3 實驗區間與建模區間的均勻性

一般好的設計, 在其實驗區間內實驗點的分佈皆有某種程度的均勻性。但在本研究中我們有實驗區間及建模區間, 而我們希望實驗點在這兩個區間皆能均勻分佈, 本章主要由這個角度來探討該如何選擇建模區間。實驗點是否均勻分佈, 有時可依據主觀的直覺來判斷, 例如比較圖3.1(a),(b) 我們可以清楚判斷 (b) 的實驗點比 (a) 均勻, 但是這樣的主觀判斷並沒有把所謂的均勻程度量化, 所以常常只要給定兩種無法由主觀判斷的佈點, 我們就可能會無法判斷何者較爲均勻, 如圖3.1(c),(d)。爲了將「均勻」程度數值化, 以對各種佈點方式的均勻程度做比較, 我們將採用均勻設計 (uniform design) 中所定義的均勻性測度, 其將於3.2節中介紹。

3.1 圓與 (r, θ) 區間上的均勻性

均勻設計的概念是希望實驗點均勻散佈在實驗區間內, 但目前我們有實驗區間及建模區間, 故我們希望此兩個區間內的實驗點皆能均勻散佈。在此我們先考慮在極座標 (R, θ) 區間上佈點, 我們希望實驗點在 (R, θ) 區間上均勻散佈, 並且希望這些實驗點透過極座標轉換後亦均勻散佈在圓上。首先令 $B_2 = \{(x, y) : x^2 + y^2 \leq 1\}$ 爲一以原點爲圓心的單位圓, 我們將使用極座標轉換, 如 (3.1) 式所示:

$$\begin{cases} x = R\cos(2\pi\theta) \\ y = R\sin(2\pi\theta) \end{cases} \tag{3.1}$$

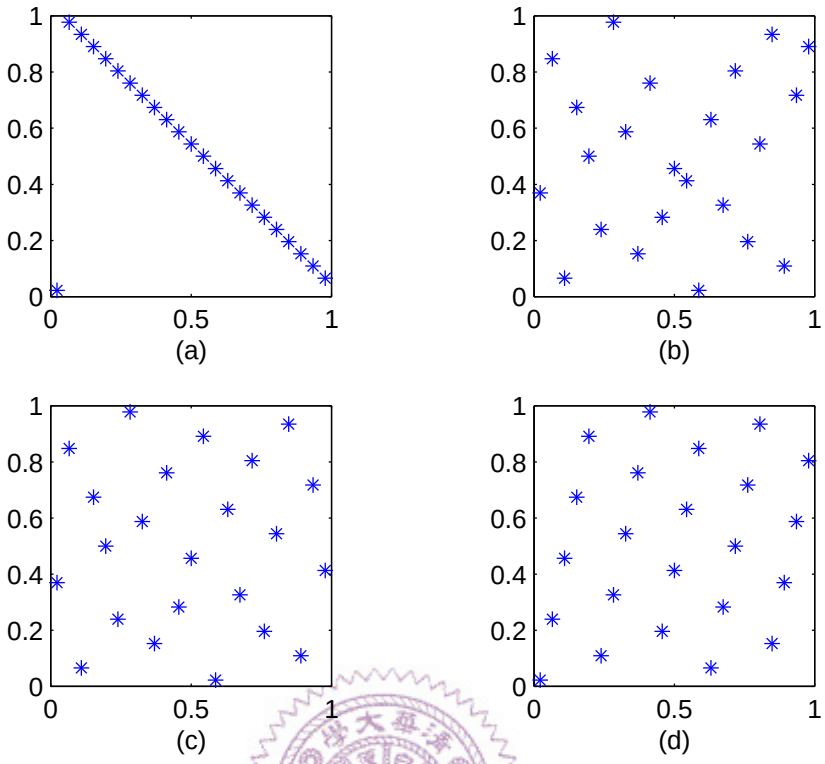

圖 3.1: 在建模區間中 4 種不同的設計

其中 $(R,\theta) \in [0,1)^2$, 以 C^2 表示。若取 $P_n = \{(R_k,\theta_k) : k = 1,...n\}$ 爲一組在 C^2 上的均勻取點, 如圖3.2(b) 所示, 將這些點代入 (3.1) 式, 接著我們可以在圓上得到 n 個點 $\{(x_k^{'}, y_k^{'}) : k = 1,...n\}$, 其中

$$\begin{aligned} x_k^{'} &= R_k \cos(2\pi\theta_k), \\ y_k^{'} &= R_k \sin(2\pi\theta_k), \quad k = 1,...,n \end{aligned} \tag{3.2}$$

很顯然這 n 個點在圓上散佈得並不均勻, 如圖3.2(a) 所示, 且集中於圓心。若在 (r,θ) 區間上畫出這些實驗點 (圖3.2(c)), 則亦可發現這些點在 (r,θ) 區間上較集中於 r 小的區域。若在圓上取 $P_n^* = \{(x_k, y_k) : k = 1,...n\}$ 爲一組在 B_2 上的均勻取點, 如圖3.2(d) 所示, 其對應到 (R,θ) 空間上的點, 如圖3.2(e) 所示, 由圖形可看出這些實

驗點呈現的趨勢是在半徑 R 小時實驗點較少, R 大時實驗點較多。若在 (r,θ) 區間上畫出這些實驗點, 我們可以看出實驗點散佈得相當均勻, 如圖3.2(f) 所示。由以上可知若在 (R,θ) 區間上均勻取點, 希望其對應到圓上亦均勻分佈是不可行的; 並且若在圓上均勻取點, 其對應到 (R,θ) 區間上的實驗點亦不是均勻分佈, 但是在 (r,θ) 區間上的實驗點則有均勻散佈的性質。以下我們將推導爲什麼會有這樣的現象。

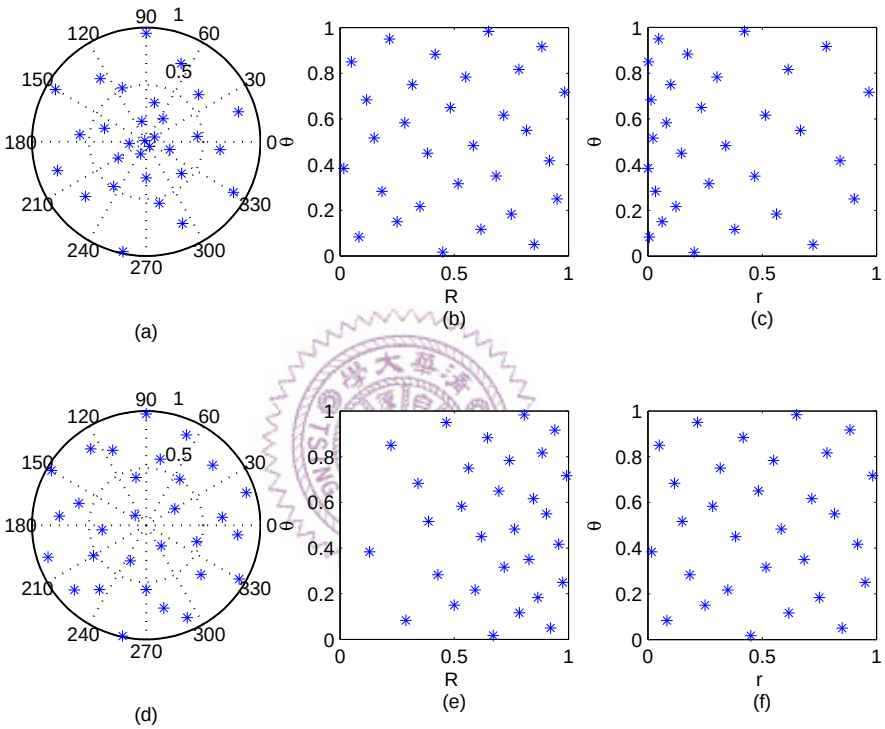

圖 3.2: 實驗點在實驗區間與建模區間上之分佈

首先假設在單位圓上的實驗點服從一個圓上的均勻分配:

$$(\boldsymbol{x},\boldsymbol{y}) \sim uniform(B_2),$$

其聯合機率 (密度) 函數爲

$$f(\boldsymbol{x},\boldsymbol{y}) = \begin{cases} \frac{1}{\pi}, & \text{if}(\boldsymbol{x},\boldsymbol{y}) \in B_2. \\ 0, & \text{otherwise.} \end{cases}$$

在圓上的點可以在 (R,θ) 區間上透過極座標公式 (3.1) 轉換得到, 如下所示:

$$\begin{cases} \boldsymbol{x} = \boldsymbol{R}\cos(2\pi\boldsymbol{\theta}) \\ \boldsymbol{y} = \boldsymbol{R}\sin(2\pi\boldsymbol{\theta}) \end{cases}$$

可表示爲

$$(\boldsymbol{x},\boldsymbol{y}) = g(\boldsymbol{R},\boldsymbol{\theta}), (\boldsymbol{R},\boldsymbol{\theta}) \in C^2$$

故其 Jacobian 爲

$$J = det[\frac{\partial g}{\partial \boldsymbol{R}\partial\boldsymbol{\theta}}] = det\begin{pmatrix} \cos(2\pi\boldsymbol{\theta}) & \sin(2\pi\boldsymbol{\theta}) \\ -2\pi\boldsymbol{R}\sin(2\pi\boldsymbol{\theta}) & 2\pi\boldsymbol{R}\cos(2\pi\boldsymbol{\theta}) \end{pmatrix} = 2\pi\boldsymbol{R}$$

故 $\boldsymbol{R}$ 與 $\boldsymbol{\theta}$ 之聯合機率函數可表示爲

$$h(\boldsymbol{R},\boldsymbol{\theta}) = f(g(\boldsymbol{x},\boldsymbol{y}))|J| = \frac{1}{\pi}\cdot 2\pi\boldsymbol{R} = 2\boldsymbol{R}, 0 \leq \boldsymbol{R},\boldsymbol{\theta} \leq 1$$

$\boldsymbol{R}$ 與 $\boldsymbol{\theta}$ 的 p.d.f 爲

$$h(\boldsymbol{R}) = \int_0^1 h(\boldsymbol{R},\boldsymbol{\theta})d\boldsymbol{\theta} = 2\boldsymbol{R}, 0 \leq \boldsymbol{R} \leq 1$$

$$h(\boldsymbol{\theta}) = \int_0^1 h(\boldsymbol{R},\boldsymbol{\theta})d\boldsymbol{R} = 1, 0 \leq \boldsymbol{\theta} \leq 1$$

因爲 $h(\boldsymbol{R},\boldsymbol{\theta}) = h(\boldsymbol{R})h(\boldsymbol{\theta})$ 可得知 $\boldsymbol{R}$ 與 $\boldsymbol{\theta}$ 彼此獨立。並且由 $(\boldsymbol{R},\boldsymbol{\theta})$ 的 p.d.f 可知在 (R,θ) 區間上的實驗點並不會均勻散佈在實驗區間內, 以下式表之

$$(\boldsymbol{R},\boldsymbol{\theta}) \not\sim uniform(C^2)$$

又 $\boldsymbol{R}$ 與 $\boldsymbol{\theta}$ 其 c.d.f 爲

$$F(\boldsymbol{R}) = \boldsymbol{R^2}, 0 \leq \boldsymbol{R} \leq 1$$

$$F(\boldsymbol{\theta}) = \boldsymbol{\theta}, 0 \leq \boldsymbol{\theta} \leq 1$$

我們可得知

$$(\boldsymbol{R^2}, \boldsymbol{\theta}) \sim uniform(C^2)$$

$$(\boldsymbol{r}, \boldsymbol{\theta}) \sim uniform(C^2), \text{ 其中 } \boldsymbol{r} = \boldsymbol{R^2}$$

由以上推導可知, 未來只需先在 (r, θ) 區間單位正方形中找到一組均勻分佈的點集合, 再將這些實驗點轉換到圓上, 則這些點在圓上亦會均勻分佈, 並可知在圓上均勻的設計在 (r, θ) 區間上也會均勻分佈。因爲我們希望實驗點在圓上有均勻的性質, 並且希望在建模區間也要有均勻的性質, 因此往後我們會將 (r, θ) 區間視爲建模區間, 而不再考慮 (R, θ) 區間。

3.2 均勻性測度

目前爲止已有數種均勻性測度被提出, 但常見的均勻設計表是採用星偏差 (star discrepancy) 或 L_2-偏差作爲均勻性度量 (Wang and Fang(1981),Fang and Wang(1994))。當我們使用偏差作爲均勻性度量時, 原點明顯扮演著特殊的地位, 一切方體 $[0, x)$ 均從原點開始。但是在定義域 $C^s = [0, 1)^s$ 單位立方體中, 這 2^s 個頂點的重要性應該是要一視同仁的。Hickernell(1998) 提出的 Centered L_p-discrepancy 可以修正偏差具有方向性的缺點。此測量法的優點是同時也把分布在高維度上的點集合投影到低維度的座標子空間上來做均勻性的判斷; 也就是說, 使用 Centered L_p-discrepancy 來當均勻設計的準則時, 不僅考慮了點集合在高維度空間中散佈的均勻性 , 同時也考慮了投影在低維度空間的均勻性。在介紹 Centered L_p-discrepancy 之前, 先引入一些

記號: 令 $x = (x_1, ..., x_s)$ 表示單位方體 C^s 中的點, 集合 $S = \{1, ..., s\}$ 表示 C^s 的座標軸, 對於 S 任一子集 u, $|u|$ 表示 u 中元素的個數, R^u 表示由 u 所表示的座標軸對應的 $|u|$ 維實數空間, C^u表 示 $|u|$ 維方體, 它是 C^s 在 R^u 上的投影, x_u 表示 x 中由 u 所代表的分量構成的 $|u|$ 維向量, $P = \{x_1, ..., x_n\}$ 爲 C^s 上的 n 個點組成的點集, P_u 表示 P 在方體 C^u上的投影。關於 Centered L_p-discrepancy 敍述如下: 對任意兩點 $x, y \in C^s$, 令 $J(x, y)$ 爲介於 x 和 y 之間的方體區域, 即 $J(x, y) = \{(t_1, \ldots, t_s) : \min(x_j, y_j) \leq t_j < \max(x_j, y_j), j = 1, \ldots, s\}$。爲了減少原點0的特殊性, 克服 $J(x) = [0, x)$ 的缺點, 令 A 爲 $[0, 1)^s$ 單位方體的 2^s 個頂點, 若此 s 維度分別由中點切割, 可將 $[0, 1)^s$ 劃分爲 2^s 個方體。例如在 $[0, 1)^2$ 單位正方形中, 若由 $x = 1/2$ 及 $y = 1/2$ 處切開, 則 $[0, 1)^2$ 可被劃分成四個方形。任取 $x \in C^s$, 它必然落入這 2^s 個方體中的一個。令 $a_x \in$ A 爲該方體的頂點, 它與 x 組成方體 $J(a_x, x)$。令 $R^c(x) = Vol(J(a_x, x)) - \frac{N(P, J(a_x, x))}{n}$, $Vol(J(a_x, x))$ 表示在 R^s 空間中 $J(a_x, x)$ 的體積, $N(P, J(a_x, x))$ 表示 P 中的點落入 $J(a_x, x)$ 的個數, $R^c(x_u)$ 爲 $R^c(x)$ 關於方體 C^u 和投影 P_u 的相對應表達式。其相對應的偏差定義如下:

$$CL_p(P) = \{ \sum_{\emptyset \subset u \subset S} \int_{C^u} |R^c(_u)|^p dx_u \}^{\frac{1}{p}} \tag{3.3}$$

稱爲 Centered L_p-discrepancy。當 $p = 2$ 時, Centered L_2-discrepancy 有下面的表達式:

$$\begin{aligned}(CL_2(P_n))^2 = (\frac{13}{12})^s - \frac{2}{n} \sum_{k=1}^{n} \prod_{j=1}^{s} (1 + \frac{1}{2}|x_{kj} - .5| - \frac{1}{2}|x_{kj} - .5|^2) + \\ \frac{1}{n^2} \sum_{k=1}^{n} \sum_{j=1}^{n} \prod_{i=1}^{s} [1 + \frac{1}{2}|x_{ki} - .5| + \frac{1}{2}|x_{ji} - .5| - \frac{1}{2}|x_{ki} - x_{ji}|]\end{aligned} \tag{3.4}$$

由此式可以看到, Centered L_2-discrepancy 在計算時對 C^s 的 2^s 個頂點一視同仁, 從而取消了原點的特殊地位。對任一 $j(1 \leq j < s)$, 如將 x_j 換成 $1 - x_j$, CD_2 保持不變, 即它具有反射不變性。在本研究中所提到的均勻性度量皆以(3.4) 式爲準則, 簡稱 CL_2。