貝氏期末報告

Wishart Distribution & Gibbs sampler for a mixture of normals

January 6, 2022

企研三 蔡德脩　統研一 程長磊

目錄

壹、研究大綱.. 2

貳、Wishart Distribution........................ 2

定義.. 2

Wishart 延伸應用 2

參、Gibbs Sampling 抽樣 3

參數設定...3

抽樣結果...3

收斂診斷...6

肆、EM 演算法 6

參數設定...6

抽樣結果...6

EM 演算法與 Gibbs 方法比較 7

伍、結論.. 8

陸、參考資料.. 9

壹、 研究大綱

本研究將針對課堂所提及的，Gibbs Sampling 模擬混和多元常態分布進行延伸，介紹 Wishart 分配以及多元常態之 Full conditional。
透過 faithful 資料集作為觀察值進行模擬，利用 Gibbs Sampling 和 EM 演算法找出最適合的母體參數，並檢定其是否收斂。

貳、 Wishart Distribution 介紹

定義

假定W為$p \times p$連續隨機矩陣，且

$$R_W = \{w \in \mathbb{R}^{p \times p}: w\ is\ symmetric\ and\ positive\ definite\}$$

則我們可以說 W 服從 Wishart distribution 且其 pdf 如下所示

$$f_W(w) = c[\det(w)]^{\frac{n}{2}-\frac{(p+1)}{2}} \exp\left(-\frac{n}{2} tr(\Sigma^{-1}w)\right),$$

$$\text{where c} = \frac{n^{\frac{n}{2}}}{2^{\frac{np}{2}}[\det(\Sigma)]^{\frac{n}{2}}\pi^{\frac{p(p-1)}{4}}\prod_{j-1}^{p}\Gamma\left(\frac{n}{2}+\frac{(1-j)}{2}\right)},$$

其中Σ為一正定且對稱矩陣，且$n > p-1$，而$\Gamma(.)$為 Gamma 函數。

Wishart 延伸應用

假使有一組資料$X_1, \dots, X_n$ 為 n 個獨立的$p \times 1$隨機向量並且每一個向量都服從平均數為 0 的多元常態分布，即$X_i \sim N(0, \Sigma)$，則我們可以得到:

$$W = \sum_{i=1}^{n} X_i X_i^T \sim W_p(\Sigma, n)$$

此外，Wishart 分配也具有以下性質:

1. For $M \sim W_p(n, \Sigma)\ and\ B_{(p \times m)}, B'MB \sim W_m(n, B'\Sigma B)$.
2. For $M \sim W_p(n, \Sigma)\ with\ \Sigma > 0,\ \Sigma^{\frac{-1}{2}} M \Sigma^{\frac{-1}{2}} \sim W_p(n, \mathbb{I}_p)$.
3. If M_i are independent $W_p(n_i, \Sigma), (i = 1, \dots, k)$,
   $then\ \sum_{i=1}^{k} M_i \sim W_p(n, \Sigma)\ where\ n = n_1 + \cdots + n_k$.
4. For $M_n \sim W_p(n, \Sigma),\ EM_n = n\Sigma$

參、 Gibbs Sampling

參數設定

先驗分配相關參數

$\Sigma_j \sim IW(m_j, \varphi_j)$ $m_j = 2$， $\varphi_j = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$,for j = 1,2

$\mu_j | \Sigma_j \sim N\left(\mu_{0j}, \frac{\Sigma_j}{\tau_{0j}}\right)$ $\mu_{0j} = [3,70]$， $\tau_{0j} = 1$ for $j = 1,2$

$\alpha_i \sim \text{Dirichlet}(a1, a2)$ $a1 = a2 = 2.5$

起始值

$\mu_{01} = [2,55]$,$\mu_{02} = [4.4,80]$

$\Sigma_{01} = \Sigma_{02} = \begin{bmatrix} 0.8 & 7 \\ 7 & 70 \end{bmatrix}$

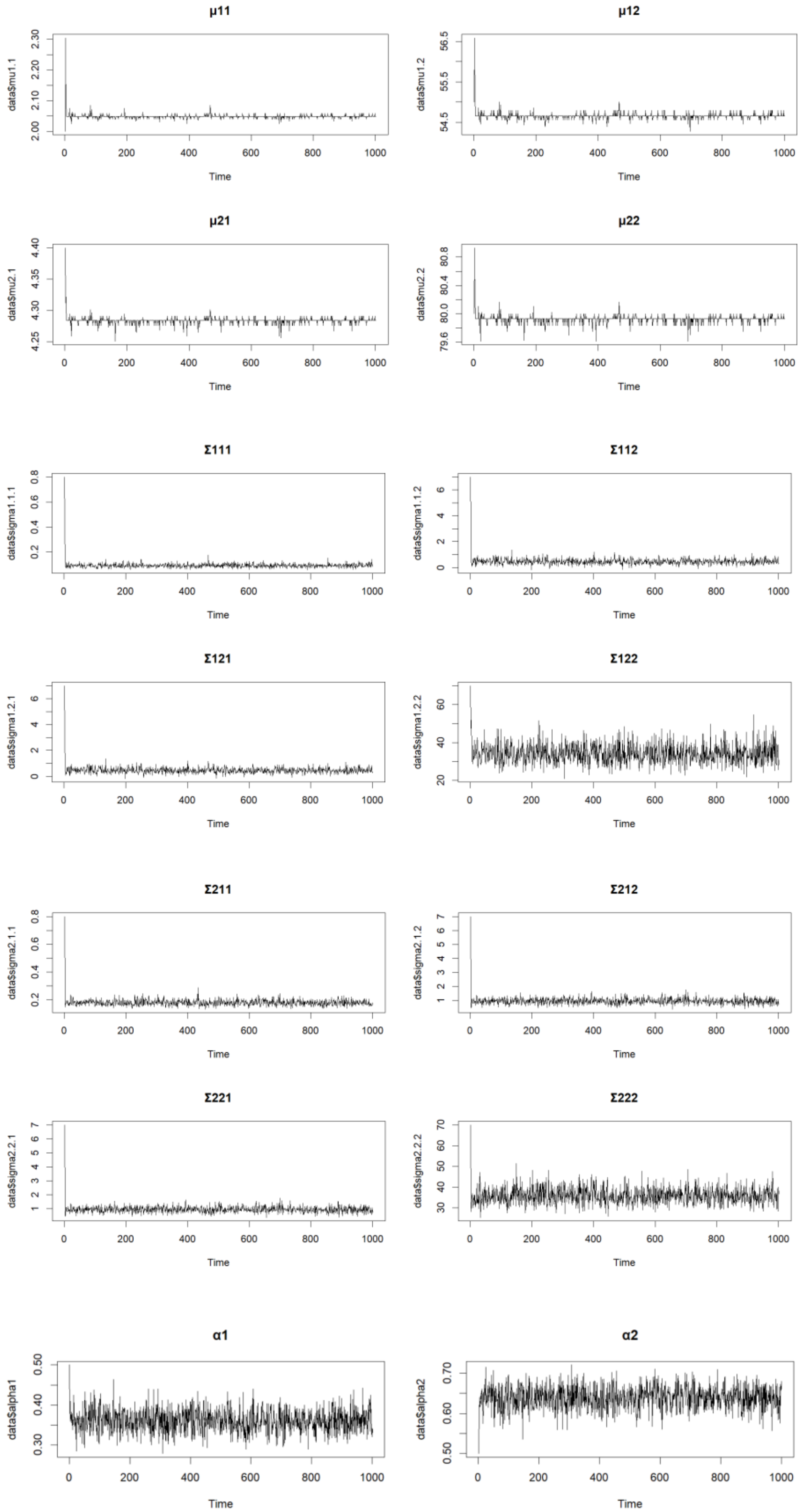
$\alpha_1 = \alpha_2 = 0.5$

抽樣結果

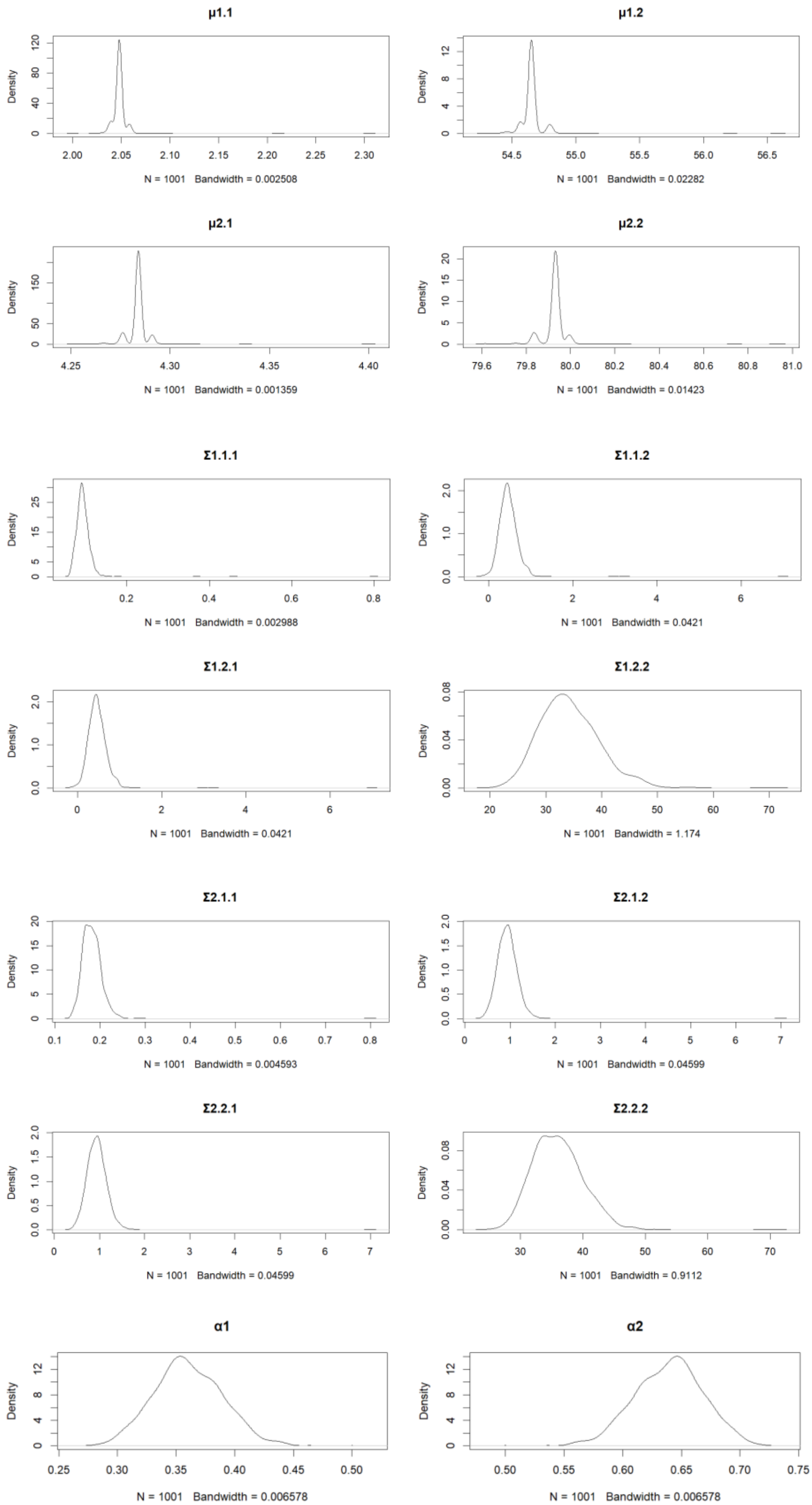
我們利用先前所設定之先驗分配參數來進行抽樣，抽取 1000 次，並且去除前 500 次避免我們的馬可夫鏈在達成收斂之前的狀態的值影響我們的計算結果，接著將後 500 次的結果取平均和變異數，結果如下。

	μ_{11}	μ_{12}	μ_{21}	μ_{22}	Σ_{111}	Σ_{112}	Σ_{121}
平均	2.047	54.651	4.283	79.921	0.091	0.445	0.445
變異數	2.49e-5	0.0034	1.75e-5	0.0022	0.0002	0.0400	0.0400
	Σ_{122}	Σ_{211}	Σ_{212}	Σ_{221}	Σ_{222}	α_1	α_2
平均	34.600	0.1816	0.9366	0.9366	36.196	0.3611	0.6388
變異數	25.476	0.0004	0.0472	0.0472	15.801	0.0008	0.0008

以下是個參數時間序列圖形與收斂狀況，初步判斷所有的鏈收斂的速度都很快，大約在抽樣 20 次後就會達成一個平穩的狀態。


μ11
data$mu1.1
μ12
data$mu1.2
μ21
data$mu2.1
μ22
data$mu2.2
Σ111
data$sigma1.1.1
Σ112
data$sigma1.1.2
Σ121
data$sigma1.2.1
Σ122
data$sigma1.2.2
Σ211
data$sigma2.1.1
Σ212
data$sigma2.1.2
Σ221
data$sigma2.2.1
Σ222
data$sigma2.2.2
α1
data$alpha1
α2
data$alpha2
Time

機率密度圖


μ1.1
N = 1001 Bandwidth = 0.002508
μ1.2
N = 1001 Bandwidth = 0.02282
μ2.1
N = 1001 Bandwidth = 0.001359
μ2.2
N = 1001 Bandwidth = 0.01423
Σ1.1.1
N = 1001 Bandwidth = 0.002988
Σ1.1.2
N = 1001 Bandwidth = 0.0421
Σ1.2.1
N = 1001 Bandwidth = 0.0421
Σ1.2.2
N = 1001 Bandwidth = 1.174
Σ2.1.1
N = 1001 Bandwidth = 0.004593
Σ2.1.2
N = 1001 Bandwidth = 0.04599
Σ2.2.1
N = 1001 Bandwidth = 0.04599
Σ2.2.2
N = 1001 Bandwidth = 0.9112
α1
N = 1001 Bandwidth = 0.006578
α2
N = 1001 Bandwidth = 0.006578
Density

收斂診斷

接著我們針對每個參數重複抽取三次得到三條馬可夫鏈，利用 Rstan 套件計算個參數的 R-hat 藉此判斷我們的結果是否收斂，由下表可以看出，每個參數的 R-hat 皆小於 1.05，因此判斷我們所設定的抽取 1000 次以及重複抽取三條鏈可以得到一個平穩且達成收斂狀態的馬可夫鏈。

	Mean	Std	2.5%	25%	50%	75%	97.5%	$\hat{R}$
μ_{11}	2.048	0.005	2.039	2.048	2.048	2.048	2.058	1.006
μ_{11}	54.652	0.062	54.567	54.653	54.653	54.653	54.798	1.006
μ_{11}	4.283	0.004	4.276	4.284	4.284	4.284	4.291	1.009
μ_{11}	79.923	0.051	79.836	79.932	79.932	79.932	79.994	1.009
Σ_{111}	0.091	0.014	0.067	0.081	0.09	0.099	0.121	1
Σ_{112}	0.471	0.183	0.107	0.351	0.466	0.59	0.86	1.007
Σ_{121}	0.471	0.183	0.107	0.351	0.466	0.59	0.86	1.007
Σ_{122}	34.535	5.035	26.064	30.973	34.185	37.561	45.806	0.999
Σ_{211}	0.181	0.021	0.146	0.166	0.181	0.195	0.226	1.001
Σ_{212}	0.942	0.222	0.552	0.775	0.937	1.077	1.43	1
Σ_{221}	0.942	0.222	0.552	0.775	0.937	1.077	1.43	1
Σ_{222}	36.249	4.041	29.087	33.499	35.892	38.265	45.078	0.998
α_1	0.361	0.03	0.307	0.34	0.359	0.381	0.417	1
α_2	0.639	0.03	0.583	0.619	0.641	0.66	0.693	1

肆、 EM 演算法

參數設定

起始值

$\mu_{01} = [2,55]$, $\mu_{02} = [4.4,80]$

$\Sigma_{01} = \Sigma_{02} = \begin{bmatrix} 0.8 & 7 \\ 7 & 70 \end{bmatrix}$

$\alpha_1 = \alpha_2 = 0.5$

抽樣結果

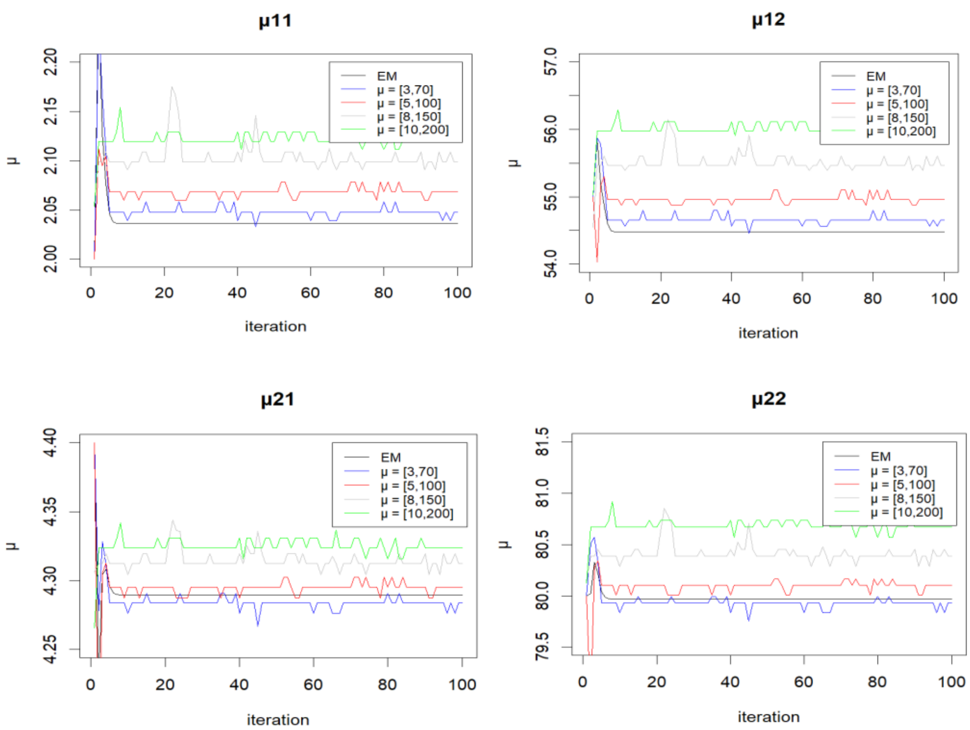
我們設定迭代次數 100 進行計算，最終得到以下結果，並且我們與前面 Gibbs Sampling 抽取結果的平均值進行比較。

	μ_{11}	μ_{12}	μ_{21}	μ_{22}	Σ_{111}	Σ_{112}	Σ_{121}
EM	2.036	54.478	4.289	79.968	0.069	0.435	0.435
Gibbs	2.047	54.651	4.283	79.921	0.091	0.445	0.445
	Σ_{122}	Σ_{211}	Σ_{212}	Σ_{221}	Σ_{222}	α_1	α_2
EM	33.697	0.169	0.941	0.941	36.046	0.356	0.644
Gibbs	34.600	0.1816	0.9366	0.9366	36.196	0.3611	0.6388

EM 演算法與 Gibbs 方法比較

在前面的結果我們可以看出，在這樣的參數設定以及所抽出的樣本進行兩種不同的方法所得出的參數估計是差不多的，但是我們知道，EM 演算法是一種單純依靠樣本推斷母體參數的方法，而 Gibbs 則在某種程度上參考了先驗分配的結果去進行的模擬，因此我們想嘗試在不同的先驗分配參數下進行模擬會不會得出不一樣的結果。首先我們針對μ參數進行調整設定四組不同的參數進行模擬。

1. $\mu_j = [3{,}70] \quad for\ \ j = 1{,}2$
2. $\mu_j = [5{,}100] \quad for\ \ j = 1{,}2$
3. $\mu_j = [8{,}150] \quad for\ \ j = 1{,}2$
4. $\mu_j = [10{,}200] \quad for\ \ j = 1{,}2$



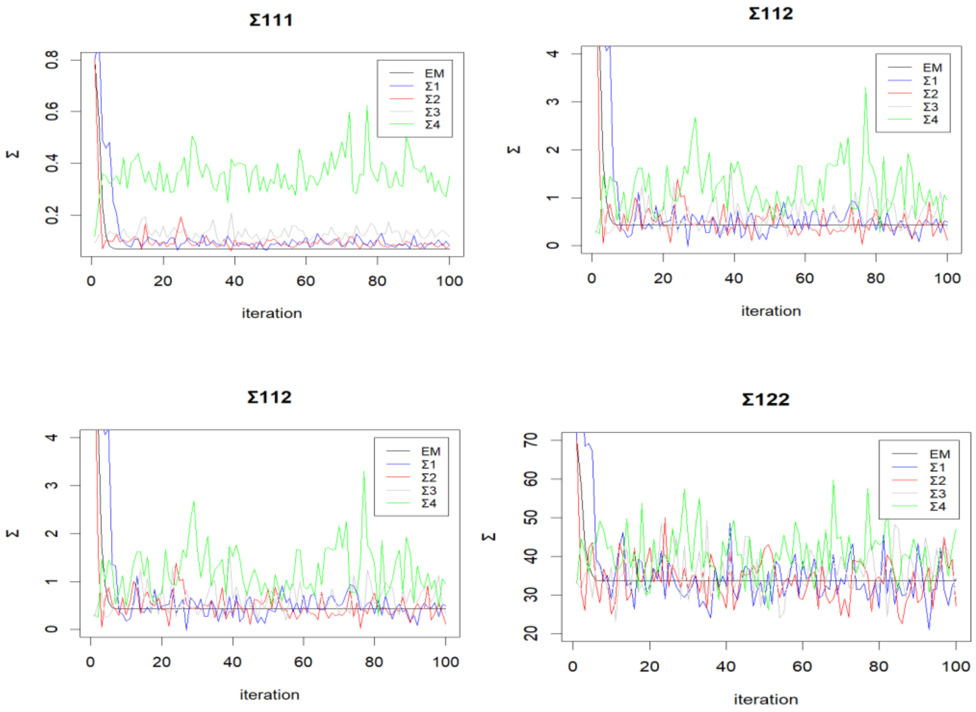
接著我們也設定四組不同的先驗分配φ_j矩陣

1. $\varphi_j = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \quad for \ \ j = 1,2$

2. $\varphi_j = \begin{bmatrix} 5 & 10 \\ 10 & 100 \end{bmatrix} \quad for \ \ j = 1,2$

3. $\varphi_j = \begin{bmatrix} 25 & 50 \\ 50 & 500 \end{bmatrix} \quad for \ \ j = 1,2$

4. $\varphi_j = \begin{bmatrix} 50 & 100 \\ 100 & 1000 \end{bmatrix} \quad for \ \ j = 1,2$



首先，我們可以看到，以 faithful 這組資料來說，使用兩種方法的收斂速度都相當的快，大約在 20 次模擬之內就能夠達成收斂狀態並且大幅往樣本平均數收斂。此外，我們也不難看出我們模擬的結果會一定程度上受到先驗分布的影響，當我們設定的參數與 faithful 這組樣本結果差距愈大的時候，利用 Gibbs 所模擬的馬可夫鏈數值會與 EM 演算法估計的參數有著愈大的差距，而且當我們設定的數值愈大，所抽出的Σ中的各項參數波動幅度更大並有著較大的變異性。

伍、 結論

本次研究利用 Gibbs Sampling 與 EM 演算法針對 faithful 資料集進行模

擬，計算並推測該資料集之母體參數，我們從研究中可以歸納出以下幾個重點。

1. Gibbs 和 EM 兩種方法收斂速度都很快，Gibbs 所生成之馬可夫鏈能在 20 次內達成平穩狀態，而 EM 所計算結果在迭代次數 20 內就可以達到理論中的最佳值。
2. 先驗分配會影響 Gibbs 的模擬結果，因此在使用這兩種方法進行參數估計時是有可能得到差異較大的結果的。
3. 即便設定的先驗分配參數值較不合理，模擬結果也會大幅度地往樣本平均進行收斂。

陸、 參考資料

- EM versus Markov chain Monte Carlo for Estimation of Hidden Markov Models: A Computational Perspective. Tobias Ryden, Bayesian Analysis(2008)
- Comparing Gibbs, EM and SEM for MAP Inference in Mixture Models. Manzil Zaheer, Michael Wick.
- https://www.stat.pitt.edu/sungkyu/course/2221Fall13/lec2.pdf
- https://www.statlect.com/probability-distributions/wishart-distribution
- https://en.wikipedia.org/wiki/Wishart_distribution