熵(Entropy)(下)

李天岩

§3. 拓樸熵(Topological Entropy)

連續性是自然界的基本屬性之一。數學上連續的概念是由拓樸來刻劃的。拓樸空間X中由所有開集生成的Borel代數相當於測度空間裡的σ-代數。拓樸空間上的連續映射相當於測度空間裡的可測變換。由此，我們可以將上節中所談的Kolmogorov熵在拓樸空間裡做相似性的定義，來描述連續映射的不確定性。在這過程中最大的困擾是：一般拓樸空間中，並沒有一個相似於測度空間裡的「測度」的度量。

假設X爲一個緊緻(compact) Hausdorff空間，$f: X\to X$爲一連續映射。由一般拓樸學知，X存在有限開覆蓋(finit open covering)。設$\bar{A}=\{A_1,\cdots\cdots,A_m\}$爲$X$的一個有限開覆蓋，$\bar{A}$中能覆蓋$X$的子集族稱爲$\bar{A}$中的子覆蓋(subcover)。我們稱$\bar{A}$中的子覆蓋爲「極小」(minimal)，如果在$\bar{A}$中沒有一個比它元素少的子覆蓋。通常用$N(\bar{A})$來代表$\bar{A}$中極小子覆蓋裡元素的個數。若把極小子覆蓋裡的每一個開集當做一個「基本事件」，大家的概率都是$\dfrac{1}{N(\bar{A})}$，我們則得到一個樣本空間。它的Shannon熵可以很輕易的算出，是$\log N(\bar{A})$。我們稱這個數目爲開覆蓋$\bar{A}$的熵，同時用符號$H(\bar{A})$來表示。

當$\bar{A}$爲X上的開覆蓋時，$f^{-1}(\bar{A})=\{f^{-1}(A): A\in\bar{A}\}$也是一個開覆蓋。若$\{A_1,\cdots\cdots,A_{N(\bar{A})}\}$是$\bar{A}$中的一個極小子覆蓋，$\{f^{-1}(A_1),\cdots\cdots,f^{-1}(A_{N(\bar{A})})\}$是$f^{-1}(\bar{A})$的一個子覆蓋，但不一定是極小。所以，

$$N(f^{-1}(\bar{A}))\leqslant N(\bar{A}) \qquad (3\text{-}1)$$

若$\bar{A}$，$\bar{B}$爲X上的二個開覆蓋，則$\bar{A}\vee\bar{B}$代表$\{A\cap B \mid A\in\bar{A}, B\in\bar{B}\}$這個開覆蓋。若$\{A_1,\cdots\cdots,A_N\}$是$\bar{A}$的一個子覆蓋$\{B_1,\cdots\cdots,B_M\}$是$\bar{B}$的一個子覆蓋，則$\{A_i\cap B_j: i=1,\cdots\cdots,N, j=1,2,\cdots\cdots,M\}$是$\bar{A}\vee\bar{B}$的一個子覆蓋。因此，$N(\bar{A}\vee\bar{B})\leqslant N(\bar{A})N(\bar{B})$。以及

$$H(\bar{A}\vee\bar{B})\leqslant H(\bar{A})+H(\bar{B}) \qquad (3\text{-}2)$$

我們還是用$\bigvee_{i=0}^{n-1} f^{-i}(\bar{A})$來代表$\bar{A}\vee f^{-1}(\bar{A})\vee\cdots\cdots\vee f^{-n+1}(\bar{A})$，同時用$N(\bar{A},f,n)$來表示覆蓋$\bigvee_{i=0}^{n-1} f^{-i}(\bar{A})$中極小子覆蓋元素的個數。相似於定義2-2′，我們定義：

定義3-1：連續映射f關於有限覆蓋$\bar{A}$的拓樸熵(topological entropy)定義爲：

$$h_{top}(f,\bar{A})$$
$$\equiv \lim_{n\to\infty}\frac{1}{n}H(\bigvee_{i=0}^{n-1}f^{-i}(\bar{A}))$$
$$\equiv \lim_{n\to\infty}\frac{1}{n}\log N(\bar{A},f,n)$$

定義3-2：緊緻Hausdorff拓樸空間X上的連續映射f的拓樸熵爲

$$h_{top}(f)=\sup\{h_{top}(f,\bar{A}):\bar{A}\text{ 爲}X\text{的有限覆蓋}\}$$

要使上述關於拓樸熵的定義合理，我們必須證明定義3-1中的極限存在。爲此我們求助於下列仍然是初等微積分的結果。

引理3-3：設實數序列$\{a_n\}_{n\geq 1}$滿足條件$a_{n+p}\leq a_n+a_p$，$\forall n,p$，則$\lim_{n\to\infty}\frac{a_n}{n}$存在且等於$\inf_n\frac{a_n}{n}$。

證明：固定$p>0$，每個$n>0$可寫成$n=kp+i$，其中$0\leqslant i<p$，則

$$\frac{a_n}{n}=\frac{a_{i+kp}}{i+kp}\leqslant\frac{a_i}{kp}+\frac{a_{kp}}{kp}\leqslant\frac{a_i}{kp}+\frac{ka_p}{kp}$$
$$=\frac{a_i}{kp}+\frac{a_p}{p}。$$

當$n\to\infty$時$k\to\infty$，故有

$$\lim_{n\to\infty}\sup\frac{a_n}{n}\leq\frac{a_p}{p}$$

因此，$\lim_{n\to\infty}\sup\frac{a_n}{n}<\inf_p\frac{a_p}{p}$。

另一方面，從不等式

$$\inf_p\frac{a_p}{p}\leq\lim_{n\to\infty}\inf\frac{a_n}{n}$$

可知，極限$\lim_{n\to\infty}\frac{a_n}{n}$存在且等於$\inf_n\frac{a_n}{n}$。

定理3-4：設X爲緊緻Hausdorff空間，$f:X\to X$連續。任給X的有限開覆蓋$\bar{A}$，極限

$$\lim_{n\to\infty}\frac{1}{n}\log N(\bar{A},f,n)$$

存在。

證明：由定義知，$\log N(\bar{A},f,n)=H(\bigvee_{i=0}^{n-1}f^{-i}(\bar{A}))$，我們令其爲$a_n$。由(3-1)，(3-2)可知

$$a_{n+p}=H(\bigvee_{i=0}^{n+p-1}f^{-i}(\bar{A}))$$
$$=H(\bigvee_{i=0}^{n-1}f^{-i}(\bar{A})\vee\bigvee_{i=n}^{n+p-1}f^{-i}(\bar{A}))$$
$$\leq H(\bigvee_{i=0}^{n-1}f^{-i}(\bar{A}))+H(f^{-n}(\bigvee_{i=0}^{p-1}f^{-i}(\bar{A})))$$
$$=a_n+a_p。$$

定理結論恰由引理3-3推得。

拓樸熵是由R. Adler，A. Konhein及M. McAndrew在1965年引進的，它是爲了研究關於拓樸共軛不變量應運而生的。拓樸共軛的定義可如下給出。

定義3-5：設$T:X\to X$和$S:Y\to Y$分別爲緊緻拓樸空間X和Y上的連續映射。若存在同胚(homeomorphism)$\phi:X\to Y$使得$\phi\circ T=S\circ\phi$，則稱T**拓樸共軛**(topologically conjugate)於S。這時，ϕ就稱爲一個共軛。

我們可以證明，拓樸熵是拓樸共軛性的一個不變量。也就是說，兩個拓樸共軛的連續映射有相同的拓樸熵，反之亦然。事實上，二個拓樸共軛的連續映射在本質上給出相同的遍歷性質，因此，拓樸熵數學地刻劃了不同共軛類的拓樸動力系統(Topological dynamical system)的性質。

我們已知拓樸空間X加上其全體開集張成的Borel σ-代數β構成一個可測空間，(X

，β），而 β 上給定任何一個概率測度 μ 就構成一個概率測度空間（X，β，μ）。X 上的任一連續映射 f 同時也為（X，β，μ）上的可測變換，因而就有一個Kolmogorov 熵隨之確定。另一方面，Borel 代數 β 上存在着衆多不同的概率測度，這樣就有了對應於不同概率測度的Kolmogorov 熵組成的數集。然而，作為連續映射 f 的拓樸熵之定義與測度無關，它是唯一確定的。我們自然發出疑問：這兩種熵有何內在聯繫？既然拓樸熵概念是由Kolmogorov 熵概念衍生而來，我們有信心認為它們的確存在着情同手足的關係。這方面的結果多采多姿。比如說，

定理3-8：設 f 為緊緻拓樸空間 X 上的連續映射，則

$$h_{\text{top}}(f)=\sup\{h_\mu(f):\text{概率測度}\ \mu\ \text{關於}\ f\ \text{遍歷}\}$$

本節拓樸熵中，空間為緊緻的假設並非必要。70年代初Dinaburg 和 Bowen 分別給出了拓樸熵的等價定義。這些定義的優越性在於它們引導了一系列關於拓樸熵和測度熵（Kolmogorov 熵）之間聯繫結果的證明。Bowen 的定義是對更廣泛的距離空間上一致連續映射族而言的。且導致了 n - 環面上自同構拓樸熵公式的幾何證明。可惜，這位在遍歷理論研究中做出突出成就的數學家剛過而立之年就與世長辭了。

§4. Boltzmann熵

熱力學中熵是一個極其重要的概念，最初由Clausius 引進。後來L. Boltzmann 在他發表在 1866 年關於氣體動力學理論的開創性工作中給出了熵的另一形式。這個熵在物理、化學的若干領域裡自始至終扮演着關鍵性的角色。可是Boltzmann 熵和我們先前定義的Kolmogorov 熵或拓樸熵並非一致。儘管如此，它們在數學的背景下，仍存在着千絲萬縷的聯繫。在這最後一節，我們將遨遊於Boltzmann 熵的數學描述。

設（X, p_1, ……, p_n）為一有限樣本空間，則其 Shannon熵為 $H(p_1,\cdots,p_n)=-\sum_{i=1}^{n}p_i\log p_i$，現設（$X$，$\Sigma$，$\mu$）為一測度空間。記 $L^1(X)$ 為定義在 X 上的Lebesgue可積函數全體。$L^1(X)$ 中滿足等式

$$\int_X f(x)\,d\mu=1$$

的非負函數 $f(x)$ 稱為密度函數，其集合記為 D。易見等式

$$\mu_f(A)=\int_A f(x)\,d\mu \qquad A\in\Sigma$$

定義了（X，Σ）上的一個概率測度，其對應的密度就是 $f(x)$。概率空間（X，Σ，μ_f）可看成是無窮樣本空間。由 Shannon 熵的啓廸，我們可以如下定義 f 的Boltzmann 熵。為此，令函數 $\eta(u)$ 定義為

$$\eta(u)=\begin{cases}-u\log u, & u>0\\ 0 & u=0\end{cases}$$

$\eta(u)$ 的圖像由圖 4-1 表示。

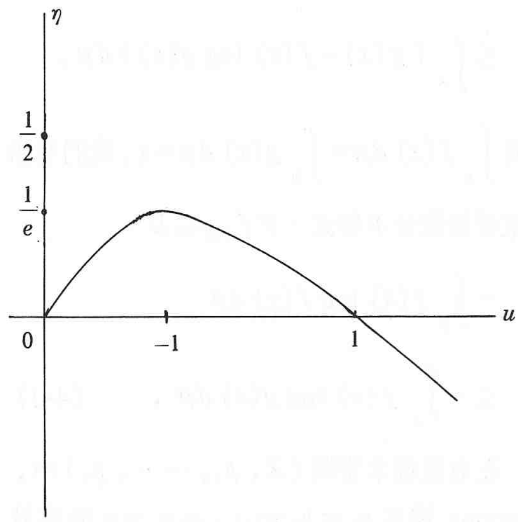

圖 4-1

定義4-1：設$f\in D$且$\eta(f)\in L'(X)$ 則f的Boltzmann 熵定義爲

$$H(f)=\int_X \eta(f(x))\,d\mu$$
$$=-\int_X f(x)\log f(x)\,d\mu$$

由$\eta(u)$定義知，$\eta'(u)=-(\log u+1)$，$\eta''(u)=-\frac{1}{u}<0$。因而η是〔$0,\infty$）上的嚴格凹函數，由Taylor 展式，任給$u,v\geqslant 0$,

$$\eta(u)=\eta(v)+\eta'(v)(u-v)+\frac{\eta''(\xi)}{2!}(u-v)^2$$
$$<\eta(v)+\eta'(v)(u-v)\text{。}$$

即，

$$-u\log u\leq -v\log v-(\log v+1)(u-v)$$

簡化之，我們便有着名的Gibbs不等式，

$$u-u\log u\leq v-u\log v\text{ 。}$$

任給函數$f,g\in D$，由Gibbs 不等式和積分的單調性，

$$\int_X (f(x)-f(x)\log f(x))\,d\mu$$
$$\leq\int_X (g(x)-f(x)\log g(x))\,d\mu\text{。}$$

由於$\int_X f(x)\,d\mu=\int_X g(x)\,d\mu=1$，我們有如下重要的積分不等式：$\forall f,g\in D$

$$-\int_X f(x)\log f(x)\,d\mu$$
$$\leq -\int_X f(x)\log g(x)\,d\mu\text{ ,}\qquad (4\text{-}1)$$

在有限樣本空間$(X,p_1,\cdots\cdots,p_n)$中，Shannon 熵在$p_1=p_2=\cdots\cdots=p_n=n$時爲最大，Boltzmann 熵在概率測度空間裡也有類似的性質。

命題4-2：

設$\mu(X)<+\infty$，則密度函數$f_0(x)\equiv\frac{1}{\mu(X)}$滿足

$$H(f_0)=\log\mu(X)=\max\{H(f):f\in D\}\text{。}$$

證明：首先易見$f_0\in D$。其次，任給$f\in D$，由不等式(4-1)，

$$H(f)=-\int_X f(x)\log f(x)\,d\mu$$
$$\leq -\int_X f(x)\log\frac{1}{\mu(X)}\,d\mu$$
$$=\log\mu(X)\int_X f(x)\,d\mu$$
$$=\log\mu(X)=H(f_0)\text{。}\qquad\#$$

爲了描述一些與Boltzmann 熵有關的條件極值問題。我們引進一些概率論常用的術語。設X爲一個隨機變量（Random Variable），即X爲某一固定樣本空間上的可測實函數。$f(x)$爲這個測度空間的密度函數，則

$$E(X)=\int_{-\infty}^{\infty} x f(x)\,dx$$

稱爲X的期望值（Expected value 或 Expectation）。而數

$$\mathrm{Var}(X)=\int_{-\infty}^{\infty}(x-E(X))^2 f(x)\,dx$$

則稱爲X的**變異數**（variance）。期望值是關於隨機變量X平均值的一個度量，變異數則表示隨機變量偏離其平均值的程度。下列性質，可以很輕易的被驗證：

(i) $E(aX+bY)=aE(X)+bE(Y)$

(ii) $\mathrm{Var}(cX)=c^2\,\mathrm{Var}(X)$

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

(iv)若X和Y「獨立（independent）」，則

$$\mathrm{Var}(X+Y)=\mathrm{Var}(X)+\mathrm{Var}(Y)。$$

設有一列獨立隨機變量 $\{X_k\}_{k\geq 1}$，$E(X_k)=m_k$，$\mathrm{Var}(X_k-m_k)=\sigma_k^2$，令

$$S_n=\sum_{k=1}^{n}(X_k-m_k)$$

則，

$$\mathrm{Var}(S_n)=\mathrm{Var}\left(\sum_{k=1}^{n}(X_k-m_k)\right)$$

$$=\sum_{k=1}^{n}\mathrm{Var}(X_k-m_k)=\sum_{k=1}^{n}\sigma_k^2$$

我們標準化 S_n，即令

$$T_n=\frac{S_n}{\sqrt{\mathrm{Var}(S_n)}}$$

則 $\qquad E(T_n)=0,\qquad \mathrm{Var}(T_n)=1$ 。

概率理論中有個非常重要的基本定理：**中心極限定理**（Central Limit Theorem）。它大概的意思是說，在漸近狀態下，通常隨機變量 T_n 的概率分佈（probability distribution）是遵循Gauss分佈規律的，也就是說，

$$\lim_{n\to\infty}P(a\leqslant T_n\leqslant b)=\frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{u^2}{2}}du$$

其中 P 爲樣本空間的概率分佈。

但是，爲什麼大家都遵循的是Gauss分佈規律，而不是其他的分佈規律呢？事實上，這和熱力學第二定律有異曲同工之妙。熱力學第二定律大致上說，自然界的規律是，一切動態系統都是在向「熵」高的方向發展。從這個角度來看，在 $E(T_n)=0$，$\mathrm{Var}(T_n)=1$ 的條件下，Gauss 分佈的確有最大的Boltzmann熵。我們用下面的命題，對這點略加說明。

記 $\overline{D}=\{f\in D:\int_{-\infty}^{\infty}xf(x)\,dx=0$，

$$\int_{-\infty}^{\infty}x^2f(x)\,dx=1\}$$

命題 4-3

設 $f_0(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$，則 $f\in D$ 且

$$H(f_0)=\max\{H(f):f\in\overline{D}\}$$

$$=\log\sqrt{2\pi}+\frac{1}{2}$$

證明：由公式 $\int_{-\infty}^{\infty}e^{-x^2}dx=\sqrt{\pi}$，易知

$$\int_{-\infty}^{\infty}f_0(x)\,dx=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\frac{x^2}{2}}dx=1。$$

即 $f_0\in D$。又由部分積分法易證

$$\int_{-\infty}^{\infty}xf_0(x)\,dx=0，以及$$

$$\int_{-\infty}^{\infty}x^2f_0(x)\,dx=1$$

即 $f_0\in\overline{D}$。任給 $f\in\overline{D}$，由不等式 (4-1)

$$H(f)=-\int_{-\infty}^{\infty}f(x)\log f(x)\,dx$$

$$\leqslant-\int_{-\infty}^{\infty}f(x)\log\left(\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\right)dx$$

$$=-\int_{-\infty}^{\infty}f(x)\left[\log\left(e^{-\frac{x^2}{2}}\right)-\log\sqrt{2\pi}\right]dx$$

$$=\log(\sqrt{2\pi})+\frac{1}{2}=H(f_0)。\qquad\#$$

類似地，記 $\overline{\overline{D}}=\{f\in D,\int_0^{\infty}xf(x)=\frac{1}{\lambda}\}$。比照上述證明，我們有

命題 4-4：

設 $f_0(x)=\lambda e^{-\lambda x}$，則 $f_0\in\overline{\overline{D}}$，且

$$H(f_0)=\max\{H(f):f\in\overline{\overline{D}}\}$$

$$=1-\log\lambda。$$

上述兩命題，可推廣到下述一般情形。設 $g\in L^{\infty}$，給定約束

$$\int_X g(x)\, f(x)\, dx = \bar{g}$$

則$H(f)$在此約束下，最大值的密度函數應爲

$$f_0(x) = e^{-rg(x)} \Big/ \int_X e^{-rg(x)}\, dx$$

其中r爲一常數。同樣，若有兩個約束

$$\int_X g_1(x)\, f(x)\, dx = \bar{g}_1 \quad 和$$

$$\int_X g_2(x)\, f(x)\, dx = \bar{g}_2$$

則密度函數

$$f_0(x) = \frac{e^{-(r_1 g_2(x) + r_2 g_2(x))}}{\int_X e^{-(r_1 g_1(x) + r_2 g_2(x))}\, dx}$$

給出了$H(f)$在這兩個約束下的最大值$H(f_0)$，其中r_1, r_2爲兩常數。更一般地，我們有

命題 4-4：

設(X, Σ, μ)爲一測度空間，非負函數$g_1, \cdots\cdots, g_m \in L^\infty(X)$及正常數$r_1, \cdots\cdots, r_m$滿足條件

$$\frac{\int_X g_i(x) \prod_{j=1}^m e^{-r_j g_j(x)}\, d\mu}{\int_X \prod_{j=1}^m e^{-r_j g_j(x)}\, d\mu} = \bar{g}_i \,,$$

$$i = 1, \cdots\cdots, m$$

則$H(f)$在約束

$$\int_X g_i(x) f(x)\, dx = \bar{g}_i \,,\ i = 1, \cdots\cdots, m$$

下，最大值的密度函數爲

$$f_0(x) = \frac{\prod_{i=1}^m e^{-r_i g_i(x)}}{\int_X \prod_{i=1}^m e^{-r_i g_i(x)}\, d\mu}$$

證明：爲簡單起見，令

$$z = \int_X \prod_{i=1}^m e^{-r_i g_i(x)}\, d\mu \text{，則}$$

$$f_0(x) = z^{-1} \prod_{i=1}^m e^{-r_i g_i(x)} \text{。不難算出，}$$

$$H(f_0) = \log z + \sum_{i=1}^m r_i \bar{g}_i \text{。}$$

任給密度函數f滿足上述約束條件，由不等式(4-1)知，

$$H(f) \le -\int_X f(x) \log\left[z^{-1} \prod_{i=1}^m e^{-r_i g_i(x)} \right] d\mu$$

$$= -\int_X f(x) \left[-\log z - \sum_{i=1}^m r_i g_i(x) \right] d\mu$$

$$= \log z + \sum_{i=1}^m r_i \bar{g}_i = H(f_0) \qquad \#$$

特別，當$m=1$時，若$g(x)$看成是系統的能量時，$f_0(x) = z^{-1} e^{-rg(x)}$恰好就是Gibbs典型分布函數，且$z = \int_X e^{-rg(x)}\, d\mu$爲其分析函數，而對應的最大熵$H(f_0) = \log z + r\bar{g}$恰好就是衆所周知的熱力學熵。

——本文作者任教於美國密西根州立大學數學系——