國立政治大學 應用數學系

碩士 學位論文

二維條件分配相容性問題之新解法

A new approach to solve the compatibility issues for two-dimensional conditional distributions

碩士班學生：郭柏辛 撰

指導教授：宋傳欽 博士

中華民國 105 年 6 月 23 日

中文摘要

給定二元隨機向量 (X, Y) 之聯合機率分配，可容易得到其條件機率分配 $f_{X|Y}$ 與 $f_{Y|X}$；反之，給定條件機率分配 $f_{X|Y}$ 與 $f_{Y|X}$，是否能獲得對應的聯合機率分配呢？條件分配相容性研究的主要內容包括：(一) 如何判斷給定的條件分配是否相容？(二) 若相容，則如何找到聯合分配？(三) 若不相容，則該如何找到最近似的聯合分配？

根據比值矩陣法的理論，檢驗比值矩陣是否為秩 1 矩陣或者有擴張秩 1 矩陣，便可得知給定的條件分配是否相容。當比值矩陣的元素皆為正值時，本文運用線性代數中之奇異值分解定理，先發展出奇異值分解法來處理條件分配相容性問題；當比值矩陣的元素非皆為正值時，接續發展出最近似秩 1 矩陣法來解決相容性問題，而最近似秩 1 矩陣法可視為奇異值分解法的延伸。在發展最近似秩 1 矩陣法時，我們利用到類 Frobenius 範數的概念，並提出了三種求解過程 (無限制條件法、Lagrange 乘數法與高維度牛頓法) 以及相關的演算法。本文詳細剖析了三種求解過程之數學流程，並輔以實際例子予以說明。

當條件機率分配不相容時，我們通常可獲得兩組近似聯合分配。如何將它們做適當的組合，也是值得探討的問題。最後，針對等加權之組合方式、權重與總誤差成反比之組合方式以及特徵向量法之組合方式進行比較分析。

關鍵字：條件機率分配、相容性、比值矩陣、特徵向量法、奇異值分解法、最近似秩 1 矩陣法、類 Frobenius 範數、Lagrange 乘數法、高維度牛頓法、最佳化法、最近似聯合分配

Abstract

Given a bivariate joint distribution of random vector (X, Y), we can easily derive the conditional probability distributions of $f_{X|Y}$ and $f_{Y|X}$. Conversely, given conditional probability distributions of $f_{X|Y}$ and $f_{Y|X}$, can we find the corresponding joint distribution? The compatibility issues of conditional distribution include: (a) how to determine whether they are compatible; (b) how to find the joint distribution if they are compatible; (c) how to find the most nearly joint distribution if they are incompatible.

Using the theory of ratio matrix approach, we can determine the given conditional probability distributions are compatible or not by checking whether their corresponding ratio matrix or the extension matrix of this ratio matrix is rank one or not. When elements of the ratio matrix are all positive, this thesis uses the singular value decomposition theorem of linear algebra to develop the singular value decomposition approach to deal with the compatibility issues. When elements of the ratio matrix are not all positive, we provide the most nearly rank one matrix approach to solve the compatibility issues. This most nearly rank one matrix approach can be considered as the extension of singular value decomposition approach. To develop the most nearly rank one matrix approach, we use the concept of semi-Frobenius norm to provide three solving methods (unconstrained method, Lagrange multiplier method, and multivariate Newton's method) with related algorithms. This thesis gives the mathematical procedure on these three solving methods in detail and uses examples to explain the compatibility issues.

When the conditional distributions are incompatible, we usually have two nearly joint distributions. It would be worth of discussing the combination of these two nearly joint distributions. Hence, this thesis compares and analyzes the compatibility issues with three different weights, which are equal, inverse proportional to the total errors, and relating to eigenvectors.

Keywords：conditional probability distribution, compatibility, ratio matrix, eigenvector approach, singular value decomposition approach, most nearly rank one matrix approach, semi-Frobenius norm, Lagrange multiplier method, multivariate Newton's method, optimization method, most nearly joint distribution

目錄

第一章　緒論

第一節　研究動機與目的

給定隨機變數 X 和 Y 的聯合機率分配，可得到 $X|Y$ 和 $Y|X$ 的條件機率分配；反之，給定 $X|Y$ 和 $Y|X$ 的條件機率分配，是否能得到 (X,Y) 的聯合機率分配呢？給定條件機率分配，相容性問題的研究主要包含：(一) 如何判斷它們是否相容？即存在 (X,Y) 的聯合機率分配，其條件機率分配與給定的相同；(二) 若相容，則該如何找到聯合機率分配？(三) 若不相容，則該依據何種評量標準找到最近似聯合分配？目前在文獻中有多種不同解決相容性問題的方法，例如：Arnold and Press (1989)、Arnold et al. (2004) 和 Song et al. (2010) 的比值矩陣法 (ratio matrix approach)、Arnold et al. (2002) 的數學規劃法 (mathematical programming approach) 和顧仲航 (2011) 的特徵向量法 (eigenvector approach) 等等。

當 X 和 Y 為有限、離散時，假設 $A=(a_{ij})$ 和 $B=(b_{ij})$ 分別為對應 $X|Y$ 和 $Y|X$ 的條件機率矩陣，$C=(c_{ij})$ 為 A 對 B 的比值矩陣，即

$$c_{ij}=\frac{a_{ij}}{b_{ij}}.$$

根據 Arnold and Press (1989)，若比值矩陣 C 的元素皆為正值，則僅需考慮 C 是否為秩 1 矩陣，即可判斷給定的條件機率分配是否相容。

而檢驗一個矩陣是否為秩 1 矩陣或近似秩 1 矩陣，可透過線性代數中的奇異值分解定理 (singular value decomposition theorem，簡稱 SVD) 來辨別，也就是對 C 做奇異值分解：

$$C = \sigma_1 \mathbf{u}_1 \mathbf{v}_1' + \sigma_2 \mathbf{u}_2 \mathbf{v}_2' + \cdots + \sigma_r \mathbf{u}_r \mathbf{v}_r',$$

其中 $rank(C) = r$，且奇異值 (singular values) $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$。若 $r = 1$，則 $rank(C) = 1$，故 A 和 B 相容。Arnold and Press (1989) 提出，若 A 和 B 相容，則存在向量 $\mathbf{u} = (u_i)$ 和 $\mathbf{v} = (v_j)$ 使得 $c_{ij} = u_i v_j$，$\forall (i,j) \in N$。令 $\boldsymbol{\tau} = (\tau_i)$，$\boldsymbol{\eta} = (\eta_j)$，其中

$$\tau_i = \frac{u_i}{\sum_{i=1}^{I} u_i}, \quad \eta_j = \frac{1}{v_j \sum_{k=1}^{J} \frac{1}{v_k}}, \tag{1.1}$$

則 $\boldsymbol{\tau}$ 和 $\boldsymbol{\eta}$ 分別為 X 和 Y 的邊際分配；再加上給定的條件機率分配，可得 (X, Y) 的聯合機率分配。若 $r \geq 2$，則 A 和 B 不相容，這時我們如何找到 (X, Y) 的近似聯合機率分配？

一般而言，比值矩陣的元素不一定皆為正，可能會出現沒定義的元素，通常以 $*$ 表示。此時我們無法直接用奇異值分解定理來判斷給定的條件機率分配是否相容，因而也就無法求出 (X, Y) 的聯合機率分配或近似聯合機率分配。將比值矩陣法的一些結果和奇異值分解定理結合起來應用，啟發本文發展出新的方法來解決前述所說的條件分配相容性問題。新的解法除了有理論基礎外，還可藉助電腦軟體輔助計算，快速解決問題。

第二節　研究架構

本文共分為六個章節，第一章為緒論，介紹本文的研究動機、目的和架構；第二章為文獻探討，回顧 Arnold and Press (1989)、Arnold et al. (2004) 和 Song et al. (2010) 的比值矩陣法、Arnold et al. (2002) 的數學規劃法和顧仲航 (2011) 的特徵向量法；第三章為奇異值分解法的介紹，說明如何利用線性代數中的奇異值分解定理來檢驗給定的條件機率分配是否相容；第四章為在不相容情況下，找出最近似秩 1 矩陣；第五章為最近似聯合分配之探討，以總誤差作為評量標準，透過實例之計算，對不同方法所求得的最近似聯合分配進行比較; 第六章為結論，歸納和整理本文的研究結果，並提出未來可繼續研究的相關議題。

第二章　文獻探討

在現今文獻中有許多種解決條件分配相容性問題的方法，本章僅對 Arnold and Press (1989)、Arnold et al. (2004) 和 Song et al. (2010) 的比值矩陣法、Arnold et al. (2002) 的數學規劃法和顧仲航 (2011) 的特徵向量法做簡單扼要的回顧與描述。這些文章和本文內容較為相關，文中將多次引用，而且本文的新方法將和他們的方法進行比較。

第一節　比值矩陣法

本節中所介紹的比值矩陣法，內容主要參考 Arnold and Press (1989)、Arnold et al. (2004) 和 Song et al. (2010)。

令有限、離散隨機變數 X 和 Y 的值域分別為 $S_X = \{1, 2, \cdots, I\}$，$S_Y = \{1, 2, \cdots, J\}$；$A = (a_{ij})$ 和 $B = (b_{ij})$ 分別為對應 $X|Y$ 和 $Y|X$ 的條件機率矩陣，即

$$a_{ij} = P(X = i|Y = j), \quad b_{ij} = P(Y = j|X = i).$$

由上述得知

$$\sum_{i=1}^{I} a_{ij} = 1, \forall j; \quad \sum_{j=1}^{J} b_{ij} = 1, \forall i.$$

A 和 B 相容，即 $X|Y$ 和 $Y|X$ 相容，定義如下：

定義 2.1. 若存在非負聯合分配 $P = (p_{ij})$，使得

$$a_{ij} = p_{ij}/p_{.j}, \quad b_{ij} = p_{ij}/p_{i.}, \quad i = 1, 2, \cdots, I, \quad j = 1, 2, \cdots, J,$$

其中

$$p_{ij} = P(X = i, Y = j), \quad p_{i.} = \sum_{j=1}^{J} p_{ij}, \quad p_{.j} = \sum_{i=1}^{I} p_{ij},$$

則 A 和 B 稱作相容 (compatible)。

給定條件機率矩陣 A 和 B，什麼情況下會滿足相容的條件呢？Arnold and Press (1989) 指出，若 A 和 B 相容，則 $N^A = N^B$，其中 $N^A = \{(i,j)|a_{ij} > 0\}$，$N^B = \{(i,j)|b_{ij} > 0\}$，也就是 $N^A = N^B$ 是 A 和 B 相容的必要條件。因此在討論 A 和 B 是否相容時，都先假設 $N^A = N^B = N$。

定義 2.2. 令

$$c_{ij} = \begin{cases} \dfrac{a_{ij}}{b_{ij}}, & \text{若 } (i,j) \in N \\ *, & \text{若 } (i,j) \notin N, \end{cases}$$

則 $C = (c_{ij})$ 為 A 對 B 的比值矩陣。

比值矩陣 C 在處理相容性問題時扮演很重要的角色。

定理 2.3. *設 C 中的元素皆為正值，則 A 和 B 相容的充要條件為 $c_{i_1,j_1}c_{i_2,j_2} = c_{i_1,j_2}c_{i_2,j_1}$，$\forall(i_1, i_2, j_1, j_2)$，其中 $c_{ij} = a_{ij}/b_{ij}$，$\forall(i,j)$，即 C 就是秩 1 矩陣。*

定義 2.4. 設 C 為 A 對 B 的比值矩陣，若 c_{ij} 皆為正值，則 C 稱為完整矩陣 (complete)；反之，則 C 稱為不完整矩陣 (incomplete)。若存在矩陣 $\overline{C}$，其所有元素皆為正值，且 $\overline{c}_{ij} = c_{ij}$，$\forall(i,j) \in N$，則 $\overline{C}$ 稱為 C 的正擴張矩陣 (positive extension matrix)。若 $\overline{C}$ 的秩為 1，則 $\overline{C}$ 稱為 C 的秩 1 正擴張矩陣 (rank one positive extension matrix，簡稱 *ROPE* 矩陣)。

當比值矩陣 C 皆為正值時，則 $\overline{C} = C$。下面為檢查 A 和 B 是否相容的重要定理。

定理 2.5. *設 A 和 B 為條件機率矩陣，C 為 A 對 B 的比值矩陣。A 和 B 相容的充要條件為 C 有 ROPE 矩陣。*

然而在比值矩陣法中，可能因為比值矩陣之結構過於複雜或是 $*$ 出現太頻繁，在實際執行定理 2.5上會產生困難，故 Song et al. (2010) 提出不可約化塊狀對角矩陣 (irreducible block diagonal matrix，簡稱 *IBD* 矩陣)。透過 *IBD* 矩陣處理相容性問題，僅需檢驗對角線上的塊狀矩陣即可，使判斷是否相容的程序會更有效率。

定義 2.6. 若比值矩陣 C 經過列或行交換後，可表示如下：

$$\left(\begin{array}{c|c} T_1 & * \\ \hline * & T_2 \end{array}\right),$$

其中塊狀矩陣 T_1 和 T_2 以外的元素皆為 $*$，則 C 稱為可約化 (reducible)；反之，C 稱為不可約化 (irreducible)。

引理 2.7. *對任意比值矩陣 C，經過列或行交換後，可表示如下：*

$$T(C) = \left(\begin{array}{c|ccccc} T_1 & * & \cdots & \cdots & * \\ \hline * & \boxed{T_2} & * & \cdots & * \\ \vdots & * & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & * \\ * & * & \cdots & * & \boxed{T_M} \end{array}\right),$$

即 C 可轉換成 IBD 矩陣，其中塊狀矩陣 $T_1, \cdots, T_M (M \geq 1)$ 為不可約化且塊狀矩陣以外的元素皆為 $*$。當 $M = 1$ 時，C 稱為不可約化。為了方便起見，我們以 $T(C) = Diag(T_1, \cdots, T_M)$，$M \geq 1$ 表示。

而在 IBD 矩陣法中，該如何判斷給定的條件機率分配是否相容呢？Song et al. (2010) 提出結果如下：

定理 2.8. *設 C 為 A 對 B 的比值矩陣，且 $T(C)=Diag(T_1,\cdots,T_M)$ 為 C 的 IBD 矩陣，則 A 和 B 相容的充要條件為 $N^A = N^B$ 且每個 $T_m(1 \leq m \leq M)$ 是秩 1 矩陣 (或有 $ROPE$ 矩陣)。*

當得知給定條件機率矩陣 A 和 B 相容時，A 和 B 所對應的聯合分配是否唯一是接下來需要探討的問題，也就是該如何檢驗聯合分配的唯一性或找出所有的聯合分配。Song et al. (2010) 以 IBD 矩陣的技巧提供唯一性的檢驗方法，定理如下：

定理 2.9. *設 C 為 A 對 B 的比值矩陣，則下列敘述等價：*

(i) 對任意 C 的 IBD 矩陣 $T(C)=Diag(T_1,\cdots,T_M)$，恆有 $M=1$。

(ii) C 為不可約化。

(iii) C 有唯一的 $ROPE$ 矩陣。

(iv) C 所對應的聯合分配是唯一。

Song et al. (2010) 也證明 C 的每一個 $ROPE$ 矩陣對應了一個聯合分配，定理如下：

定理 2.10. *設條件機率矩陣 A 和 B 相容，且 C 為 A 對 B 的比值矩陣。令 $\mathfrak{C}$ 為 C 之所有 $ROPE$ 矩陣的集合，$\mathfrak{F}$ 為所有聯合分配 (A 和 B 為其條件分配) 的集合，則映射 $H:\mathfrak{C} \to \mathfrak{F}$ 定義為 $H(\overline{C}) = f$，並 H 為雙射 (bijection)，其中 $\overline{C} = \boldsymbol{u}\boldsymbol{v}'$，且 $f(x_i,y_j) = b_{ij}\dfrac{u_i}{u_\cdot}$，$u. = \sum_{i=1}^{I} u_i$，$\forall(i,j) \in N$。*

倘若條件機率矩陣 A 和 B 相容，但 A 和 B 所對應的聯合分配不唯一，也就是比值矩陣 C 為可約化或 $ROPE$ 矩陣不唯一，Song et al. (2010) 提出一個有效的方法來找出所有聯合分配，敘述如下：

定理 2.11. *設條件機率矩陣 A 和 B 相容，且 C 為 A 對 B 的比值矩陣。令 $T(C) = ECF = Diag(T_1, \cdots, T_M)$ 為 C 的 IBD 矩陣，其中 E 和 F 為排列矩陣 (permutation matrix)，且 $\overline{T}_m = \boldsymbol{u}_m \boldsymbol{v}'_m$ 為 $T_m (1 \leq m \leq M)$ 的 ROPE 矩陣。對任意 $\boldsymbol{k} = (k_1, \cdots, k_M)$，其中 $k_m > 0$，令 $\boldsymbol{u}'_{\boldsymbol{k}} = (k_1 \boldsymbol{u}'_1, \cdots, k_M \boldsymbol{u}'_M)$，$E' \boldsymbol{u}_{\boldsymbol{k}} = \boldsymbol{p}_{\boldsymbol{k}}$，$f_{\boldsymbol{k}}(x_i, y_j) = b_{ij} p_{\boldsymbol{k}i} / p_{\boldsymbol{k}+}$，$\forall (i,j) \in N$，其中 $\boldsymbol{p}'_{\boldsymbol{k}} = (p_{\boldsymbol{k}1}, \cdots, p_{\boldsymbol{k}I})$ 且 $p_{\boldsymbol{k}+} = \sum_{i=1}^{I} p_{\boldsymbol{k}i}$。則 $\{f_{\boldsymbol{k}} | \boldsymbol{k} = (k_1, \cdots, k_M), k_m > 0, 1 \leq m \leq M\}$ 為所有聯合分配 (C 為其比值矩陣) 的集合。*

若 $\boldsymbol{k}^* = \boldsymbol{k} / (\sum_{m=1}^{M} k_m)$，則 $f_{\boldsymbol{k}^*} = f_{\boldsymbol{k}}$。故所有可能聯合機率分配可表示如下：

$$\{f_{\boldsymbol{k}} | \boldsymbol{k} = (k_1, \cdots, k_M), \quad \sum_{m=1}^{M} k_m = 1, \quad k_m > 0, \quad 1 \leq m \leq M\}.$$

第二節　數學規劃法

當條件機率矩陣 A 和 B 不相容時，則該依據何種評量標準找到最近似聯合分配？我們希望能訂出測量不相容程度的方式和求得對應的最近似聯合分配。Arnold et al. (2002) 提出 ϵ-相容 (ϵ-compatability) 的概念作為條件機率矩陣不相容程度的評量標準，並以數學規劃法找出不相容程度的值和最近似聯合分配。Arnold et al. (2002) 測量不相容程度的選擇策略如下，其中 $W = (w_{ij}) \geq 0$ 為權重矩陣，用來顯示聯合機率值 p_{ij} 在不同位置之精準度的相對重要性：

選擇策略 (一) 給定條件機率矩陣 $A = (a_{ij})$ 和 $B = (b_{ij})$，求聯合機率矩陣 $P = (p_{ij})$，其中 $\sum_{(i,j)\in N} p_{ij} = 1$，使得

$$
\begin{aligned}
|p_{ij} - a_{ij}p_{.j}| &\leq \epsilon w_{ij}, \quad \forall (i,j) \in N; \\
|p_{ij} - b_{ij}p_{i.}| &\leq \epsilon w_{ij}, \quad \forall (i,j) \in N.
\end{aligned}
$$

選擇策略 (二) 給定條件機率矩陣 $A = (a_{ij})$ 和 $B = (b_{ij})$，求向量 $\boldsymbol{\tau} = (\tau_i) \geq 0$ 和 $\boldsymbol{\eta} = (\eta_j) \geq 0$，其中 $\sum_i \tau_i = 1$，$\sum_j \eta_j = 1$，使得

$$
|a_{ij}\eta_j - b_{ij}\tau_i| \leq \epsilon w_{ij}, \quad \forall (i,j) \in N.
$$

選擇策略 (三) 給定條件機率矩陣 $A = (a_{ij})$ 和 $B = (b_{ij})$，求邊際機率向量 $\boldsymbol{\tau} = (\tau_i) \geq 0$，其中 $\sum_i \tau_i = 1$， 使得

$$
\left| a_{ij} \sum_{k=1}^{I} b_{kj}\tau_k - b_{ij}\tau_i \right| \leq \epsilon w_{ij}, \quad \forall (i,j) \in N.
$$

定義 2.12. 給定條件機率矩陣 A 和 B，則 A 和 B 是 ϵ-相容的充要條件為 ϵ 是在上述任一選擇策略下有解之最小值。

由定義 (2.12) 得知，A 和 B 相容的充要條件為它們是 0-相容。

當聯合機率值 p_{ij} 在每個位置的重要性皆相同時，可令 $w_{ij} = 1$，$\forall(i,j)$。此外，Arnold et al. (2002) 亦提出卡方測量的概念，敘述如下：

選擇策略 (四) 給定條件機率矩陣 $A = (a_{ij})$ 和 $B = (b_{ij})$，求聯合機率矩陣 $P = (p_{ij})$，其中 $\sum_{(i,j)\in N} p_{ij} = 1$，使得

$$\sum_{(i,j)\in N} \frac{(b_{ij}p_{i.} - a_{ij}p_{.j})^2}{a_{ij}p_{.j}} \leq \epsilon.$$

選擇策略 (五) 給定條件機率矩陣 $A = (a_{ij})$ 和 $B = (b_{ij})$，求向量 $\boldsymbol{\tau} = (\tau_i) \geq 0$ 和 $\boldsymbol{\eta} = (\eta_j) \geq 0$，其中 $\sum_i \tau_i = 1$，$\sum_j \eta_j = 1$，使得

$$\sum_{(i,j)\in N} \frac{(a_{ij}\eta_j - b_{ij}\tau_i)^2}{b_{ij}\tau_i} \leq \epsilon.$$

選擇策略 (六) 給定條件機率矩陣 $A = (a_{ij})$ 和 $B = (b_{ij})$，求邊際機率向量 $\boldsymbol{\tau} = (\tau_i) \geq 0$，其中 $\sum_i \tau_i = 1$，使得

$$\sum_{(i,j)\in N} \frac{\left(a_{ij}\sum_{k=1}^{I} b_{kj}\tau_k - b_{ij}\tau_i\right)^2}{b_{ij}\tau_i} \leq \epsilon.$$

各選擇策略下，測量不相容程度的聯合機率矩陣 $P^{(1)}$ 和 $P^{(2)}$，定義如下：

表 2.1 各選擇策略下測量不相容程度的聯合機率矩陣 $P^{(1)}$ 和 $P^{(2)}$

選擇策略	$P^{(1)}$	$P^{(2)}$
(一) 和 (四)	$a_{ij}p_{.j}$	$b_{ij}p_{i.}$
(二) 和 (五)	$b_{ij}\tau_i$	$a_{ij}\eta_j$
(三) 和 (六)	$b_{ij}\tau_i$	$a_{ij}\sum_{k=1}^{I} b_{kj}\tau_k$

選擇策略 (一)、(四) 是考量 P 和 $P^{(1)}$、$P^{(2)}$ 的差異，選擇策略 (二)、(三)、(五)、(六) 是考量 $P^{(1)}$ 與 $P^{(2)}$ 間的差異。在不同選擇策略下，可以 ϵ 為目標函數，利用作業研究中的數學規劃法，求得不相容程度的最小值 ϵ。採取選擇策略 (一)∼(三) 時，限制條件是線性的；而採取選擇策略 (四)∼(六) 時，限制條件是非線性的。詳細內容請參閱 Arnold et al. (2002)。

當條件機率矩陣 A 和 B 不相容時，如何求出適當的近似聯合分配？在選擇策略 (一)、(四) 時，可直接算出近似聯合分配 $P = (p_{ij})$；而在其他選擇策略時，可用 $(P^{(1)} + P^{(2)})/2$ 作為近似聯合分配。

第三節　特徵向量法

本節介紹如何以特徵向量法來處理相容性問題，內容主要參考顧仲航 (2011)。如同比值矩陣法在討論條件機率矩陣 A 和 B 是否相容時，都先假設 $N^A = N^B = N$。

若 A 和 B 相容，也就是 $X|Y$ 和 $Y|X$ 相容，則 X 和 Y 的邊際分配 $\boldsymbol{\tau}$ 和 $\boldsymbol{\eta}$ 必定滿足

$$diag(\boldsymbol{\tau}) \cdot B = A \cdot diag(\boldsymbol{\eta}), \tag{2.1}$$

其中 $diag(\boldsymbol{\tau})$ 和 $diag(\boldsymbol{\eta})$ 分別為以 $\boldsymbol{\tau}$ 和 $\boldsymbol{\eta}$ 作為對角元素的對角矩陣，而 $diag(\boldsymbol{\tau}) \cdot B = A \cdot diag(\boldsymbol{\eta})$ 就是 (X, Y) 的聯合分配。在判斷 A 和 B 是否相容時，只需檢驗是否有邊際分配 $\boldsymbol{\tau}$ 和 $\boldsymbol{\eta}$ 滿足式子 (2.1)。

當 A 和 B 相容時，邊際分配 $\boldsymbol{\tau}$ 和 $\boldsymbol{\eta}$ 也會滿足下面結果：

$$\begin{cases} A\boldsymbol{\eta} = \boldsymbol{\tau} \\ B'\boldsymbol{\tau} = \boldsymbol{\eta}, \end{cases} \tag{2.2}$$

根據式子 (2.2)，可得

$$\begin{cases} AB'\boldsymbol{\tau} = \boldsymbol{\tau} \\ B'A\boldsymbol{\eta} = \boldsymbol{\eta}. \end{cases} \tag{2.3}$$

定義 2.13. 給定向量 $\mathbf{u} = (u_1, \cdots, u_n)'$，若 $u_i > 0, \forall i$，則 $\mathbf{u}$ 稱為正向量；$\mathbf{u}$ 的長度以 $\|\mathbf{u}\|_1$ 表示，定義為 $\|\mathbf{u}\|_1 = \sum\limits_{i=1}^{n} |u_i|$。

由式子 (2.3) 可得知邊際分配 $\boldsymbol{\tau}$ 和 $\boldsymbol{\eta}$ 分別為 AB' 和 $B'A$ 對應於特徵值 1 的單位長正特徵向量。在特徵向量法中，僅需檢驗 AB' 和 $B'A$ 對應於特徵值 1 的單位長正特徵向量是否滿足式子 (2.1)，即可判斷 A 和 B 是否相容。因此不需要檢驗所有單位長正向量，只需利用上述過程便能有效地找到滿足式子 (2.1) 的邊際分配。我們將以上結果整理成定理如下：

定理 2.14. *給定條件機率矩陣 A 和 B，則 A 和 B 相容的充要條件為 AB' 和 $B'A$ 分別存在對應於特徵值 1 之單位長正特徵向量 $\boldsymbol{\tau}$ 和 $\boldsymbol{\eta}$ 滿足式子 (2.1)。*

在考量近似聯合分配優劣時，特徵向量法考慮了條件分配的誤差和邊際分配的誤差之加總當作總誤差，並以此作為不相容程度的評量標準，說明如下：

若 $P = (p_{ij})$ 為近似聯合分配，則總誤差為 $e = e_1 + e_2$，其中

$$
\begin{aligned}
e_1 &= \sum_i \sum_j (\frac{p_{ij}}{p_{.j}} - a_{ij})^2 + \sum_i \sum_j (\frac{p_{ij}}{p_{i.}} - b_{ij})^2, \\
e_2 &= \sum_i (\boldsymbol{\tau}_i - (AB'\boldsymbol{\tau})_i)^2 + \sum_j (\boldsymbol{\eta}_j - (B'A\boldsymbol{\eta})_j)^2; \\
\boldsymbol{\tau} &= (\tau_1, \cdots, \tau_I), \quad \tau_i = P_{i.}, \quad i = 1, 2, \cdots, I, \\
\boldsymbol{\eta} &= (\eta_1, \cdots, \eta_J), \quad \eta_j = P_{.j}, \quad j = 1, 2, \cdots, J.
\end{aligned}
$$

而總誤差可表示如下

$$
\begin{aligned}
e =& e_1 + e_2 \\
=& \sum_i \sum_j (\frac{p_{ij}}{p_{.j}} - a_{ij})^2 + \sum_i \sum_j (\frac{p_{ij}}{p_{i.}} - b_{ij})^2 \\
&+ \sum_i (\boldsymbol{\tau}_i - (AB'\boldsymbol{\tau})_i)^2 + \sum_j (\boldsymbol{\eta}_j - (B'A\boldsymbol{\eta})_j)^2.
\end{aligned}
$$

顧仲航 (2011) 考慮 $P^{(1)}$ 和 $P^{(2)}$ 的凸組合，即 $P = \lambda P^{(1)} + (1-\lambda)P^{(2)}, 0 \leq \lambda \leq 1$，其中 $P^{(1)} = diag(\boldsymbol{\tau}) \cdot B$，$P^{(2)} = A \cdot diag(\boldsymbol{\eta})$，且 $\boldsymbol{\tau}$ 和 $\boldsymbol{\eta}$ 滿足式子 (2.3)；並證明了 $\tau_i = P_{i.}$ 和 $\eta_j = P_{.j}$ 成立，使得 $e_2 = 0$。

同時，顧仲航推導出，當 $\lambda = \dfrac{v}{u+v}$ 時，總誤差 e 有最小值 $\dfrac{uv}{u+v}$，此時最近似聯合分配為

$$\hat{P} = \frac{v}{u+v}P^{(1)} + \frac{u}{u+v}P^{(2)}, \tag{2.4}$$

其中

$$u = \sum_i \sum_j \left(\frac{p_{ij}^{(1)} - p_{ij}^{(2)}}{\eta_j} \right)^2, \quad v = \sum_i \sum_j \left(\frac{p_{ij}^{(1)} - p_{ij}^{(2)}}{\tau_i} \right)^2.$$

最後，顧仲航以實例說明，由特徵向量法和 Arnold et al. (2002) 的選擇策略 (一)~(三) 所求得的最近似聯合分配做比較。雖然特徵向量法之條件分配的誤差並非最小，但邊際分配的誤差為 0。若以總誤差 e 作為評量最近似聯合分配優劣的標準，整體來看，特徵向量法所推導出之最近似聯合分配，可視為使總誤差為最小的方法，詳細結果可參閱顧仲航 (2011)。

第三章　奇異值分解法

檢驗條件機率矩陣 A 和 B 是否相容，即檢驗比值矩陣 C 是否為秩 1 矩陣。若 C 是秩 1 矩陣，則 A 和 B 相容；若 C 不是秩 1 矩陣，則 A 和 B 不相容。然而檢驗一個矩陣是否為秩 1 矩陣或近似秩 1 矩陣的程度，可透過奇異值分解定理來判斷。本章介紹如何利用奇異值分解定理對比值矩陣 C 做分解來處理條件機率分配相容性問題，並以此為基礎發展出奇異值分解法。在使用該法時，需假設條件機率矩陣 A 和 B 的元素皆為正值，即比值矩陣 C 的元素皆為正值。

第一節　奇異值分解定理

本節將介紹奇異值分解定理，主要內容參考 David Poloe (2006) 和周志成 (2016)。

定理 3.1. *設 C 為 $I \times J$ 矩陣，且 $rank(C) = r$，則奇異值分解具有*

$$C = U\Sigma V'$$

的形式，其中 U 為 $I \times I$ 正交矩陣 (orthogonal matrix)，V 為 $J \times J$ 正交矩陣，即 $U^{-1} = U'$ 和 $V^{-1} = V'$。而 Σ 為 $I \times J$ 矩陣，其中奇異值 $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$，則 Σ 可表示如下：

$$\Sigma = \left(\begin{array}{c|c} D & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right), \quad D = \begin{pmatrix} \sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_r \end{pmatrix}.$$

奇異值分解定理除了前述形式外，還有類似譜分解 (spectral decomposition) 的表達方式，定理如下：

定理 3.2. *設 C 為 $I \times J$ 矩陣，且 $rank(C) = r$，則 C 可表示成 r 個秩 1 矩陣相加，即*

$$C = \sigma_1 \mathbf{u}_1 \mathbf{v}_1' + \sigma_2 \mathbf{u}_2 \mathbf{v}_2' + \cdots + \sigma_r \mathbf{u}_r \mathbf{v}_r', \tag{3.1}$$

其中 $\mathbf{u}_1, \cdots, \mathbf{u}_r$ 和 $\mathbf{v}_1, \cdots, \mathbf{v}_r$ 分別為定理 3.1 中之正交矩陣 U 和 V 的前 r 個行向量，且 $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ 為奇異值。

由上述定理可以得知，如果 I 和 J 皆很大且 r 遠小於 I 和 J 時，則矩陣 C 可僅用少部分的行向量、列向量和分塊即可表示，也就是透過奇異值分解的好處是可以大幅減少儲存量。

除了奇異值分解定理，也介紹其在運算上經常使用的性質，命題如下：

命題 3.3. *假設 C 為 $I \times J$ 矩陣，$rank(C) = r$，$\mathbf{u}_1, \cdots, \mathbf{u}_r$ 和 $\mathbf{v}_1, \cdots, \mathbf{v}_r$ 分別為定理 3.1 中之正交矩陣 U 和 V 的前 r 個行向量，且 $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ 為奇異值，則下列性質成立：*

(i) $CC'\mathbf{u}_l = \sigma_l^2 \mathbf{u}_l, l = 1, 2, \cdots, r, \quad CC'\mathbf{u}_l = \mathbf{0}, l = r+1, r+2, \cdots, I;$
$C'C\mathbf{v}_l = \sigma_l^2 \mathbf{v}_l, l = 1, 2, \cdots, r, \quad C'C\mathbf{v}_l = \mathbf{0}, l = r+1, r+2, \cdots, J.$

(ii) $C\mathbf{v}_l = \sigma_l \mathbf{u}_l, \quad C'\mathbf{u}_l = \sigma_l \mathbf{v}_l, \quad l = 1, 2, \cdots, r.$

定義 3.4. 設 $M = (m_{ij})$ 為 $I \times J$ 矩陣，則 M 的 Frobenius 範數 (frobenius norm) 定義如下：

$$\|M\|_F = \sqrt{\sum_{i=1}^{I} \sum_{j=1}^{J} |m_{ij}|^2} = \sqrt{trace(MM')}.$$

下面定理 3.5和推論 3.9是低秩近似法的特殊情況，也就是當低秩矩陣的秩為 1 的時候。一般情形的證明請參考 Markovsky, Ivan (2011) 的低秩近似法 (Low rank approximation)，並在此提供本文的證明方法。

定理 3.5. *設 C 為 $I \times J$ 比值矩陣，$rank(C) = r$，且 C 的奇異值分解如式子 (3.1) 所示。則 $\min_{\mathbf{x},\mathbf{y}} \|C - \mathbf{x}\mathbf{y}'\|_F^2 = \sigma_2^2 + \cdots + \sigma_r^2$，其中 $\mathbf{x} \in \mathbb{R}^I$，$\mathbf{y} \in \mathbb{R}^J$。*

Proof.

$$
\begin{aligned}
&\|C - \mathbf{x}\mathbf{y}'\|_F^2 \\
&= tr[(C - \mathbf{x}\mathbf{y}')'(C - \mathbf{x}\mathbf{y}')] \\
&= tr\Big[(\sigma_1\mathbf{u}_1\mathbf{v}_1' + \sigma_2\mathbf{u}_2\mathbf{v}_2' + \cdots + \sigma_r\mathbf{u}_r\mathbf{v}_r' - \mathbf{x}\mathbf{y}')'(\sigma_1\mathbf{u}_1\mathbf{v}_1' + \sigma_2\mathbf{u}_2\mathbf{v}_2' + \cdots + \sigma_r\mathbf{u}_r\mathbf{v}_r' - \mathbf{x}\mathbf{y}')\Big] \\
&= tr\Big[\sigma_1^2\mathbf{v}_1\mathbf{v}_1' - 2\sigma_1\mathbf{v}_1\mathbf{u}_1'\mathbf{x}\mathbf{y}' + \sigma_2^2\mathbf{v}_2\mathbf{v}_2' - 2\sigma_2\mathbf{v}_2\mathbf{u}_2'\mathbf{x}\mathbf{y}' + - \cdots + \sigma_r^2\mathbf{v}_r\mathbf{v}_r' - 2\sigma_r\mathbf{v}_r\mathbf{u}_r'\mathbf{x}\mathbf{y}' + \mathbf{y}\mathbf{x}'\mathbf{x}\mathbf{y}'\Big] \\
&= \sigma_1^2 - 2\sigma_1 tr(\mathbf{v}_1\mathbf{u}_1'\mathbf{x}\mathbf{y}') + \sigma_2^2 - 2\sigma_2 tr(\mathbf{v}_2\mathbf{u}_2'\mathbf{x}\mathbf{y}') + - \cdots + \sigma_r^2 - 2\sigma_r tr(\mathbf{v}_r\mathbf{u}_r'\mathbf{x}\mathbf{y}') + tr(\mathbf{y}\mathbf{x}'\mathbf{x}\mathbf{y}').
\end{aligned}
$$

為了方便起見，令 $\triangle = \|C - \mathbf{x}\mathbf{y}'\|_F^2$，接著對 $\triangle$ 的變數 $\mathbf{x}$ 和 $\mathbf{y}$ 做偏微分，可得

$$\frac{\partial\triangle}{\partial\mathbf{x}} = -2\sigma_1\mathbf{u}_1\mathbf{v}_1'\mathbf{y} - 2\sigma_2\mathbf{u}_2\mathbf{v}_2'\mathbf{y} - \cdots - 2\sigma_r\mathbf{u}_r\mathbf{v}_r'\mathbf{y} + 2\mathbf{x}(\mathbf{y}'\mathbf{y}).$$

令 $\frac{\partial\triangle}{\partial\mathbf{x}} = 0$，得到

$$\mathbf{x}\|\mathbf{y}\|^2 = \sigma_1\mathbf{u}_1\mathbf{v}_1'\mathbf{y} + \sigma_2\mathbf{u}_2\mathbf{v}_2'\mathbf{y} + \cdots + \sigma_r\mathbf{u}_r\mathbf{v}_r'\mathbf{y} = C\mathbf{y},$$

即

$$\mathbf{x} = C\mathbf{y}/\|\mathbf{y}\|^2. \tag{3.2}$$

同理，由 $\frac{\partial\triangle}{\partial\mathbf{y}} = 0$ 可得

$$\mathbf{y} = C'\mathbf{x}/\|\mathbf{x}\|^2. \tag{3.3}$$

由式子 (3.2) 和 (3.3) 得到

$$\|\mathbf{y}\|^2\mathbf{x} = C\mathbf{y} = CC'\mathbf{x}/\|\mathbf{x}\|^2,$$

即

$$CC'\mathbf{x} = \|\mathbf{x}\|^2\|\mathbf{y}\|^2\mathbf{x}.$$

同理，可得

$$C'C\mathbf{y} = \|\mathbf{x}\|^2\|\mathbf{y}\|^2\mathbf{y}.$$

從上面敘述得知，$\|\mathbf{x}\|^2\|\mathbf{y}\|^2$ 為 CC' 和 $C'C$ 的特徵值，且 $\mathbf{x}$ 和 $\mathbf{y}$ 分別為 CC' 和 $C'C$ 的特徵向量。令 $\mathbf{x} = d_1\mathbf{u}_1 + d_2\mathbf{u}_2 + \cdots + d_I\mathbf{u}_I$ 和 $\mathbf{y} = e_1\mathbf{v}_1 + e_2\mathbf{v}_2 + \cdots + e_J\mathbf{v}_J$。經由命題 3.3知道 $CC'\mathbf{u}_l = \mathbf{0}$，$l = r+1, r+2, \cdots, I$ 與 $C'C\mathbf{v}_l = \mathbf{0}$，$l = r+1, r+2, \cdots, J$，因此可將 $\mathbf{x}$ 和 $\mathbf{y}$ 改寫如下:

$$\mathbf{x} = d_1\mathbf{u}_1 + d_2\mathbf{u}_2 + \cdots + d_r\mathbf{u}_r, \quad \mathbf{y} = e_1\mathbf{v}_1 + e_2\mathbf{v}_2 + \cdots + e_r\mathbf{v}_r.$$

透過式子 (3.2) 得到

$$\begin{aligned} C\mathbf{y} &= (\sigma_1\mathbf{u}_1\mathbf{v}_1' + \sigma_2\mathbf{u}_2\mathbf{v}_2' + \cdots + \sigma_r\mathbf{u}_r\mathbf{v}_r')(e_1\mathbf{v}_1 + e_2\mathbf{v}_2 + \cdots + e_r\mathbf{v}_r) \\ &= \sigma_1 e_1\mathbf{u}_1 + \sigma_2 e_2\mathbf{u}_2 + \cdots + \sigma_r e_r\mathbf{u}_r, \end{aligned} \tag{3.4}$$

又

$$\begin{aligned} C\mathbf{y} &= \|\mathbf{y}\|^2\mathbf{x} \\ &= \|\mathbf{y}\|^2(d_1\mathbf{u}_1 + d_2\mathbf{u}_2 + \cdots + d_r\mathbf{u}_r). \end{aligned} \tag{3.5}$$

比較式子 (3.4) 和 (3.5) 可得

$$\sigma_l e_l = \|\mathbf{y}\|^2 d_l,\ l = 1, 2, \cdots, r. \tag{3.6}$$

同理，得到

$$\sigma_l d_l = \|\mathbf{x}\|^2 e_l,\ l = 1, 2, \cdots, r. \tag{3.7}$$

由式子 (3.6) 和 (3.7) 可得

$$\sigma_l^2 = \|\mathbf{x}\|^2 \|\mathbf{y}\|^2,\ l = 1, 2, \cdots, r. \tag{3.8}$$

重新計算 $\triangle$ 中的 $tr(\mathbf{v}_l \mathbf{u}_l' \mathbf{x}\mathbf{y}')$ 和 $tr(\mathbf{y}\mathbf{x}'\mathbf{x}\mathbf{y}')$，可得到它們分別為 $d_l e_l$ 和 σ_l^2，$l = 1, 2, \cdots, r$，所以 $\triangle$ 可改寫如下:

$$\begin{aligned}\triangle &= \sigma_1^2 - 2\sigma_1 tr(\mathbf{v}_1 \mathbf{u}_1' \mathbf{x}\mathbf{y}') + \sigma_2^2 - 2\sigma_2 tr(\mathbf{v}_2 \mathbf{u}_2' \mathbf{x}\mathbf{y}') + - \cdots + \sigma_r^2 - 2\sigma_r tr(\mathbf{v}_r \mathbf{u}_r' \mathbf{x}\mathbf{y}') + tr(\mathbf{y}\mathbf{x}'\mathbf{x}\mathbf{y}') \\ &= (\sigma_1^2 - 2\sigma_1 d_1 e_1) + (\sigma_2^2 - 2\sigma_2 d_2 e_2) + \cdots + (\sigma_r^2 - 2\sigma_r d_r e_r) + \|\mathbf{x}\|^2 \|\mathbf{y}\|^2.\end{aligned}$$

從式子 (3.8) 知道，只會有一個 l，使得 d_l 和 e_l 同時不為 0。假定為 k，使得 $\mathbf{x} = d_k \mathbf{u}_k$ 和 $\mathbf{y} = e_k \mathbf{v}_k$。從上述式子得到

$$\triangle = (\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_{k-1}^2 + \sigma_{k+1}^2 + \cdots + \sigma_r^2) + (\sigma_k^2 - 2\sigma_k d_k e_k) + \sigma_k^2, \tag{3.9}$$

經由計算可得

$$\|C - \mathbf{x}\mathbf{y}'\|_F^2 = \sigma_1^2 + \sigma_2^2 + \cdots + (\sigma_k - d_k e_k)^2 + \cdots + \sigma_r^2. \tag{3.10}$$

由式子 (3.9) 和 (3.10) 推得

$$2\sigma_k^2 - 2\sigma_k d_k e_k = (\sigma_k - d_k e_k)^2,$$

即

$$\sigma_k = d_k e_k. \tag{3.11}$$

由式子 (3.11)，可將式子 (3.10) 改寫如下：

$$\begin{aligned}\triangle &= (\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_{k-1}^2 + \sigma_{k+1}^2 + \cdots + \sigma_r^2) + 2\sigma_k^2 - 2\sigma_k d_k e_k \\ &= \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_{k-1}^2 + \sigma_{k+1}^2 + \cdots + \sigma_r^2.\end{aligned}$$

綜合以上結果，即 $\mathbf{x}$ 和 $\mathbf{y}'$ 的乘積具有 $d_l e_l \mathbf{u}_l \mathbf{v}_l'$，$l = 1, 2, \cdots, r$ 的形式。可得當 $l = 1$ 時，$\triangle$ 會有最小值，也就是當 $\mathbf{x} = d_1 \mathbf{u}_1$ 和 $\mathbf{y} = e_1 \mathbf{v}_1$ 時，C 和 $\mathbf{xy}'$ 的誤差平方和最小。

$\square$

定理 3.6. *設 C 為 $I \times J$ 比值矩陣，$rank(C) = r$。令 $\mathbf{x} = \sum_{i=1}^{I} d_i \mathbf{u}_i$，$\mathbf{y} = \sum_{j=1}^{J} e_j \mathbf{v}_j$，其中 $0 \le d_i \le 1$，$\sum_{i=1}^{I} d_i = 1$；$0 \le e_j \le 1$，$\sum_{j=1}^{J} e_j = 1$。且 $D = \mathbf{x}\mathbf{y}'$，則 $\|C - D\|_F^2 = \sum_{l=1}^{r} \sigma_l^2 - 2\sum_{l=1}^{r} \sigma_l d_l e_l + \sum_{i=1}^{I}\sum_{j=1}^{J} d_i^2 e_j^2$。*

Proof.

$$
\begin{aligned}
\|C - \mathbf{x}\mathbf{y}'\|_F^2 &= \| \sum_{l=1}^{r} \sigma_l \mathbf{u}_l \mathbf{v}_l' - (\sum_{i=1}^{I} d_i \mathbf{u}_i)(\sum_{j=1}^{J} e_j \mathbf{v}_j)' \|_F^2 \\
&= \| \sum_{l=1}^{r} \sigma_l \mathbf{u}_l \mathbf{v}_l' - \sum_{i=1}^{I}\sum_{j=1}^{J} d_i e_j \mathbf{u}_i \mathbf{v}_j' \|_F^2 \\
&= tr\left[(\sum_{l=1}^{r} \sigma_l \mathbf{u}_l \mathbf{v}_l' - \sum_{i=1}^{I}\sum_{j=1}^{J} d_i e_j \mathbf{u}_i \mathbf{v}_j')(\sum_{l=1}^{r} \sigma_l \mathbf{u}_l \mathbf{v}_l' - \sum_{i=1}^{I}\sum_{j=1}^{J} d_i e_j \mathbf{u}_i \mathbf{v}_j')' \right] \\
&= \sum_{l=1}^{r} \sigma_l^2 - 2\sum_{l=1}^{r} \sigma_l d_l e_l + \sum_{i=1}^{I}\sum_{j=1}^{J} d_i^2 e_j^2.
\end{aligned}
$$

□

推論 3.7. *在定理 3.6中，若 $d_2, \cdots, d_I, e_2, \cdots, e_J$ 都很接近 0，則 $\|C - D\|_F^2 \approx (\sigma_1 - d_1 e_1)^2 + \sigma_2^2 + \cdots + \sigma_r^2$。*

在定理 3.6中，令 $D = \rho \mathbf{u}_l \mathbf{v}_l'$，即為 $D = \mathbf{x}\mathbf{y}'$ 的特殊情況，推得結果如下：

定理 3.8. *設 C 為 $I \times J$ 比值矩陣，$rank(C) = r$，且 C 的奇異值分解如式子 (3.1) 所示。令 $D = \rho \mathbf{u}_l \mathbf{v}_l'$，則 $\|C - D\|_F^2 = \sigma_1^2 + \cdots + \sigma_{l-1}^2 + (\sigma_l - \rho)^2 + \sigma_{l+1}^2 + \cdots \sigma_r^2$。*

Proof.

因 $D = \rho \mathbf{u}_l \mathbf{v}_l'$，故 $\rho = d_l e_l$，且 $\forall i \neq l$，$d_i = 0$，$e_i = 0$。所以

$$\begin{aligned}
\|C - D\|_F^2 &= \sum_{l=1}^{r} \sigma_l^2 - 2\sigma_l d_l e_l + d_l^2 e_l^2 \\
&= \sigma_1^2 + \cdots + \sigma_{l-1}^2 + (\sigma_l - d_l e_l)^2 + \sigma_{l+1}^2 + \cdots + \sigma_r^2 \\
&= \sigma_1^2 + \cdots + \sigma_{l-1}^2 + (\sigma_l - \rho)^2 + \sigma_{l+1}^2 + \cdots + \sigma_r^2.
\end{aligned}$$

□

推論 3.9. *$\|C - \sigma_1 \mathbf{u}_1 \mathbf{v}_1'\|_F^2 = \sigma_2^2 + \cdots + \sigma_r^2$，即 $\sigma_1 \mathbf{u}_1 \mathbf{v}_1'$ 是 C 的最近似秩 1 矩陣。*

當 A 和 B 不相容時，怎麼知道找到的近似秩 1 矩陣有多接近比值矩陣呢？通常我們會選擇其最大奇異值佔所有奇異值總和比值是否接近 1，即 σ_1 相對其他奇異值比較大，而 $\sigma_2, \cdots, \sigma_r$ 比較小。

第二節　奇異值分解法

本節主要介紹奇異值分解法來解決二維條件分配相容性問題。利用比值矩陣法的結果以及奇異值分解定理，我們有下面的定理：

定理 3.10. *設 $A = (a_{ij})$ 和 $B = (b_{ij})$ 為 $I \times J$ 條件機率矩陣，$C = (c_{ij})$ 為 $I \times J$ 比值矩陣，$rank(C) = r$。令 C 的奇異值分解為*

$$C = \sigma_1 \mathbf{u}_1 \mathbf{v}_1' + \sigma_2 \mathbf{u}_2 \mathbf{v}_2' + \cdots + \sigma_r \mathbf{u}_r \mathbf{v}_r',$$

則 A 和 B 相容的充要條件為 $r = 1$。

推論 3.11. *設 $A = (a_{ij})$ 和 $B = (b_{ij})$ 為 $I \times J$ 條件機率矩陣，$C = (c_{ij})$ 為 $I \times J$ 比值矩陣，$rank(C) = r$。令 $\mathbf{u}_1 = (u_{11} \cdots u_{I1})'$，$\mathbf{v}_1 = (v_{11} \cdots v_{J1})'$，則 A 和 B 相容的充要條件為 $\sigma_1 = \sum\limits_{j=1}^{J} \frac{1}{v_{j1}} \Big/ \sum\limits_{i=1}^{I} u_{i1}$ 且 $\sigma_2 = \cdots = \sigma_r = 0$。*

Proof.

設 A 和 B 相容，則 $C = \sigma_1 \mathbf{u}_1 \mathbf{v}_1'$，其中

$$\mathbf{u}_1 = \begin{pmatrix} u_{11} \\ \vdots \\ u_{I1} \end{pmatrix}, \quad \mathbf{v}_1 = \begin{pmatrix} v_{11} \\ \vdots \\ v_{J1} \end{pmatrix},$$

則 $\boldsymbol{\tau} = (\tau_i)$，$\boldsymbol{\eta} = (\eta_j)$ 分別是 X，Y 的邊際分配，其中

$$\tau_i = \frac{u_{i1}}{\sum\limits_{i=1}^{I} u_{i1}}, \quad \eta_j = \frac{1}{v_{j1} \sum\limits_{k=1}^{J} \frac{1}{v_{k1}}}.$$

根據 Arnold and Press (1989)，

$$\begin{aligned} c_{ij} &= \tau_i \frac{1}{\eta_j} \\ &= \left(\frac{u_{i1}}{\sum\limits_{i=1}^{I} u_{i1}} \right) \left(v_{j1} \sum_{k=1}^{J} \frac{1}{v_{k1}} \right) \\ &= \frac{\sum\limits_{k=1}^{J} \frac{1}{v_{k1}}}{\sum\limits_{i=1}^{I} u_{i1}} u_{i1} v_{j1} \\ &= \frac{\sum\limits_{k=1}^{J} \frac{1}{v_{k1}}}{\sum\limits_{i=1}^{I} u_{i1}} \frac{c_{ij}}{\sigma_1}, \end{aligned}$$

故 $\sigma_1 = \sum\limits_{j=1}^{J} \frac{1}{v_{j1}} \Big/ \sum\limits_{i=1}^{I} u_{i1}$。

□

若 A 和 B 相容，即 $r = 1$，則存在向量 $\mathbf{u} = (u_i)$ 和 $\mathbf{v} = (v_j)$ 使得 $c_{ij} = u_i v_j$，$\forall (i,j) \in N$，即 $C = \mathbf{u}\mathbf{v}'$。利用向量 $\mathbf{u}$ 和 $\mathbf{v}$ 可得 X 和 Y 的邊際分配 $\boldsymbol{\tau} = (\tau_i)$ 和 $\boldsymbol{\eta} = (\eta_j)$，其中 τ_i，η_j 如式子 (1.1) 所定義；再加上給定的條件機率分配，可得 (X, Y) 的聯合機率分配 $P = (p_{ij})$，也就是

$$p_{ij} = \tau_i b_{ij} = \eta_j a_{ij}. \tag{3.12}$$

若 A 和 B 不相容，即 $r \geq 2$，如何求出近似聯合機率分配。由向量 $\mathbf{u}_1$ 和 $\mathbf{v}_1$，利用式子 (1.1) 找出 X 和 Y 的近似邊際分配；再加上給定的條件機率分配，利用式子 (3.12) 得到兩組 (X, Y) 的近似聯合機率分配 $P^{(1)}$ 和 $P^{(2)}$。針對上述兩組近似聯合機率分配，我們該如何組合才能得到更好的近似聯合機率分配 P 呢？最簡單的組合方式就是權重平均相加，並稱其為等加權法，即

$$P = (P^{(1)} + P^{(2)})/2. \tag{3.13}$$

而更好的組合方式則放在第五章討論。

當 A 和 B 不相容且比值矩陣 C 有 $*$ 時，可藉由矩形法則，把出現 $*$ 的地方補起來，得到擴張矩陣 $\overline{C}$，但補法不唯一；接著，如同比值矩陣的元素皆為正值的情況，可對 $\overline{C}$ 做奇異值分解，以辨別 $\overline{C}$ 近似秩 1 矩陣的程度。同時利用最大奇異值所對應的向量 $\mathbf{u}_1$ 和 $\mathbf{v}_1$，找出近似聯合分配。奇異值分解法的好處是可快速地判斷條件機率矩陣 A 和 B 是否相容；若不相容，則可根據上述過程找到近似聯合機率分配。

第三節　實例分析

本節主要以實例說明，如何以奇異值分解法來處理相容性問題。

首先，我們考慮比值矩陣 C 的元素皆為正值且相容的例子：

例題 1. (相容) 給定條件機率矩陣 A 和 B，

$$A = \begin{pmatrix} \frac{3}{8} & \frac{3}{8} & \frac{3}{8} \\ \frac{3}{8} & \frac{3}{8} & \frac{3}{8} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{pmatrix}, \quad B = \begin{pmatrix} \frac{2}{7} & \frac{2}{7} & \frac{3}{7} \\ \frac{2}{7} & \frac{2}{7} & \frac{3}{7} \\ \frac{2}{7} & \frac{2}{7} & \frac{3}{7} \end{pmatrix},$$

A 對 B 的比值矩陣 C 為

$$C = \begin{pmatrix} \frac{21}{16} & \frac{21}{16} & \frac{7}{8} \\ \frac{21}{16} & \frac{21}{16} & \frac{7}{8} \\ \frac{7}{8} & \frac{7}{8} & \frac{7}{12} \end{pmatrix}.$$

接著對比值矩陣 C 做奇異值分解，可得

$$\begin{aligned} C &= \sigma_1 \mathbf{u}_1 \mathbf{v}_1' \\ &= 3.2083 \begin{pmatrix} 0.6396 \\ 0.6396 \\ 0.4264 \end{pmatrix} \begin{pmatrix} 0.6396 & 0.6396 & 0.4264 \end{pmatrix}. \end{aligned}$$

透過奇異值分解法，得到 $rank(C) = 1$，滿足定理 3.10，故 A 和 B 相容。

利用向量 $\mathbf{u}_1$ 和 $\mathbf{v}_1$，透過式子 (1.1) 可得 X，Y 的邊際分配配 $\boldsymbol{\tau}$ 和 $\boldsymbol{\eta}$，即

$$\boldsymbol{\tau} = \begin{pmatrix} 0.3750 \\ 0.3750 \\ 0.2500 \end{pmatrix}, \quad \boldsymbol{\eta} = \begin{pmatrix} 0.2857 \\ 0.2857 \\ 0.4286 \end{pmatrix};$$

再加上給定的條件機率分配，根據式子 (3.12) 可得 (X, Y) 的聯合機率分配 P，即

$$P = \begin{pmatrix} 0.1071 & 0.1071 & 0.1607 \\ 0.1071 & 0.1071 & 0.1607 \\ 0.0714 & 0.0714 & 0.1071 \end{pmatrix}.$$

接著，我們考慮比值矩陣 C 的元素皆為正值且不相容的例子：

例題 2. (不相容) 給定條件機率矩陣 A 和 B，

$$A = \begin{pmatrix} \frac{1}{7} & \frac{1}{2} & \frac{3}{7} & \frac{1}{7} \\ \frac{2}{7} & \frac{1}{4} & \frac{1}{7} & \frac{2}{7} \\ \frac{4}{7} & \frac{1}{4} & \frac{3}{7} & \frac{4}{7} \end{pmatrix}, \quad B = \begin{pmatrix} \frac{1}{6} & \frac{1}{6} & \frac{1}{2} & \frac{1}{6} \\ \frac{2}{7} & \frac{1}{7} & \frac{2}{7} & \frac{2}{7} \\ \frac{1}{3} & \frac{1}{12} & \frac{1}{4} & \frac{1}{3} \end{pmatrix},$$

A 對 B 的比值矩陣 C 為

$$C = \begin{pmatrix} \frac{6}{7} & 3 & \frac{6}{7} & \frac{6}{7} \\ 1 & \frac{7}{4} & \frac{1}{2} & 1 \\ \frac{12}{7} & 3 & \frac{12}{7} & \frac{12}{7} \end{pmatrix}.$$

接著對比值矩陣 C 做奇異值分解，可得

$$\begin{aligned} C &= U\Sigma V' \\ &= U \begin{pmatrix} 5.7868 & 0 & 0 & 0 \\ 0 & 0.8402 & 0 & 0 \\ 0 & 0 & 0.3740 & 0 \end{pmatrix} V', \end{aligned}$$

其中

$$U=\begin{pmatrix}0.5663 & 0.8078 & -0.1635\\ 0.3937 & -0.0909 & 0.9148\\ 0.7241 & -0.5824 & -0.3695\end{pmatrix},\quad V=\begin{pmatrix}0.3664 & -0.4723 & 0.3777 & 0.7071\\ 0.7880 & 0.6156 & 0.0053 & 0\\ 0.3324 & -0.4182 & -0.8453 & 0\\ 0.3664 & -0.4723 & 0.3777 & -0.7071\end{pmatrix}.$$

由定理 3.10，故 A 和 B 不相容。利用向量 $\mathbf{u}_1$ 和 $\mathbf{v}_1$，透過式子 (1.1) 得到 X，Y 的近似邊際分配 $\boldsymbol{\tau}$ 和 $\boldsymbol{\eta}$，

$$\boldsymbol{\tau}=\begin{pmatrix}0.3363\\ 0.2338\\ 0.4300\end{pmatrix},\quad \boldsymbol{\eta}=\begin{pmatrix}0.2803\\ 0.1303\\ 0.3090\\ 0.2803\end{pmatrix};$$

再加上給定的條件機率分配，根據式子 (3.12) 可得兩組 (X,Y) 的近似聯合機率分配 $P^{(1)}$ 和 $P^{(2)}$，

$$P^{(1)}=\begin{pmatrix}0.0560 & 0.0560 & 0.1681 & 0.0560\\ 0.0668 & 0.0334 & 0.0668 & 0.0668\\ 0.1433 & 0.0358 & 0.1075 & 0.1433\end{pmatrix},\quad P^{(2)}=\begin{pmatrix}0.0400 & 0.0652 & 0.1324 & 0.0400\\ 0.0801 & 0.0326 & 0.0441 & 0.0801\\ 0.1602 & 0.0326 & 0.1324 & 0.1602\end{pmatrix},$$

按式子 (3.13) 將 $P^{(1)}$ 和 $P^{(2)}$ 組合起來得下列近似聯合分配：

$$P=\begin{pmatrix}0.0480 & 0.0606 & 0.1503 & 0.0480\\ 0.0734 & 0.0330 & 0.0555 & 0.0734\\ 0.1518 & 0.0342 & 0.1200 & 0.1518\end{pmatrix}. \tag{3.14}$$

最後我們考慮比值矩陣 C 的元素有 $*$ 且不相容的例子：

例題 3. (不相容) 給定條件機率矩陣 A 和 B，

$$A = \begin{pmatrix} \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & 0 \\ \frac{1}{2} & \frac{2}{3} & 0 & \frac{5}{6} \\ 0 & 0 & \frac{3}{4} & \frac{1}{6} \end{pmatrix}, \quad B = \begin{pmatrix} \frac{1}{2} & \frac{1}{3} & \frac{1}{6} & 0 \\ \frac{1}{4} & \frac{1}{3} & 0 & \frac{5}{12} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix},$$

A 對 B 的比值矩陣 C 為

$$C = \begin{pmatrix} 1 & 1 & \frac{3}{2} & * \\ 2 & 2 & * & 2 \\ * & * & \frac{3}{2} & \frac{1}{3} \end{pmatrix}.$$

此時無法直接用奇異值分解定理，藉由矩形法則 (請參閱 Arnold and Press (1989))，把 $*$ 的位置補滿，得到擴張矩陣 $\overline{C}$ 如下，但可能還有其他補法，即

$$\overline{C_1} = \begin{pmatrix} 1 & 1 & \frac{3}{2} & 1 \\ 2 & 2 & 3 & 2 \\ 1 & 1 & \frac{3}{2} & \frac{1}{3} \end{pmatrix}, \quad \overline{C_2} = \begin{pmatrix} 1 & 1 & \frac{3}{2} & 1 \\ 2 & 2 & 3 & 2 \\ \frac{1}{3} & \frac{1}{3} & \frac{3}{2} & \frac{1}{3} \end{pmatrix}, \quad \overline{C_3} = \begin{pmatrix} 1 & 1 & \frac{3}{2} & \frac{1}{3} \\ 2 & 2 & 3 & 2 \\ \frac{1}{3} & \frac{1}{3} & \frac{3}{2} & \frac{1}{3} \end{pmatrix};$$

其中 $\overline{C_1}$、$\overline{C_2}$ 和 $\overline{C_3}$ 其最大奇異值佔所有奇異值總和的百分比分別為 89.60%、87.96% 和 80.88%，所以我們選擇 $\overline{C_1}$ 去逼近比值矩陣 C。

如同比值矩陣的元素皆為正值的情況，透過奇異值分解定理來檢驗 $\overline{C_1}$ 是否為秩 1 矩陣或近似秩 1 矩陣。

接著對擴張矩陣 $\overline{C_1}$ 做奇異值分解，可得

$$\begin{aligned}\overline{C_1} =&\sigma_1\mathbf{u}_1\mathbf{v}_1' + \sigma_2\mathbf{u}_2\mathbf{v}_2'\\ =&5.5045\begin{pmatrix}0.4159\\0.8318\\0.3675\end{pmatrix}\begin{pmatrix}0.4446 & 0.4446 & 0.6668 & 0.4001\end{pmatrix}\\ &+0.5583\begin{pmatrix}-0.1643\\-0.3287\\0.9300\end{pmatrix}\begin{pmatrix}0.1941 & 0.1941 & 0.2911 & -0.9165\end{pmatrix}.\end{aligned}$$

由定理 3.10，故 A 和 B 不相容。利用向量 $\mathbf{u}_1$ 和 $\mathbf{v}_1$，透過式子 (1.1) 可得 X，Y 的近似邊際分配 $\boldsymbol{\tau}$ 和 $\boldsymbol{\eta}$，

$$\boldsymbol{\tau} = \begin{pmatrix}0.2575\\0.5150\\0.2275\end{pmatrix},\quad \boldsymbol{\eta} = \begin{pmatrix}0.2647\\0.2647\\0.1765\\0.2941\end{pmatrix};$$

再加上給定的條件機率分配，根據式子 (3.12) 可得兩組 (X,Y) 的近似聯合機率分配 $P^{(1)}$ 和 $P^{(2)}$，

$$P^{(1)} = \begin{pmatrix}0.1287 & 0.0858 & 0.0429 & 0\\0.1287 & 0.1717 & 0 & 0.2146\\0 & 0 & 0.1138 & 0.1138\end{pmatrix},\quad P^{(2)} = \begin{pmatrix}0.1323 & 0.0882 & 0.0441 & 0\\0.1323 & 0.1765 & 0 & 0.2451\\0 & 0 & 0.1323 & 0.0490\end{pmatrix}.$$

仿照例題 2，按式子 (3.13) 將 $P^{(1)}$ 和 $P^{(2)}$ 組合起來得下列近似聯合分配：

$$P = \begin{pmatrix} 0.1305 & 0.0870 & 0.0435 & 0 \\ 0.1305 & 0.1741 & 0 & 0.2299 \\ 0 & 0 & 0.1230 & 0.0814 \end{pmatrix}.$$

在第五章中，我們會討論如何將 $P^{(1)}$ 和 $P^{(2)}$ 組合起來的較好方法。

第四章　最近似秩 1 矩陣法

當比值矩陣 C 的元素有 $*$ 時，我們無法直接用奇異值分解定理來判斷給定的條件機率分配是否相容。Song et al. (2010) 提出 IBD 矩陣的技巧來解決問題。本章中根據定理 3.5以及類 Frobenius 範數的概念，我們提出求比值矩陣 C 之最近似秩 1 矩陣的方法。由類 Frobenius 範數的值是否為 0，我們就可判斷條件機率矩陣是否相容；同時由所求出的最近似秩 1 矩陣，我們可獲得聯合機率分配或近似聯合機率分配。這個方法我們稱為最近似秩 1 矩陣法，可視為奇異值分解法之延伸。

第一節　最近似秩 1 矩陣

設 X 和 Y 為有限、離散隨機變數，A 和 B 分別為 $X|Y$ 和 $Y|X$ 的條件機率矩陣，C 為 A 對 B 的比值矩陣。

若 A 和 B 相容，則

$$\begin{aligned}
c_{ij} &= \frac{a_{ij}}{b_{ij}} \\
&= \frac{P(X=i|Y=j)}{P(Y=j|X=i)} \\
&= \frac{P(X=i)}{P(Y=j)} \\
&= P(X=i)\left[\frac{1}{P(Y=j)}\right],
\end{aligned}$$

即

$$C = \mathbf{x}\mathbf{z}',$$

其中

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_I \end{pmatrix}, x_i = P(X = i); \quad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_J \end{pmatrix}, y_j = P(Y = j); \quad \mathbf{z} = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_J \end{pmatrix}, z_j = \frac{1}{y_j}.$$

值得注意的是 C 的秩是 1，$\sum_{i=1}^{I} x_i = 1$，$\sum_{j=1}^{J} y_j = \sum_{j=1}^{J} \frac{1}{z_j} = 1$。

對一般的比值矩陣 C，我們的目的是先求 C 的最近似秩 1 矩陣，即求 $\mathbf{x}$，$\mathbf{z}$ 使得 $\|C - \mathbf{x}\mathbf{z}'\|_F^2$ 最小，且滿足下列條件

$$\begin{cases} 1'\mathbf{x} = 1 \\ 1'\mathbf{y} = 1, \quad y_j = \dfrac{1}{z_j}, \quad j = 1, \cdots, J, \end{cases} \tag{4.1}$$

其中

$$\|C - \mathbf{x}\mathbf{z}'\|_F^2 = \sum\sum_{(i,j)\in N}(c_{ij} - x_i z_j)^2, \quad N = \{(i,j)|c_{ij} > 0\}.$$

當 C 裡面的元素都為正值時，則 $\|C - \mathbf{x}\mathbf{z}'\|_F$ 就是 Frobenius 範數；當 C 裡面的元素有 $*$ 時，則 $\|C - \mathbf{x}\mathbf{z}'\|_F$ 為類 Frobenius 範數。若 $\mathbf{x}$，$\mathbf{z}$ 滿足條件式子 (4.1)，且求出來的 Frobenius 範數最小，則 $\mathbf{x}\mathbf{z}'$ 稱為最近似秩 1 矩陣。最近似秩 1 矩陣法的好處是在比值矩陣 C 的元素有 $*$ 時，不會因為矩形法則的補法不唯一或 $*$ 出現太頻繁而造成計算上的困難，根據相同步驟可幫助我們找到近似聯合分配。

第二節　無限制條件法

我們首先分兩種情形來找最近似秩 1 矩陣。(一) 先不管限制條件，即先不考慮式子 (4.1) 情形下，導出求最近似秩 1 矩陣的相關數學式子；(二) 考慮限制條件 (4.1) 下，利用 Lagrange 乘數法導出求最近似秩 1 矩陣的相關數學式子。為了方便起見，這兩種方法分別稱為無限制條件法與 Lagrange 乘數法。針對這兩種方法，我們也分別提供了求最近似秩 1 矩陣的演算法。在本節中，我們先考慮無限制條件法。

我們的目的是找出比值矩陣 C 的最近似秩 1 矩陣，即找到 $\mathbf{x}$，$\mathbf{z}$ 使得 $\|C-\mathbf{x}\mathbf{z}'\|_F^2$ 最小，其中 $\mathbf{x}$ 和 $\mathbf{z}$ 滿足式子 (4.1)。為了方便處理起見，將符號 $\mathbf{y}$ 和 $\mathbf{z}$ 交換，並重新敘述問題如下：

求 $\mathbf{x}$，$\mathbf{y}$ 使 $L=\|C-\mathbf{x}\mathbf{y}'\|_F^2=\sum\sum_{(i,j)\in N}(c_{ij}-x_iy_j)^2$ 最小，且滿足下列條件

$$\begin{cases} 1'\mathbf{x}=1 \\ 1'\mathbf{z}=1, \quad z_j=\frac{1}{y_j}, \quad j=1,\cdots,J. \end{cases} \tag{4.2}$$

L 對 $\mathbf{x}$ 和 $\mathbf{y}$ 做偏微分，可得

$$\frac{\partial L}{\partial \mathbf{x}}=-2C^*\mathbf{y}+2diag\left(\sum_{(1,j)\in N}y_j^2,\cdots,\sum_{(I,j)\in N}y_j^2\right)\mathbf{x}.$$
$$\frac{\partial L}{\partial \mathbf{y}}=-2(C^*)'\mathbf{x}+2diag\left(\sum_{(i,1)\in N}x_i^2,\cdots,\sum_{(i,J)\in N}x_i^2\right)\mathbf{y}.$$

其中 C^* 是 $I*J$ 矩陣，將 C 中的元素 $*$ 以 0 來取代，而其他正值元素保持不變。令 $\frac{\partial L}{\partial \mathbf{x}}=0$ 和 $\frac{\partial L}{\partial \mathbf{y}}=0$，得下列關係式：

$$\mathbf{x} = diag\left(1 \Big/ \sum_{(1,j)\in N} y_j^2, \cdots, 1 \Big/ \sum_{(I,j)\in N} y_j^2\right)(C^*\mathbf{y})$$

$$= \begin{bmatrix} \left(\sum_{(1,j)\in N} c_{1j}y_j\right) \Big/ \sum_{(1,j)\in N} y_j^2 \\ \left(\sum_{(2,j)\in N} c_{2j}y_j\right) \Big/ \sum_{(2,j)\in N} y_j^2 \\ \vdots \\ \left(\sum_{(I,j)\in N} c_{Ij}y_j\right) \Big/ \sum_{(I,j)\in N} y_j^2 \end{bmatrix}, \tag{4.3}$$

$$\mathbf{y} = diag\left(1 \Big/ \sum_{(i,1)\in N} x_i^2, \cdots, 1 \Big/ \sum_{(i,J)\in N} x_i^2\right)((C^*)'\mathbf{x})$$

$$= \begin{bmatrix} \left(\sum_{(i,1)\in N} c_{i1}x_i\right) \Big/ \sum_{(i,1)\in N} x_i^2 \\ \left(\sum_{(i,2)\in N} c_{i2}x_i\right) \Big/ \sum_{(i,2)\in N} x_i^2 \\ \vdots \\ \left(\sum_{(i,J)\in N} c_{iJ}x_i\right) \Big/ \sum_{(i,J)\in N} x_i^2 \end{bmatrix}. \tag{4.4}$$

給定任何兩個正向量 $\mathbf{x} = (x_1, x_2, \cdots, x_I)'$ 和 $\mathbf{y} = (y_1, y_2, \cdots, y_J)'$，對其做下列修正，可以滿足式子 (4.2)。

$$\mathbf{x}^* = \mathbf{x} \Big/ \sum_{i=1}^{I} x_i, \quad \mathbf{y}^* = \left(\sum_{j=1}^{J} \frac{1}{y_j} \right) \mathbf{y}. \tag{4.5}$$

求最近似秩 1 矩陣之演算法 (一)

1. 選定初值 $(\mathbf{x}^{(0)}, \mathbf{y}^{(0)})$。
2. 將 $\mathbf{y}^{(0)}$ 和 $\mathbf{x}^{(0)}$ 分別代入式子 (4.3) 和 (4.4)，再把計算結果 $\mathbf{x}$ 和 $\mathbf{y}$ 代入式子 (4.5) 做修正，得 $\mathbf{x}^{(1)}$ 和 $\mathbf{y}^{(1)}$，並計算 $L_1 = \|C - \mathbf{x}^{(1)}\mathbf{y}^{(1)'}\|_F^2$。
3. 從 $(\mathbf{x}^{(1)}, \mathbf{y}^{(1)})$ 開始，重複步驟 2，得 $(\mathbf{x}^{(2)}, \mathbf{y}^{(2)})$ 以及 $L_2 = \|C - \mathbf{x}^{(2)}\mathbf{y}^{(2)'}\|_F^2$。依此類推，可得 $(\mathbf{x}^{(3)}, \mathbf{y}^{(3)})$ 以及 $L_3 = \|C - \mathbf{x}^{(3)}\mathbf{y}^{(3)'}\|_F^2 \cdots$。
4. 在第 $n+1$ 步驟時，若 $|L_{n+1} - L_n| \leq 10^{-5}$，則停止計算，然後利用 $\mathbf{x}^{(n)}$ 和 $\mathbf{y}^{(n)}$ 來決定最近似秩 1 矩陣。

第三節 Lagrange 乘數法

本節中，在限制條件 (4.2) 下，考慮如何利用 Lagrange 乘數法來幫助我們求得最近似秩 1 矩陣。

利用 Lagrange 乘數法，我們令

$$
\begin{aligned}
L^* &= L + \lambda_1(1'\mathbf{x} - 1) + \lambda_2(1'\mathbf{z} - 1) + \sum_{j=1}^{J} \rho_j(y_j z_j - 1) \\
&= \|C - \mathbf{x}\mathbf{y}'\|_F^2 + \lambda_1(1'\mathbf{x} - 1) + \lambda_2(1'\mathbf{z} - 1) + \sum_{j=1}^{J} \rho_j(y_j z_j - 1) \\
&= \sum\sum\nolimits_{(i,j)\in N}(c_{ij} - x_i y_j)^2 + \lambda_1(1'\mathbf{x} - 1) + \lambda_2(1'\mathbf{z} - 1) + \sum_{j=1}^{J} \rho_j(y_j z_j - 1).
\end{aligned}
$$

L^* 對 $\mathbf{x}$、$\mathbf{y}$、$\mathbf{z}$、λ_1、λ_2 和 $\boldsymbol{\rho}$ 做偏微分，可得

$$
\frac{\partial L^*}{\partial \mathbf{x}} = -2C^*\mathbf{y} + 2diag\left(\sum_{(1,j)\in N} y_j^2, \cdots, \sum_{(I,j)\in N} y_j^2\right)\mathbf{x} + \lambda_1 \underset{\sim}{1}. \tag{4.6}
$$

$$
\frac{\partial L^*}{\partial \mathbf{y}} = -2(C^*)'\mathbf{x} + 2diag\left(\sum_{(i,1)\in N} x_i^2, \cdots, \sum_{(i,J)\in N} x_i^2\right)\mathbf{y} + diag\,(\rho_1, \cdots, \rho_J)\,\mathbf{z}. \tag{4.7}
$$

$$
\frac{\partial L^*}{\partial \mathbf{z}} = \lambda_2 \underset{\sim}{1} + diag\,(\rho_1, \cdots, \rho_J)\,\mathbf{y}. \tag{4.8}
$$

$$
\frac{\partial L^*}{\partial \lambda_1} = 1'\mathbf{x} - 1. \tag{4.9}
$$

$$
\frac{\partial L^*}{\partial \lambda_2} = 1'\mathbf{z} - 1. \tag{4.10}
$$

$$
\frac{\partial L^*}{\partial \boldsymbol{\rho}} = diag\,(y_1, \cdots, y_J)\,\mathbf{z} - 1. \tag{4.11}
$$

其中 C^* 是 $I * J$ 矩陣，將 C 中的元素 $*$ 以 0 來取代，而其他正值元素保持不變，$\boldsymbol{\rho} = (\rho_1, \rho_2, \cdots, \rho_J)'$。

令 $\dfrac{\partial L^*}{\partial \mathbf{x}} = 0$、$\dfrac{\partial L^*}{\partial \mathbf{y}} = 0$、$\dfrac{\partial L^*}{\partial \mathbf{z}} = 0$、$\dfrac{\partial L^*}{\partial \lambda_1} = 0$、$\dfrac{\partial L^*}{\partial \lambda_2} = 0$ 和 $\dfrac{\partial L^*}{\partial \boldsymbol{\rho}} = 0$，得下列關係式：

$$-2C^*\mathbf{y} + 2diag\left(\sum_{(1,j)\in N} y_j^2, \cdots, \sum_{(I,j)\in N} y_j^2\right)\mathbf{x} + \lambda_1 \underset{\sim}{1} = 0. \tag{4.12}$$

$$-2(C^*)'\mathbf{x} + 2diag\left(\sum_{(i,1)\in N} x_i^2, \cdots, \sum_{(i,J)\in N} x_i^2\right)\mathbf{y} + diag\left(\rho_1, \cdots, \rho_J\right)\mathbf{z} = 0. \tag{4.13}$$

$$\lambda_2 \underset{\sim}{1} + diag\left(\rho_1, \cdots, \rho_J\right)\mathbf{y} = 0. \tag{4.14}$$

$$1'\mathbf{x} - 1 = 0. \tag{4.15}$$

$$1'\mathbf{z} - 1 = 0. \tag{4.16}$$

$$diag\left(y_1, \cdots, y_J\right)\mathbf{z} - 1 = 0. \tag{4.17}$$

由式子 (4.12) 和 (4.13) 得到

$$\mathbf{x} = diag\left(1 \Big/ \sum_{(1,j)\in N} y_j^2, \cdots, 1 \Big/ \sum_{(I,j)\in N} y_j^2\right)\left(C^*\mathbf{y} - \frac{\lambda_1}{2}\underset{\sim}{1}\right)$$

$$= \begin{bmatrix} \left(\sum_{(1,j)\in N} c_{1j}y_j - \frac{\lambda_1}{2}\right) \Big/ \sum_{(1,j)\in N} y_j^2 \\ \left(\sum_{(2,j)\in N} c_{2j}y_j - \frac{\lambda_1}{2}\right) \Big/ \sum_{(2,j)\in N} y_j^2 \\ \vdots \\ \left(\sum_{(I,j)\in N} c_{Ij}y_j - \frac{\lambda_1}{2}\right) \Big/ \sum_{(I,j)\in N} y_j^2 \end{bmatrix}, \tag{4.18}$$

$$\mathbf{y} = diag\left(1 \Big/ \sum_{(i,1)\in N} x_i^2, \cdots, 1 \Big/ \sum_{(i,J)\in N} x_i^2\right)\left((C^*)'\mathbf{x} - \frac{1}{2}diag\left(\rho_1, \cdots, \rho_J\right)\mathbf{z}\right)$$

$$= \begin{bmatrix} \left(\sum_{(i,1)\in N} c_{i1}x_i - \frac{\rho_1 z_1}{2}\right) \Big/ \sum_{(i,1)\in N} x_i^2 \\ \left(\sum_{(i,2)\in N} c_{i2}x_i - \frac{\rho_2 z_2}{2}\right) \Big/ \sum_{(i,2)\in N} x_i^2 \\ \vdots \\ \left(\sum_{(i,J)\in N} c_{iJ}x_i - \frac{\rho_J z_J}{2}\right) \Big/ \sum_{(i,J)\in N} x_i^2 \end{bmatrix}. \tag{4.19}$$

將式子 (4.12) 左乘 $\mathbf{x}'$、式子 (4.13) 左乘 $\mathbf{y}'$ 和式子 (4.14) 左乘 $\mathbf{z}'$，可得

$$-2\mathbf{x}'C^*\mathbf{y} + 2\mathbf{x}'diag\left(\sum_{(1,j)\in N} y_j^2, \cdots, \sum_{(I,j)\in N} y_j^2\right)\mathbf{x} + \lambda_1\mathbf{x}'\underset{\sim}{1} = 0, \tag{4.20}$$

$$-2\mathbf{y}'(C^*)'\mathbf{x} + 2\mathbf{y}'diag\left(\sum_{(i,1)\in N} x_i^2, \cdots, \sum_{(i,J)\in N} x_i^2\right)\mathbf{y} + \mathbf{y}'diag\left(\rho_1, \cdots, \rho_J\right)\mathbf{z} = 0, \tag{4.21}$$

$$\lambda_2\mathbf{z}'\underset{\sim}{1} + \mathbf{z}'diag\left(\rho_1, \cdots, \rho_J\right)\mathbf{y} = 0. \tag{4.22}$$

由式子 (4.20) 和 (4.21) 得到

$$\sum_{j=1}^{J}\rho_j = \lambda_1, \tag{4.23}$$

由式子 (4.22) 和 (4.23) 可得

$$\sum_{j=1}^{J}\rho_j = \lambda_1 = -\lambda_2. \tag{4.24}$$

將式子 (4.12) 左乘 $\underset{\sim}{1}'$，得到

$$-2\underset{\sim}{1}'C^*\mathbf{y} + 2\underset{\sim}{1}'diag\left(\sum_{(1,j)\in N} y_j^2, \cdots, \sum_{(I,j)\in N} y_j^2\right)\mathbf{x} + \lambda_1\underset{\sim}{1}'\underset{\sim}{1} = 0, \tag{4.25}$$

由式子 (4.25) 可得

$$\begin{aligned}\lambda_1 &= \frac{2\left(\underset{\sim}{1}'C^*\mathbf{y} - \underset{\sim}{1}'diag\left(\sum_{(1,j)\in N} y_j^2, \cdots, \sum_{(I,j)\in N} y_j^2\right)\mathbf{x}\right)}{\underset{\sim}{1}'\underset{\sim}{1}} \\ &= \frac{2\left[\sum\sum_{(i,j)\in N} c_{ij}y_j - \sum\sum_{(i,j)\in N} x_i y_j^2\right]}{I}.\end{aligned} \tag{4.26}$$

根據上述過程，Lagrange 乘數法求 $\mathbf{x}$、$\mathbf{y}$ 解的演算法如下：

求最近似秩 1 矩陣之演算法 (二)

1. 選定初值 $(\mathbf{x}^{(0)}, \mathbf{y}^{(0)})$。
2. 將 $\mathbf{x}^{(0)}$ 和 $\mathbf{y}^{(0)}$ 代入式子 (4.26)，再把計算結果 $\lambda_1^{(0)}$ 代入式子 (4.24)，可得 $\lambda_2^{(0)}$。
3. 將 $\mathbf{y}^{(0)}$ 和 $\lambda_2^{(0)}$ 代入式子 (4.14)，再把 $\mathbf{y}^{(0)}$ 代入式子 (4.17)，可分別得到 $\boldsymbol{\rho}^{(0)}$ 和 $\mathbf{z}^{(0)}$。
4. 將 $\mathbf{y}^{(0)}$ 和 $\lambda_1^{(0)}$ 代入式子 (4.18)，可得 $\mathbf{x}$。
5. 將 $\mathbf{x}^{(0)}$、$\mathbf{z}^{(0)}$ 和 $\boldsymbol{\rho}^{(0)}$ 代入式子 (4.19)，可得 $\mathbf{y}$。
6. 將計算結果 $\mathbf{x}$ 和 $\mathbf{y}$ 代入式子 (4.5) 做修正，得 $\mathbf{x}^{(1)}$ 和 $\mathbf{y}^{(1)}$，並計算 $L_1 = \|C - \mathbf{x}^{(1)}\mathbf{y}^{(1)'}\|_F^2$。
7. 從 $(\mathbf{x}^{(1)}, \mathbf{y}^{(1)})$ 開始，重複步驟 $2 \sim 6$，得 $(\mathbf{x}^{(2)}, \mathbf{y}^{(2)})$ 以及 $L_2 = \|C - \mathbf{x}^{(2)}\mathbf{y}^{(2)'}\|_F^2$。依此類推，可得 $(\mathbf{x}^{(3)}, \mathbf{y}^{(3)})$ 以及 $L_3 = \|C - \mathbf{x}^{(3)}\mathbf{y}^{(3)'}\|_F^2 \cdots$。
8. 在第 $n+1$ 步驟時，若 $|L_{n+1} - L_n| \leq 10^{-5}$，則停止計算，然後利用 $\mathbf{x}^{(n)}$ 和 $\mathbf{y}^{(n)}$ 來決定最近似秩 1 矩陣。

第四節　高維度牛頓法

除了上述兩種演算法，本節也提供非線性求解常使用的方法，即牛頓法，主要內容參考周志成 (2016)。牛頓法的優點為：若初值的選擇不錯，且 Jocobian 矩陣為非奇異，則牛頓法所產生的數列會收斂，即可找到不錯的近似解；牛頓法的缺點為：收斂結果與否會受到初值選擇的影響；當 Jocobian 矩陣為奇異，則迭代過程中斷。

在第二節中，利用 Lagrange 乘數法時，$\mathbf{x}$，$\mathbf{y}$ 需滿足式子 (4.2)。在此我們將其改寫如下：

$$\begin{cases} \sum_{i=1}^{I} x_i = 1 \\ \sum_{j=1}^{J} \frac{1}{y_j} = 1. \end{cases} \tag{4.27}$$

並將 L^* 改寫為

$$L^{**} = \sum\sum_{(i,j)\in N}(c_{ij} - x_i y_j)^2 + \lambda_1(\sum_{i=1}^{I} x_i - 1) + \lambda_2(\sum_{j=1}^{J} \frac{1}{y_j} - 1),$$

以減少未知數個數。

L^{**} 對 $\mathbf{x}$、$\mathbf{y}$、λ_1、λ_2 做偏微分，可得

$$\frac{\partial L^{**}}{\partial \mathbf{x}} = 2\left[diag\left(\sum_{(1,j)\in N} y_j^2, \cdots, \sum_{(I,j)\in N} y_j^2\right)\mathbf{x} - C^*\mathbf{y}\right] + \lambda_1 \underset{\sim}{1}. \tag{4.28}$$

$$\frac{\partial L^{**}}{\partial \mathbf{y}} = 2\left[diag\left(\sum_{(i,1)\in N} x_i^2, \cdots, \sum_{(i,J)\in N} x_i^2\right)\mathbf{y} - (C^*)'\mathbf{x}\right] - \lambda_2 \begin{pmatrix} 1/y_1^2 \\ \vdots \\ 1/y_J^2 \end{pmatrix}. \tag{4.29}$$

$$\frac{\partial L^{**}}{\partial \lambda_1} = \sum_{i=1}^{I} x_i - 1. \tag{4.30}$$

$$\frac{\partial L^{**}}{\partial \lambda_2} = \sum_{j=1}^{J} \frac{1}{y_j} - 1. \tag{4.31}$$

其中 C^* 是 $I * J$ 矩陣，將 C 中的元素 $*$ 以 0 來取代，而其他正值元素保持不變。

令 $\frac{\partial L^{**}}{\partial \mathbf{x}} = 0$、$\frac{\partial L^{**}}{\partial \mathbf{y}} = 0$、$\frac{\partial L^{**}}{\partial \lambda_1} = 0$ 和 $\frac{\partial L^{**}}{\partial \lambda_2} = 0$，得下列關係式：

$$2\left[diag\left(\sum_{(1,j)\in N} y_j^2, \cdots, \sum_{(I,j)\in N} y_j^2\right)\mathbf{x} - C^*\mathbf{y}\right] + \lambda_1 \underset{\sim}{1} = 0. \tag{4.32}$$

$$2\left[diag\left(\sum_{(i,1)\in N} x_i^2, \cdots, \sum_{(i,J)\in N} x_i^2\right)\mathbf{y} - (C^*)'\mathbf{x}\right] - \lambda_2 \begin{pmatrix} 1/y_1^2 \\ \vdots \\ 1/y_J^2 \end{pmatrix} = 0. \tag{4.33}$$

$$\sum_{i=1}^{I} x_i - 1 = 0. \tag{4.34}$$

$$\sum_{j=1}^{J} \frac{1}{y_j} - 1 = 0. \tag{4.35}$$

將式子 (4.32) 左乘 $\mathbf{x}'$ 與式子 (4.33) 左乘 $\mathbf{y}'$，並利用式子 (4.34) 和 (4.35)，得到

$$2\mathbf{x}' diag\left(\sum_{(1,j)\in N} y_j^2, \cdots, \sum_{(I,j)\in N} y_j^2\right)\mathbf{x} - 2\mathbf{x}'C^*\mathbf{y} + \lambda_1 = 0, \tag{4.36}$$

$$2\mathbf{y}' diag\left(\sum_{(i,1)\in N} x_i^2, \cdots, \sum_{(i,J)\in N} x_i^2\right)\mathbf{y} - 2\mathbf{y}'(C^*)'\mathbf{x} - \lambda_2 = 0. \tag{4.37}$$

由式子 (4.36) 和 (4.37)，可得

$$\lambda_1 = -\lambda_2 = 2\left[\mathbf{x}'C^*\mathbf{y} - \mathbf{x}'diag\left(\sum_{(1,j)\in N} y_j^2, \cdots, \sum_{(I,j)\in N} y_j^2\right)\mathbf{x}\right]. \tag{4.38}$$

將式子 (4.38) 代回式子 (4.32) 和 (4.33) ，得到

$$\mathbf{F} = \begin{pmatrix} f_1 \\ f_2 \end{pmatrix} = \mathbf{0}, \tag{4.39}$$

其中

$$\begin{aligned} f_1 = {} & 2\left[diag\left(\sum_{(1,j)\in N} y_j^2, \cdots, \sum_{(I,j)\in N} y_j^2\right)\mathbf{x} - C^*\mathbf{y}\right] \\ & + 2\left[\mathbf{x}'C^*\mathbf{y} - \mathbf{x}'diag\left(\sum_{(1,j)\in N} y_j^2, \cdots, \sum_{(I,j)\in N} y_j^2\right)\mathbf{x}\right]\underset{\sim}{1}, \end{aligned} \tag{4.40}$$

$$\begin{aligned} f_2 = {} & 2\left[diag\left(\sum_{(i,1)\in N} x_i^2, \cdots, \sum_{(i,J)\in N} x_i^2\right)\mathbf{y} - (C^*)'\mathbf{x}\right] \\ & + 2\left[\mathbf{x}'C^*\mathbf{y} - \mathbf{x}'diag\left(\sum_{(1,j)\in N} y_j^2, \cdots, \sum_{(I,j)\in N} y_j^2\right)\mathbf{x}\right]\begin{pmatrix} 1/y_1^2 \\ \vdots \\ 1/y_J^2 \end{pmatrix}. \end{aligned} \tag{4.41}$$

將 $f_1 = 0$ 左乘 $\mathbf{x}'$ 與 $f_2 = 0$ 左乘 $\mathbf{y}'$，可分別得到

$$\sum_{i=1}^{I} x_i = 1, \quad \sum_{j=1}^{J} \frac{1}{y_j} = 1.$$

即為式子 (4.34) 和 (4.35)。因此可把求解式子 (4.32)∼ (4.35) 簡化成求解式子 (4.32) 和 (4.33) 即可。

接著欲解方程式 $\mathbf{F}=0$。我們先求 Jocobian 矩陣，即

$$\mathbf{J}=\begin{pmatrix} \frac{\partial f_1}{\partial \mathbf{x}} & \frac{\partial f_1}{\partial \mathbf{y}} \\ \frac{\partial f_2}{\partial \mathbf{x}} & \frac{\partial f_2}{\partial \mathbf{y}} \end{pmatrix}, \tag{4.42}$$

其中

$$\frac{\partial f_1}{\partial \mathbf{x}} = 2\,diag\left(\sum_{(1,j)\in N} y_j^2,\cdots,\sum_{(I,j)\in N} y_j^2\right) + 2\,\underset{\sim}{1}\left[(C^*\mathbf{y})' - 2(diag\left(\sum_{(1,j)\in N} y_j^2,\cdots,\sum_{(I,j)\in N} y_j^2\right)\mathbf{x})'\right], \tag{4.43}$$

$$\frac{\partial f_1}{\partial \mathbf{y}} = 2\left(2(\mathbf{x}\mathbf{y}')_{(i,j)\in N} - C^*\right) + 2\,\underset{\sim}{1}\left[\mathbf{x}'C^* - 2(diag\left(\sum_{(i,1)\in N} x_i^2,\cdots,\sum_{(i,J)\in N} x_i^2\right)\mathbf{y})'\right], \tag{4.44}$$

$$\frac{\partial f_2}{\partial \mathbf{x}} = 2\left(2(\mathbf{y}\mathbf{x}')_{(i,j)\in N} - (C^*)'\right) + 2\begin{pmatrix} 1/y_1^2 \\ \vdots \\ 1/y_J^2 \end{pmatrix}\left[(C^*\mathbf{y})' - 2(diag\left(\sum_{(1,j)\in N} y_j^2,\cdots,\sum_{(I,j)\in N} y_j^2\right)\mathbf{x})'\right], \tag{4.45}$$

$$\begin{aligned}\frac{\partial f_2}{\partial \mathbf{y}} = {} & 2\,diag\left(\sum_{(i,1)\in N} x_i^2,\cdots,\sum_{(i,J)\in N} x_i^2\right) + 2\left[\begin{pmatrix} 1/y_1^2 \\ \vdots \\ 1/y_J^2 \end{pmatrix}\mathbf{x}'C^* - 2diag\left(\frac{\mathbf{x}'C^*\mathbf{y}}{y_1^3},\cdots,\frac{\mathbf{x}'C^*\mathbf{y}}{y_J^3}\right)\right] \\ & - 4\left[\begin{pmatrix} 1/y_1^2 \\ \vdots \\ 1/y_J^2 \end{pmatrix}\mathbf{y}'diag\left(\sum_{(i,1)\in N} x_i^2,\cdots,\sum_{(i,J)\in N} x_i^2\right) - diag\left(\frac{\sum\sum_{(i,j)\in N} x_i^2 y_j^2}{y_1^3},\cdots,\frac{\sum\sum_{(i,j)\in N} x_i^2 y_j^2}{y_J^3}\right)\right]\end{aligned} \tag{4.46}$$

若第 n 次迭代得 $(\mathbf{x}^{(n)}, \mathbf{y}^{(n)})$，欲求 $(\mathbf{x}^{(n+1)}, \mathbf{y}^{(n+1)})$。令誤差向量 $\delta^{(n)} = -(\mathbf{J}^{(n)})^{-1}\mathbf{F}^{(n)}$，則由解方程式

$$\mathbf{J}^{(n)}\delta^{(n)} = -\mathbf{F}^{(n)}, \tag{4.47}$$

可得 $\delta^{(n)}$。最終，透過下列式子 (4.48) 得到 $\mathbf{x}^{(n+1)}$ 和 $\mathbf{y}^{(n+1)}$，即

$$\begin{pmatrix} \mathbf{x}^{(n+1)} \\ \mathbf{y}^{(n+1)} \end{pmatrix} = \begin{pmatrix} \mathbf{x}^{(n)} \\ \mathbf{y}^{(n)} \end{pmatrix} + \delta^{(n)}. \tag{4.48}$$

根據上述過程，高維度牛頓法求 $\mathbf{x}$、$\mathbf{y}$ 解的演算法如下：

求最近似秩 1 矩陣之演算法 (三)

1. 選定初值 $(\mathbf{x}^{(0)}, \mathbf{y}^{(0)})$。
2. 將 $(\mathbf{x}^{(0)}, \mathbf{y}^{(0)})$ 代入式子 (4.39) 和 (4.42)，再把計算結果 $\mathbf{F}^{(0)}$ 和 $\mathbf{J}^{(0)}$ 代入式子 (4.47) 可得 $\delta^{(0)}$。
3. 將 $\delta^{(0)}$ 代入式子 (4.48)，再把計算結果 $\mathbf{x}$ 和 $\mathbf{y}$ 代入式子 (4.5) 做修正，得 $\mathbf{x}^{(1)}$ 和 $\mathbf{y}^{(1)}$，並計算 $L_1 = \|C - \mathbf{x}^{(1)}\mathbf{y}^{(1)\prime}\|_F^2$。
4. 從 $(\mathbf{x}^{(1)}, \mathbf{y}^{(1)})$ 開始，重複步驟 $2 \sim 3$，得 $(\mathbf{x}^{(2)}, \mathbf{y}^{(2)})$ 以及 $L_2 = \|C - \mathbf{x}^{(2)}\mathbf{y}^{(2)\prime}\|_F^2$。依此類推，可得 $(\mathbf{x}^{(3)}, \mathbf{y}^{(3)})$ 以及 $L_3 = \|C - \mathbf{x}^{(3)}\mathbf{y}^{(3)\prime}\|_F^2 \cdots$。
5. 在第 $n+1$ 步驟時，若 $|L_{n+1} - L_n| \leq 10^{-5}$，則停止計算，然後利用 $\mathbf{x}^{(n)}$ 和 $\mathbf{y}^{(n)}$ 來決定最近似秩 1 矩陣。

第五節　實例探討

本節主要以實例說明，如何以秩 1 矩陣法解相容性問題。

例題 4. (例題 1續) A 對 B 的比值矩陣 C 為

$$C=\begin{pmatrix}\frac{21}{16}&\frac{21}{16}&\frac{7}{8}\\\frac{21}{16}&\frac{21}{16}&\frac{7}{8}\\\frac{7}{8}&\frac{7}{8}&\frac{7}{12}\end{pmatrix}.$$

以下三個演算法的初值 $\mathbf{x}^{(0)}$，$\mathbf{y}^{(0)}$ 皆是以均勻分配來決定，即

$$\mathbf{x}^{(0)}=\begin{pmatrix}\frac{1}{3}\\\frac{1}{3}\\\frac{1}{3}\end{pmatrix},\quad \mathbf{y}^{(0)}=\begin{pmatrix}3\\3\\3\end{pmatrix}.$$

透過**演算法 (一)**，在 $n=2$ 時，可得 Frobenius 範數的值為 0；由**演算法 (二)**，在 $n=6$ 時，得到 Frobenius 範數的值為 0；最後，經由**演算法 (三)**，在 $n=11$ 時，求得 Frobenius 範數的值為 0。

各種演算法停止計算時的 $\mathbf{x}^{(n)}$ 和 $\mathbf{y}^{(n)}$ 皆為

$$\mathbf{x}^{(n)}=\begin{pmatrix}0.3750\\0.3750\\0.2500\end{pmatrix},\quad \mathbf{y}^{(n)}=\begin{pmatrix}3.5000\\3.5000\\2.3333\end{pmatrix}.$$

由於 Frobenius 範數的值為 0，故 A 和 B 相容。若利用 Matlab 在限制條件下求函數最小值的指令，所得 $\mathbf{x}$ 和 $\mathbf{y}$ 的解和上面一樣，且 Frobenius 範數的值也為 0。將 $\mathbf{x}$ 和 $\mathbf{y}$ 代入式子 (4.5)，可得 X 和 Y 的邊際分配；再加上給定的條件機率分配，得到 (X,Y) 的聯合分配，如例題 1 所示。

例題 5. (例題 2續) A 對 B 的比值矩陣 C 為

$$C = \begin{pmatrix} \frac{6}{7} & 3 & \frac{6}{7} & \frac{6}{7} \\ 1 & \frac{7}{4} & \frac{1}{2} & 1 \\ \frac{12}{7} & 3 & \frac{12}{7} & \frac{12}{7} \end{pmatrix}.$$

經由 Matlab 之最佳化計算，得 Frobenius 範數的最小值為 0.8459，此時的 $\mathbf{x}$ 和 $\mathbf{y}$ 滿足式子 (4.5)，即

$$\mathbf{x} = \begin{pmatrix} 0.3363 \\ 0.2337 \\ 0.4300 \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} 3.5672 \\ 7.6787 \\ 3.2351 \\ 3.5672 \end{pmatrix},$$

且最近似秩 1 矩陣為

$$\begin{pmatrix} 1.1996 & 2.5823 & 1.0880 & 1.1996 \\ 0.8337 & 1.7945 & 0.7560 & 0.8337 \\ 1.5339 & 3.3018 & 1.3911 & 1.5339 \end{pmatrix}.$$

以下三個演算法的初值 $\mathbf{x}^{(0)}$，$\mathbf{y}^{(0)}$ 皆是以均勻分配來決定，即

$$\mathbf{x}^{(0)} = \begin{pmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{pmatrix}, \quad \mathbf{y}^{(0)} = \begin{pmatrix} 4 \\ 4 \\ 4 \\ 4 \end{pmatrix}.$$

透過**演算法 (一)**，在 $n = 6$ 時，可得 Frobenius 範數的值為 0.8459，此時的 $\mathbf{x}^{(n)}$ 和 $\mathbf{y}^{(n)}$ 為

$$\mathbf{x}^{(n)} = \begin{pmatrix} 0.3363 \\ 0.2338 \\ 0.4300 \end{pmatrix}, \quad \mathbf{y}^{(n)} = \begin{pmatrix} 3.5673 \\ 7.6720 \\ 3.2362 \\ 3.5673 \end{pmatrix}.$$

由**演算法 (二)**，在 $n = 590$ 時，得到 Frobenius 範數的值為 1.0300，此時的 $\mathbf{x}^{(n)}$ 和 $\mathbf{y}^{(n)}$ 為

$$\mathbf{x}^{(n)} = \begin{pmatrix} 0.3330 \\ 0.2289 \\ 0.4381 \end{pmatrix}, \quad \mathbf{y}^{(n)} = \begin{pmatrix} 3.5585 \\ 8.3361 \\ 3.1445 \\ 3.5585 \end{pmatrix}.$$

最後，經由**演算法 (三)**，在 $n = 20$ 時，求得 Frobenius 範數的值為 0.8459，此時的 $\mathbf{x}^{(n)}$ 和 $\mathbf{y}^{(n)}$ 為

$$\mathbf{x}^{(n)} = \begin{pmatrix} 0.3363 \\ 0.2336 \\ 0.4300 \end{pmatrix}, \quad \mathbf{y}^{(n)} = \begin{pmatrix} 3.5672 \\ 7.6863 \\ 3.2339 \\ 3.5672 \end{pmatrix}.$$

因 Frobenius 範數的最小值不為 0，故 A 和 B 不相容。由**演算法 (一)** 計算出的 $\mathbf{x}^{(n)}$ 和 $\mathbf{y}^{(n)}$，再加上給定的條件機率分配，根據式子 (3.12) 可得兩組 (X, Y) 的近似聯合機率分配 $P^{(1)}$ 和 $P^{(2)}$，

$$P^{(1)} = \begin{pmatrix} 0.0560 & 0.0560 & 0.1681 & 0.0560 \\ 0.0668 & 0.0334 & 0.0668 & 0.0668 \\ 0.1433 & 0.0358 & 0.1075 & 0.1433 \end{pmatrix}, \quad P^{(2)} = \begin{pmatrix} 0.0400 & 0.0652 & 0.1324 & 0.0400 \\ 0.0801 & 0.0326 & 0.0441 & 0.0801 \\ 0.1602 & 0.0326 & 0.1324 & 0.1602 \end{pmatrix},$$

按式子 (3.13) 將 $P^{(1)}$ 和 $P^{(2)}$ 組合起來得下列近似聯合分配：

$$P = \begin{pmatrix} 0.0480 & 0.0606 & 0.1503 & 0.0480 \\ 0.0734 & 0.0330 & 0.0555 & 0.0734 \\ 0.1518 & 0.0342 & 0.1200 & 0.1518 \end{pmatrix}.$$

重複上述過程，**演算法 (三)** 和 Matlab 最佳化法也都能得到近似的聯合分配，因**演算法 (一)**、**演算法 (三)** 和 Matlab 最佳化法所求得的 Frobenius 範數最小值都為 0.8459，故它們可獲得的近似聯合分配幾乎相同。

此外，我們也藉此例題說明推論 3.7。經計算後可得，

$$\mathbf{x} = -0.5938\mathbf{u}_1 + 0.0001\mathbf{u}_3, \quad \mathbf{y} = -9.7406\mathbf{v}_1 - 0.0047\mathbf{v}_2 - 0.0009\mathbf{v}_3,$$

其中 $\mathbf{u}_1$、$\mathbf{u}_3$、$\mathbf{v}_1$、$\mathbf{v}_2$ 和 $\mathbf{v}_3$ 如例題 2 中正交矩陣 U 和 V 的行向量所示。觀察發現 d_3、e_2 和 e_3 幾乎很接近 0，根據推論 3.7，

$$\begin{aligned}
\|C - \mathbf{x}\mathbf{y}'\|_F^2 &\approx (\sigma_2^2 + \sigma_3^2) + (\sigma_1 - d_1 e_1)^2 \\
&= (0.840238^2 + 0.373955^2) + (5.786801 - 0.593809 * 9.740591)^2 \\
&= 0.845841 + 0.002750^2 \\
&= 0.845848.
\end{aligned}$$

若利用 $\rho\mathbf{u}_1\mathbf{v}_1'$ 作為逼近 C 的秩 1 矩陣，即僅用 $\mathbf{u}_1$ 與 $\mathbf{v}_1$ 來求近似聯合分配，則 $\|C - \rho\mathbf{u}_1\mathbf{v}_1'\|_F^2 = (\sigma_1 - \rho)^2 + \sigma_2^2 + \sigma_3^2 = 0.845874$，其中 $\rho = \sum_{j=1}^{J} \frac{1}{v_{j1}} \Big/ \sum_{i=1}^{I} u_{i1} = 5.780992$。然而，$\|C - \mathbf{x}\mathbf{y}'\|_F^2$ 真正的最小值為 0.845857，故 $\rho\mathbf{u}_1\mathbf{v}_1'$ 不是最近似的秩 1 矩陣。同時，上述的秩 1 矩陣 $\mathbf{x}\mathbf{y}'$ 比 $\rho\mathbf{u}_1\mathbf{v}_1'$ 還好。

例題 6. (不相容) 給定條件機率矩陣 A 和 B，

$$A=\begin{pmatrix}\frac{3}{22}&0&\frac{6}{31}&\frac{2}{11}&0\\ \frac{7}{22}&0&0&\frac{4}{11}&\frac{1}{4}\\ \frac{6}{11}&\frac{9}{22}&\frac{11}{31}&0&\frac{3}{4}\\ 0&\frac{13}{22}&\frac{14}{31}&\frac{5}{11}&0\end{pmatrix},\quad B=\begin{pmatrix}\frac{1}{4}&0&\frac{5}{12}&\frac{1}{3}&0\\ \frac{1}{3}&0&0&\frac{8}{21}&\frac{2}{7}\\ \frac{2}{7}&\frac{3}{14}&\frac{11}{42}&0&\frac{5}{21}\\ 0&\frac{13}{42}&\frac{1}{3}&\frac{5}{14}&0\end{pmatrix},$$

A 對 B 的比值矩陣 C 為

$$C=\begin{pmatrix}\frac{6}{11}&*&\frac{72}{155}&\frac{6}{11}&*\\ \frac{21}{22}&*&*&\frac{21}{22}&\frac{7}{8}\\ \frac{21}{11}&\frac{21}{11}&\frac{42}{31}&*&\frac{63}{20}\\ *&\frac{21}{11}&\frac{42}{31}&\frac{14}{11}&*\end{pmatrix}.$$

經由 Matlab 最佳化計算，得類 Frobenius 範數的最小值為 0.3319，此時的 $\mathbf{x}$ 和 $\mathbf{y}$ 滿足式子 (4.5)，即

$$\mathbf{x}=\begin{pmatrix}0.1227\\0.1506\\0.3798\\0.3470\end{pmatrix},\quad \mathbf{y}=\begin{pmatrix}5.2294\\5.3010\\3.8295\\4.2741\\7.9977\end{pmatrix},$$

且最近似秩 1 矩陣為

$$\begin{pmatrix}0.6414&0.6502&0.4697&0.5242&0.9810\\0.7875&0.7982&0.5767&0.6436&1.2043\\1.9859&2.0131&1.4543&1.6231&3.0372\\1.8146&1.8395&1.3289&1.4832&2.7753\end{pmatrix}.$$

和例題 5相同，以下三個演算法的初值 $\mathbf{x}^{(0)}$，$\mathbf{y}^{(0)}$ 皆是以均勻分配來決定，即

$$\mathbf{x}^{(0)} = \begin{pmatrix} \frac{1}{4} \\ \frac{1}{4} \\ \frac{1}{4} \\ \frac{1}{4} \end{pmatrix}, \quad \mathbf{y}^{(0)} = \begin{pmatrix} 5 \\ 5 \\ 5 \\ 5 \\ 5 \end{pmatrix}.$$

透過**演算法 (一)**，在 $n = 7$ 時，可得類 Frobenius 範數的值為 0.3380，此時的 $\mathbf{x}^{(n)}$ 和 $\mathbf{y}^{(n)}$ 為

$$\mathbf{x}^{(n)} = \begin{pmatrix} 0.1185 \\ 0.1503 \\ 0.3856 \\ 0.3456 \end{pmatrix}, \quad \mathbf{y}^{(n)} = \begin{pmatrix} 5.1859 \\ 5.3448 \\ 3.8052 \\ 4.3084 \\ 7.9890 \end{pmatrix}.$$

由**演算法 (二)**，在 $n = 8$ 時，得到類 Frobenius 範數的值為 1.0013，此時的 $\mathbf{x}^{(n)}$ 和 $\mathbf{y}^{(n)}$ 為

$$\mathbf{x}^{(n)} = \begin{pmatrix} 0.1037 \\ 0.1576 \\ 0.4206 \\ 0.3181 \end{pmatrix}, \quad \mathbf{y}^{(n)} = \begin{pmatrix} 5.3354 \\ 5.5064 \\ 3.7605 \\ 4.0066 \\ 8.6617 \end{pmatrix}.$$

最後，經由**演算法 (三)**，在 $n = 7$ 時，求得類 Frobenius 範數的值為 1.5334，此時的 $\mathbf{x}^{(n)}$ 和 $\mathbf{y}^{(n)}$ 為

$$\mathbf{x}^{(n)} = \begin{pmatrix} 0.1564 \\ 0.2081 \\ 0.3791 \\ 0.2565 \end{pmatrix}, \quad \mathbf{y}^{(n)} = \begin{pmatrix} 4.7959 \\ 5.2213 \\ 4.5204 \\ 4.6903 \\ 6.0406 \end{pmatrix}.$$

因**演算法 (一)**、**演算法 (二)** 和**演算法 (三)** 的計算結果受到初值很大的影響，和初值的選擇有很大的關係。我們透過電腦軟體以均勻分配的亂數來產生 100 次不同的初值 $\mathbf{x}^{(0)}$ 和 $\mathbf{y}^{(0)}$，且將 $\mathbf{x}^{(0)}$，$\mathbf{y}^{(0)}$ 代入式子 (4.5) 做修正。計算結果顯示**演算法 (一)** 在 100 次中有 1 次獲得類 Frobenius 範數的最小值 0.3335；**演算法 (二)** 在 100 次中也有 1 次獲得類 Frobenius 範數的最小值 0.3319；**演算法 (三)** 在 100 次中大約有 30 次獲得類 Frobenius 範數的最小值 0.3319。從上述結果我們可以知道，在三個演算法中，**演算法 (三)** 有較多的次數找到類 Frobenius 範數的最小值，也就是其計算結果比較不會受到初值選擇的影響。

我們取**演算法 (三)** 中的一組初值為例，即

$$\mathbf{x}^{(0)} = \begin{pmatrix} 0.2859 \\ 0.1609 \\ 0.4517 \\ 0.1016 \end{pmatrix}, \quad \mathbf{y}^{(0)} = \begin{pmatrix} 7.2827 \\ 9.7381 \\ 6.3210 \\ 2.4015 \\ 5.3941 \end{pmatrix}.$$

經由**演算法 (三)**，在 $n = 12$ 時，求得類 Frobenius 範數的值為 0.3319，此時的 $\mathbf{x}^{(n)}$ 和 $\mathbf{y}^{(n)}$ 為

$$\mathbf{x}^{(n)} = \begin{pmatrix} 0.1227 \\ 0.1506 \\ 0.3798 \\ 0.3470 \end{pmatrix}, \quad \mathbf{y}^{(n)} = \begin{pmatrix} 5.2294 \\ 5.3010 \\ 3.8295 \\ 4.2741 \\ 7.9977 \end{pmatrix}.$$

所獲得的計算結果和 Matlab 最佳化計算結果幾乎一致。

因類 Frobenius 範數的最小值不為 0，故 A 和 B 不相容。利用計算出的 $\mathbf{x}^{(n)}$ 和 $\mathbf{y}^{(n)}$，再加上給定的條件機率分配，根據式子 (3.12) 可得兩組 (X,Y) 的近似聯合機率分配 $P^{(1)}$ 和 $P^{(2)}$，

$$P^{(1)} = \begin{pmatrix} 0.0307 & 0 & 0.0511 & 0.0409 & 0 \\ 0.0502 & 0 & 0 & 0.0574 & 0.0430 \\ 0.1085 & 0.0814 & 0.0995 & 0 & 0.0904 \\ 0 & 0.1074 & 0.1157 & 0.1239 & 0 \end{pmatrix},$$

$$P^{(2)} = \begin{pmatrix} 0.0261 & 0 & 0.0505 & 0.0425 & 0 \\ 0.0608 & 0 & 0 & 0.0851 & 0.0313 \\ 0.1043 & 0.0772 & 0.0927 & 0 & 0.0938 \\ 0 & 0.1115 & 0.1179 & 0.1063 & 0 \end{pmatrix},$$

按式子 (3.13) 將 $P^{(1)}$ 和 $P^{(2)}$ 組合起來得下列近似聯合分配：

$$P = \begin{pmatrix} 0.0284 & 0 & 0.0508 & 0.0417 & 0 \\ 0.0555 & 0 & 0 & 0.0713 & 0.0372 \\ 0.1064 & 0.0793 & 0.0961 & 0 & 0.0921 \\ 0 & 0.1095 & 0.1168 & 0.1151 & 0 \end{pmatrix}.$$

重複上述過程，Matlab 最佳化法也能得到和上述結果幾乎一致的近似聯合分配。

第五章　最近似聯合分配

當條件機率矩陣 A 和 B 不相容時，我們可獲得兩組近似聯合分配。如何將兩組近似聯合分配組合起來而得到更好的近似聯合分配，是本章要探討的內容。

第一節　最近似聯合分配

在第四章中，我們已經討論過在式子 (4.2) 的限制條件下，如何求 $\mathbf{x}$，$\mathbf{y}$ 使得 $\|C - \mathbf{x}\mathbf{y}'\|_F^2$ 最小。利用向量 $\mathbf{x}$ 和 $\mathbf{y}$，透過式子 (1.1) 可得 X，Y 的邊際分配配 (τ_i) 和 (η_j)；再加上已知的條件機率矩陣 A 和 B，可分別得到近似聯合分配矩陣 $P^{(1)} = (P_{ij}^{(1)}) = (\tau_i b_{ij})$ 和 $P^{(2)} = (P_{ij}^{(2)}) = (\eta_j a_{ij})$。令 P 為 $P^{(1)}$ 和 $P^{(2)}$ 的凸組合，即

$$P = \lambda P^{(1)} + (1-\lambda)P^{(2)}, \quad 0 \le \lambda \le 1.$$

使總誤差為最小的 P，即為最近似聯合分配。總誤差的定義如下：

定義 5.1. 給定一個聯合機率分配 $P = (p_{ij})$，則 P 的總誤差 E 定義如下：

$$E = \|P_{X|Y} - A\|_F^2 + \|P_{Y|X} - B\|_F^2,$$

其中 $P_{X|Y}$ 和 $P_{Y|X}$ 分別是 P 的條件機率矩陣。

在總誤差的標準下，欲獲得最近似聯合分配，牽涉到解一個很複雜的方程式，在實際應用上並不可行。我們可以用下列方式將 $P^{(1)}$ 和 $P^{(2)}$ 組合起來，即

$$\begin{aligned} P &= \frac{\frac{1}{E_1}}{\frac{1}{E_1}+\frac{1}{E_2}}P^{(1)} + \frac{\frac{1}{E_2}}{\frac{1}{E_1}+\frac{1}{E_2}}P^{(2)} \\ &= \frac{E_2}{E_1+E_2}P^{(1)} + \frac{E_1}{E_1+E_2}P^{(2)}, \end{aligned} \tag{5.1}$$

其中

$$\begin{aligned} E_1 &= \sum_{i=1}^{I}\sum_{j=1}^{J}\left(\frac{p_{ij}^{(1)}}{p_{.j}^{(1)}} - a_{ij}\right)^2 + \sum_{i=1}^{I}\sum_{j=1}^{J}\left(\frac{p_{ij}^{(1)}}{p_{i.}^{(1)}} - b_{ij}\right)^2, \\ E_2 &= \sum_{i=1}^{I}\sum_{j=1}^{J}\left(\frac{p_{ij}^{(2)}}{p_{.j}^{(2)}} - a_{ij}\right)^2 + \sum_{i=1}^{I}\sum_{j=1}^{J}\left(\frac{p_{ij}^{(2)}}{p_{i.}^{(2)}} - b_{ij}\right)^2. \end{aligned}$$

上述結果得知，$P^{(1)}$ 和 $P^{(2)}$ 所佔的權重會和其條件機率矩陣的誤差成反比，即誤差大的權重小，誤差小的權重大，期望以凸組合 (5.1) 的組合方式找到更好的近似聯合分配。

定理 5.2. $E_1 = \sum_{i=1}^{I}\sum_{j=1}^{J}\left(\frac{p_{ij}^{(1)} - p_{ij}^{(2)}}{\eta_j}\right)^2$，$E_2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\left(\frac{p_{ij}^{(1)} - p_{ij}^{(2)}}{\tau_i}\right)^2$，*即凸組合 (5.1) 的組合結果與顧仲航推導出的最近似聯合分配完全相同。*

Proof.

由 $P^{(1)}$ 和 $P^{(2)}$ 的定義可得

$$P^{(1)} = (p_{ij}^{(1)}) = (b_{ij}\tau_i), \quad P^{(2)} = (p_{ij}^{(2)}) = (a_{ij}\eta_j). \tag{5.2}$$

以及

$$p_{i.}^{(1)} = \sum_j p_{ij}^{(1)} = \tau_i, \quad p_{.j}^{(2)} = \sum_i p_{ij}^{(2)} = \eta_j. \tag{5.3}$$

由式子 (5.2) 和 (5.3) 得到

$$\frac{p_{ij}^{(1)}}{p_{i.}^{(1)}} = \frac{p_{ij}^{(1)}}{\tau_i} = b_{ij}, \quad \frac{p_{ij}^{(2)}}{p_{.j}^{(2)}} = \frac{p_{ij}^{(2)}}{\eta_j} = a_{ij}. \tag{5.4}$$

透過式子 (5.4) 可得

$$\begin{aligned}
E_1 &= \sum_{i=1}^{I}\sum_{j=1}^{J}\left(\frac{p_{ij}^{(1)}}{p_{.j}^{(1)}} - a_{ij}\right)^2 + \sum_{i=1}^{I}\sum_{j=1}^{J}\left(\frac{p_{ij}^{(1)}}{p_{i.}^{(1)}} - b_{ij}\right)^2 \\
&= \sum_{i=1}^{I}\sum_{j=1}^{J}\left(\frac{p_{ij}^{(1)}}{p_{.j}^{(1)}} - a_{ij}\right)^2, \\
E_2 &= \sum_{i=1}^{I}\sum_{j=1}^{J}\left(\frac{p_{ij}^{(2)}}{p_{.j}^{(2)}} - a_{ij}\right)^2 + \sum_{i=1}^{I}\sum_{j=1}^{J}\left(\frac{p_{ij}^{(2)}}{p_{i.}^{(2)}} - b_{ij}\right)^2 \\
&= \sum_{i=1}^{I}\sum_{j=1}^{J}\left(\frac{p_{ij}^{(2)}}{p_{i.}^{(2)}} - b_{ij}\right)^2.
\end{aligned}$$

故可推得凸組合 (5.1) 的結果與顧仲航 (2011) 提出的 $\hat{P}$ 完全相同。

□

根據定理 5.2，在顧仲航 (2011) 所定義的總誤差下，即考慮了條件機率矩陣的誤差和邊際分配的誤差。他推導出的最近似聯合分配 $\hat{P}$ 如式子 (2.4) 所示，且 $\hat{P}$ 和凸組合 (5.1) 的組合方式相同，故這印證了以凸組合 (5.1) 來求近似聯合分配是合理的，並稱其為權重與總誤差成反比法。

第二節　各種組合方式的比較

本節主要以實例說明，如何透過凸組合 (5.1) 的方式來找到近似聯合分配，並對不同的組合方式進行比較。

例題 7. (例題 5續) 考量條件機率矩陣 A 和 B，

由 Matlab 最佳化法延續例題 5所得兩組近似聯合分配 $P^{(1)}$ 和 $P^{(2)}$ ，

$$P^{(1)} = \begin{pmatrix} 0.0561 & 0.0561 & 0.1682 & 0.0561 \\ 0.0668 & 0.0334 & 0.0668 & 0.0668 \\ 0.1433 & 0.0358 & 0.1075 & 0.1433 \end{pmatrix}, \quad P^{(2)} = \begin{pmatrix} 0.0400 & 0.0651 & 0.1325 & 0.0400 \\ 0.0801 & 0.0326 & 0.0442 & 0.0801 \\ 0.1602 & 0.0326 & 0.1325 & 0.1602 \end{pmatrix};$$

再藉由權重與總誤差成反比法可得近似聯合分配，即

$$P = \begin{pmatrix} 0.0460 & 0.0618 & 0.1457 & 0.0460 \\ 0.0752 & 0.0329 & 0.0525 & 0.0752 \\ 0.1539 & 0.0338 & 0.1232 & 0.1539 \end{pmatrix},$$

其中 $\|P_{X|Y} - A\|_F^2 = 0.0054$ 和 $\|P_{Y|X} - B\|_F^2 = 0.0086$，並總誤差 E 為 0.0140。

由特徵向量法延續例題 5所得兩組近似聯合分配 $P^{(1)}$ 和 $P^{(2)}$ ，

$$P^{(1)} = \begin{pmatrix} 0.0466 & 0.0466 & 0.1399 & 0.0466 \\ 0.0670 & 0.0335 & 0.0670 & 0.0670 \\ 0.1619 & 0.0405 & 0.1214 & 0.1619 \end{pmatrix}, \quad P^{(2)} = \begin{pmatrix} 0.0394 & 0.0603 & 0.1407 & 0.0394 \\ 0.0787 & 0.0302 & 0.0469 & 0.0787 \\ 0.1575 & 0.0302 & 0.1407 & 0.1575 \end{pmatrix};$$

再藉由凸組合 (2.4) 可得最近似聯合分配，即

$$P = \begin{pmatrix} 0.0419 & 0.0555 & 0.1404 & 0.0419 \\ 0.0746 & 0.0313 & 0.0540 & 0.0746 \\ 0.1590 & 0.0338 & 0.1339 & 0.1590 \end{pmatrix},$$

其中 $\|P_{X|Y} - A\|_F^2 = 0.0042$ 和 $\|P_{Y|X} - B\|_F^2 = 0.0077$，並總誤差 E 為 0.0119。

我們將例題 7的計算結果整理於表 5.1，並加上等加權法的近似聯合分配如式子 (3.14) 所示。從表 5.1能發現等加權法雖是簡易的方式，但其總誤差也最大；權重與總誤差成反比法的結果和特徵向量法的最近似聯合分配差異不大，故也可說是不錯的組合方法。

各種組合方式的計算結果如下：

表 5.1 各種組合方式的權重、近似聯合分配和總誤差

組合方式	權重 λ	$P = \lambda P^{(1)} + (1-\lambda)P^{(2)}$	E
等加權法	0.5000	$\begin{pmatrix} 0.0480 & 0.0606 & 0.1503 & 0.0480 \\ 0.0734 & 0.0330 & 0.0555 & 0.0734 \\ 0.1518 & 0.0342 & 0.1200 & 0.1518 \end{pmatrix}$	0.0152
權重與總誤差成反比法	0.3709	$\begin{pmatrix} 0.0460 & 0.0618 & 0.1457 & 0.0460 \\ 0.0752 & 0.0329 & 0.0525 & 0.0752 \\ 0.1539 & 0.0338 & 0.1232 & 0.1539 \end{pmatrix}$	0.0140
特徵向量法	0.3543	$\begin{pmatrix} 0.0419 & 0.0555 & 0.1404 & 0.0419 \\ 0.0746 & 0.0313 & 0.0540 & 0.0746 \\ 0.1590 & 0.0338 & 0.1339 & 0.1590 \end{pmatrix}$	0.0119

例題 8. (例題 6續) 考量條件機率矩陣 A 和 B，

A 對 B 的比值矩陣 C 為

$$C = \begin{pmatrix} \frac{6}{11} & * & \frac{72}{155} & \frac{6}{11} & * \\ \frac{21}{22} & * & * & \frac{21}{22} & \frac{7}{8} \\ \frac{21}{11} & \frac{21}{11} & \frac{42}{31} & * & \frac{63}{20} \\ * & \frac{21}{11} & \frac{42}{31} & \frac{14}{11} & * \end{pmatrix},$$

藉由矩形法則，把 $*$ 的位置補滿，得到擴張矩陣 $\overline{C}$ 如下：

$$\overline{C} = \begin{pmatrix} \frac{6}{11} & \frac{6}{11} & \frac{72}{155} & \frac{6}{11} & \frac{1}{2} \\ \frac{21}{22} & \frac{21}{22} & \frac{21}{31} & \frac{21}{22} & \frac{7}{8} \\ \frac{21}{11} & \frac{21}{11} & \frac{42}{31} & \frac{14}{11} & \frac{63}{20} \\ \frac{21}{11} & \frac{21}{11} & \frac{42}{31} & \frac{14}{11} & \frac{63}{20} \end{pmatrix},$$

接著對擴張矩陣 $\overline{C}$ 做奇異值分解，可得

$$\begin{aligned} \overline{C} &= U\Sigma V' \\ &= U \begin{pmatrix} 6.7911 & 0 & 0 & 0 & 0 \\ 0 & 0.7262 & 0 & 0 & 0 \\ 0 & 0 & 0.0624 & 0 & 0 \end{pmatrix} V', \end{aligned}$$

其中

$$U=\begin{pmatrix} 0.1634 & 0.4864 & 0.8583 & 0 \\ 0.2798 & 0.8114 & -0.5131 & 0 \\ 0.6690 & -0.2291 & 0.0025 & 0.7071 \\ 0.6690 & -0.2291 & 0.0025 & -0.7071 \end{pmatrix},$$

$$V=\begin{pmatrix} 0.4286 & 0.2276 & -0.1937 & -0.4381 & 0.7315 \\ 0.4286 & 0.2276 & -0.1937 & -0.5134 & -0.6807 \\ 0.3060 & 0.2134 & 0.9278 & 0 & 0 \\ 0.3032 & 0.6291 & -0.2447 & 0.6717 & -0.0358 \\ 0.6687 & -0.6746 & -0.0654 & 0.3053 & -0.0163 \end{pmatrix}.$$

利用向量 $\mathbf{u}_1$ 和 $\mathbf{v}_1$，透過式子 (1.1) 來得到 X，Y 的近似邊際分配 $\boldsymbol{\tau}$ 和 $\boldsymbol{\eta}$，

$$\boldsymbol{\tau}=\begin{pmatrix} 0.0917 \\ 0.1571 \\ 0.3756 \\ 0.3756 \end{pmatrix},\quad \boldsymbol{\eta}=\begin{pmatrix} 0.1833 \\ 0.1833 \\ 0.2567 \\ 0.2591 \\ 0.1175 \end{pmatrix};$$

再加上給定的條件機率分配，根據式子 (3.12) 可得兩組 (X,Y) 的近似聯合機率分配 $P^{(1)}$ 和 $P^{(2)}$，

$$P^{(1)}=\begin{pmatrix} 0.0229 & 0 & 0.0382 & 0.0306 & 0 \\ 0.0524 & 0 & 0 & 0.0598 & 0.0449 \\ 0.1073 & 0.0805 & 0.0984 & 0 & 0.0894 \\ 0 & 0.1163 & 0.1252 & 0.1341 & 0 \end{pmatrix},$$

$$P^{(2)} = \begin{pmatrix} 0.0250 & 0 & 0.0497 & 0.0471 & 0 \\ 0.0583 & 0 & 0 & 0.0942 & 0.0294 \\ 0.1000 & 0.0750 & 0.0911 & 0 & 0.0881 \\ 0 & 0.1083 & 0.1159 & 0.1178 & 0 \end{pmatrix}.$$

按式子 (3.13) 將 $P^{(1)}$ 和 $P^{(2)}$ 組合起來得到近似聯合分配，其中 $\|P_{X|Y} - A\|_F^2 = 0.0125$ 和 $\|P_{Y|X} - B\|_F^2 = 0.0116$，並總誤差 E 為 0.0241。

由 Matlab 最佳化法延續例題 6所得兩組近似聯合分配 $P^{(1)}$ 和 $P^{(2)}$ ，

$$P^{(1)} = \begin{pmatrix} 0.0307 & 0 & 0.0511 & 0.0409 & 0 \\ 0.0502 & 0 & 0 & 0.0574 & 0.0430 \\ 0.1085 & 0.0814 & 0.0995 & 0 & 0.0904 \\ 0 & 0.1074 & 0.1157 & 0.1239 & 0 \end{pmatrix},$$

$$P^{(2)} = \begin{pmatrix} 0.0261 & 0 & 0.0505 & 0.0425 & 0 \\ 0.0608 & 0 & 0 & 0.0851 & 0.0313 \\ 0.1043 & 0.0772 & 0.0927 & 0 & 0.0938 \\ 0 & 0.1115 & 0.1179 & 0.1063 & 0 \end{pmatrix};$$

再藉由權重與總誤差成反比法可得近似聯合分配，其中 $\|P_{X|Y} - A\|_F^2 = 0.0063$ 和 $\|P_{Y|X} - B\|_F^2 = 0.0102$，並總誤差 E 為 0.0165。

由特徵向量法延續例題 6所得兩組近似聯合分配 $P^{(1)}$ 和 $P^{(2)}$ ，

$$P^{(1)} = \begin{pmatrix} 0.0293 & 0 & 0.0488 & 0.0390 & 0 \\ 0.0598 & 0 & 0 & 0.0683 & 0.0512 \\ 0.1083 & 0.0812 & 0.0993 & 0 & 0.0902 \\ 0 & 0.1005 & 0.1082 & 0.1159 & 0 \end{pmatrix},$$

$$P^{(2)} = \begin{pmatrix} 0.0269 & 0 & 0.0496 & 0.0406 & 0 \\ 0.0628 & 0 & 0 & 0.0812 & 0.0354 \\ 0.1076 & 0.0743 & 0.0909 & 0 & 0.1061 \\ 0 & 0.1074 & 0.1157 & 0.1015 & 0 \end{pmatrix};$$

再藉由凸組合 (2.4) 可得最近似聯合分配，其中 $\|P_{X|Y} - A\|_F^2 = 0.0044$ 和 $\|P_{Y|X} - B\_F^2 = 0.0085$，並總誤差 E 為 0.0128。

我們將例題 8的計算結果整理於表 5.2。從表 5.2能發現特徵向量法最優，權重與總誤差成反比法其次。

各種組合方式的計算結果如下：

表 5.2　各種組合方式的權重、近似聯合分配和總誤差

組合方式	權重 λ	$P = \lambda P^{(1)} + (1-\lambda)P^{(2)}$	E
等加權法	0.5000	$\begin{pmatrix} 0.0240 & 0 & 0.0440 & 0.0388 & 0 \\ 0.0553 & 0 & 0 & 0.0770 & 0.0371 \\ 0.1037 & 0.0777 & 0.0947 & 0 & 0.0888 \\ 0 & 0.1123 & 0.1206 & 0.1260 & 0 \end{pmatrix}$	0.0241
權重與總誤差成反比法	0.4087	$\begin{pmatrix} 0.0280 & 0 & 0.0508 & 0.0419 & 0 \\ 0.0565 & 0 & 0 & 0.0738 & 0.0361 \\ 0.1060 & 0.0789 & 0.0954 & 0 & 0.0924 \\ 0 & 0.1098 & 0.1170 & 0.1135 & 0 \end{pmatrix}$	0.0165
特徵向量法	0.3388	$\begin{pmatrix} 0.0277 & 0 & 0.0493 & 0.0401 & 0 \\ 0.0618 & 0 & 0 & 0.0768 & 0.0407 \\ 0.1079 & 0.0767 & 0.0937 & 0 & 0.1007 \\ 0 & 0.1050 & 0.1132 & 0.1064 & 0 \end{pmatrix}$	0.0128

第六章　結論

綜合前面研究結果，可得到以下結論：

1. 藉由線性代數中的奇異值分解定理，我們發展出奇異值分解法處理二維條件分配相容性問題。

2. 設 A 和 B 為正值條件機率矩陣，C 是比值矩陣，且 $rank(C) = r$。令 C 的奇異值分解為
$$C = \sigma_1 \mathbf{u}_1 \mathbf{v}_1' + \sigma_2 \mathbf{u}_2 \mathbf{v}_2' + \cdots + \sigma_r \mathbf{u}_r \mathbf{v}_r',$$
則 A 和 B 相容的充要條件為 $r = 1$。

3. 設 A 和 B 為正值條件機率矩陣，C 是比值矩陣。當 A 和 B 不相容時，我們可由 C 的奇異值分解判斷出 C 近似秩 1 矩陣的程度；並由奇異值分解定理中的 $\mathbf{u}_1$ 和 $\mathbf{v}_1$ 獲得近似聯合分配。

4. 假設條件機率矩陣 A 和 B 中有元素 0 時，我們利用類 Frobenius 範數導出求比值矩陣 C 的最近似秩 1 矩陣之方法。當類 Frobenius 範數值為 0 時，A 和 B 相容，並可利用生成最近似秩 1 矩陣 (即 C 的 $ROPE$ 矩陣) 的兩個行向量來獲得聯合分配；當類 Frobenius 範數值不為 0 時，A 和 B 不相容，此時可利用生成最近似秩 1 矩陣的兩個行向量來獲得近似聯合分配。這個方法我們稱為最近似秩 1 矩陣法，且其可視為奇異值分解法的延伸。

5. 對最近似秩 1 矩陣法，我們提出三種求解方式 (無限制條件法、Lagrange 乘數法與高維度牛頓法) 以及對應的演算法，並和 Matlab 最佳化法進行實例比較。這三種演算法的計算結果常受初值選擇的影響。

6. 當條件機率矩陣 A 和 B 不相容時，通常可獲得兩組近似的聯合分配，我們對等加權法、權重與總誤差成反比法與特徵向量法三種方法進行實例比較。發現特徵向量法最優，權重與總誤差成反比法其次。

參考文獻

[1] Arnold, B. C. and Press, S. J. (1989), Compatible conditional distributions. *Journal of the American Statistical Association*, 84, 152-156.

[2] Arnold, B. C., Castillo, E., and Sarabia, J. M. (2002), Exact and near compatibility of discrete conditional distributions. *Computational Statistics & Data Analysis*, 40, 231-252.

[3] Arnold, B. C., Castillo, E., and Sarabia, J. M. (2004), Compatibility of partial or complete conditional probability specifications. *Journal of Statistical Planning and Inference*, 123, 133-159.

[4] David Poole (2006), *Linear algebra: a modern introduction*. Brooks/Cole, Cengage Learning.

[5] Markovsky, Ivan (2011), *Low rank approximation: algorithms, implementation, applications*. Springer Science & Business Media.

[6] Song, C. C., Li, L. A., Chen, C. H., Jiang, T. J., and Kuo, K. L. (2010), Compatibility of finite discrete conditional distributions. *Statistical Sinica*, 20, 423-440.

[7] 周志成 (2016)，奇異值分解，檢索日期：2016 年 6 月 3 日，檢自:https://ccjou.wordpress.com/2009/09/01/

[8] 周志成 (2016)，牛頓法，檢索日期：2016 年 6 月 3 日，檢自:https://ccjou.wordpress.com/2013/07/08/

[9] 顧仲航 (2011)，*以特徵向量法解條件分配相容性問題*，國立政治大學應用數學系碩士論文。