

概率與資訊中的分析引理 及相關定理證明

——兼談中國大陸一道高考壓軸題的數學背景

徐瀝泉 · 張建良

0. 引言

概率論與資訊學中導出某些隨機變量的重要的分佈函數時屢次要用到數學分析中的結論。本文將詳細而嚴格地論證兩條分析引理，它們是產生卜瓦松 (Poisson) 分佈，常態分佈 ($N(x, \sigma)$) 和熵與資訊中申農定理 (1948年由美國工程師 C. E. Shannon 的開創性論文 A mathematical theory of communication, Bell System Tech., 27(1948), 379-423, 623-656 發表後確立) 的重要機制。兼談 2001 年中國數學高考壓軸題 (大陸卷) 的數學背景。

1. 引理 I — 產生卜瓦松分佈和常態分佈的重要機制

1.1. 引理 I 與卜瓦松分佈

引理 I: 若 $f(x)$ 是嚴格單調函數和連續函數且 $\forall x, y \in R, f(x) \cdot f(y) = f(x+y)$ 成立, 則 $f(x) = a^x, a > 0$ 且 $a \neq 1$ 。

證明: $f(x) = f\left(\frac{x}{2} + \frac{x}{2}\right) = f\left(\frac{x}{2}\right) \cdot f\left(\frac{x}{2}\right) = f^2\left(\frac{x}{2}\right) \geq 0$,

同理 $f(2x) = f^2(x), f(3x) = f^3(x), \dots, f(nx) = f^n(x), n \in N$ 。

令 $x = \frac{1}{n}$ 得 $f(1) = f^n\left(\frac{1}{n}\right)$ 即 $f^{\frac{1}{n}}(1) = f\left(\frac{1}{n}\right)$ 或 $f\left(\frac{1}{n}\right) = f^{\frac{1}{n}}(1)$ 。

設 $f(1) = a \Rightarrow f\left(\frac{1}{n}\right) = a^{\frac{1}{n}}$ 顯然 $a \geq 0$, 下面我們來詳細討論 a 的取值範圍:

$\therefore f(0) = f(0+0) = f^2(0), \therefore f(0) = 0$ 或 $f(0) = 1$ 。

若 $f(0) = 0$ 引理的結論成立, 則有 $a^0 = 0$ 這樣的 a 不存在, 故 $f(0) \neq 0$;

若 $f(0) = 1$ 引理的結論成立, 則有 $a^0 = 1$, 則 a 可取 0 以外的任何數。但由題設中的單

調性, 當 $f(x)$ 是增函數時 $\because 1 > 0, \therefore f(1) > f(0)$ 即 $a > 1$; 當 $f(x)$ 是減函數時 $\because 1 > 0, \therefore f(1) < f(0)$ 即 $a < 1$, 由此可以確定 $a > 0$ 且 $a \neq 1$ 。

從上可知 $\forall m \in N, f\left(m \cdot \frac{1}{n}\right) = f^m\left(\frac{1}{n}\right)$ 即 $f\left(\frac{m}{n}\right) = a^{\frac{m}{n}}$ 。

由 m 和 n 的任意性可知對一切有理數 $\alpha, f(\alpha) = a^\alpha$ 成立, 再利用單調性和連續性可證對無理數也成立, 從而引理 I 獲證。

下面我們來看所謂的卜瓦松分佈或常態分佈。若研究對象滿足:

- (1) 過程的平穩性 (即它的概率規律不隨時間的推移而改變);
- (2) 獨立增量性 (無後效性, 即在互不相交的時間間隔內過程進行的相互獨立性);
- (3) 普通性 (在同一時間瞬間事件出現有兩次或兩次以上的情形實際上是不可能的)

則這些隨機變量皆服從卜瓦松分佈:

$$p_k(x) = \frac{(\lambda x)^{kx}}{k!} e^{-\lambda x}, \quad k = 0, 1, 2, \dots$$

其中 x 為時間間隔, k 為在區間 $[x_0, x_0 + x]$ 內事件出現的次數, 證略 (詳見參考文獻 [2], pp. 97-100), 其初始值: 上式中 $k = 0$ 時 $p_0(x) = e^{-\lambda x}$ 的由來必須要用到引理 I, 限於篇幅這裡也不再詳述 (請參見文 [1])。

作為引理 I 的一個系 (Corollary), 接著我們來看 2001 年中國大陸高考數學試卷的一道壓軸題 (第 22 題) 表述:

設 $f(x)$ 是定義在 R 上的偶函數, 其圖像關於直線 $x = 1$ 對稱, $\forall x_1, x_2 \in \left[0, \frac{1}{2}\right]$ 都有 $f(x_1 + x_2) = f(x_1) \cdot f(x_2)$ 且 $f(1) = a > 0$ 。

- (1) 求 $f\left(\frac{1}{2}\right)$ 及 $f\left(\frac{1}{4}\right)$;
- (2) 證明 $f(x)$ 是周期函數;
- (3) 記 $a_n = f\left(2n + \frac{1}{2n}\right)$ 求 $\lim(\ln a_n)$ 。

事實上, 在指數函數 $y = a^x$ 的圖像上我們截取 $x \in [0, 1]$ 的一段, 再作出關於 y 軸和直線 $x = 1$ 都對稱的圖像, 因為它定義在 R 上, 如此反復構作下去, 就是把 $y = a^{|x|}, x \in [-1, 1]$, 的圖像以 2 為周期延拓到整個上半平面上。

這就是該試題的高等數學背景。

1.2. 引理 I 的推廣與常態分佈

對於指數函數的運算性質 $a^x \cdot a^y = a^{x+y}$ 這是人所共知的, 但是反過來具有此運算性質的函數是否也一定是指數函數, 這正是引理 I 所解決的問題。引理 I 還具有下面推廣了的形式:

引理 I 的推廣: 若 $f(x)$ 和 $g(y)$ ($x \geq 0, y \geq 0$) 是不恆等於零的連續函數, 且

$$\forall x, y, \quad f(x)g(y) = h(x+y) \text{ 成立} \quad (\Delta)$$

則 $f(x) = ka^x$, 這裡 $a > 0$ 及 k 都是常數。

證明: 用反證法易知 $f(0) \neq 0, g(0) \neq 0$, 在 (Δ) 中分別令 $x = z, y = 0$ 和 $y = z, x = 0$ 可得 $f(z)g(0) = f(0)g(z) = h(z)$ 兩邊同除以 $f(0)g(0)$ 得

$$\frac{f(z)}{f(0)} = \frac{g(z)}{g(0)} = \frac{h(z)}{f(0)g(0)} = p(z)$$

這裡我們把它們的比值記為 $p(z)$, 則可推得

$$f(x) = f(0)p(x), \quad g(y) = g(0)p(y), \quad h(x+y) = f(0)g(0)p(x+y)$$

把它們代入 (Δ) 得 $p(x)p(y) = p(x+y)$ 由引理知 $p(x) = a^x, \therefore f(x) = f(0)a^x$ 證畢。

它正是導出常態分佈密度函數 $p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2}$ 必須要用到的重要條件之一。我們來看一個實例:

射擊問題中彈落點的分佈: 在射擊問題中, 以射擊目標中心為原點, 作平面直角坐標系, 則射擊過程中彈落點的座標 (x, y) 是一個二維隨機變量。假設它滿足:

- (i) x 和 y 具有連續的密度函數 $p_1(x)$ 和 $p_2(y)$;
- (ii) x 和 y 相互獨立;
- (iii) (x, y) 的密度函數在點 (x, y) 的值僅與它到原點的距離 $r = \sqrt{x^2 + y^2}$ 有關, 則由此引理可證 x 和 y 均服從常態分佈, 證明過程如下:

證明: 記 (x, y) 的密度函數為 $p(x, y)$, 由假設 (iii) 知 $p(x, y) = q(\sqrt{x^2 + y^2})$ 又由 (i) 與 (ii) 知 $p(x, y) = p_1(x)p_2(y) = q(\sqrt{x^2 + y^2})$, 令 $p_1(x) = f(x^2), p_2(y) = g(y^2), q(\sqrt{x^2 + y^2}) = h(x^2 + y^2)$, 則有 $f(x^2)g(y^2) = h(x^2 + y^2)$, 故由上述引理可得 $f(x^2) = ka^{x^2}$ 。由於 $p_1(x)$ 是 $(-\infty, +\infty)$ 上的密度函數, 不妨設 $a = \frac{1}{e^{\frac{1}{2\sigma_1^2}}}$ ($\sigma_1 > 0$), 由

$$\int_{-\infty}^{+\infty} p_1(x)dx = 1 \Rightarrow \int_{-\infty}^{+\infty} ke^{-\frac{1}{2}\left(\frac{x}{\sigma_1}\right)^2} = k \cdot \sqrt{2\pi}\sigma_1 = 1 \Rightarrow k = \frac{1}{\sqrt{2\pi}\sigma_1}$$

於是 $p_1(x) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{1}{2}\left(\frac{x}{\sigma_1}\right)^2}$; 同理 $p_2(y) = \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{1}{2}\left(\frac{y}{\sigma_2}\right)^2}$ 。

這就證明了 x 與 y 皆服從常態分佈 (詳見 [2]第 141 頁 line -4 至第 142 頁 line-8)。

2. 引理 II 與申農定理

由上述引理 I 和它的推廣及其應用, 我們自然會聯想到對數函數是否也有類似的結果呢? 無獨有偶, 我們先看下面的引理 II, 它是用來證明申農定理的基礎。

2.1. 引理 II

若 $f(x)$ 是單調函數或連續函數, 且 $\forall x, y \in R^+$ 成立 $f(xy) = f(x) + f(y)$; 則

$$f(x) = c \log_a x, \quad a > 0, a \neq 1, c > 0$$

其中 a, c 是常數。

證明與引理 I 相似, 讀者不妨自行練習一下。

2.2. 申農定理的直覺與證明

定理: 若用 $h(\alpha)$ 表示隨機試驗 α 的不肯定性程度之大小, α 只有有限個不相容的結果 $A_i (i = 1, 2, \dots, n)$, 其相應的概率為 $p(A_i)$, 為方便起見有時也簡記為 p_i , 則當且僅當:

- (1) $h(\alpha)$ 是 $p(A_i)$ 的連續函數;
- (2) 對有 n 個等可能性結果的試驗, $h(\alpha)$ 是 n 的單調上升函數;
- (3) 一個試驗可分解為兩個相繼的試驗, 且未分之前的 $h(\alpha)$ 是既分之後的 $h(\alpha)$ 的加權和。則 $h(\alpha)$ 具有下列形式:

$$h(\alpha) = -c \sum_{i=1}^n p(A_i) \log_a p(A_i), \quad \text{其中 } c > 0 \text{ 委為常數}$$

定理中的係數 c 的選擇取決於度量單位, 常用的度量單位有二進位和十進位兩種, 前者取以 2 為底的對數, 後者取常用對數。其單位分別叫做“比特”和“哈特利”; 也有取自然對數的, 其單位稱之為“奈特”。我們稱

$$h(\alpha) = - \sum p(A_i) \lg p(A_i)$$

為某隨機事件的熵, 這裡約定用十進位單位。

比如有三位射手, 他們的射擊情況分別用 $\alpha_i (i = 1, 2, 3)$ 來描述:

$$\alpha_1 : \begin{bmatrix} A_1 & \bar{A}_1 \\ 0.5 & 0.5 \end{bmatrix} \quad \alpha_2 : \begin{bmatrix} A_2 & \bar{A}_2 \\ 0.9 & 0.1 \end{bmatrix} \quad \alpha_3 : \begin{bmatrix} A_3 & \bar{A}_3 \\ 0.7 & 0.3 \end{bmatrix}$$

其中 A 表示擊中目標, \bar{A} 表示未擊中目標, 下面的數是相應的概率。則由申農定理, 甲、乙、丙三人射擊試驗相應的熵分別為:

$$h(\alpha_1) = -0.5 \lg 0.5 - 0.5 \lg 0.5 \approx 0.3010$$

$$h(\alpha_2) = -0.9 \lg 0.9 - 0.1 \lg 0.1 \approx 0.1412$$

$$h(\alpha_3) = -0.7 \lg 0.7 - 0.3 \lg 0.3 \approx 0.2653$$

擊中目標的不肯定性程度甲最大，丙次之，乙最小，完全與直觀相符。

在我們正式證明定理之前，讓我們對定理中的題設條件再作一些直觀的說明與解釋。條件 (1) 說明 $p(A_i)$ 的微小變化不會引起 $h(\alpha)$ 巨大變化，同時也只有連續函數才便於數學上的處理；條件 (2) 說明我們考慮的試驗是一種特殊的試驗，這種試驗只有 n 個結果，且每一個結果出現的概率均為 $1/n$ ，我們把這種試驗稱之為有 n 個等概結果的試驗，在這種試驗中 $h(\alpha)$ 只是關於 n 的函數，並且當 n 增大時，也即試驗有更多個可能的結果時，其相應的不肯定性程度也隨之增大；條件 (3) 則比較難以理解，它涉及到把一個試驗分解為兩個試驗，這時未分之前的 $h(\alpha)$ 是既分之後的 $h(\alpha)$ 的加權和。還是先讓我們來看一個具體的例子：

考慮有 6 個等概結果的試驗

$$\alpha : \begin{bmatrix} A_i \\ p(A_i) \end{bmatrix}, \quad p(A_i) = \frac{1}{6}, \quad i = 1, 2, \dots, 6$$

即

$$\alpha : \begin{bmatrix} A_1 & A_2 & A_3 & A_4 & A_5 & A_6 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{bmatrix}$$

其不肯定性程度 $h(\alpha) = \left[\frac{1}{6} \ \frac{1}{6} \ \frac{1}{6} \ \frac{1}{6} \ \frac{1}{6} \ \frac{1}{6} \right]$ 不妨簡記為 $h(6)$ 。

(注：對 n 個等概結果的試驗，其不肯定性程度也可簡記為 $h(n)$)。我們認為它也可以先通過試

驗 $\alpha_0 : \begin{bmatrix} A_1 \cup A_2 \cup A_3 & A_4 \cup A_5 \cup A_6 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$ 把 α 分解為 2 個有 3 個等概結果的試驗

$$\alpha_1 : \begin{bmatrix} A_1 & A_2 & A_3 \\ \frac{1}{6}/\frac{1}{2} & \frac{1}{6}/\frac{1}{2} & \frac{1}{6}/\frac{1}{2} \end{bmatrix} \quad \alpha_2 : \begin{bmatrix} A_4 & A_5 & A_6 \\ \frac{1}{6}/\frac{1}{2} & \frac{1}{6}/\frac{1}{2} & \frac{1}{6}/\frac{1}{2} \end{bmatrix} \quad \text{即}$$

$$\alpha_1 : \begin{bmatrix} A_1 & A_2 & A_3 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} \quad \alpha_2 : \begin{bmatrix} A_4 & A_5 & A_6 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} \quad \text{其相應的不肯定性程度為}$$

$h(\alpha_0) = \left(\frac{1}{2}, \frac{1}{2} \right)$, $h(\alpha_1) = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3} \right) = h(\alpha_2)$, 並分別簡記為 $h(2)$, $h(3)$, $h(3)$ 。顯然 α_0 出現的概率是 1, α_1 和 α_2 出現的概率都是 1/2, 故

$$h(\alpha) = h(\alpha_0) + \frac{1}{2}h(\alpha_1) + \frac{1}{2}h(\alpha_2), \quad h(6) = h(2 \cdot 3) = h(2) + 2 \cdot \frac{1}{2}h(3).$$

當然，我們也可以把試驗 α 分解為另兩個相繼的試驗。第一個試驗以 $4/6$ 的概率出現 $t_1 = \bigcup_{i=1}^4 A_i$, 和以 $2/6$ 的概率出現 $t_2 = \bigcup_{i=5}^6 A_i$; 而第二個試驗則是在出現結果 t_1 的基礎上考察它出

現 $A_1 \sim A_4$ 中的哪一個, 或在出現 t_2 的基礎上到底出現 $A_5 \sim A_6$ 中的哪一個, 即

$$\alpha'_0 : \left(\begin{array}{c} t_1 = \bigcup_{i=1}^4 A_i \\ \frac{4}{6} \end{array} \right) t_2 = \bigcup_{i=5}^6 A_i \quad \alpha'_1 : \left(\begin{array}{cccc} A_1 & A_2 & A_3 & A_4 \\ \frac{1}{6}/\frac{4}{6} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{array} \right) \alpha'_2 : \left(\begin{array}{cc} A_5 & A_6 \\ \frac{1}{6}/\frac{2}{6} & \frac{1}{2} \end{array} \right),$$

其中 α'_1 出現的概率是 $4/6$, α'_2 出現的概率是 $2/6$, 當然 α'_0 出現的概率仍為 1 。故

$$\begin{aligned} h(\alpha) &= h\left(\frac{4}{6}, \frac{2}{6}\right) + \frac{4}{6}h(\alpha'_1) + \frac{2}{6}h(\alpha'_2), \text{ 即} \\ h\left(\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}\right) &= h\left(\frac{4}{6}, \frac{2}{6}\right) + \frac{4}{6}h\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right) + \frac{2}{6}h\left(\frac{1}{2}, \frac{1}{2}\right) \text{ 或簡記為} \\ h(6) &= h\left(\frac{4}{6}, \frac{2}{6}\right) + \frac{4}{6}h(4) + \frac{2}{6}h(2). \text{ 由此可求得} \\ h\left(\frac{4}{6}, \frac{2}{6}\right) &= h(6) - \frac{4}{6}h(4) - \frac{2}{6}h(2). \end{aligned}$$

這後面的一個關係式正是我們所需要的。有了這些直觀的解釋和印象之後, 我們不難完成下面申農定理的證明。

證明: 首先證有 n 個等概結果的試驗 $\alpha : \left(\begin{array}{cccc} A_1 & A_2 & \cdots & A_n \\ \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \end{array} \right)$

記 $h(\alpha) = h\left(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}\right)$ 或簡記為 $h(n)$, 它表示隨機試驗 α 的不肯定性程度之大小, 由題設它是關於 n 的單調上升函數。

現對於有 mn 個等概結果的試驗, 可以把它分解為有 m 個有 n 個等概結果的試驗, 若滿足條件 (3) 應有 $h(mn) = h(m) + m \cdot \frac{1}{m}h(n) = h(m) + h(n)$, 由引理 II

$$h(n) = h\left(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}\right) = c \log_a n \quad (*)$$

其次, 當 $p(t_i)$ 是 $(0, 1]$ 中的任何有理數時, 即隨機試驗 A_0 只有有限個不相容的結果 t_i , 它們分別以概率 $p(t_i)$ (簡記為 p_i) 出現。

$$A_0 : \left(\begin{array}{cccc} t_1 & t_2 & \cdots & t_n \\ p_1 & p_2 & \cdots & p_n \end{array} \right) \text{ (注: 對應於上述直觀例子中的 } \left(\begin{array}{cc} t_1 & t_2 \\ \frac{4}{6} & \frac{2}{6} \end{array} \right))$$

其中 $p_i (i = 1, 2, \dots, n)$ 為 $(0, 1]$ 中的任何有理數, 不妨記 $p_i = \frac{n_i}{\sum_{i=1}^n n_i}$ 。由概率定義 $p_i \in (0, 1]$ 且 $\sum_{i=1}^n p_i = 1$, 因此上式的定義是合理的。如所知, 任一有理數皆可化為分數 $\frac{n}{m} (m, n \in N)$ 。現在的情形下 $m \geq n$, 不妨設 $m - n = k (k = 0, 1, \dots)$, 則 $\frac{n}{m} = \frac{n}{n+k}$ 。現取 $n = n_i$, $k = \sum_{i=1}^n n_i$, 考慮一個有 $\sum_{i=1}^n n_i$ 個 (上述直觀例子中有 $n_1 + n_2 = 4 + 2 = 6$ 個) 等概結果的試驗 A , 而這個試驗又可以看作兩個相繼的試驗, 其中第一個試驗以概率 p_i (上面直觀例

子中的 $p_1 = \frac{4}{6}, p_2 = \frac{2}{6}$) 出現結果 t_i (上述直觀例子中的 t_1 和 t_2); 第二個試驗則是在出現 t_i 的基礎上, 考察它出現 n_i (上面直觀例子中的 $n_1 = 4, n_2 = 2$) 個等概結果中的哪一個 (上面直觀例子中 $n_1 = 4$ 個等概結果 $\bigcup_{i=1}^4 A_i$ 或 $n_2 = 2$ 個等概結果 $\bigcup_{i=5}^6 A_i$ 中的哪一個)。因此按條件 (3) 應有 $h(A) = h(p_1, p_2, \dots, p_n) + \sum_{i=1}^n p_i h(A_i)$ 其中

$$h(A) = h\left(\frac{1}{\sum_{i=1}^n n_i}, \frac{1}{\sum_{i=1}^n n_i}, \dots, \frac{1}{\sum_{i=1}^n n_i}\right)$$

$$h(A_0) = h(p_1, p_2, \dots, p_n) = h\left(\frac{n_1}{\sum_{i=1}^n n_i}, \frac{n_1}{\sum_{i=1}^n n_i}, \dots, \frac{n_1}{\sum_{i=1}^n n_i}\right)$$

$$\begin{aligned} \sum_{i=1}^n p_i h(A_i) &= \frac{n_1}{\sum_{i=1}^n n_i} h\left(\frac{1}{n_1}, \frac{1}{n_1}, \dots, \frac{1}{n_1}\right) + \frac{n_2}{\sum_{i=1}^n n_i} h\left(\frac{1}{n_2}, \frac{1}{n_2}, \dots, \frac{1}{n_2}\right) + \dots \\ &\quad + \frac{n_n}{\sum_{i=1}^n n_i} h\left(\frac{1}{n_n}, \frac{1}{n_n}, \dots, \frac{1}{n_n}\right) \end{aligned}$$

其中, $\frac{1}{n_i} = \frac{\frac{1}{\sum_{i=1}^n n_i}}{p_i} = \frac{1}{\sum_{i=1}^n n_i} \cdot \frac{\sum_{i=1}^n n_i}{n_i}$ 。由 (*) 得

$$c \log_a \sum_{i=1}^n n_i = h(p_1, p_2, \dots, p_n) + c \sum_{i=1}^n p_i \log_a n_i \text{ 於是}$$

$$\begin{aligned} h(p_1, p_2, \dots, p_n) &= c \log_a \sum_{i=1}^n n_i - c \sum_{i=1}^n p_i \log_a n_i = c \left(\log_a \sum_{i=1}^n n_i - \sum_{i=1}^n p_i \log_a n_i \right) \\ &= c \left[\sum_{i=1}^n p_i \left(\log_a \sum_{i=1}^n n_i - \log_a n_i \right) \right] = c \left[\sum_{i=1}^n p_i \left(\log_a \frac{\sum_{i=1}^n n_i}{n_i} \right) \right] \\ &= c \left[\sum_{i=1}^n p_i (\log_a (p_i)^{-1}) \right] = -c \sum_{i=1}^n p_i \log_a p_i \end{aligned}$$

$$\text{即, } h(p_1, p_2, \dots, p_n) = -c \sum_{i=1}^n p_i \log_a p_i$$

最後, 對任意的 p_i 可用有理數來逼近它, 這由條件 (1) 可以確保, 從而申農定理獲證。

參考文獻

1. 徐瀝泉, 略論隨機性, 「數學傳播」, 第25卷4期, 民90年12月。
2. 復旦大學編, 概率論, 第一冊, 高等教育出版社, 1979年4月第一版。

— 本文作者任職於中國無錫市教育研究中心 —