

高效率的實驗設計：四元代數碼設計

中央研究院統計科學所 潘建興

一、前言

在現今科技發達的世界中，實驗的結構越來越複雜，實驗中投入的資源需要面對越來越大的競爭，而實驗的成本也隨之不斷上升。在這種缺乏資源的情況下，一個有高效率的實驗設計可以減少大量的資源浪費。其實，當從一些傳統上會使用的實驗設計的設定裡小心地作些微調，也許能夠得出一些更高效率的設計，而這所謂的高效率是基於估計的精準度或能夠估計的數量上來衡量的。

在現代的科學研究和工業實驗裡，研究人員通常需要同時估計多個因子對於反應值的效應。在所有實驗設計的選項中，完全因子設計有足夠的自由度來考慮所有因子的主效用和因子之間的所有交互效用，但是使用該設計的實驗往往會執行了一些來自不必要的因子組合的測試。如果想要節省實驗成本的話，研究人員通常會選擇使用部份因子設計。一個好的部份因子設計可以讓研究人員在相對少量測試的實驗中研究相對大量的因子對反應值的作用。

部份因子設計（簡稱 FFD）主要分為兩大類：正規部份因子設計（簡稱 RFFD）是一類可以用定義關係式來建構的 FFD，而非正規部份因子設計（簡稱 NRFFD）則是那些不能用定義關係式來建構的 FFD。由於 RFFD 在結構上和分析上的簡便，它在過去的幾十年間已經成為科學研究和工業應用實驗中標準的設計[1, 2]。可是，NRFFD 在過去二十年中獲得統計研究人員特別的關注。雖然在 NRFFD 裡大多數是較難分析的部份別名結構，它們在估計能力上（例如解析度、投射度等）要比相同大小的 RFFD 要好，而這也是研究人員近年來對這類設計湧現濃厚興趣的動機。[3]對 NRFFD 的發展有全面的檢討。

二、部份別名結構的重要性

相較於 RFFD 的完全別名結構，NRFFD 的

結構大多數是部份別名結構。因為主效用和一些交互效用的混淆並不完整，分析 NRFFD 中的部份別名結構要比完整別名結構困難得多。可是，NRFFD 在估計能力上的優勢完全體現在[4]一文中兩個毒理學的實驗例子上。這兩個例子顯示：(1)在不需要犧牲實驗的估計能力下，NRFFD 比 RFFD 更節省實驗成本；(2)在投入相同的實驗成本時，NRFFD 藉由部份別名結構來提供比 RFFD 更多交互效用的估計結果。

總而言之，得益於部份別名結構的存在，NRFFD 能夠同時估計部份混淆的主效用和交互效用。雖然如此，研究人員必須要在分析前先將交互作用從主效用中解放出來。[5]和[6]指出在分析時忘記了解放步驟的話，會引致錯誤的分析結果，包括在分析結果中：(1)忽略了真正重要的作用；(2)加入了偽重要的作用；(3)對估計的作用建議錯誤的方向或水平。[6]以三個分析化學中的實驗例子來演示這些潛在的陷阱如何提供錯誤的分析結果，以致整個程序在使用推薦的設定後無法達到反應值的最佳化。

三、實驗的建構方法和設計結構

NRFFD 至今無法在科學研究和工業實驗裡被廣泛採用的主要障礙是它沒有如 RFFD 一般簡易的設計結構。近年來的研究在兩水平的 NRFFD 中發展出一套使用四元代數碼來建構 NRFFD 的方法，而建構出來的實驗設計被稱為四元代數碼設計（簡稱 QCD）。QCD 基本結構的研究源於[7]一文。以下將簡介在該文中的一些符號定義和概念。

一個四元代數碼包括了屬於 Z_4 的四個水平：0、1、2 或 3。以 $G = (V, I_n)$ 來代表一個 $n \times (n+p)$ 的衍生矩陣，其中 $V = (\vec{v}_1, \dots, \vec{v}_p)$ 是一個包括 p 個長度為 n 的向量的 Z_4 矩陣，而 I_n 是一個 $n \times n$ 的單位矩陣。當將 G 的每一行在 Z_4 之下進行線性組合後，會得出一個四元線性代數碼 C 。然後透過 Gray 映射將每一個 C 矩陣中的四元代

數碼變換成兩個二次元的影子，而 Gray 映射的規則如下：

$$\phi: 0 \rightarrow (0,0), 1 \rightarrow (0,1), 2 \rightarrow (1,1), 3 \rightarrow (1,0).$$

這些二次元的影子能夠集結成一個二水平的矩陣 D 。由此所得到的 D 將是一個含有 2^{2n} 行和 $2n + 2p$ 列的二水平 QCD。由於在 G 矩陣中 I_n 的部份會產生一個 $2^{2n} \times 2n$ 的完全因子設計，QCD 的設計特性將取決於 G 矩陣中 V 部份的設定。

Gray 映射可以用一條三角函數來表示[8]。它能夠將繁複的設計特性計算簡化成三角恆等式的操作問題。由於當中涉及眾多技術性的概念和引導，本文將省略此部份的內容。

以上介紹的方法只能建構一些含有 2 的雙數次方的行和雙數的列。建構其他維度的 QCD 需要一些附加的步驟。在[9]提及《列的刪除》，這步驟可以從原本的 QCD 中減少一列而導出含有單數列的 QCD。在[7]提及《列的分支》，這步驟可以從原本的 QCD 中減少一列和總行數一半的行，而導出的 QCD 含有單數列和 2 的單數次方的行。在這些步驟中，選擇不同的列來作刪除或分支會導出不同設計特性的 QCD。

四、設計特性和估計能力

把 QCD 設計的特性可以描述成一條簡單的矩陣方程式 $K = CF$ [10]。 K 是一系列的 k -方程式，它和 RFFD 中的字長意義相似，用以計算在一個由一些不完全交絡的作用所組成的集合裡作用的總數目。 F 是一個記錄在 G 矩陣的 V 部份裡各種不同的四元代數碼出現的頻率，而 C 是一個用來形容在 F 裡的不同組合在各 k -方程式中的比重。

各種類別的 QCD 有其唯一的 C 和 F ，從而計算出用來描述設計特性的 K 矩陣。舉例來說，在[7]指出，四分之一份的 QCD（從完全因子設計中選取其行總數四分之一的行所組成的 NRFFD）中只有兩個 k -方程式。因此，這類 QCD 可以被定義 $K = (k_1, k_2)'$ ， $F = (f_0, f_1, f_2, f_3)'$ 和

$$C = \begin{pmatrix} 0 & 1 & 2 & 1 \\ 0 & 2 & 0 & 2 \end{pmatrix}.$$

在 F 向量中，不同數目的組合頻率會導出不同設

計特性的 QCD。某些 F 可以導出擁有很好估計能力的 QCD，但對 F 作些微的修改後，有可能會導出無法自主作用中解放交互作用的 QCD。因此， F 向量中組合頻率的選擇對 QCD 的設計特性攸關重要。[7]提供了能導出不同大小的四分之一份 QCD 所需要的 F 向量中組合頻率，而[9]和[11]分別對八分之一份，十六分之一份，三十二分之一份和六十四分之一份的 QCD 提供建議的 F 向量中組合頻率。

五、到底四元代數碼設計有多好？

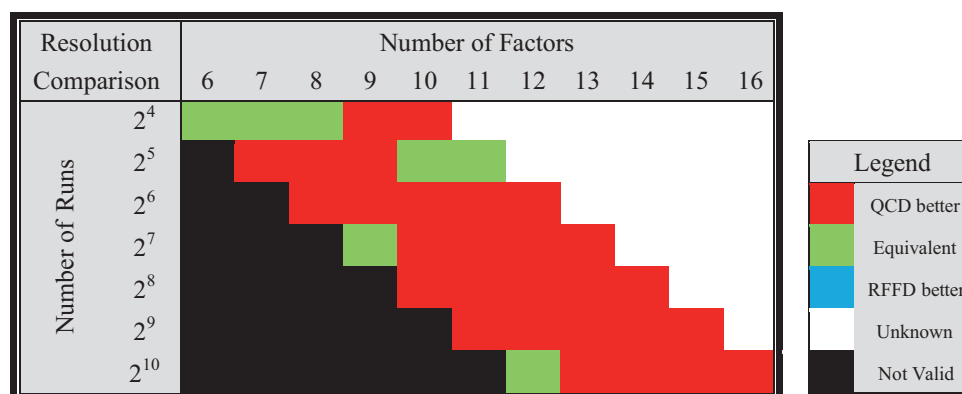
很多文獻中曾提出過不同種類和方向的標準來定論一個實驗設計的好壞。[7-11]主要使用以下三種標準：廣義解析度[12]、廣義最小偏差[13]和投射度[14]。本文只考慮廣義解析度（簡稱 GR）的比較以作演示。

如果以 GR 來衡量實驗設計的好壞的話，簡單來說，越高的 GR 代表越好的估計能力，也就是說，追求的目標就是擁有最好估計能力的設計。圖一顯示了以 GR 為標準時最佳的 QCD 和最佳的 RFFD 在不同維度的比較。在圖中紅色的格子代表最佳的 QCD 擁有比最佳的 RFFD 更高的 GR，而相反的情況以藍色的格子代表。綠色的格子代表最佳的 QCD 和最佳的 RFFD 擁有相同的 GR。白色的格子代表該維度的最佳 QCD 並未在已出版的文章中發表，或是仍然在調查中。黑色的格子則代表該維度的最佳 QCD 並不存在，它們有些是二分之一份 NRFFD 和完全因子設計，又或者是重複設定測試的實驗。圖一中的比較清楚地顯示相較於相同大小的最佳 RFFD，最佳的 QCD 擁有相等或更好的估計能力。雖然在已知的大維度裡，有非常少部分的最佳 QCD 擁有遜於最佳 RFFD 的 GR，但是只要在本文介紹的建構方法上作一點修改後，便可以得到比最佳 RFFD 更高 GR 的 QCD。

圖一的比較止於 16 個因子和 1024 行的設計，但在文獻中可以找到更大的設計的比較結果。此外，文獻中也包括了最佳的 QCD 和最佳的 RFFD 在以廣義最小偏差和投射度為標準的比較結果。

六、結語

本文簡要地解釋了當要設計一個科學研究



圖一 以廣義解析度為標準，最佳的 QCD 和最佳的 RFFD 的比較

或工業實驗時，相較於傳統的 RFFD，為何 NRFFD 會是一個更好的選擇。雖然 NRFFD 擁有相對較好的設計特性，但是由於它缺乏如 RFFD 一般簡易的設計結構，至今為止它無法被廣泛採用。因此，本文介紹了一種被稱為 QCD 的全新類型 NRFFD。QCD 能夠以類似 RFFD 的簡易方法來建構，而它的設計特性則由 K 矩陣所主宰。因此，QCD 能夠有系統地被建構，也能夠用基礎理論來研究。以 GR 為標準下的特性比較發現，最佳的 QCD 擁有與最佳的 RFFD 相等，或者是更好的估計能力。因此，本文建議在科學研究和工業實驗的研究人員在設計實驗時，可以考慮使用本文介紹的 QCD。

參考文獻

- [1] C. F. J. Wu and M. Hamada, *Experiments: Planning, Analysis and Parameter Design Optimization*. New York: Wiley (2000).
- [2] R. Mukerjee and C. F. J. Wu, *A Modern Theory of Factorial Designs*, New York: Springer (2006).
- [3] H. Xu, F. K. H. Phoa and W. K. Wong, *Statistics Surveys*, **3**, 18 (2009).
- [4] F. K. H. Phoa, W. K. Wong and H. Xu, *J. Chemometr.*, **23**, 545 (2009).
- [5] M. Hamada and C. F. J. Wu, *J. Qual. Technol.*, **24**, 130 (1992).
- [6] F. K. H. Phoa, H. Xu and W. K. Wong, *Food Chem. Toxicol.*, **47**, 2183 (2009).
- [7] F. K. H. Phoa and H. Xu, *Ann. Statist.*, **37**, 2561 (2009).
- [8] R. Zhang, F. K. H. Phoa, R. Mukerjee and H. Xu, *Ann. Statist.*, **39**, 931 (2011).
- [9] F. K. H. Phoa, R. Mukerjee and H. Xu, *J. Statist. Plann. Inference*, **142**, 1073 (2012).
- [10] F. K. H. Phoa, *Ann. Statist.*, **40**, 3161 (2012).
- [11] F. K. H. Phoa, H. W. Chen and S. C. Lin, *under review* (2013+).
- [12] L. Y. Deng and B. Tang, *Statist. Sinica*, **9**, 1071 (1999).
- [13] B. Tang and L. Y. Deng, *Ann. Statist.*, **27**, 1914 (1999).
- [14] G. E. P. Box and J. Tyssedal, *Biometrika*, **83**, 950 (1996).