 成功機率為 p 。令 $\overline{X}_n$ 表首 n 項之平均, 此亦為首 n 次試驗成功數之比率。即

$$\overline{X}_1 = X_1, \quad \overline{X}_2 = \frac{X_1 + X_2}{2}, \quad \overline{X}_3 = \frac{X_1 + X_2 + X_3}{3}, \dots$$

(2.3)式指出, 只要取夠大的 n , 便可使 $\overline{X}_n$ 與 p 之差異不超過一很小的正數 ε 之機率任意接近1。要注意的是, 一般而言, $\overline{X}_n$ 並未剛好等於 p , 而且我們只是很有信心(而非絕對有把握) n 很大時, $\overline{X}_n$ 與 p 之差不超過一給定的值。

以上討論便引出了下述定義。

定義2.1 設 $X_1, X_2, \dots$ 為一數列之隨機變數, 若存在一隨機變數 X , 使得

$$(2.7) \quad \lim_{n \rightarrow \infty} P(|X_n - X| < \varepsilon) = 1, \forall \varepsilon > 0,$$

則稱 $\{X_n, n \geq 1\}$ **機率收斂**(converges in probability)至 X , 以 $X_n \xrightarrow[n \rightarrow \infty]{p} X$

表之。

(2.7)式等價於

$$\begin{aligned}
 (2.8) \quad & \lim_{n \rightarrow \infty} P(|X_n - X| \geq \varepsilon) = 0, \forall \varepsilon > 0 \\
 & \Leftrightarrow \lim_{n \rightarrow \infty} P(|X_n - X| > \varepsilon) = 0, \forall \varepsilon > 0 \\
 & \Leftrightarrow \lim_{n \rightarrow \infty} P(|X_n - X| \leq \varepsilon) = 1, \forall \varepsilon > 0.
 \end{aligned}$$

注意在定義2.1中的隨機變數 $X_1, X_2, \dots$ 不一定要i.i.d.。又由例2.3知, 對 $\forall x \in R$, 隨機樣本之經驗分佈函數 $\{F_n(x), n \geq 1\}$, 機率收斂至 $F(x)$ 。

定理2.1 弱大數法則 (Weak law of large numbers). 設 $X_1, X_2, \dots$ 為i.i.d.

之隨機變數, 且 $E(X) = \mu$ 存在。則樣本平均 $\overline{X}_n \xrightarrow[n \rightarrow \infty]{p} \mu$ 。

證明. 本定理的證明此處略去。但當假設變異數 $E(X^2) = \sigma^2$ 亦存在, 則便可藉助第三章(8.20)式的柴比雪夫不等式, 證明便很容易: 由於 $E(\overline{X}_n) = \mu$, $\text{Var}(\overline{X}_n) = \sigma^2/n$, 故當 $n \rightarrow \infty$ 時, 對 $\forall \varepsilon > 0$,

$$P(|\overline{X}_n - \mu| \geq \varepsilon) \leq \frac{\text{Var}(\overline{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \rightarrow 0.$$

依定義知 $\overline{X}_n \xrightarrow[n \rightarrow \infty]{p} \mu$ 。

弱大數法則是一優美的結果, 而只要變異數存在, 證明便很簡單。此結果指出, 在很一般的條件下(期望值存在), 隨機樣本的樣本平均機率收斂至期望值。所以樣本平均當樣本數很大時, 在“某種意義下”與期望值很接近。更明確地說, 當 n 很大時, $\overline{X}_n$ 很接近 μ 的機率會很大。如果 X 的變異數存在, 那就看得更清楚:

$$\text{Var}(\overline{X}_n) = \frac{1}{n} \text{Var}(X) \xrightarrow[n \rightarrow \infty]{} 0.$$

但若期望值不存在, 則樣本平均的變異, 就不見得會隨著樣本數之增大而變得很小。見底下例2.4。

有弱大數法則，顯然也應有強大數法則(Strong law of large numbers)。的確有，不過本書不擬討論。另外，定理2.1的條件尚可再弱些，而 $\{X_n, n \geq 1\}$ 也可以不用假設獨立，此處我們也不多討論。

例2.4 設 $X_1, X_2, \dots$ 為i.i.d., 以 $\mathcal{C}(\theta, a)$ 為共同分佈。由第二章(3.39)式, $E(e^{itX}) = e^{-a|t|+it\theta}$, $t \in R$ 。因此

$$\begin{aligned} E(e^{it\bar{X}_n}) &= E(e^{i(t/n)\sum_{j=1}^n X_j}) \\ &= \prod_{j=1}^n E(e^{i(t/n)X_j}) \\ &= (E(e^{i(t/n)X}))^n \\ &= (e^{-a|t/n|+i(t/n)\theta})^n \\ &= e^{-a|t|+it\theta} \\ &= E(e^{itX})。 \end{aligned}$$

即對 $\forall n \geq 1$, $\bar{X}_n \stackrel{d}{=} X$ 。這是一特殊的現象，樣本平均與任一樣本的分佈仍相同！這是柯西分佈之一特性：採用樣本平均 $\bar{X}_n$ 與單一樣本 X 的“效果”一模一樣(什麼效果?)，不論樣本數為何。

例2.5 設 $X_1, X_2, \dots$ 為i.i.d.之隨機變數，且 $E(X) = \mu$, $\text{Var}(X) = \sigma^2$ 皆存在。則樣本變異數 S_n^2 是否會機率收斂至 σ^2 ?

解. 利用柴比雪夫不等式，得 $\forall \varepsilon > 0$,

$$P(|S_n^2 - \sigma^2| \geq \varepsilon) \leq \frac{\text{Var}(S_n^2)}{\varepsilon^2}。$$

故 $S_n^2 \xrightarrow[n \rightarrow \infty]{p} \sigma^2$ 之一充分條件為 $\text{Var}(S_n^2) \xrightarrow[n \rightarrow \infty]{} 0$ 。當然這只是充分條件，不表此條件不成立時， S_n^2 便不機率收斂至 σ^2 。事實上即使沒有 $\text{Var}(S_n^2) \xrightarrow[n \rightarrow \infty]{} 0$ 的條件，此結果仍成立。我們留至第七章例5.3再證明。證明當然不能再藉助柴比雪夫不等式。

大數法則告訴我們，對一組隨機樣本 $X_1, \dots, X_n$ ，母體之期望值及變異數分別以 μ 及 σ^2 表之，則不論母體之分佈為何，當 $n \rightarrow \infty$ 時，樣本平均 $\bar{X}_n$ 皆機率收斂至 μ 。還有沒有其他結果呢？當 n 很大時，不論母體之分佈為何， $\bar{X}_n$ 的分佈是否會收斂至某分佈？

我們先給底下定義。這是我們要介紹的另一種收斂方式。如同定義2.1，其中 $X_1, \dots, X_n$ 亦不須假設為i.i.d.。

定義2.2 設 $X_1, X_2, \dots$ 為一數列之隨機變數， X_n 之分佈函數以 $F_{X_n}(x)$ 表之。若存在一隨機變數 X ，以 $F(x)$ 為分佈函數，使得

$$(2.9) \quad \lim_{n \rightarrow \infty} F_{X_n}(x) = F(x),$$

對函數 $F(\cdot)$ 之每個連續點 x 成立，則稱 $\{X_n, n \geq 1\}$ 分佈收斂(converges in distribution)至 X ，以 $X_n \xrightarrow[n \rightarrow \infty]{d} X$ 表之；或說 $\{X_n, n \geq 1\}$ 分佈收斂至以 F 為分佈函數之分佈。 F 又稱為 $\{X_n, n \geq 1\}$ 之極限分佈(limiting distribution)。

要注意的是，分佈收斂其實是分佈函數的收斂，而非隨機變數的收斂。此因可以有無限多個隨機變數有同一分佈函數 $F(x)$ 。所以在定義2.2中，當 $X_n \xrightarrow[n \rightarrow \infty]{d} X$ ，並不是說 X_n 與 X 有多接近，而是說它們的分佈函數會接近(隨著 n 之增大)。此點我們稍後會再說明。因此也常有下述這類寫法：

$$X_n \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(0, 1), \quad Y_n \xrightarrow[n \rightarrow \infty]{d} \mathcal{P}(\lambda)。$$

附帶一提，類似地，有時會以 $P(\mathcal{N}(0, 1) \leq x)$ 表 $\mathcal{N}(0, 1)$ 分佈之分佈函數。其他分佈也一樣。

例2.6 設 X_n 有 $\mathcal{B}(n, p_n)$ 分佈， $n \geq 1$ ，且 $\lim_{n \rightarrow \infty} np_n = \lambda$ ， $0 < \lambda < \infty$ 。因

$$F_{X_n}(x) = \sum_{k=0}^{[x]} P(X_n = k),$$

而由第二章定理2.2,

$$\lim_{n \rightarrow \infty} P(X_n = k) = \frac{e^{-\lambda} \lambda^k}{k!}, k = 0, 1, \dots。$$

故對 $\forall x \in R$,

$$\begin{aligned} \lim_{n \rightarrow \infty} F_{X_n}(x) &= \lim_{n \rightarrow \infty} \sum_{k=0}^{[x]} P(X_n = k) = \sum_{k=0}^{[x]} \lim_{n \rightarrow \infty} P(X_n = k) \\ &= \sum_{k=0}^{[x]} \frac{e^{-\lambda} \lambda^k}{k!} = F(x), k = 0, 1, \dots。 \end{aligned}$$

其中 F 為 $\mathcal{P}(\lambda)$ 之分佈函數, 又若 $i < 0$, 則定義 $\sum_{k=0}^i e^{-\lambda} \lambda^k / k! = 0$, 因此由定義2.2, $X_n \xrightarrow[n \rightarrow \infty]{d} \mathcal{P}(\lambda)$ 。

例2.7 設 X_1 有 $\mathcal{N}(0, 1)$ 分佈, $X_2 = -X_1$, $X_3 = X_1$, $X_4 = -X_1, \dots$, 即 $X_{2n} = -X_1$, $X_{2n+1} = X_1, \forall n \geq 1$ 。顯然 X_n 皆有 $\mathcal{N}(0, 1)$ 分佈。取 $X = X_1$, 仍有 $\mathcal{N}(0, 1)$ 分佈。因 $\forall n \geq 1, F_{X_n}(x) = \Phi(x)$, 故 $X_n \xrightarrow[n \rightarrow \infty]{d} X$ 。但 n 再大時, X_n 的值與 X 都不會接近: 當 X 的值為5時, X_n 交替地取5及-5。

對期望值為 μ , 變異數為 σ^2 之隨機樣本, 令 Z_n 如(1.5)式, 則由例1.3, 當母體有常態分佈, Z_n 有 $\mathcal{N}(0, 1)$ 分佈, 故若 Z_n 有一共同的極限分佈存在, 此分佈必為常態。這就是著名的中央極限定理。

定理2.2 (中央極限定理). 設 $X_1, X_2, \dots$ 為i.i.d.之隨機變數, 且 $E(X) = \mu$, $\text{Var}(X) = \sigma^2$ 皆存在。令 $F_n(x)$ 表 $\sqrt{n}(\bar{X}_n - \mu)/\sigma$ 之分佈函數。則

$$(2.10) \quad \lim_{n \rightarrow \infty} F_n(x) = \Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du, x \in R。$$

即 $\sqrt{n}(\bar{X}_n - \mu)/\sigma$ 分佈收斂至 $\mathcal{N}(0, 1)$ 分佈, 亦可以 $\sqrt{n}(\bar{X}_n - \mu)/\sigma \xrightarrow[n \rightarrow \infty]{d}$
 $\mathcal{N}(0, 1)$ 表之。

中央極限定理的證明我們略去。此定理指出適當地常態化後, i.i.d. 隨機變數的和, 即

$$(2.11) \quad \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} = \sum_{i=1}^n \frac{X_i - \mu}{\sigma\sqrt{n}},$$

會趨近至標準常態分佈。在很多實際的情況中, 通常樣本數 $n \geq 30$, 就已適用, 即誤差已不會太大。甚至可以有更一般的結果: 在較高等的機率論中, 有 $X_1, X_2, \dots$ 之分佈不一定要相同之中央極限定理的版本。

中央極限定理解釋了為什麼常態分佈可用來作為很多實際現象的模式。諸如身高、智商、量測的誤差, 都是很多(假設有 n 個)微小的獨立的量之和, 因此其總和 S_n , “差不多”可以常態分佈 $\mathcal{N}(n\mu, n\sigma^2)$ 當作近似分佈, 其中 μ, σ^2 分別為每一微小的量之期望值及變異數。

由於

$$(2.12) \quad \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{S_n - n\mu}{\sigma\sqrt{n}},$$

所以在定理2.2, 我們亦可說

$$(2.13) \quad \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(0, 1),$$

或

$$(2.14) \quad \frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(0, 1)。$$

例2.8 設 $X_1, X_2, \dots$ 為 i.i.d. 以 $\mathcal{P}(\lambda)$ 為其共同分佈。因 $E(X) = \text{Var}(X) = \lambda$, 故

$$\frac{\sqrt{n}(\bar{X}_n - \lambda)}{\sqrt{\lambda}} \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(0, 1)。$$

下例可與上例相對應。

例2.9 設 X_n 有 χ_n^2 分佈。底下我們證明

$$(2.15) \quad \frac{X_n}{n} \xrightarrow[n \rightarrow \infty]{p} 1,$$

$$(2.16) \quad \frac{X_n - n}{\sqrt{2n}} \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(0, 1)。$$

此為 χ_n^2 分佈之性質。

首先由第三章定理2.4-(iv),

$$X_n \stackrel{d}{=} \sum_{i=1}^n Y_i,$$

其中 $Y_i, i = 1, \dots, n$, 為i.i.d., 以 χ_1^2 為共同分佈。因 $E(Y_i) = 1, \text{Var}(Y_i) = 2$, 故由大數法則

$$\frac{X_n}{n} = \frac{\sum_{i=1}^n Y_i}{n} \xrightarrow[n \rightarrow \infty]{p} 1 = E(Y),$$

再由中央極限定理

$$\frac{X_n - n}{\sqrt{2n}} = \frac{\sum_{i=1}^n Y_i - nE(Y)}{\sqrt{n}\sqrt{\text{Var}(Y)}} \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(0, 1)。$$

對 $\forall 0 < p < 1$, 令 $\chi_{p,n}^2$ 為 χ_n^2 分佈之第 p 分位數。即 $\chi_{p,n}^2$ 滿足 $P(\chi_n^2 \leq \chi_{p,n}^2) = p$ 。則由(2.16)式得, 對 $\forall x > 0$,

$$(2.17) \quad \begin{aligned} P(\chi_n^2 \leq x) &= P\left(\frac{\chi_n^2 - n}{\sqrt{2n}} \leq \frac{x - n}{\sqrt{2n}}\right) \\ &\doteq P\left(\mathcal{N}(0, 1) \leq \frac{x - n}{\sqrt{2n}}\right) = \Phi\left(\frac{x - n}{\sqrt{2n}}\right)。 \end{aligned}$$

其次如前令 z_p 表 $\mathcal{N}(0, 1)$ 分佈之第 p 分位數。則上式導致 n 很大時,

$$(2.18) \quad \chi_{p,n}^2 \doteq n + z_p \sqrt{2n}。$$

χ_n^2 分佈之機率值即可藉助(2.17)或(2.18)式得到。例如,

$$P(\chi_{30}^2 \leq 40) \doteq \Phi\left(\frac{40 - 30}{\sqrt{2 \cdot 30}}\right) \doteq \Phi(1.291) \doteq 0.9017。$$

又

$$\chi_{30,0.95}^2 \doteq 30 + z_{0.95} \sqrt{2 \cdot 30} \doteq 30 + 1.645 \sqrt{60} \doteq 42.74。$$

由附表5, 此二值分別約為0.89及43.77, 差異不太大。

例2.10 設 $X_1, X_2, \dots$ 為 i.i.d. 以 $\text{Ber}(p)$ 為其共同分佈, $0 < p < 1$ 。因 $E(X) = p$, $\text{Var}(X) = p(1-p)$, 故

$$(2.19) \quad \frac{\sum_{i=1}^n X_i - np}{\sqrt{np(1-p)}} \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(0, 1)。$$

由於 $\sum_{i=1}^n X_i$ 有 $\mathcal{B}(n, p)$ 分佈, 上式指出當樣本數 n 夠大時, 二項分佈可以常態分佈作為近似分佈。此結果又稱**棣美弗-拉普拉斯定理**(De Moivre-Laplace theorem), 這可說是機率論裡, 第一個關於極限的定理, 當然也是中央極限定理的一個特例。通常只要 np 及 $n(1-p)$ 皆大於 5, 此近似便夠精確。若 np 或 $n(1-p)$ 太小, 便不太適合以常態分佈來當做二項分佈之近似, 產生的誤差可能過大。實際應用時, 若採用**連續性的更正**(continuity correction)會更精確。即對任二非負整數 k 及 j ,

$$(2.20) \quad \begin{aligned} P(j \leq S_n \leq k) &= P(j - \frac{1}{2} \leq S_n \leq k + \frac{1}{2}) \\ &= P(\frac{j - np - 1/2}{\sqrt{np(1-p)}} \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq \frac{k - np + 1/2}{\sqrt{np(1-p)}}) \\ &\doteq \Phi(\frac{k - np + 1/2}{\sqrt{np(1-p)}}) - \Phi(\frac{j - np - 1/2}{\sqrt{np(1-p)}})。 \end{aligned}$$

當然如果 $1/2$ 與 $\sqrt{np(1-p)}$ 相比很小, 忽略 $1/2$ 便不致造成太大的影響。

例如, $n = 400, p = 1/2$, 則 $np = 200, \sqrt{np(1-p)} = 10$, 故得

$$\begin{aligned} &P(190 \leq \sum_{i=1}^{400} X_i \leq 210) \\ &= P(\frac{190 - 200 - 0.5}{10} \leq \frac{\sum_{i=1}^{400} X_i - 200}{10} \leq \frac{210 - 200 + 0.5}{10}) \\ &= P(-1.01 \leq \frac{\sum_{i=1}^{400} X_i - 200}{10} \leq 1.01) \\ &\doteq \Phi(1.01) - \Phi(-1.01) \\ &\doteq 0.6876。 \end{aligned}$$

上述推導顯示, 投擲一公正的銅板 400 次, 則中央極限定理告訴我們, 所得之正面數與期望次數 200 之差距, 不超出 10 之機率約為 0.6826。但若

想以二項分佈求

$$\sum_{i=190}^{210} \binom{400}{i} \left(\frac{1}{2}\right)^{400},$$

明顯地困難不少。

讀者也可自行計算, 投擲一公正的銅板, 要得到正面數有65個以上, 是很難發生的(機率約0.0013), 更不要說是得到95個正面了。

有時佐以其他極限的定理, 會使中央極限定理更有用。底下即著名的史拉斯基定理(Slutsky's theorem), 證明在此略去。

定理2.3 設 $X_n \xrightarrow[n \rightarrow \infty]{d} X$, $Y_n \xrightarrow[n \rightarrow \infty]{p} a$, a 為一常數。則

$$(i) X_n + Y_n \xrightarrow[n \rightarrow \infty]{d} X + a;$$

$$(ii) X_n - Y_n \xrightarrow[n \rightarrow \infty]{d} X - a;$$

$$(iii) X_n Y_n \xrightarrow[n \rightarrow \infty]{d} aX;$$

$$(iv) X_n / Y_n \xrightarrow[n \rightarrow \infty]{d} X/a, \text{ 只要 } a \neq 0。$$

例2.11 北銀樂透彩之中獎機率(指中頭獎至普獎)約為 $p = 0.02906$ (見第二章例2.13)。不過因投注者可自由選號, 所以每期實際中獎率均有波動。令 N 表某期賣出的張數, 則該期總共中獎數 X 有 $B(N, p)$ 分佈。一般而言 N 均很大, Np 及 $N(1-p)$ 皆大於5也不成問題, 因此由例2.10,

$$Z = \frac{X - Np}{\sqrt{Np(1-p)}}$$

有近似的 $\mathcal{N}(0, 1)$ 分佈。因此

$$P(|Z| \leq 2) = P\left(\left|\frac{X - Np}{\sqrt{Np(1-p)}}\right| \leq 2\right) \doteq 0.9544。$$

同理

$$P(|Z| \leq 3) \doteq 0.9974。$$

$|Z|$ 值要超過3之機率可說相當小，更不要說 $|Z|$ 值超過10了。現以第30期為例，該期 $N = 13,210,954$ 。而

$$\left| \frac{X - 13,210,954 \cdot 0.02906}{\sqrt{13,210,954 \cdot 0.02906 \cdot 0.97094}} \right| \leq 2$$

$$\Leftrightarrow 382,689.25 \leq X \leq 385,131.39。$$

即約有0.9544之機率，該期中獎數應介於382,689與385,131間。我們將該期各獎之實際中獎數列於表2.1，相加後得總中獎數為365,568，明顯地偏低。對應的 Z 值約為

$$Z = \frac{365,568 - 13,210,954 \cdot 0.02906}{\sqrt{13,210,954 \cdot 0.02906 \cdot 0.97094}}$$

$$\doteq -30.042。$$

對 $\mathcal{N}(0, 1)$ 分佈，會得到如此小的 Z 值，其機率是很低的。這顯示該期樂透彩填注之號碼不夠隨機(前30期 Z 值最大約為298，最小約為-238，但無一介於 ± 3 之間)。若是電腦選號較不會有此現象，但人腦“隨意”選號，則常有這種現象發生。此概念在第十章我們會再討論。

表2.1 第30期樂透彩各獎中獎數

頭獎	貳獎	參獎	肆獎	普獎
0	13	389	21,455	343,711

例2.12 假設進行某項化學實驗的量測，而得到30筆觀測值 $X_1, \dots, X_{30}$ 。設 $X_1, \dots, X_{30}$ 為i.i.d.，期望值 μ ，且依過去經驗，變異數 $\sigma^2 = 6$ 。欲以樣本平均 $\overline{X}_{30}$ 來估計 μ ，試給出誤差超過1之機率的近似值。

解.由中央極限定理

$$P(|\overline{X}_{30} - \mu| > 1) = P\left(\left|\frac{\overline{X}_{30} - \mu}{\sigma/\sqrt{n}}\right| > \frac{1}{\sqrt{6}/\sqrt{30}}\right)$$

$$\doteq P(|Z| > 2.236) \doteq 0.02532。$$

雖然一次觀測之標準差為 $\sqrt{6} \doteq 2.45$ ，但當樣本數為30，樣本平均要偏離真實值一單位之機率並不大。這是為什麼很多時候，以樣本平均來估計真實值。

如果是利用柴比雪夫不等式, 則得

$$P(|\bar{X}_{30} - \mu| > 1) \leq \frac{\sigma^2/30}{1^2} = \frac{6}{30} = 0.2。$$

雖然這是一“正確的”不等式, 而由中央極限定理所得到的 $P(|\bar{X}_{30} - \mu| > 1)$ 只是一近似值, 但由 0.02532 約只為 0.2 之八分之一倍, 從實用的觀點, 顯然會採前者。

例2.13 由定理2.3-(iii)(取 $Y_n \equiv a$)知, 在定理2.2中,

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow[n \rightarrow \infty]{d} \sigma\mathcal{N}(0, 1) = \mathcal{N}(0, \sigma^2)。$$

必須一提的定理還有不少。如**機率收斂會導致分佈收斂**, 即

$$(2.21) \quad X_n \xrightarrow[n \rightarrow \infty]{p} X \Rightarrow X_n \xrightarrow[n \rightarrow \infty]{d} X。$$

但若存在一常數 a , 使得 $P(X = a) = 1$, 則二者等價。即

$$(2.22) \quad X_n \xrightarrow[n \rightarrow \infty]{p} a \Leftrightarrow X_n \xrightarrow[n \rightarrow \infty]{d} a。$$

又設 g 為一連續函數, 則若 $X_n \xrightarrow[n \rightarrow \infty]{d} X$, 便有 $g(X_n) \xrightarrow[n \rightarrow \infty]{d} g(X)$ (習題第15題給出本結果之一些應用); 若 $X_n \xrightarrow[n \rightarrow \infty]{p} X$, 便有 $g(X_n) \xrightarrow[n \rightarrow \infty]{p} g(X)$ 。底下定理也常會用到。

定理2.4 設 b 為一常數, 而 $n \rightarrow \infty$ 時, $a_n \rightarrow \infty$, 且 $a_n(Z_n - b) \xrightarrow[n \rightarrow \infty]{d} X$ 。又設 g 為一可微的實函數, 且 g' 在 b 連續。則

$$(2.23) \quad a_n(g(Z_n) - g(b)) \xrightarrow[n \rightarrow \infty]{d} g'(b)X。$$

例2.14 在例2.5我們曾指出: 若 $\text{Var}(S_n^2) \rightarrow 0$, 則 $S_n^2 \xrightarrow[n \rightarrow \infty]{p} \sigma^2$ 。因此(習題第15題) $\sigma/S_n \xrightarrow[n \rightarrow \infty]{p} 1$ 。在中央極限定理裡, 若只知期望值 μ , 而不知變異

數 σ^2 之值, 則可以 S_n 取代 σ :

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} = \frac{\sigma}{S_n} \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(0, 1)。$$

弱大數法則是說 n 很大時, $\bar{X}_n - \mu$ 會機率收斂至0。但若將 $\bar{X}_n - \mu$ 放大, 即乘上 $\sqrt{n}$, 便趨近至常態分佈。 $\sqrt{n}$ 是一適當的倍數, 若乘上位階(order)小於 $\sqrt{n}$ 者, 仍很小, 若乘上位階大於 $\sqrt{n}$ 者, 會使其過大(趨近至無限大)。如果以(2.6)式來看, $S_n - n\mu$ 除以 n 會機率收斂至0, 不再隨機了。但若除以 $\sqrt{n}$, 即得 $(S_n - n\mu)/\sqrt{n}$, 便趨近至一仍為隨機的分佈。

大數法則可用來支持**頻率對機率的解釋**(frequentistic interpretation of probability)。某事件發生的機率為 p 是什麼意思? 假設此事件是可以重複的觀測(見例2.1), 如反覆地投擲一銅板, 令

$$X_i = \begin{cases} 1, & \text{第} i \text{次出現正面,} \\ 0, & \text{其他。} \end{cases}$$

又設 $X_1, \dots, X_n$ 為i.i.d.之隨機變數, 令 $p = E(X)$ 。則

$$\bar{X}_n = \frac{X_1 + \dots + X_n}{n}$$

表出現正面的相對頻率。弱大數法則指出, n 很大時, 此相對頻率有相當大的機率會接近 p 。換句話說, 若投擲數夠多, $\bar{X}_n$ “差不多”可當做 p 之一很好的近似值。以這種頻率的觀點來解釋機率是很自然的。當然對不能重複觀測的事件, 如氣象局宣佈: 明日降雨機率為0.3, 就很難清楚解釋機率為0.3的意義了。