弱大數法則，Berustein多項式及Weierstrass逼近定理

劉豐哲

Weierstrass 逼近定理敍述如下：如果 $f(a)$ 是〔0，1〕上的連續函數，則給定 $\varepsilon>0$，可以找到一個多項式 $P(x)$ 使得 $|f(x)-P(x)|<\varepsilon$ 對所有〔0，1〕上的點 x 成立。換言之，可以找到一序列多項式 $\{P_n(x)\}$，便得 $P_n(x)$ 均勻地趨近到 $f(x)$。

Weierstrass逼近定理是數學上很基本而重要的定理，因此它有好幾種證明；到目前爲止，筆者所知道的最直接的證明是經由 Berstein 多項式去證明的。本文將介紹這種證明；它和機率論中的弱大數法則有關，因此特別有興味。

1. 設 $0<p<1$，並令 $0<q=1-p<1$。想像有一枚銅板，當我們隨意地丟它的時候，正面出現的機率爲 p。假如我們逐次地丟這枚銅板，即得到一序列的試驗，每次試驗的結果或爲正面出現或爲反面出現，我們除了它們出現的機率分別爲 p 和 q 外，無法預測那一面出現。如果正面出現時用 1 表示，反面出現時用 0 表示，則這一序列的試驗可用一序列的隨機變數 $\{\Omega_n(\omega)\}$ 表示，其中 ω 是樣品空間的點，$\Omega_n(\omega)=1$ 或 0，且 $\Omega_n=1$ 的那些樣品所成的事件的機率爲 p，$\Omega_n=0$ 的那些樣品所成事件的機率爲 q

。如果每次試驗之後，銅板的特性沒有改變（即出現正面和反面的機率仍然維持爲 p 和 q ），那麼用機率論的語言來敍述的話，就是 $\{\Omega_n(\omega)\}$ 是獨立隨機變數序列。

首先我們用一個具體的方法來表現獨立隨機變數序列 $\{\Omega_n(\omega)\}$ 。我們取半閉的空間〔 0 ，1 ）爲樣品空間，並逐次定義隨機變數 $\{\Omega_n(\omega)\}$ 如下：令 $I_1^{(1)}=[0,p)$ ，$I_2^{(1)}=[p,1)$ ，並視 $\omega\in I_1^{(1)}$ 或 $I_2^{(1)}$ 而 $\Omega_1(\omega)=1$ 或 0 ；如果 $I=[a,b)$ 是〔0，1）的子區間，則令 $I'=[a,c)$ ，$I''=[c,b)$ 其中 $(c-a):(b-c)=p:q$ ；現令 $I_1^{(2)}=I_1^{(1)\prime}$ ，$I_2^{(2)}=I_1^{(1)\prime\prime}$ ，$I_3^{(2)}=I_2^{(1)\prime}$ ，$I_4^{(2)}=I_2^{(1)\prime\prime}$ ，並視 $\omega\in I_1^{(1)\prime}\cup I_2^{(1)\prime}$ 或 $I_1^{(1)\prime\prime}\cup I_2^{(1)\prime\prime}$ 而令 $\Omega_2(\omega)=1$ 或 0 ；如此繼續做下去，假設 $I_1^{(k)}$ ，$I_2^{(k)}$ ，……，$I_{2^k}^{(k)}$ 及 $\Omega_k(\omega)$ 已有定義，而令 $I_{2j-1}^{(k+1)}=I_j^{(k)\prime}$ ，$I_{2j}^{(k+1)}=I_j^{(k)\prime\prime}$ ，$j=1,2,\cdots\cdots,2^k$ ，並視 $\omega\in I_1^{(k)\prime}\cup I_2^{(k)\prime}\cup\cdots\cdots\cup I_{2^k}^{(k)\prime}$ 或 $I_1^{(k)\prime\prime}\cup\cdots\cdots\cup I_{2^k}^{(k)\prime\prime}$ 而令 $\Omega_{k+1}(\omega)=1$ 或 0。如此，對應於每個自然數 n ，我們把〔0，1）分解成 $I_1^{(n)}$ ，$I_2^{(n)}$ ，………，$I_{2^n}^{(n)}$ ，並定義一函數 Ω_n 如下：$\Omega_n(\omega)=1$ 或 0 ，視 $\omega\in I_j^{(n)}$ 中之 j 爲奇數或偶數而定。我們所取定的樣品空間爲〔0，1），而我們所考慮的函數列 $\{\Omega_n\}$ 所取的值僅爲 0 和 1 ，而且 $\{\Omega_n=1\}$ 及 $\{\Omega_n=0\}$ 的事件皆由〔0，1）的半閉子區間的聯集所組成，因此，此後我們討論中所出現之事件皆由〔0，1）之半閉區間所組成，這種事件的機率就定爲組成此事件的區間的長度和。此後我們用 $P(E)$ 來表示事件 E 的機率，特別 $P(\Omega_n=1)=p$ ，$n=1,2,\cdots\cdots$。由 $\{\Omega_n\}$ 的定義，如果 n ，m 爲任二不相等的自然數，則在 $\{\Omega_n=1\}$ 或 $\{\Omega_n=0\}$ 上 $\{\Omega_m=1\}$ 和 $\{\Omega_m=0\}$ 的區間和之比仍爲 $p:q$ ，即和 $\{\Omega_m=1\}$ 和 $\{\Omega_m=0\}$ 在全樣品空間〔1，0）上的機率比一樣，換言之 Ω_m 和 Ω_n 是互爲獨立的。事實上，也不難看出：如果 m ，n_1 ，n_2 ，……，n_k ，爲互異之自然數，則 $\{\Omega_{n_1}=\alpha_1,\cdots\cdots,\Omega_{n_k}=\alpha_k\}$ 發生時，$\Omega_m=1$ 或 0 之條件機率仍分別爲 r 和 q ，其中每個 α_1 或爲 1 或爲 0 ，也就是說 $\{\Omega_n\}$ 是個獨立隨機變數序列。因此我們在樣品空間〔0，1）上構造出來的函數列 $\{\Omega_n\}$ 事實上可視爲本節一開頭所考慮的試驗序列的數學模型。此後，我們固定這個模型，而且我們所考慮的隨機變數 f 都是在〔0，1）分割成有限個區間的每個區間上爲常數，因此它的期望 $E(f)$ 就是它的積分，即

$$E(f)=\int_0^1 f(\omega)\,d\omega \text{ 。}$$

如果 f ，g 是兩個隨機變數，t 爲任意實數，則

$$0\leq E((f+tg)^2)=E(f^2)+2tE(fg)+t^2E(g^2)$$

由於上式右邊爲 t 之二次式，並且恒 ≥ 0 ，我們得到如下的不等式

$$[E(fg)]^2\leq E(f^2)E(g^2) \text{ ，即}$$
$$|E(fg)|\leq\sqrt{E(f^2)}\cdot\sqrt{E(g^2)}$$

此不等式稱爲 Schwarz 不等式。

相對於每個整數 n ，考慮隨機變數

$$\frac{\Omega_1+\cdots\cdots\cdots\cdots+\Omega_n}{n}\equiv\Phi_n$$

此隨機變數即爲前面 n 次試驗中正面出現之機率。直覺上，我們認爲它應該和 p 很接近；事實上，我們有下面的弱大數法則：

定理 1：對任何 $\epsilon>0$ ，下面式子成立

$$\lim_{n\to\infty}P(|\Phi_n-p|\geq\varepsilon)=0$$

證明定理 1 之前，先證明 Chebyshev 不等式：

補題：設 f 爲一隨機變數，並令 $V(f)=E([f-E(f)]^2)$ ，則

$$P(\{|f-E(f)|\geq\epsilon\})\leq\frac{1}{\epsilon^2}V(f)$$

證明：令$A_\epsilon=\{|f-Ef|\geq\epsilon\}$，並定義隨機變數$g_\epsilon$如下：

$$g_\epsilon(\omega)=1 \text{ 如果 } \omega\in A_\epsilon$$
$$=0 \text{ 如果 } \omega\notin A_\epsilon$$

則

$$\epsilon P(\{|f-E(f)|\geq\epsilon\})\leq E(|f-E(f)|\cdot g_\epsilon)$$
$$\leq\sqrt{E(g_\epsilon^2)}\cdot\sqrt{E([f-E(f)]^2)}$$
$$=\sqrt{P(\{|f-E(f)|\geq\epsilon\})}\cdot\sqrt{V(f)}$$

上式我們用到Schwarz不等式。將上式兩邊平方，再除以$\epsilon^2\cdot P(\{|f-E(f)|\geq\epsilon\})$，就證明了補題。

定理1之證明： 首先注意到$E(\Phi_n)=\frac{1}{n}\{E(\Omega_1)+\cdots\cdots+E(\Omega_n)\}=p$。由補題知道

$$P(|\Phi_n-p|\geq\epsilon)\leq\frac{1}{\epsilon^2}V(\Phi_n)\text{。}$$

現在來計算$V(\Phi_n)$：

$$V(\Phi_n)=E([\Phi_n-p]^2)$$
$$=\frac{1}{n^2}E([\sum_{j=1}^n(\Omega_j-p)]^2)$$
$$=\frac{1}{n^2}E(\sum_{j=1}^n[\Omega_j-p]^2$$
$$+\sum_{j\neq k=1}^n[\Omega_j-p][\Omega_k-p]);$$

但是注意到$j\neq k$時Ω_j-p和Ω_k-p是互為獨立的，因此（見附錄）

$$E([\Omega_j-p][\Omega_k-p])$$
$$=E(\Omega_j-p)\cdot E(\Omega_j-p)$$
$$=0$$

所以

$$V(\Phi_n)=\frac{1}{n^2}E(\sum_{j=1}^n[\Omega_1-p]^2)$$
$$=\frac{1}{n^2}\sum_{j=1}^n V(\Omega_j)$$
$$=\frac{1}{n^2}\cdot n\cdot p(1-p)$$
$$=\frac{1}{n}p(1-p)$$

因為$V(\Omega_j)=p(1-p)$，$j=1,2,\cdots\cdots,n$。把前面的結果合起來，即得

$$(*)\qquad P(|\Phi_n-p|\geq\epsilon)\leq\frac{1}{\epsilon^2}V(\Phi_n)$$
$$=\frac{1}{\epsilon^2 n}p(1-p)$$

由此式即證得定理1。

2. 設f為$[0,1]$上的連續函數，Φ_n，$n=1,2,\cdots\cdots$為如上節所定義者，則由Φ_n之定義及$\{\Omega_n\}$為獨立的隨機變數序列這個事實，我們不難說服自己如下的等式：

$$E(f\circ\Phi_n)$$
$$=E(f(\Phi_n))$$
$$=\int_0^1 f(\Phi_n(\omega))d\omega$$
$$=\sum_{k=0}^n f(\frac{k}{n})\binom{n}{k}p^k(1-p)^{n-k}$$

其中$\binom{n}{k}$為從n個不同物件中取k個不同物件的組合數。將上式的右邊視為p之多項式時，稱為f的第n個Bernstein多項式，用$B_n(f;p)$示之。

我們來證明$B_n(f;p)$均勻地收斂到$f(p)$。當$p=0$或1時$B_n(f;p)$分別為$f(0)$及$f(1)$，因此僅需要考慮$0<p<1$，所以我們可用上面所構造的數學模型，特別由上節的$(*)$式我們得到

$$(**)\qquad \sup_{0<p<1}P(|\Phi_n-p|\geq\delta)$$
$$\leq\sup_{0<p<1}\frac{1}{\delta^2 n}p(1-p)\leq\frac{1}{4\delta^2 n}$$

設$\epsilon>0$為給定。由於f在$[0,1]$上均勻連續，可得$\delta>0$使得當$0\leq s,t\leq1$，$|s-t|<\delta$時$|f(s)-f(t)|<\frac{1}{2}\epsilon$。對此

δ，由（$**$）式可得 n_0 使得 $n \geq n_0$ 時，對任何 $0 < p < 1$ 皆有

$$(\Delta) \qquad P(|\Phi_n - p| \geq \delta) \leq \frac{\epsilon}{4M}$$

其中 $M = \max_{0 \leq t \leq 1} |f(t)|$。現設 $n \geq n_0$，$0 < r < 1$，並令

$$A_1 = \{\omega \in [0,1) : |\Phi_n(\omega) - p| \geq \delta\},$$

$$A_2 = \{\omega \in [0,1) : |\Phi_n(\omega) - p| < \delta\}。$$

則

$$|B_n(f;p) - f(p)| = |E(f_0\Phi_n) - f(p)|$$
$$\leq E(|f_0\Phi_n - f(p)|)$$
$$= \int_{A_1} |f(\Phi_n(\omega) - f(p)| d\omega$$
$$+ \int_{A_2} |f(\Phi_n(\omega) - f(p)| d\omega;$$

很明顯地，

$$\int_{A_1} |f(\Phi_n(\omega)) - f(p)| d\omega \leq 2MP(A_1)$$
$$= 2MP(|\Phi_n - p| \geq \delta) \leq 2M \cdot \frac{\epsilon}{4M} = \frac{1}{2}\epsilon$$

上面我們用了 (Δ) 式；另外，由 δ 的取法

$$\int_{A_2} |f(\Phi_n(\omega)) - f(p)| d\omega$$
$$< \frac{1}{2}\epsilon \; P(A_2) \leq \frac{1}{2}\epsilon \;。$$

總結起來，我們證明了當 $n \geq n_0$ 時

$$|B_n(f;p) - f(p)| < \epsilon \;, \quad 0 < p < 1$$

這樣我們就證得了。

定理 2：$\lim_{n\to\infty} B_n(f;p) = f(p)$，並且是均勻的；即 Bernstein 多項式 $\sum_{k=0}^{n} f(\frac{k}{n})\binom{n}{k} p^k \cdot (1-p)^{n-k}$ 均勻地收斂到 $f(p)$。

這個定理是Weierstrass逼近定理的構造式定理，因爲它事實上寫出能均勻收斂到 $f(p)$ 的多項式序列。

3.附錄 我們來證明：如果 f，g 是互相獨立的的隨機變數，則 $E((f-\alpha)(g-\beta)) = E(f-\alpha) \cdot E(g-\beta)$，其中 α、β 爲兩個常數。注意到，我們所考慮的隨機變數只取有限個值，因此我們只證明這種特殊情形，雖然這個結論一般成立。設 f 取 $\alpha_1, \cdots\cdots, \alpha_n$ 等 n 個值，g 取 $\beta_1, \cdots\cdots, \beta_m$ 等 m 個值，則

$$E(fg) = \sum_{i=1}^{n}\sum_{j=1}^{m} \alpha_i \beta_j P(\{f=\alpha_i \;; g=\beta_j\}),$$

其中 $\{f=\alpha_i \;; g=\beta_j\} = \{\omega \in [0,1) : f(\omega)=\alpha_i \;, g(\omega)=\beta_j\}$；由於 f 和 g 互爲獨立，因此

$$P(\{f=\alpha_i \;; g=\beta_j\})$$
$$= P(\{f=\alpha_i\}) \cdot P(\{g=\beta_j\}),$$

從而

$$E(fg) = \sum_{i=1}^{n}\sum_{j=1}^{m} \alpha_i \beta_j P(f=\alpha_i) \cdot P(g=\beta_j)$$
$$= \left(\sum_{i=1}^{n} \alpha_i P(f=\alpha_i)\right) \cdot \left(\sum_{j=1}^{m} \beta_j P(g=\beta_j)\right)$$
$$= E(f) \cdot E(g)$$

現在來計算 $E((f-\alpha)(g-\beta))$：

$$E((f-\alpha)(g-\beta))$$
$$= E(fg - \alpha g - \beta f + \alpha\beta)$$
$$= E(f)E(g) - \alpha E(g) - \beta E(f) + \alpha\beta$$
$$= E(f-\alpha) \cdot E(g-\beta),$$

這正是我們所聲稱的。

——本文作者為中央研究院數學所研究員，本刊主編。