

數學建模講義

第 4 講：經驗模型、插值與樣條

Kenneth Sok Kin Cheng

修訂日期：2026 年 8 月 20 日

學習目標

完成本課後，讀者應能：

1. 區分插值、平滑與外推；
2. 建立並解釋兩個觀測值之間的線性插值式；
3. 以拉格朗日形式建立二次插值多項式；
4. 利用差分表辨認簡單多項式結構；
5. 解釋為何精確的高次插值式在數據點之間可能表現欠佳；
6. 比較分段線性插值與自然三次樣條，並選擇合適的經驗方法。

1 先有數據、後有公式

第 3 講先提出模型族，再由數據估計參數。然而，在許多實際情況中，我們尚未掌握可信機制，可能只有若干時間或位置的測量，並需要估計它們之間的合理數值。這便引出經驗模型：主要由觀測數據建立的數學表述。

引入活動：曲線應否通過每個點？

[10 分鐘]

課室感測器每五分鐘記錄一次二氧化碳濃度：

時間 (min)	0	5	10	15	20
CO ₂ (ppm)	612	648	701	694	760

二人一組討論：

1. 你會如何估計第 12 分鐘的濃度？
2. 估計曲線應否通過每個記錄值？
3. 在這裏，測量雜訊與完全吻合觀測值，哪一項較重要？
4. 用同一方法預測第 60 分鐘是否有充分理據？

建模要點：模型用途與數據可靠性決定精確插值是否可取。

插值、平滑與外推

- 插值建立與指定數據值相符的函數，用以估計觀測範圍內的數值。
- 平滑容許與含雜訊數據有少量偏差，以揭示底層趨勢。
- 外推預測觀測範圍外的數值，依賴單憑數據無法核證的假設。

若觀測含雜訊，或曲線在觀測值之間的行為不合理，插值式即使在數學上精確，在科學上仍

可能不可信。

模型檢視：完全吻合不等於可信

若溫度計只能準確至最接近的整數度，強制複雜曲線通過每個報告值，可能是在模擬捨入雜訊而非溫度。反之，若數據是電腦圖形中的精確幾何坐標，通過每個點可能不可或缺。方法必須由數據生成過程指導。

2 線性插值與分段線性模型

假設兩個可信觀測值為 (x_0, y_0) 與 (x_1, y_1) ，且 $x_0 < x_1$ 。對兩者之間的點 x ，定義相對位置

$$\lambda = \frac{x - x_0}{x_1 - x_0}, \quad 0 \leq \lambda \leq 1.$$

線性插值式為

$$L(x) = (1 - \lambda)y_0 + \lambda y_1 = y_0 + \frac{y_1 - y_0}{x_1 - x_0}(x - x_0).$$

此公式是加權平均： x 越接近 x_0 ， y_0 的權重越大。

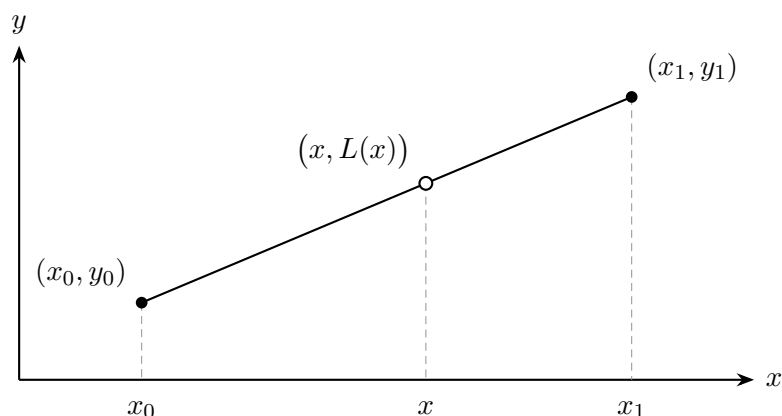


圖 1: 插值點位於連接兩個觀測值的線段上；其橫向位置決定 y_0 與 y_1 的權重。

例題 1：估計兩次讀數之間的溫度

某加熱實驗在 $t = 4$ 分鐘時記錄 38.0°C ，在 $t = 10$ 分鐘時記錄 62.0°C 。估計 $t = 7.5$ 分鐘時的溫度。

相對位置為

$$\lambda = \frac{7.5 - 4}{10 - 4} = \frac{3.5}{6}.$$

因此

$$T(7.5) = 38.0 + \frac{3.5}{6}(62.0 - 38.0) = 38.0 + 14.0 = \boxed{52.0^\circ\text{C}}.$$

此估計假設溫度在這六分鐘內以近似固定速率改變，並非聲稱整個加熱過程都是線性的。

有多個觀測值時，可用線段連接每組相鄰點。所得分段線性插值式連續且易於計算，但斜率會在數據點突然改變。

引導練習：建立分段模型

[10分鐘]

某圖書館記錄已佔用座位數：

$$(0, 20), \quad (2, 44), \quad (5, 68),$$

其中 x 為開館後的小時數。

1. 寫出 $0 \leq x \leq 2$ 上的線性公式。
2. 寫出 $2 \leq x \leq 5$ 上的線性公式。
3. 估計 $x = 1.5$ 與 $x = 3.5$ 時的佔用數。
4. 比較兩段斜率。斜率改變在情境中代表甚麼？

檢查點：插值還是外推？

對 $t = 4$ 與 $t = 10$ 分鐘的溫度觀測，把每項要求分類：

1. 估計 $T(8)$ ；
2. 估計 $T(15)$ ；
3. 估計一小時後的最高溫度。

解釋為何只有一項要求可單憑線性插值回答。

3 通過三點的二次插值

兩點確定一條直線；三個 x 坐標互異的點確定唯一一個次數不超過二的多項式。拉格朗日形式可直接建立此多項式。

對三點 (x_0, y_0) 、 (x_1, y_1) 與 (x_2, y_2) ，

$$P_2(x) = y_0 L_0(x) + y_1 L_1(x) + y_2 L_2(x),$$

其中

$$L_0(x) = \frac{(x - x_1)(x - x_2)}{(x_0 - x_1)(x_0 - x_2)},$$

$$L_1(x) = \frac{(x - x_0)(x - x_2)}{(x_1 - x_0)(x_1 - x_2)},$$

$$L_2(x) = \frac{(x - x_0)(x - x_1)}{(x_2 - x_0)(x_2 - x_1)}.$$

每個基多項式均滿足 $L_i(x_i) = 1$ ，並在另外兩個節點取零。

例題 2：二次插值式

建立通過下列三點的多項式：

$$(0, 1), \quad (1, 3), \quad (2, 2).$$

基函數為

$$L_0(x) = \frac{(x - 1)(x - 2)}{2}, \quad L_1(x) = -x(x - 2), \quad L_2(x) = \frac{x(x - 1)}{2}.$$

因此

$$\begin{aligned} P_2(x) &= L_0(x) + 3L_1(x) + 2L_2(x) \\ &= -\frac{3}{2}x^2 + \frac{7}{2}x + 1. \end{aligned}$$

當 $x = 1.5$ 時，

$$P_2(1.5) = \boxed{2.875}.$$

代入可運算核證 $P_2(0) = 1$ 、 $P_2(1) = 3$ 與 $P_2(2) = 2$ 完全成立。

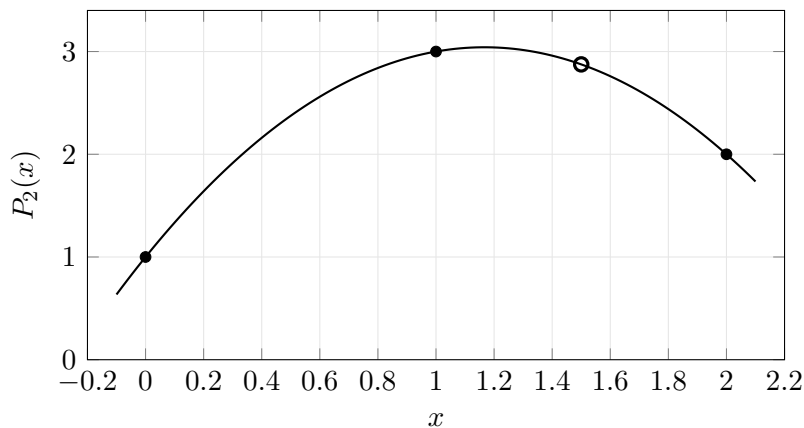


圖 2: 實線曲線為 $P_2(x) = -1.5x^2 + 3.5x + 1$ 。實心點是插值節點；空心點是 $x = 1.5$ 時的估計。

模型檢視：點與點之間的曲線代表甚麼？

插值條件唯一確定二次曲線，卻不能證明真實過程為二次。不同物理機制可以產生相同三個測量，卻在其間取不同數值。插值提供方便曲線；科學解釋仍須依據情境。

4 以差分表作診斷

對等距的 x 值，逐階差分可揭示低次多項式結構：

- 一階差分固定，提示線性模型；
- 二階差分固定，提示二次模型；
- 三階差分固定，提示三次模型。

對完全符合多項式的數據，此結果精確成立；對含雜訊觀測，則只近似成立。

例題 3：辨認二次模式

考慮

x	0	1	2	3	4
y	2	5	10	17	26

差分表為

y	2	5	10	17	26
Δy		3	5	7	9
$\Delta^2 y$			2	2	2

二階差分固定，故數據落在二次曲線上。當間距為一，首項係數是固定二階差分的一半，故

$a = 1$ 。寫成

$$y = x^2 + bx + c$$

再利用 $y(0) = 2$ 與 $y(1) = 5$ ，得 $c = 2$ 及 $b = 2$ 。因此

$$y = x^2 + 2x + 2.$$

差分表能做與不能做的事

差分表是診斷工具，而非真理證書。對測量數據，高階差分往往會放大雜訊。此方法最適合從準確且等距的數據辨認簡單模式。對不等距節點，應使用均差；附錄 A 提供選讀介紹。

模式還是雜訊？

[8 分鐘]

對每個數列計算足夠階數的差分，以提出簡單模型。

1. 4, 9, 16, 25, 36

2. 3.0, 6.1, 11.9, 21.2, 34.0

對第二個數列，解釋為何「近似二次」比「完全二次」更有理據。

5 高次多項式為何可能失效

給定 $n + 1$ 個相異點，存在唯一一個次數不超過 n 且通過所有點的多項式。此定理只保證存在性，並不保證良好行為。次數增加時，全域多項式可在節點之間劇烈振盪，尤其接近端點之處。

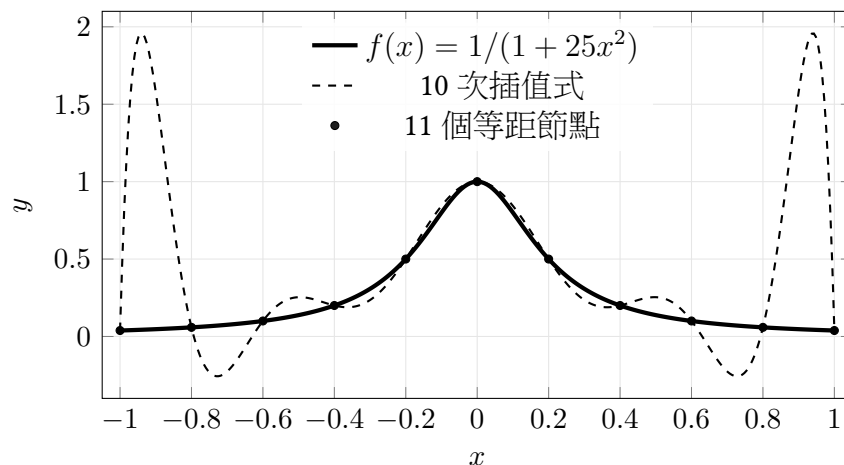


圖 3: 多項式吻合每個節點，卻在節點之間嚴重振盪；這稱為龍格現象。

模型檢視：零殘差陷阱

在插值節點上，每個殘差都是零。因此，只在這些節點計算的 SSE、MAE 與 RMSE 亦為零，即使曲線在節點之間表現欠佳。評估模型必須考慮觀測之間的行為、對數據擾動的敏感度，以及曲線的預定用途。

可行補救方法包括：減少自由度、平滑含雜訊數據、選擇分佈較佳的節點，或以分段低次多項式取代單一全域多項式。

6 分段插值與三次樣條

樣條在每對相鄰節點之間使用不同多項式。線性樣條以直線段連接各點；三次樣條使用三次多項式片段，並施加平滑連接條件。

對三次樣條 S ：

1. S 通過每個數據點；
2. 相鄰片段在每個內部節點取相同值；
3. 一階導數相同，因此沒有尖角；
4. 二階導數相同，因此曲率平滑改變。

自然三次樣條另在兩個端點滿足 $S'' = 0$ 。

例題 4：通過三點的自然樣條

建立通過下列三點的自然三次樣條：

$$(0, 0), \quad (1, 1), \quad (2, 0).$$

令 $M_i = S''(x_i)$ 。自然端點條件給出 $M_0 = M_2 = 0$ 。當等距 $h = 1$ 時，唯一的內部樣條方程為

$$M_0 + 4M_1 + M_2 = 6(y_2 - 2y_1 + y_0) = -12,$$

故 $M_1 = -3$ 。所得樣條為

$$S(x) = \begin{cases} -\frac{1}{2}x^3 + \frac{3}{2}x, & 0 \leq x \leq 1, \\ 1 - \frac{3}{2}(x-1)^2 + \frac{1}{2}(x-1)^3, & 1 \leq x \leq 2. \end{cases}$$

在 $x = 0.5$ 時，樣條給出 $S(0.5) = 0.6875$ ，線性插值則給出 0.5。兩者都不會自動成為「正確」答案；較平滑的三次估計隱含對斜率變化更強的假設。

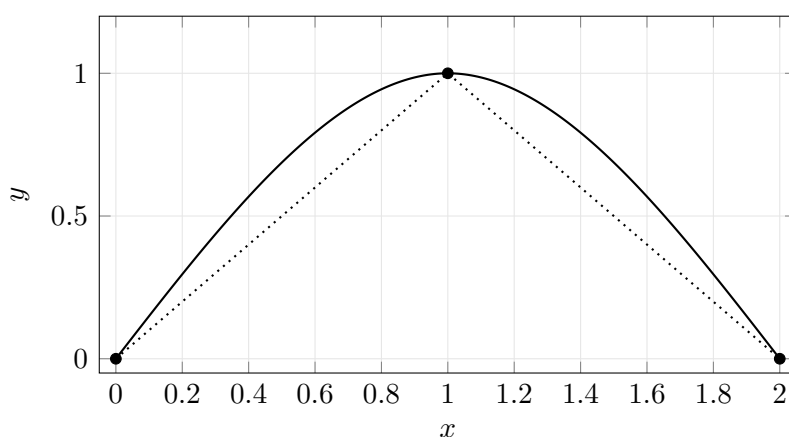


圖 4：點線為線性樣條，實線曲線為自然三次樣條，實心點為節點。兩者插值同一組數據，但只有三次樣條在 $x = 1$ 處的斜率與曲率連續。

7 選擇經驗方法

方法	優點	主要注意事項
線性插值	透明、局部、易於計算	節點有尖角；假設每段速率固定
二次或低次插值	以少量點捕捉曲率	小型數據集不能證明多項式機制
高次全域多項式	通過每個點	振盪、敏感且外推表現差
低次平滑模型	處理含雜訊數據並揭示趨勢	不會完全重現每個觀測
三次樣條	局部、平滑，在多個可信點之間準確	可能過衝；邊界條件很重要

檢查點：哪項性質最重要？

在勘測點測量道路剖面，用以估計車輛運動。你會選擇線性樣條還是三次樣條？說明所選方法的一項優點與一項風險。

模型檢視：三次樣條不會自動保持形狀

三次樣條平滑且通常穩定，但可能在數據點之間過衝，甚至在建模量必須非負時產生負值。對單調或有界數據，專門保持形狀的樣條可能更合適。

方法選擇工作坊

[12分鐘]

為每種情況選擇合適方法，並用一至兩句話論證。

1. 以精確坐標定義製成部件的邊緣。
 2. 每日需求數據含大量隨機變異，目標是揭示每週趨勢。
 3. 溫度感測器在兩個可靠讀數之間失靈兩分鐘。
 4. 十個可信測量描述平滑的道路高程剖面。
 5. 小組希望用五年數據預測其後十年。
- 對最後一種情況，解釋為何任何插值方法單獨使用都不足夠。

實用經驗建模流程

1. 釐清任務是插值、平滑還是外推。
2. 繪畫原始數據，檢查單位、次序、缺失值與離群值。
3. 決定應把觀測視為精確還是含雜訊。
4. 從符合目的的最簡單局部方法開始。
5. 檢查點與點之間的曲線，而非只檢查觀測點。
6. 測試移除或擾動一個數據點時的敏感度。
7. 明確陳述有效區間與任何形狀約束。

8 課堂總結

應掌握的六個觀念

1. 插值估計數據範圍內的數值；外推需要額外假設。
2. 線性插值是兩個觀測值的加權平均。
3. 三個相異節點確定唯一二次曲線，卻未必確定真實機制。
4. 差分表可從等距數據揭示簡單多項式模式。
5. 插值節點上的零殘差不能保證節點之間表現良好。
6. 樣條以平滑、局部、低次片段取代一個不穩定的高次多項式。

檢查點：課末小結

[2分鐘]

完成以下兩句：

1. 「當_____時，我會選擇插值而非平滑。」
2. 「通過每個點的曲線仍可能欠佳，因為_____。」

9 練習題

核心練習

1. 某感測器在 $t = 3$ 時讀數為 18.2 單位，在 $t = 9$ 時為 25.4 單位。
 - (a) 用線性插值估計 $t = 5.5$ 時的數值。
 - (b) 陳述計算背後的假設。
 - (c) 解釋為何不應自動在 $t = 15$ 時使用同一公式。
2. 建立通過 $(0, 2)$ 、 $(2, 6)$ 與 $(5, 3)$ 的分段線性插值式。計算 $x = 1$ 、 $x = 3$ 及 $x = 4.5$ 時的值，並指出斜率改變之處。
3. 利用拉格朗日形式求通過 $(0, 2)$ 、 $(1, 1)$ 與 $(3, 5)$ 的二次插值多項式。計算 $x = 2$ 時的值，並運算核證三個插值條件。
4. 為下列數列建立差分表：

$7, 12, 19, 28, 39, 52.$

 提出多項式次數，並求以 $n = 0, 1, 2, \dots$ 表示的公式。
5. 某 12 次多項式完全通過 13 個實驗測量。提出三項理由，說明為何低次平滑模型或樣條仍可能更可信。
6. 對三點 $(0, 0)$ 、 $(1, 2)$ 與 $(2, 0)$ ：
 - (a) 寫出線性樣條；
 - (b) 描述三次樣條所需的連續性質；
 - (c) 不計算完整三次樣條，預測其圖像在 $x = 1$ 附近會有何不同。
7. 判斷主要任務是插值、平滑，還是兩者皆非：
 - (a) 估計兩次觀測之間某個未監測小時的降雨量；

- (b) 從含雜訊的每月銷售找出長期趨勢；
- (c) 利用截至 2025 年的記錄預測 2070 年人口；
- (d) 從精確控制點重建電腦繪製曲線。

練習核對與提示

1. 線性插值給出 21.2 單位；它假設 $t = 3$ 至 $t = 9$ 之間的變化率近似固定，不適用於觀測區間之外。
2. 兩段分別為 $2 + 2x$ ($0 \leq x \leq 2$) 及 $8 - x$ ($2 \leq x \leq 5$)；所求值為 4、5 及 3.5，斜率在 $x = 2$ 改變。
3. $P(x) = x^2 - 2x + 2$ ，故 $P(2) = 2$ ；直接代入可核對全部三個節點。
4. 常數二階差分為 2，支持二次式 $y_n = n^2 + 4n + 7$ 。
5. 相關問題包括節點之間的振盪、雜訊放大、端點行為、敏感度及留出資料表現。
6. 線性樣條先為 $2x$ ，再為 $4 - 2x$ 。三次樣條及其一階、二階導數均連續，故在 $x = 1$ 附近把折角平滑化。
7. 分類依次為插值、平滑、兩者皆非（長距離外推），以及插值。

建模挑戰

某學校每五分鐘記錄走廊入口的人流，但有數個讀數缺失，部分記錄值亦可能含點算誤差。撰寫一頁經驗建模簡報，內容包括：

1. 重建缺失數據的目的；
 2. 哪些觀測視為可靠或含雜訊；
 3. 比較分段線性插值、三次樣條與平滑；
 4. 一幅診斷圖或一項診斷計算；
 5. 結果的有效區間；
 6. 一項敏感度測試及一項限制。
- 論證質素比使用最複雜方法更重要。

A 選讀延伸：牛頓均差

對不等距節點，以均差取代普通差分：

$$f[x_i, x_{i+1}] = \frac{f(x_{i+1}) - f(x_i)}{x_{i+1} - x_i},$$

$$f[x_i, \dots, x_{i+k}] = \frac{f[x_{i+1}, \dots, x_{i+k}] - f[x_i, \dots, x_{i+k-1}]}{x_{i+k} - x_i}.$$

牛頓插值多項式為

$$P_n(x) = f[x_0] + f[x_0, x_1](x - x_0) + \dots + f[x_0, \dots, x_n] \prod_{j=0}^{n-1} (x - x_j).$$

與拉格朗日形式不同，加入新數據點時可有效率地擴展牛頓形式。

不等距數據

對 $(0, 1)$ 、 $(1, 3)$ 與 $(3, 7)$ ，

$$f[0, 1] = 2, \quad f[1, 3] = 2,$$

因此

$$f[0, 1, 3] = \frac{2 - 2}{3 - 0} = 0.$$

故 $P_2(x) = 1 + 2x$ ，說明三點雖然間距不等，實際上仍共線。

B 選讀延伸：一般自然樣條系統

令 $x_0 < \cdots < x_n$ 、 $h_i = x_{i+1} - x_i$ 且 $M_i = S''(x_i)$ 。對自然三次樣條， $M_0 = M_n = 0$ ，內部二階導數滿足

$$h_{i-1}M_{i-1} + 2(h_{i-1} + h_i)M_i + h_iM_{i+1} = 6 \left(\frac{y_{i+1} - y_i}{h_i} - \frac{y_i - y_{i-1}}{h_{i-1}} \right).$$

這是三對角線性系統，因此即使樣條含許多節點，亦可有效率地計算。此公式有助實作，但不屬於兩小時核心課堂內容。

模型審核卡

在信任或傳達模型之前先作審核

目的	模型要回答甚麼問題或支援甚麼決策？	資料	哪些輸入屬觀測、推導或模擬資料？
假設	哪些簡化可能改變答案？	單位	量綱與尺度是否一致？
運算 / 程式核證	方程、程式、限制與極限情況是否已核對？	模型效度驗證	是否用獨立證據檢驗預定用途下的表現？
不確定性	測量、參數、求解器及模型結構誤差如何影響結論？	倫理	誰受益、誰承擔風險，公平準則是甚麼？
可重現性	是否記錄資料來源、版本、種子與轉換？	溝通	主張是否有條件、可追溯且與決策相關？

延伸閱讀

- 插值法使用指南（SciPy）
- 《微積分》第一冊（OpenStax）

來源、資料脈絡與授權

本講義依循 SIAM / COMAP GAIMME 報告對反覆建模與評核的觀點。競賽相關說明於 2026 年 8 月 20 日依據官方 COMAP HiMCM 規則核對；提交前仍須查閱當時最新規則。除非另有明確來源，課堂資料與數值情境均屬合成或模擬教學例子，不構成任何真實機構的實證主張。第三方參考資料沿用各自條款。Kenneth Sok Kin Cheng 的原創教材採用 CC BY-NC-SA 4.0 授權。

學生作答指引與檢核準則

封閉式計算題請列出定義、代入、單位及至少一項獨立檢查；本課的完整例題可作方法提示，但須自行完成數值。開放建模題按五項檢核：目的與輸出是否明確、資料與來源是否交代、假設與適用範圍是否合理、運算／程式核證與模型效度驗證是否可追蹤，以及不確定性、限制與建議是否互相一致。每項須附具體證據，而非只寫「已完成」。