國 立 政 治 大 學 應 用 數 學 系

數 學 教 學 碩 士 在 職 專 班

碩士學位論文

2x2 列聯表模型下 MLE 與 MPLE 之比較

The comparison between MLE and MPLE under two-by-two contingency table models

碩專班學生：郭名展 撰

指導教授：姜志銘 博士

宋傳欽 博士

中華民國 103 年 1 月 8 日

中文摘要

Arnold and Strauss (1991) 探討 2x2 列聯表中的 3 個方格 (cell) 有相同機率 θ 的問題，他們比較了參數 θ 的最大概似估計值與最大擬概似估計值，發現參數 θ 的最大概似估計值與最大擬概似估計值是不相同的。在本論文中，我們將 2x2 列聯表中的 3 個方格的參數值 (機率值)，從限制為相同 θ，放寬為成某種比例，並證明了在一般情況下參數 θ 的最大概似估計值與最大擬概似估計值也不相同。我們也提出一些使參數θ的最大概似估計值及最大擬概似估計值相同的特殊條件，諸如三個方格內的觀察值跟機率值成比例或格子內的觀察值有某些特定值。本論文也透過電腦模擬的結果，發現最大概似估計式較最大擬概似估計式來得精確，而且當參數 θ 在參數空間之中點附近時，最大概似估計值與最大擬概似估計值的差異為最大。

關鍵詞：列聯表、最大概似估計、最大擬概似估計

Abstract

Arnold and Strauss (1991) study the cases that three of the four cells in the 2x2 contingency table have the same cell probability θ. In particular, Arnold and Strauss (1991) compare the maximum likelihood estimate (MLE) and maximum pseudolikelihood estimate (MPLE) of the parameter θ. They find that MLE and MPLE of the parameter are not the same. In this thesis, we relax the assumptions so that those three cell probabilities may not be the same and each is proportional to a parameter θ. We find that, in general, MLE's of θ are still not the same as MPLE's of θ. Some special cases that make MLE the same as MPLE are also given. We also find, through computer simulations, that MLE's are accurate than MPLE's and that the difference between MLE and MPLE is getting larger when the parameter θ is closer to the midpoint of its space.

Keywords: contingency table; maximum likelihood estimate; maximum pseudolikelihood estimate

目次

1. 簡介

1.1 研究動機

計數 (count) 是常見的統計資料型態之一，通常在進行此類資料的統計分析時，皆會假設資料來自某種母體分配。然而在各種母體分配中皆有一些未知而我們有興趣的母體特徵值（即參數），這些參數不僅決定了母體的分佈情形，也決定了樣本資料可能發生的機率，因此如何利用搜集到的資料來有效正確地估計未知的母體參數一直是我們想努力研究的議題。

何謂估計 (estimation)？估計主要是利用樣本的資訊，來對母體的未知參數進行推估，其中視需求分為兩種：第一種為點估計式 (point estimator)，利用隨機樣本求得的資料代入估計式得到一個估計值(point estimate)，這樣的方式就稱為是點估計 (point estimation)，例如：當研究者收集完資料後，對資料計算其平均數，並利用求得的平均數來推估母體平均數，這就是所謂的點估計；第二種方式為信賴區間(confidence interval)，利用樣本資訊所求出的點估計式，並根據其抽樣分配 (sampling distribution)的性質建構一包含未知參數的區間，稱為區間估計 (interval estimation)。由於區間估計是以點估計為基礎，因此探討何謂「最好」的點估計式便是統計學者所關注的議題。

我們可以利用許多不同的方法去尋找參數的點估計式，例如：最大概

似估計法 (method of maximum likelihood estimation)、動差法 (method of moments)、最小平方法 (method of the least squares)，以及上述條件不適一般情形下而衍生出的估計法還有最大擬概似估計法(method of maximum pseudolikelihood estimation)、邊際概似估計法(method of marginal maximum likelihood)和一般化最小平方估計法 (method of generalized least squares)。這些方法找出的估計式常不相同，所以研究者就必須藉由一些性質與準則來比較這些估計式的優劣。

一般來說，統計學者認為較佳的點估計式必須具備許多良好的性質，其中最重要的兩個性質為不偏性 (unbiasedness)與有效性 (efficiency)，然而何謂不偏性？其定義如下：若點估計式的抽樣分配之期望值等於其欲估計之母體參數時，就稱這個點估計式具備不偏性。另一個重要的性質是點估計式必須具備有效性，通常估計母體參數的點估計式會有很多個，甚至無窮多個，何者比較好可透過變異數之大小來比較而決定；某點估計式具備有效性，是指此一不偏點估計式的變異數比任何其它不偏之點估計式的變異數還要小。

統計學用期望值與變異數衡量準和精，估計式本身的期望值若等於真正母體的參數，那就是不偏；估計式本身的變異數夠小，則愈有效。若估計式非不偏，則為求有效，統計學家用均方差 (mean square errors, 簡稱 MSE) 以判斷估計式的精確度，當 MSE 愈小，表示所

有估計值平均來說較接近母體參數，所以 MSE 愈小愈好。

根據 Arnold and Press (1989) 提出檢驗兩個條件分配是否滿足相容條件的理論結果，我們可以得知，當兩個條件分配相容時，在某些狀況下我們可以透過條件分配推得其聯合分配。此時，我們可以透過兩種估計法來估計其母體參數，也就是最大概似估計法以及最大擬概似估計法。

最大概似估計法是常用的統計估計方法，它用概似函數 (likelihood function) 來估計密度函數中的參數。然而，在有些狀況下我們無法完整表示出概似函數，因此也就無法求得最大概似估計式 (maximum likelihood estimator, 簡稱 MLE)。例如收集到的資訊當中，只提供了一些條件密度函數 (conditional density function)，但卻無法由它們求出聯合密度函數 (joint density function)，因此無法使用最大概似估計法來估計未知參數。此時我們可以去計算這些條件密度函數的乘積，將乘出來的結果視為參數的函數，就是所謂的擬概似函數 (pseudolikelihood function)，而最大擬概似估計值 (maximum pseudolikelihood estimate, 簡稱 MPLE) 就是使擬概似函數最大的參數值。

Strauss and Ikeda (1990)首先探討指數族分配的擬概似估計法。van Duijn, Gile and Handcock (2009) 進一步對指數族隨機圖模型下的

最大擬概似估計值與最大概似估計值作比較。而 Arnold and Strauss (1991) 則指出在特殊 2x2 列聯表模型下最大擬概似估計值與最大概似估計值的解不同。本論文將透過更一般性的 2x2 二維列聯表，來進行參數的最大概似估計與最大擬概似估計的理論分析及模擬探討，並比較兩者的差異。

1.2 研究目的

在有些簡單的例子中，MLE 和 MPLE 這兩種參數估計值可能會一樣。例如在 Arnold and Strauss (1991) 的論文中曾提到，當二維常態分配中的兩個均數及兩個變異數為已知，但相關係數為未知時，因為兩個邊際分配與相關係數無關，可證得相關係數的 MLE 和 MPLE 會相同。然而 Arnold and Strauss (1991) 也舉了以下的例子來說明 MLE 和 MPLE 是不相同的。

設 X 及 Y 是與 2x2 列聯表有關的兩個變數，且每一變數的值僅為 0 或 1，令他們在第 i 列第 j 行的聯合機率值為 $f(i,j)=P(X=i,Y=j)$，其中 $f(i,j)$ 和未知參數 θ 有關。又全部 n 個觀察值中，落在第 i 列第 j 行的個數為 n_{ij}，因此 $n=n_{00}+n_{01}+n_{10}+n_{11}$。我們可將前述之符號整理成表 1-1 的列聯表型式如下：

表 1-1：2x2 列聯表模型(格中分別為列聯表機率值 $f(i,j)$ 及觀察值 n_{ij})

Y \ X	0	1
0	$f(0,0)$ n_{00}	$f(0,1)$ n_{01}
1	$f(1,0)$ n_{10}	$f(1,1)$ n_{11}

在 $f(0,0)=\theta$ 、 $f(1,0)=\theta$ 、 $f(0,1)=\theta$ 、 $f(1,1)=1-3\theta$ 的情況下，Arnold and Strauss (1991) 比較了 θ 的 MLE 和 MPLE，發現這兩種估計值不相同。

本論文在第 3 章中首先將推廣 Arnold and Strauss (1991) 的結果，把表 1-1 中格子(0,0)、(0,1)以及(1,0)的機率從 1:1:1 改為 1:a:b，其中 $a>0$ 、 $b>0$ 為已知，且 $0\le(1+a+b)\theta\le1$ ，如表 1-2 所示。

表 1-2：2x2 列聯表模型(其中三個格子機率比為 1:a:b)

Y \ X	0	1
0	θ n_{00}	$a\theta$ n_{01}
1	$b\theta$ n_{10}	$1-\theta-a\theta-b\theta$ n_{11}

也就是說，我們想要透過表 1-2 來解決以下的問題：

問題 3.1 在表 1-2 的模型下，參數 θ 之 MLE 以及 MPLE 為何？是否相同？

由問題 3.1 所求得的結果，發現參數 θ 的 MLE 及 MPLE 看似不相同，而且其估計式非常複雜，因此在表 1-3 之特殊情況下進一步探討前述的問題。

表 1-3：2x2 列聯表模型(其中三個格子機率比為 1:2:3)

X \ Y	0	1
0	θ n_{00}	2θ n_{01}
1	3θ n_{10}	$1-6\theta$ n_{11}

我們提出下面第二個問題：

問題 3.2 當 2x2 列聯表模型如表 1-3 時，推導出參數 θ 的 MLE 與 MPLE，並且討論在什麼情況下 MLE 與 MPLE 會相同？

由問題 3.2 所求得的結果，發現在大部分的情況下，參數 θ 的 MLE 與 MPLE 不相同。因此在表 1-3 的模型下，進一步透過電腦模擬資料的方式，來實際比較 MLE 與 MPLE 的差異性，亦即處理以下

之問題 3.3。

問題 3.3 在表 1-3 列聯表的模型下，θ 的 MLE 與 MPLE 的差異如何隨著 θ 值的改變而有所變化？θ 的 MLE 與 MPLE 哪一個估計方式較精確？

1.3 研究架構

本研究第 1 章簡述研究背景與動機。第二章介紹 MLE 以及 MPLE 之基本理論，接著第 3 章將以理論推導的方式，完整的回答問題 3.1 以及問題 3.2，並且再以電腦模擬的方式，回答問題 3.3。最後在第 4 章作結論。

2. 最大概似估計法與最大擬概似估計法

2.1 最大概似估計法

一般母體之參數 θ 皆未知，但若從母體隨機抽出一組樣本，再利用此組樣本去找到一個 θ 的值 $\hat{\theta}$，使這組樣本發生之可能性為最大，則此 $\hat{\theta}$ 值稱為 θ 的最大概似估計值。

更進一步的解釋如下：

設 $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ 為抽自母體(密度函數為 $f_{\mathbf{X}}(\mathbf{x}, \theta)$)的 m 維隨機樣本，

則此 n 個隨機向量 $(\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_n)$ 的聯合密度函數為：

$$f(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n; \theta) = f_{\mathrm{X}}(\mathbf{x}_1; \theta) \cdot f_{\mathrm{X}}(\mathbf{x}_2; \theta) .... f_{\mathrm{X}}(\mathbf{x}_n; \theta)$$

$$= \prod_{i=1}^{n} f_{\mathrm{X}}(\mathbf{x}_i; \theta)$$

【定義 2.1】將聯合密度函數 $f(\mathbf{x}_1, \mathbf{x}_2 ..., \mathbf{x}_n; \theta)$ 視為 θ 的函數 $L(\theta)$，則稱 $L(\theta)$ 為概似函數。

【定義 2.2】設 $L(\theta)$ 為概似函數，若 $\hat{\theta}(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n)$ 使 $L(\theta)$ 為最大時，則 $\hat{\theta}(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n)$ 稱為參數 θ 的最大概似估計值。而 $\hat{\theta}(\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_n)$ 稱為參數 θ 之最大概似估計式。

無論是最大概似估計式或是最大概似估計值，為了方便起見皆以 MLE 來表示。

一般在算 MLE 時，要先了解概似函數 $L(\theta)$ 是否可以微分，若函數 $L(\theta)$ 可以微分則依照微積分求最大值之方法，即可快速求出 MLE，但若此函數 $L(\theta)$ 不可微分或不易微分，則可利用數值分析方法，以求出 MLE。

若概似函數 $L(\theta)$ 可微分，則一般求算 MLE 的步驟如下：

先找概似函數 $L(\theta) = f\left(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n; \theta\right) = \prod_{i=1}^{n} f_{\mathrm{X}}(\mathbf{x}_i, \theta)$

令 $\frac{d \log L(\theta)}{d\theta} = 0$，解 θ，可獲得 $\hat{\theta}$。

則 $\hat{\theta}$ 即為 θ 之 MLE。

2.2 最大擬概似估計法

【定義 2.3】令前節的第 i 個 m 維隨機向量為 $\mathbf{X}_i=(\mathrm{X}_{i,1},\mathrm{X}_{i,2},...,\mathrm{X}_{i,m})$ 且 $\mathbf{X}_{i,-j}=(\mathrm{X}_{i,1},...,\mathrm{X}_{i,j-1},\mathrm{X}_{i,j+1},...,\mathrm{X}_{i,m})$。若給定 $\mathrm{X}_{i,j}\big|\mathbf{X}_{i,-j}$ 的條件密度函數 $f_{ij}(x_{ij}\big|\mathbf{x}_{i,-j})$, $i=1,2,...,n$, $j=1,2,...,m$ ，則 θ 的擬概似函數為

$$PL(\theta)=\prod_{i=1}^{n}\prod_{j=1}^{m}f_{ij}(x_{ij}\big|\mathbf{x}_{i,-j})$$

【定義 2.4】若 $\hat{\theta}(\mathbf{x}_1,\mathbf{x}_2,...,\mathbf{x}_n)$ 使 $PL(\theta)$ 最大，則稱 $\hat{\theta}$ 是 θ 的最大擬概似估計值，而 $\hat{\theta}(\mathbf{X}_1,\mathbf{X}_2,...,\mathbf{X}_n)$ 稱為最大擬概似估計式 (maximum pseudolikelihood estimator，簡稱 MPLE)。

無論是最大擬概似估計式或是最大擬概似估計值，為了方便起見皆以 MPLE 來表示。

若 $PL(\theta)$ 可微分，則一般求 θ 的 MPLE 之步驟如下：

先建立擬概似函數 $PL(\theta)=\prod_{i=1}^{n}\prod_{j=1}^{m}f_{ij}(x_{ij}\big|\mathbf{x}_{i,-j})$

令 $\dfrac{d\log\mathrm{PL}(\theta)}{d\theta}=0$ ，解 θ 得 $\hat{\theta}$，則 $\hat{\theta}$ 即為 θ 之 MPLE。

3.2x2 列聯表模型中參數之 MLE 和 MPLE 的探討

3.1 理論分析

本章主要探討二維列聯表模型下參數的估計，因此前章所介紹的

m 維隨機向量，可以將之限制為二維隨機向量，亦即 $m=2$。為簡化符號，原先第二章的第 i 個隨機向量 $\mathbf{X}_i=(\mathrm{X}_{i,1},\mathrm{X}_{i,2})$ 在本章將以 $(\mathrm{X}_i,\mathrm{Y}_i)$ 來取代。

首先，分析 MLE 和 MPLE 的關係。因為

$$f_{X|Y}(x|y)=\frac{f_{X,Y}(x,y)}{f_Y(y)}$$

且

$$\begin{aligned}PL&=\prod_{i=1}^{n} f_{X|Y}(x_i|y_i)\cdot f_{Y|X}(y_i|x_i)\\&=\prod_{i=1}^{n}\frac{f_{X,Y}(x_i,y_i)}{f_Y(y_i)}\cdot\frac{f_{X,Y}(x_i,y_i)}{f_X(x_i)}\\&=\prod_{i=1}^{n}\frac{f^2{}_{X,Y}(x_i,y_i)}{f_X(x_i)\cdot f_Y(y_i)}\end{aligned}$$

故

$$\log PL=2\sum_{i=1}^{n}\log f_{X,Y}\left(x_i,y_i\right)-\sum_{i=1}^{n}\left(\log f_X(x_i)+\log f_Y(y_i)\right)$$

所以

$$\begin{aligned}\frac{d}{d\theta}\log PL(\theta)&=\sum_{i=1}^{n}\frac{d}{d\theta}2\log f_{X,Y}\left(x_i,y_i\right)-\sum_{i=1}^{n}\left(\frac{d}{d\theta}\log f_X(x_i)+\frac{d}{d\theta}\log f_Y(y_i)\right)\\&=2\frac{d}{d\theta}\log L(\theta)-\sum_{i=1}^{n}\left(\frac{d}{d\theta}\log f_X(x_i)+\frac{d}{d\theta}\log f_Y(y_i)\right)\end{aligned}$$

因此我們有下列結論：

定理一：若 X 及 Y 的邊際分配與參數 θ 無關，則 θ 的 MLE 與 MPLE

相同。

以下將試著回答**問題 3.1**：

問題 3.1 在表 1-2 的模型下，參數 θ 之 MLE 以及 MPLE 為何？是否

相同？

先求 θ 的 MLE，步驟如下：

先找概似函數：$L(\theta)=\theta^{n_{00}}(a\theta)^{n_{01}}(b\theta)^{n_{10}}(1-\theta-a\theta-b\theta)^{n_{11}}$

令 $\frac{d\log L(\theta)}{d\theta}=0$，解 θ，可獲得 θ 的 MLE，令為 $\hat{\theta}_1$。

對概似函數取自然對數：

$$\log\mathrm{L}(\theta)=n_{00}\log\theta+n_{01}\log a\theta+n_{10}\log b\theta+n_{11}\log(1-a\theta-b\theta-\theta)$$

上式對 θ 微分，並令最後結果為 0：

$$\frac{d\log L(\theta)}{d\theta}=\frac{n_{00}}{\theta}+\frac{n_{01}}{\theta}+\frac{n_{10}}{\theta}+\frac{n_{11}(-a-b-1)}{1-a\theta-b\theta-\theta}=0$$

解上述等式之 θ，可得：

$$\hat{\theta}_1=\frac{n-n_{11}}{(a+b+1)n} \tag{3.1.1}$$

則 $\hat{\theta}_1$ 即為 θ 之 MLE。

其次求 θ 的 MPLE 並令為 $\hat{\theta}_2$，過程如下：

先求 $\log PL(\theta)$

在表 1-2 的模型下，我們可以分別推導出 $X|Y$ 以及 $Y|X$ 的條件密度函

數如下：

$$f(x|y)=\left[\left(\frac{\theta}{\theta+b\theta}\right)^{1-x}\left(\frac{b\theta}{\theta+b\theta}\right)^{x}\right]^{1-y}\left[\left(\frac{a\theta}{a\theta+(1-\theta-a\theta-b\theta)}\right)^{1-x}\left(\frac{1-\theta-a\theta-b\theta}{a\theta+(1-\theta-a\theta-b\theta)}\right)^{x}\right]^{y}$$

及

$$f(y|x)=\left[\left(\frac{\theta}{\theta+a\theta}\right)^{1-y}\left(\frac{a\theta}{\theta+a\theta}\right)^{y}\right]^{1-x}\left[\left(\frac{b\theta}{b\theta+(1-\theta-a\theta-b\theta)}\right)^{1-y}\left(\frac{1-\theta-a\theta-b\theta}{b\theta+(1-\theta-a\theta-b\theta)}\right)^{y}\right]^{x}$$

假設(x_i, y_i), $i=1,2,...,n$, 為隨機樣本的值，則

$$\log PL=\sum_{i=1}^{n}[\log f_{X|Y}(x_i|y_i)+\log f_{Y|X}(y_i|x_i)]$$

$$=n_{00}\log\left(\frac{\theta}{\theta+b\theta}\right)+n_{10}\log\left(\frac{b\theta}{\theta+b\theta}\right)+n_{01}\log\left(\frac{a\theta}{a\theta+(1-\theta-a\theta-b\theta)}\right)+$$

$$n_{00}\log\left(\frac{\theta}{\theta+a\theta}\right)+n_{01}\log\left(\frac{a\theta}{\theta+a\theta}\right)+n_{11}\log\left(\frac{1-\theta-a\theta-b\theta}{a\theta+(1-\theta-a\theta-b\theta)}\right)+$$

$$n_{10}\log\left(\frac{b\theta}{b\theta+(1-\theta-a\theta-b\theta)}\right)+n_{11}\log\left(\frac{1-\theta-a\theta-b\theta}{b\theta+(1-\theta-a\theta-b\theta)}\right)$$

上式對 θ 微分可得：

$$n_{01}\left[\frac{1}{\theta}+\frac{b+1}{1-b\theta-\theta}\right]+n_{11}\left[\frac{-a-b-1}{1-a\theta-b\theta-\theta}+\frac{b+1}{1-b\theta-\theta}\right]+n_{10}\left[\frac{1}{\theta}+\frac{a+1}{1-a\theta-\theta}\right]+$$
$$n_{11}\left[\frac{-a-b-1}{1-a\theta-b\theta-\theta}+\frac{a+1}{1-a\theta-\theta}\right]$$

並令其為 0，得：

$$\frac{n_{01}+n_{10}}{\theta}+\frac{bn_{01}+bn_{11}+n_{01}+n_{11}}{1-b\theta-\theta}+\frac{an_{10}+an_{11}+n_{10}+n_{11}}{1-a\theta-\theta}+\frac{-2an_{11}-2bn_{11}-2n_{11}}{1-a\theta-b\theta-\theta}=0$$

左邊通分並整理後，得分子為 θ 的二次式，其各項係數如下：

常數項係數：$n_{01}+n_{10}$

θ項係數：$-2an_{01}-bn_{01}-2n_{01}-an_{10}-2bn_{10}-2n_{10}-an_{11}-bn_{11}$

θ^2項係數：$(a^2+ab+2a+b+1)n_{01}+(b^2+ab+a+2b+1)n_{10}+(a^2+b^2+a+b)n_{11}$

令分子為 0，並解 θ 後得

$\hat{\theta}_2=$

$$\frac{2n_{01}+2n_{10}+2an_{01}+an_{10}+bn_{01}+an_{11}+2bn_{10}+bn_{11}}{2(n_{01}+n_{10}+2an_{01}+an_{10}+bn_{01}+an_{11}+2bn_{10}+bn_{11}+a^2n_{01}+a^2n_{11}+b^2n_{10}+b^2n_{11}+abn_{01}+abn_{10})}$$

$$\pm\frac{\sqrt{a^2{n_{10}}^2-2a^2n_{10}n_{11}+a^2{n_{11}}^2+2abn_{01}n_{10}+6abn_{01}n_{11}+6abn_{10}n_{11}+2ab{n_{11}}^2+b^2{n_{01}}^2-2b^2n_{01}n_{11}+b^2{n_{11}}^2}}{2(n_{01}+n_{10}+2an_{01}+an_{10}+bn_{01}+an_{11}+2bn_{10}+bn_{11}+a^2n_{01}+a^2n_{11}+b^2n_{10}+b^2n_{11}+abn_{01}+abn_{10})} \quad (3.1.2)$$

根據上述推導結果，經由回答**問題 3.1**，可得以下之結論：

定理二：在表 1-2 模型下，θ 的 MLE $\hat{\theta}_1$ 與 MPLE $\hat{\theta}_2$ 分別為(3.1.1)及(3.1.2)。

由(3.1.1)以及(3.1.2)，$\hat{\theta}_1$ 與 $\hat{\theta}_2$ 是不相同的。

註：隨著 a 、b 值的改變，$\hat{\theta}_2$ 的兩個值中，我們預測有一個會不符合參數 θ 值的範圍，即 $0\le\theta\le\frac{1}{1+a+b}$；事實上在表 1-3 的特殊模型下，即 $a=2$ 及 $b=3$ 時，我們將透過回答**問題 3.2** 來證明 $\hat{\theta}_2$ 的值只有一個。

問題 3.2 當 2x2 列聯表模型如表 1-3 時，推導出參數 θ 的 MLE 與 MPLE，並且討論在什麼情況下 MLE 與 MPLE 會相同？

利用(3.1.1)式的結果，將 $a=2$ 、$b=3$ 代入，即可求出 θ 之 MLE：

$$\hat{\theta}_1=\frac{n-n_{11}}{6n} \quad (3.1.3)$$

同理利用(3.1.2)式的結果將$a=2$、$b=3$代入，即可求出θ之MPLE：

$$\hat{\theta}_2=\frac{9n_{01}+10n_{10}+5n_{11}}{12(3n_{01}+4n_{10}+3n_{11})}\pm\frac{\sqrt{9n_{01}{}^2+12n_{01}n_{10}+18n_{01}n_{11}+4n_{10}{}^2+28n_{10}n_{11}+25n_{11}{}^2}}{12(3n_{01}+4n_{10}+3n_{11})}$$

（3.1.4）

在(3.1.4)式中等號右邊"+"或"−"的地方，我們只能夠選擇"−"，證明過程如下：

令

$$A=\frac{9n_{01}+10n_{10}+5n_{11}}{12(3n_{01}+4n_{10}+3n_{11})}$$

$$B=\frac{\sqrt{9n_{01}{}^2+12n_{01}n_{10}+18n_{01}n_{11}+4n_{10}{}^2+28n_{10}n_{11}+25n_{11}{}^2}}{12(3n_{01}+4n_{10}+3n_{11})}$$

因為

$$\mathrm{B}\geq\frac{\sqrt{25n_{11}{}^2}}{12(3n_{01}+4n_{10}+3n_{11})}=\frac{5n_{11}}{12(3n_{01}+4n_{10}+3n_{11})}$$

所以

$$\hat{\theta}_2=A+B\geq A+\frac{5n_{11}}{12(3n_{01}+4n_{10}+3n_{11})}$$

因此

$$\hat{\theta}_2\geq A+\frac{5n_{11}}{12(3n_{01}+4n_{10}+3n_{11})}=\frac{9n_{01}+10n_{10}+10n_{11}}{12(3n_{01}+4n_{10}+3n_{11})}>\frac{6n_{01}+8n_{10}+6n_{11}}{12(3n_{01}+4n_{10}+3n_{11})}=\frac{1}{6}$$

又$0\leq\theta\leq\frac{1}{6}$，故（3.1.4）式子中等號右邊的"+"不合。

接著討論在什麼狀況下，會有 $MLE = MPLE$ 的現象。

(1)當 $n_{00} : n_{01} : n_{10} = 1:2:3$ 時，$MLE = MPLE$

證明過程如下：

令 $n_{00} = r$ 、 $n_{01} = 2r$ 、 $n_{10} = 3r$ 並代入(3.1.3)式，得：

$$\hat{\theta}_1 = \frac{n-(n-r-2r-3r)}{6n} = \frac{6r}{6n} = \frac{r}{n}$$

再將 $n_{00} = r$ 、 $n_{01} = 2r$ 、 $n_{10} = 3r$ 代入（3.1.4）式，得：

$$\hat{\theta}_2 = \frac{18r+5n}{36n} - \frac{\sqrt{25n^2 - 180rn + 324r^2}}{36n} = \frac{18r+5n-(5n-18r)}{36n} = \frac{r}{n}$$

故 $MLE = MPLE$

(2)當 $n_{11} = 0$ 時，$MLE = MPLE$

證明過程如下：

由(3.1.3)式得 $\hat{\theta}_1 = \frac{1}{6}$

由(3.1.4)式得

$$\hat{\theta}_2 = \frac{9n_{01} + 10n_{10} - \sqrt{9n_{01}{}^2 + 12n_{01}n_{10} + 4n_{10}{}^2}}{12(3n_{01} + 4n_{10})} = \frac{9n_{01} + 10n_{10} - (3n_{01} + 2n_{10})}{36n_{01} + 48n_{10}} = \frac{1}{6}$$

故 $MLE = MPLE$

(3)當 $n_{00} = n_{01} = n_{10} = 0$ 時，$MLE = MPLE$

證明過程如下：

由(3.1.3)式得 $\hat{\theta}_1 = \frac{n - n_{11}}{6n} = 0$

由(3.1.4)式得 $\hat{\theta}_2 = \dfrac{5n_{11} - \sqrt{25n_{11}}}{36n_{11}} = 0$

故 $MLE = MPLE$

經上述推導過程，可得以下結果:

定理三：在表 1-3 模型下，θ 的 MLE $\hat{\theta}_1$ 與 MPLE $\hat{\theta}_2$ 分別為(3.1.3)及(3.1.4)。

另外，在以下三種狀況下，可得 $\hat{\theta}_1 = \hat{\theta}_2$：

(1)$n_{00} : n_{01} : n_{10} = 1:2:3$,

(2)$n_{11} = 0$,

(3)$n_{00} = n_{01} = n_{10} = 0$ 。

3.2 模擬比較

在表 1-3 的模型下，從區間[0,1/6]中選擇不同的參數值 θ，透過統計軟體 Matlab 模擬出 4 個格子中的觀察值 n_{00}、n_{01}、n_{10}、n_{11}。觀察 MLE 與 MPLE 這兩種估計結果的差異情況，並藉由圖表的呈現找出差異變化的趨勢。本節中就是要回答以下的問題：

問題 3.3 在表 1-3 列聯表的模型下，θ 的 MLE 與 MPLE 的差異如何隨著 θ 值的改變而有所變化？θ 的 MLE 與 MPLE 哪一個估計方式較精確？

在 θ 值分別為 0 、1/60 、3/60 、5/60 、7/60 、9/60 、10/60 時，模擬出 n=50 筆觀察值，再代入(3.1.3)式與(3.1.4)式中可分別得 MLE $\hat{\theta}_1$ 及 MPLE $\hat{\theta}_2$。如此重複模擬 3 萬次，得 3 萬筆 $\hat{\theta}_1$ 的值與 3 萬筆 $\hat{\theta}_2$ 的值。經計算後可得這些 $\hat{\theta}_1$ 的平均數 $\bar{\hat{\theta}}_1$ 、$\hat{\theta}_2$ 的平均 $\bar{\hat{\theta}}_2$ 、$\hat{\theta}_1$ 的變異數、$\hat{\theta}_2$ 的變異數、$\hat{\theta}_1$ 的均方差以及 $\hat{\theta}_2$ 的均方差，如表 3-1、3-2 所示。又 $\hat{\theta}_1$ 與 $\hat{\theta}_2$ 的變異數折線圖、$\hat{\theta}_1$ 與 $\hat{\theta}_2$ 的均方差折線圖如圖 3-1、3-2 所示。

表 3-1： 模擬 3 萬次而得 MLE $\hat{\theta}_1$ 及 MPLE $\hat{\theta}_2$ 之均數及變異數彙整表。

θ 值	$\bar{\hat{\theta}}_1$	$\bar{\hat{\theta}}_2$	$\hat{\theta}_1$ 的變異數	$\hat{\theta}_2$ 的變異數
0	0	0		
1/60 （0.016666）	0.016635	0.016547	0.0000499192	0.0000585537
3/60 (0.05)	0.049956	0.049819	0.00011677	0.0001324458
5/60 (0.083333)	0.083231	0.083094	0.0001383234	0.0001527626
7/60 (0.1166666)	0.116667	0.116524	0.000115062	0.0001228046
9/60 (0.15)	0.14999	0.149928	0.0000497027	0.0000512955
10/60 (0.1666666)	10/60	10/60		

備註：當 $\theta = 0$ 時，可由**定理三**得 $\hat{\theta}_1 = \hat{\theta}_2 = 0$；當 $\theta = \frac{10}{60}$ 時，可由**定理三**得

$$\hat{\theta}_1 = \hat{\theta}_2 = \frac{1}{6}$$

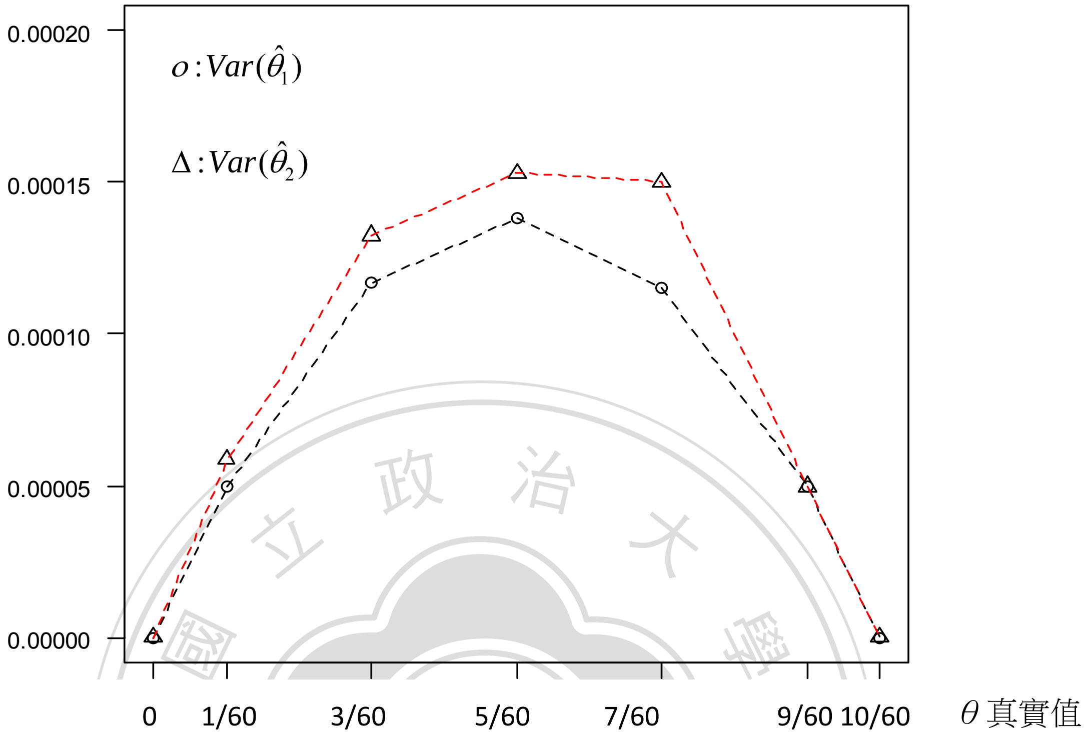

圖 3-1：$\hat{\theta}_1$ 和 $\hat{\theta}_2$ 之變異數比較折線圖

由圖 3-1 可以看出在 θ=7/60 區間附近的時候，$\hat{\theta}_1$ 的變異數和 $\hat{\theta}_2$ 的變異數差距最大。並且 θ 不論在任何值，$\hat{\theta}_2$ 的變異數都大於 $\hat{\theta}_1$ 的變異數，由此可知 MLE 的方法較 MPLE 的方法為佳。

表 3-2： 模擬三萬次而得 MLE $\hat{\theta}_1$ 及 MPLE $\hat{\theta}_2$ 之均數及均方差彙整表。

θ 值	$\bar{\hat{\theta}}_1$	$\bar{\hat{\theta}}_2$	$\hat{\theta}_1$ 的均方差	$\hat{\theta}_2$ 的均方差
0	0	0		
1/60（0.016666）	0.016635	0.016547	0.0000499185	0.000058566
3/60 (0.05)	0.049956	0.049819	0.0001167681	0.0001324742
5/60 (0.083333)	0.083231	0.083094	0.0001383293	0.0001528146
7/60 (0.1166666)	0.116667	0.116524	0.0001150581	0.0001228208
9/60 (0.15)	0.14999	0.149928	0.0000497011	0.000051299
10/60 (0.1666666)	10/60	10/60		

備註：當 $\theta=0$ 時，可由定理三得 $\hat{\theta}_1=\hat{\theta}_2=0$；當 $\theta=\frac{10}{60}$ 時，可由定理三得

$$\hat{\theta}_1=\hat{\theta}_2=\frac{1}{6}$$

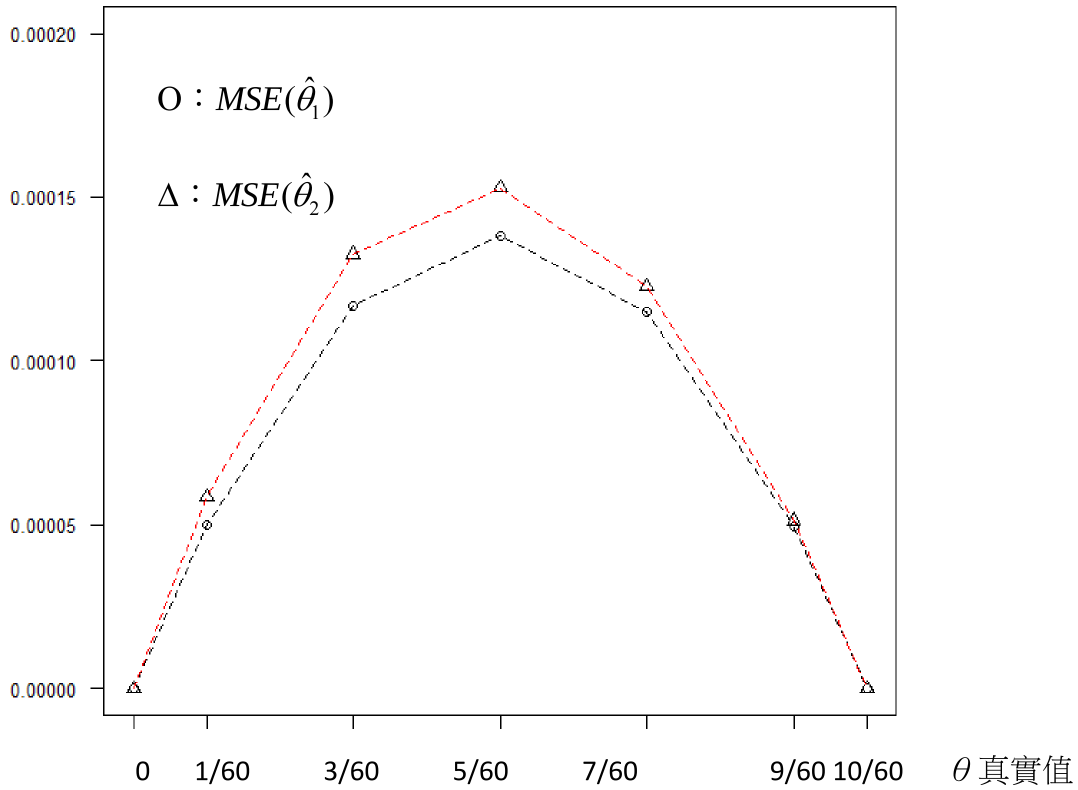

圖 3-2：$\hat{\theta}_1$ 和 $\hat{\theta}_2$ 之均方差比較折線圖

由圖 3-2 可以看出在 θ=3/60 區間附近的時候，$\hat{\theta}_1$ 的均方差和 $\hat{\theta}_2$ 的均方差差距最大。並且 θ 不論在任何值，$\hat{\theta}_2$ 的均方差都大於 $\hat{\theta}_1$ 的均方差。由此可進一步發現 MLE 的方法較 MPLE 的方法為佳。

令三萬筆兩估計值差的標準差為 $S_{\theta_1-\theta_2}$ ，利用中央極限定理，$\theta_1-\theta_2$ 的 95%信賴區間可表示為 $(\bar{\hat{\theta}}_1-\bar{\hat{\theta}}_2)\pm 1.96 s_{\theta_1-\theta_2}$ ，其中 $\theta_1=E(\hat{\theta}_1)$ 、$\theta_2=E(\hat{\theta}_2)$ 。經計算並整理後得表 3-3 之結果。

表 3-3：模擬三萬次而得 $\theta_1-\theta_2$ 的 95%信賴區間彙整表。

θ 值	$\bar{\hat{\theta}}_1$	$\bar{\hat{\theta}}_2$	$\bar{\hat{\theta}}_1 - \bar{\hat{\theta}}_2$	$\theta_1-\theta_2$ 的 95%信賴區間
0	0	0	0	
1/60（0.016666）	0.016635	0.016547	0.00008777	$[0.00005288, 0.000122675]$
3/60 (0.05)	0.049956	0.049819	0.000137525	$[0.000090692, 0.000184357]$
5/60 (0.083333)	0.083231	0.083094	0.00027289	$[0.00022905, 0.00031672]$
7/60 (0.1166666)	0.116667	0.116524	0.00014273	$[0.00011051, 0.00017495]$
9/60 (0.15)	0.14999	0.149928	0.000062767	$[0.00004917, 0.000076363]$
10/60 (0.1666666)	10/60	10/60	0	

備註：當 $\theta=0$ 時，可由**定理三**得 $\hat{\theta}_1=\hat{\theta}_2=0$；當 $\theta=\frac{10}{60}$ 時，可由**定理三**得

$$\hat{\theta}_1=\hat{\theta}_2=\frac{1}{6}$$

當 θ 值改變時，$\bar{\hat{\theta}}_1 - \bar{\hat{\theta}}_2$ 值對應變化的折線圖如圖 3-3

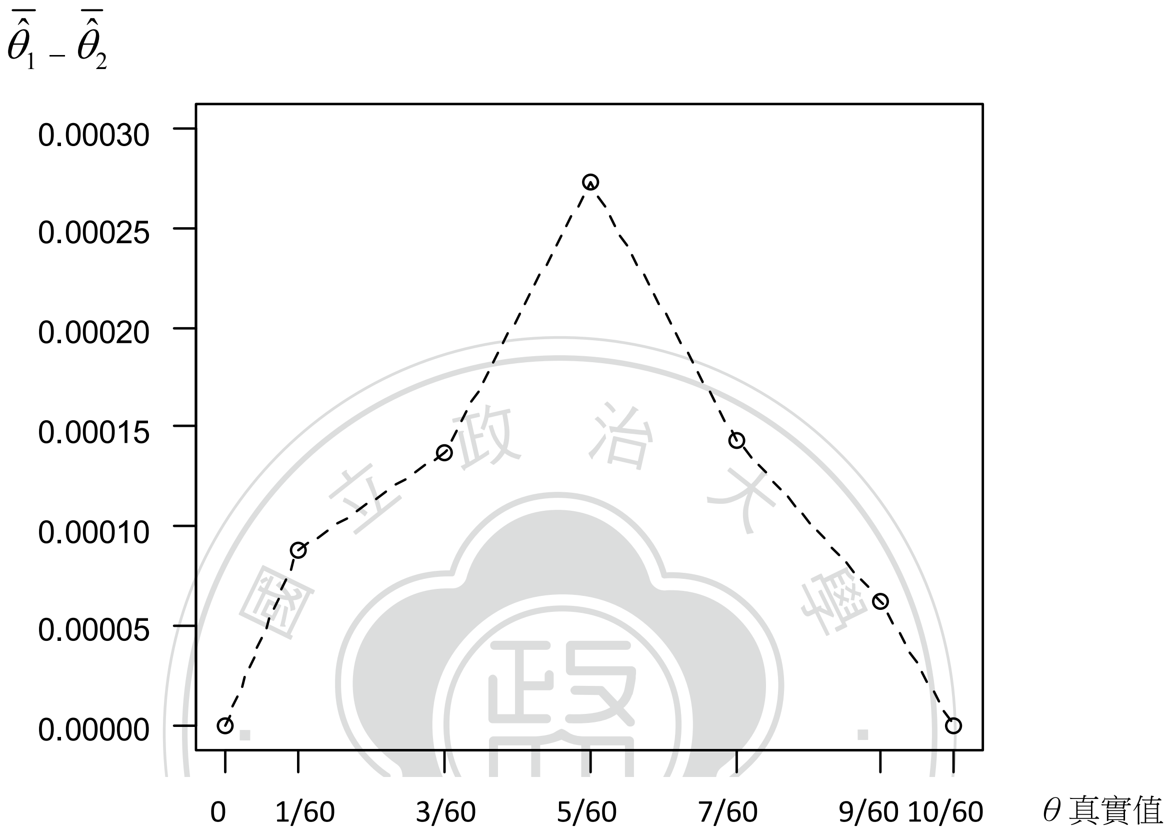

圖 3-3：θ 值變動時，$\bar{\hat{\theta}}_1 - \bar{\hat{\theta}}_2$ 值對應變動的折線圖。

由圖 3-3 可以看出，在 θ=5/60 區間附近時，$\bar{\hat{\theta}}_1 - \bar{\hat{\theta}}_2$ 有最大值並且圖形往兩側遞減。

根據表 3-1 至表 3-3 及圖 3-1 至 3-3 的結果，可以歸納出以下的結論：由表 3-1 可以看出 $\bar{\hat{\theta}}_1$ 的值比 $\bar{\hat{\theta}}_2$ 的值要接近真正的 θ 值，故 MLE 比 MPLE 估計的較準；由表 3-1、3-2(或圖 3-1、3-2)可以看出，無論比較 $\hat{\theta}_1$ 和 $\hat{\theta}_2$ 之變異數或者均方差，MLE 皆優於 MPLE；由圖 3-3 可看出在 θ=5/60 區間附近，$\bar{\hat{\theta}}_1$ 和 $\bar{\hat{\theta}}_2$ 的差異最大。

4. 結論

將 Arnold and Strauss (1991) 探討的 2x2 列聯表中，3 個方格內之機率由限制為相同的參數 θ，放寬為某種比例後，本文證明了 θ 的 MLE 及 MPLE 仍不相同，也證明了當這三個方格內的觀察值成某種比例或有某些特定值時，MLE 及 MPLE 會相同。本文也透過電腦模擬，發現 MLE 較 MPLE 來得精確，而且當參數 θ 在其空間中點附近，MLE 與 MPLE 的差異最大。最後再透過圖表可看出，MPLE 是偏估計，而且通常會低估。

參考文獻

Arnold, B. C. and Press, S. J. (1989)."Compatible Conditional Distributions." *Journal of the American Statistical Association* 84(405): 152-156.

Arnold, B. C. and Strauss, D (1991). "Pseudolikelihood Estimation: Some Examples" *Journal of the Indian Journal of Statistics* 53(2): 233-243.

Strauss, D. and Ikeda. M. (1990). "Pseudolikelihood Estimation for Social Networks."*Journal of the American Association* 85(409):204-212

van Duijn, M. A., Gile, K. J., and Handcock , M. S. (2009). "A Framework for the Comparison of Maximum Pseudo Likelihood and Maximum Likelihood Estimation of Exponential Family Random Graph Models." *Social Networks* 31(1):52-62

附錄

在表 1-3 模型下且 n=50，利用 Matlab 數學軟體模擬 30000 次之程式碼 [每次模擬觀察值為n_{00}、n_{01}、n_{10}、n_{11}，因此將有 3 萬筆 (n_{00}, n_{01}, n_{10}, n_{11})的資料，其中$n = n_{00} + n_{01} + n_{10} + n_{11}$=50。] 如下：

（1）θ=1/60

```
a=1/60;
b=randsample(1:4,50,'true',[a*[1:3],1-6*a]);
histc(randsample(1:4,50,'true',[a*[1:3],1-6*a]),1:4)
fori=1:30000
bi=histc(randsample(1:4,50,'true',[a*[1:3],1-6*a]),1:4)
C(i,:)=[bi]
end
```

（2）θ=3/60

```
a=3/60;
b=randsample(1:4,50,'true',[a*[1:3],1-6*a]);
histc(randsample(1:4,50,'true',[a*[1:3],1-6*a]),1:4)
fori=1:30000
bi=histc(randsample(1:4,50,'true',[a*[1:3],1-6*a]),1:4)
C(i,:)=[bi]
end
```

（3）θ=5/60

```
a=5/60;
b=randsample(1:4,50,'true',[a*[1:3],1-6*a]);
histc(randsample(1:4,50,'true',[a*[1:3],1-6*a]),1:4)
fori=1:30000
bi=histc(randsample(1:4,50,'true',[a*[1:3],1-6*a]),1:4)
C(i,:)=[bi]
end
```

（4） θ=7/60

```
a=7/60;
b=randsample(1:4,50,'true',[a*[1:3],1-6*a]);
histc(randsample(1:4,50,'true',[a*[1:3],1-6*a]),1:4)
fori=1:30000
```

```
bi=histc(randsample(1:4,50,'true',[a*[1:3],1-6*a]),1:4)
C(i,:)=[bi]
end
```

（5）θ=9/60

```
a=9/60;
b=randsample(1:4,50,'true',[a*[1:3],1-6*a]);
histc(randsample(1:4,50,'true',[a*[1:3],1-6*a]),1:4)
fori=1:30000
bi=histc(randsample(1:4,50,'true',[a*[1:3],1-6*a]),1:4)
C(i,:)=[bi]
end
```