#1.

- 用 MC-step (Monte Carlo)取代 E-step.

由於

$$Q(\theta|\theta^{(n)}) = E_{\theta^{(n)}}(\log f(\boldsymbol{Y}|\theta)|\boldsymbol{x}) \quad ,where\ \ \boldsymbol{x} = (y_1, y_2, y_3 + y_4) = (38,34,125)$$

$$= E_{\theta^{(n)}}(\log h(\boldsymbol{Y})|\boldsymbol{x}) + 38\log\left(\frac{1}{2} - \frac{\theta}{2}\right) + 34\log\left(\frac{\theta}{4}\right) + \log\left(\frac{\theta}{4}\right)E_{\theta^{(n)}}(\boldsymbol{Y_3}|\boldsymbol{x})$$

因此我們要用 Monte Carlo 估計$E_{\theta^{(n)}}(\boldsymbol{Y_3}|\boldsymbol{x}) = \int_0^{125} y f_{Y_3|\boldsymbol{x}}(y)dy,$

$\Rightarrow$ 1. 生成$U_1, U_2, \dots, U_k \sim U(0,1)$

2. 令$T_i = 125U_i \ \sim U(0,125)\ \ , i = 1, \dots, k$

3. 估計量 $\hat{I} = \frac{1}{k}\sum_{i=1}^{k}\frac{T_i f_{Y_3|\boldsymbol{x}}(T_i)}{f_T(T_i)}$

- M-step

$\max_\theta Q(\theta|\theta^{(n)}) \Leftrightarrow \max_\theta W(\theta|\theta^{(n)})$,

$$where\ \ W(\theta|\theta^{(n)}) = 38\log\left(\frac{1}{2} - \frac{\theta}{2}\right) + 34\log\left(\frac{\theta}{4}\right) + \log\left(\frac{\theta}{4}\right)\hat{I}$$

這裡用 Newton Method 來求$W(\theta|\theta^{(n)})$最大值

$\Rightarrow\ \ \theta_{m+1} = \theta_m - \frac{w'(\theta|\theta^{(n)})}{w''(\theta|\theta^{(n)})}$,直到滿足$|\theta_{m+1} - \theta_m| < 10^{-5}$.

$\Rightarrow\ \ \theta^{(n+1)} = \arg\max_\theta W(\theta|\theta^{(n)})$

- ***Simulation:***

Initial value $\theta^{(0)} = 0.5$,stopping rule: $|\theta^{(n+1)} - \theta^{(n)}| < 10^{-7}$.

Case1:

在 MC-step 時，取 k=1000，發現很難達成停止條件 $|\theta^{(n+1)} - \theta^{(n)}| < 10^{-7}$.

因為 Monte Carlo 估計$E_{\theta^{(n)}}(\boldsymbol{Y_3}|\boldsymbol{x})$本身就存在誤差，所以考慮透過增加樣本數方式減少誤差。

Case2:

在 MC-step 時，第 n 步取 $k^{(n+1)} = k^{(n)} + 1000$,$k^{(0)} = 0$，

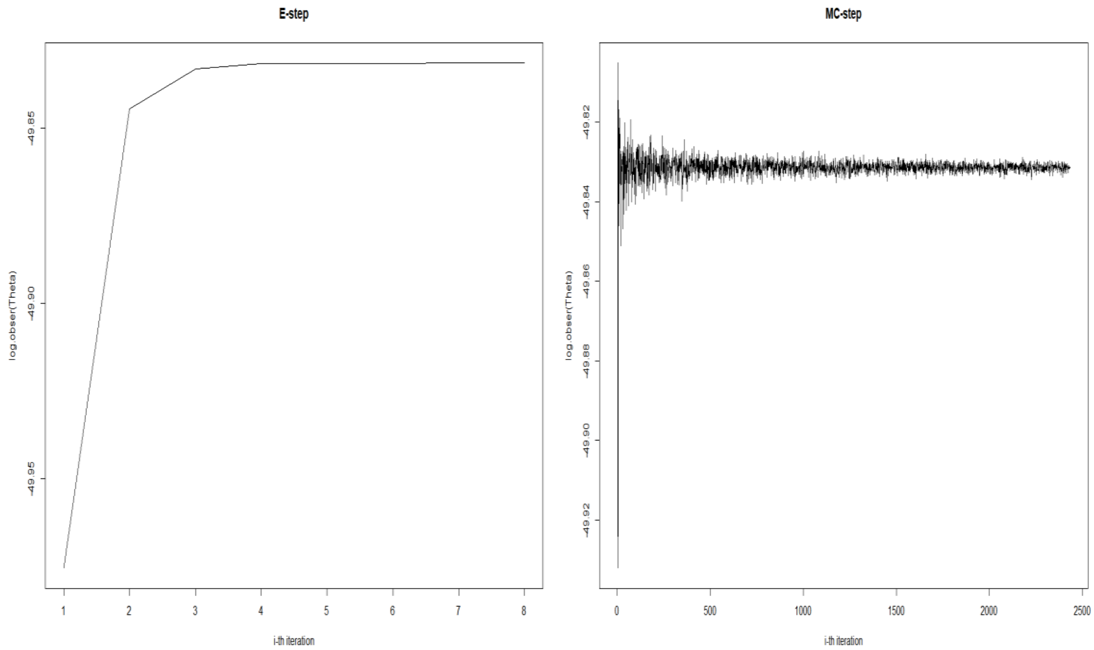
則當 k=2434000，滿足停止條件 $|\theta^{(n+1)} - \theta^{(n)}| < 10^{-7}$, $\hat{\theta}_{MC} = 0.6267990521$.

Note:

Observed $\mathbf{x} = (y_1, y_2, y_3 + y_4) \ \ \Leftrightarrow\ \ \mathbf{x} = (y_1, y_2)$,$n = y_1 + y_2 + y_3 + y_4 = 197$

$\Rightarrow\ \ \mathbf{X} = (Y_1, Y_2) \sim \text{multinomial}(72, \frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2})$

$\Rightarrow$ loglikelihood function: $\log f_{\mathbf{X}}(\mathbf{x}|\theta^{(n)}) = \log f_{Y_1,Y_2}(y_1, y_2|\theta^{(n)})$



左圖為 E-step, y 軸為$\log f_{\boldsymbol{X}}(\mathbf{x}|\theta^{(n)})$, x 軸為第 n 次疊代.
右圖為 MC-step, y 軸為$\log f_{\boldsymbol{X}}(\mathbf{x}|\theta^{(n)})$, x 軸為第 n 次疊代.
觀察到我們用 MC-step 取代 E-step 時，它的值由於 monte carlo method 關係上下跳動，故無法保持理論上的性質: monotonical property for log-likelihood function.
然而，隨著 monte carlo method 中 k 的增加(用愈多的隨機變數去估計)，可以發現它是會收斂的，
$\hat{\theta}_{MC} = 0.6267990521$ (c.f. $\hat{\theta}_E = 0.6268214841$)

※2: $X_1, X_2, \dots, X_n \overset{iid}{\sim} \pi(x \mid \theta)$, where $\pi(x \mid \theta) = \sum_{j=1}^{3} P_j \phi_{\sigma_j}(x-\mu_j)$

$, \sum_{j=1}^{3} P_j = 1 \quad , \quad \phi_\sigma(x-\mu) = \frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right) \sim N(\mu, \sigma^2)$

latent variable

let $Z_{ij} = \begin{cases} 1 & \text{if } x_i \text{ comes from } \phi_{\sigma_j}(x-\mu_j) \\ 0 & \text{o.w.} \end{cases}$. $\quad i = 1, \dots, n \quad j = 1, 2, 3$

s.t. $\quad Z_{n\times 3} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ \vdots & \vdots & \vdots \end{pmatrix}_{n \times 3}$

j.p.d.f of complete data $(\underset{\sim}{x}, \underset{\sim}{z})$

$$f(\underset{\sim}{x}, \underset{\sim}{z} \mid \theta) = \prod_{i=1}^{n}\left(\prod_{j=1}^{3}\left(P_j\, \phi_{\sigma_j}(x_i-\mu_j)\right)^{Z_{ij}}\right) \quad , \theta = (P_1, P_2, P_3, \mu_1, \mu_2, \mu_3, \sigma_1, \sigma_2, \sigma_3)$$

$\Rightarrow$ loglikelihood : $\log f(\underset{\sim}{x}, \underset{\sim}{z} \mid \theta)$

$$= \sum_{i=1}^{n}\left(Z_{i1} \log P_1 \phi_{\sigma_1}(x_i-\mu_1) + Z_{i2}\log P_2 \phi_{\sigma_2}(x_i-\mu_2) + Z_{i3}\log P_3 \phi_{\sigma_3}(x_i-\mu_3)\right)$$

E - step :

note that $E_{\theta^{(m)}}(Z_{ij} \mid \underset{\sim}{x}) = P_{\theta^{(m)}}(Z_{ij} = 1 \mid \underset{\sim}{x})$

$$= \frac{P_j^{(m)} \phi_{\sigma_j^{(m)}}(x_i - \mu_j^{(m)})}{\sum_{j=1}^{3} P_j^{(m)} \phi_{\sigma_j^{(m)}}(x_i - \mu_j^{(m)})}$$

$$= Z_{ij}^{(m)} \quad , \quad i = 1, 2, \dots, n \quad j = 1, 2, 3$$

$\Longrightarrow Q(\theta \mid \theta^{(m)}) = E_{\theta^{(m)}}\left(\log f(\underset{\sim}{x}, \underset{\sim}{z} \mid \theta) \mid \underset{\sim}{x}\right)$

$$= \sum_{i=1}^{n} Z_{i1}^{(m)} \log P_1 \phi_{\sigma_1}(x_i-\mu_1) + \sum_{i=1}^{n} Z_{i2}^{(m)} \log P_2 \phi_{\sigma_2}(x_i-\mu_2) + \sum_{i=1}^{n} Z_{i3}^{(m)} \log P_3 \phi_{\sigma_3}(x_i-\mu_3)$$

where $\phi_{\sigma_j}(x_i - \mu_j) = \frac{1}{\sqrt{2\pi}\sigma_j} e^{-\frac{(x_i-\mu_j)^2}{2\sigma_j^2}} \quad , j = 1, 2, 3$

M-step :

$$\theta^{(m+1)} = \underset{\theta}{\arg\max}\ Q(\theta \mid \theta^{(m)}) \quad , \theta = (p_1, p_2, p_3, \mu_1, \mu_2, \mu_3, \sigma_1, \sigma_2, \sigma_3)$$

$$\Rightarrow \begin{cases} \dfrac{\partial Q(\theta \mid \theta^{(m)})}{\partial p_j} = 0 \\ \dfrac{\partial Q(\theta \mid \theta^{(m)})}{\partial \mu_j} = 0 \\ \dfrac{\partial Q(\theta \mid \theta^{(m)})}{\partial \sigma_j} = 0 \end{cases} \quad , j = 1, 2, 3 .$$

① for μ_j , $j = 1, 2, 3$.

$$\frac{\partial Q(\theta \mid \theta^{(m)})}{\partial \mu_1} = \frac{\partial}{\partial \mu_1} \left(\sum_{i=1}^{n} z_{i1}^{(m)} \cdot \left(- \frac{(x_i - \mu_1)^2}{2\sigma_1^2} \right) \right) = 0$$

$$\Rightarrow \sum_{i=1}^{n} z_{i1}^{(m)} \frac{1}{2\sigma_1^2} \cdot 2 \cdot (x_i - \mu_1) = 0$$

$$\Rightarrow \mu_1^{(m+1)} = \frac{\sum_{i=1}^{n} z_{i1}^{(m)} x_i}{\sum_{i=1}^{n} z_{i1}^{(m)}}$$

Similarly , $\mu_2^{(m+1)} = \dfrac{\sum_{i=1}^{n} z_{i2}^{(m)} x_i}{\sum_{i=1}^{n} z_{i2}^{(m)}}$, $\mu_3^{(m+1)} = \dfrac{\sum_{i=1}^{n} z_{i3}^{(m)} x_i}{\sum_{i=1}^{n} z_{i3}^{(m)}}$

② for σ_j^2 , $j = 1, 2, 3$.

$$\frac{\partial Q(\theta \mid \theta^{(m)})}{\partial \sigma_1} = \frac{\partial}{\partial \sigma_1} \left(\sum_{i=1}^{n} z_{i1}^{(m)} \left(\log \frac{1}{\sigma_1} + - \frac{(x_i - \mu_1)^2}{2\sigma_1^2} \right) \right) = 0$$

$$\Rightarrow \sum_{i=1}^{n} z_{i1}^{(m)} \left(\sigma_1 \cdot \frac{-1}{\sigma_1^2} + \frac{(x_i - \mu_1)^2}{\sigma_1^3} \right) = 0$$

$$\overset{\times \sigma_1^3}{\Rightarrow} \sum_{i=1}^{n} z_{i1}^{(m)} \left(-\sigma_1^2 + (x_i - \mu_1)^2 \right) = 0$$

$$\Rightarrow \sigma_1^{2\,(m+1)} = \frac{\sum_{i=1}^{n} z_{i1}^{(m)} (x_i - \mu_1^{(m+1)})^2}{\sum_{i=1}^{n} z_{i1}^{(m)}}$$

Similarly , $\sigma_2^{2\,(m+1)} = \dfrac{\sum_{i=1}^{n} z_{i2}^{(m)} (x_i - \mu_2^{(m+1)})^2}{\sum_{i=1}^{n} z_{i2}^{(m)}}$

$$\sigma_3^{2\,(m+1)} = \frac{\sum_{i=1}^{n} z_{i3}^{(m)} (x_i - \mu_3^{(m+1)})^2}{\sum_{i=1}^{n} z_{i3}^{(m)}}$$

③ for P_j, $j=1,2,3$.

note: $\max_{P_j} Q(\theta \mid \theta^{(m)})$ with $\sum_{j=1}^{3} P_j = 1$

$\overset{\text{Largrange Multiple}}{\Longleftrightarrow}$ $\max_{P_j} Q(\theta \mid \theta^{(m)}) + \lambda\left(\sum_{j=1}^{3} P_j - 1\right)$ —— (*)

Thus, $\dfrac{\partial (*)}{\partial P_1} = \dfrac{\sum_{i=1}^{n} Z_{i1}^{(m)}}{P_1} + \lambda = 0$

$$\Rightarrow P_1^{(m+1)} = \frac{\sum_{i=1}^{n} Z_{i1}^{(m)}}{-\lambda}$$

Also, have $P_2^{(m+1)} = \dfrac{\sum_{i=1}^{n} Z_{i2}^{(m)}}{-\lambda}$, $P_3^{(m+1)} = \dfrac{\sum_{i=1}^{n} Z_{i3}^{(m)}}{-\lambda}$

since $\sum_{j=1}^{3} P_j^{(m+1)} = 1$ $\Rightarrow$ $\dfrac{\sum_{i=1}^{n}\left(Z_{i1}^{(m)} + Z_{i2}^{(m)} + Z_{i3}^{(m)}\right)}{-\lambda} = 1$

$$\Rightarrow \lambda = -n$$

That is, $P_j^{(m+1)} = \dfrac{\sum_{i=1}^{n} Z_{ij}^{(m)}}{n}$, $j=1,2,3$.

Simulation:

true parameter $\theta_0 = (\underset{\sim}{P_0}, \underset{\sim}{\mu_0}, \underset{\sim}{\sigma_0^2})$

with $\underset{\sim}{P_0} = (P_1, P_2, P_3) = (0.2, 0.3, 0.5)$

$\underset{\sim}{\mu_0} = (\mu_1, \mu_2, \mu_3) = (-3, 0, 2)$

$\underset{\sim}{\sigma_0^2} = (\sigma_1^2, \sigma_2^2, \sigma_3^2) = (0.64, 0.36, 0.25)$

$\Rightarrow$ case 1: $\theta^{(0)} = \theta_0$ (起始值即為真實參數值), $n=2000$

$\Rightarrow \theta^{(1)} = (\underset{\sim}{P^{(1)}}, \underset{\sim}{\mu^{(1)}}, \underset{\sim}{\sigma^{(1)^2}})$

with $\underset{\sim}{P^{(1)}} = (0.2, 0.3, 0.4)$

$\underset{\sim}{\mu^{(1)}} = (0.4077481, 0.4077481, 0.4077481)$

$\underset{\sim}{\sigma^{(1)^2}} = (3.9921538, 3.9921538, 3.9921538)$

$\Rightarrow$ 接著 $\theta^{(2)}, \theta^{(3)}, \theta^{(4)}, \ldots$ 的值皆沒有改變.

case 2： 設 $\theta^{(0)}$ 為其他值，但皆和 case 1 一樣

$\theta^{(m+1)}$ 沒有透過 $\theta^{(m)}$ 更新

即使將樣本數 $n=2000 \xrightarrow[\text{加}]{\text{增}} n=20000$，結果也一樣.

3: Apply Metropolis-Hasting algorithm with $g(\cdot|x) \sim U(x-\varepsilon, x+\varepsilon)$ to simulate data from $\pi(\cdot) \sim N(0,1)$.

(i)

- Proposal distribution:

$$g(y|x) \sim U(x-\varepsilon, x+\varepsilon)$$

$$\text{where} \quad g(y|x) = \begin{cases} \frac{1}{2\varepsilon} & , y \in (x-\varepsilon, x+\varepsilon) \\ 0 & , \quad o.w. \end{cases}$$

Acceptance probability:

$$\alpha(x,y) = \min\{ \frac{\pi(y)g(x|y)}{\pi(x)g(y|x)}, 1 \}$$

- Metropolis-Hasting algorithm:

Step 1: start with $X^{(0)} = x^{(0)}$ s.t. $\pi(x^{(0)}) > 0$.

Step 2: generate $y \sim g(\cdot|x^{(m)}) \sim U(x^{(m)} - \varepsilon, x^{(m)} + \varepsilon)$.

Step 3: compute

$$\alpha(x^{(m)}, y) = \min\{ \frac{\pi(y)g(x^{(m)}|y)}{\pi(x^{(m)})g(y|x^{(m)})}, 1 \}$$

$$= \begin{cases} \exp\left(-\frac{y^2}{2} + \frac{x^{(m)\,2}}{2}\right) & , |x^{(m)}| \le |y| \\ 1 & , o.w. \end{cases}$$

Step 4: generate $u \sim U(0,1)$

If $u \le \alpha(x^{(m)}, y)$, then set $x^{(m+1)} = y$.

Otherwise, set $x^{(m+1)} = x^{(m)}$.

Repeat Step 2~4, we have $x^{(1)}, x^{(2)}, \ldots, x^{(n)} \sim N(0,1)$

- Simulation:

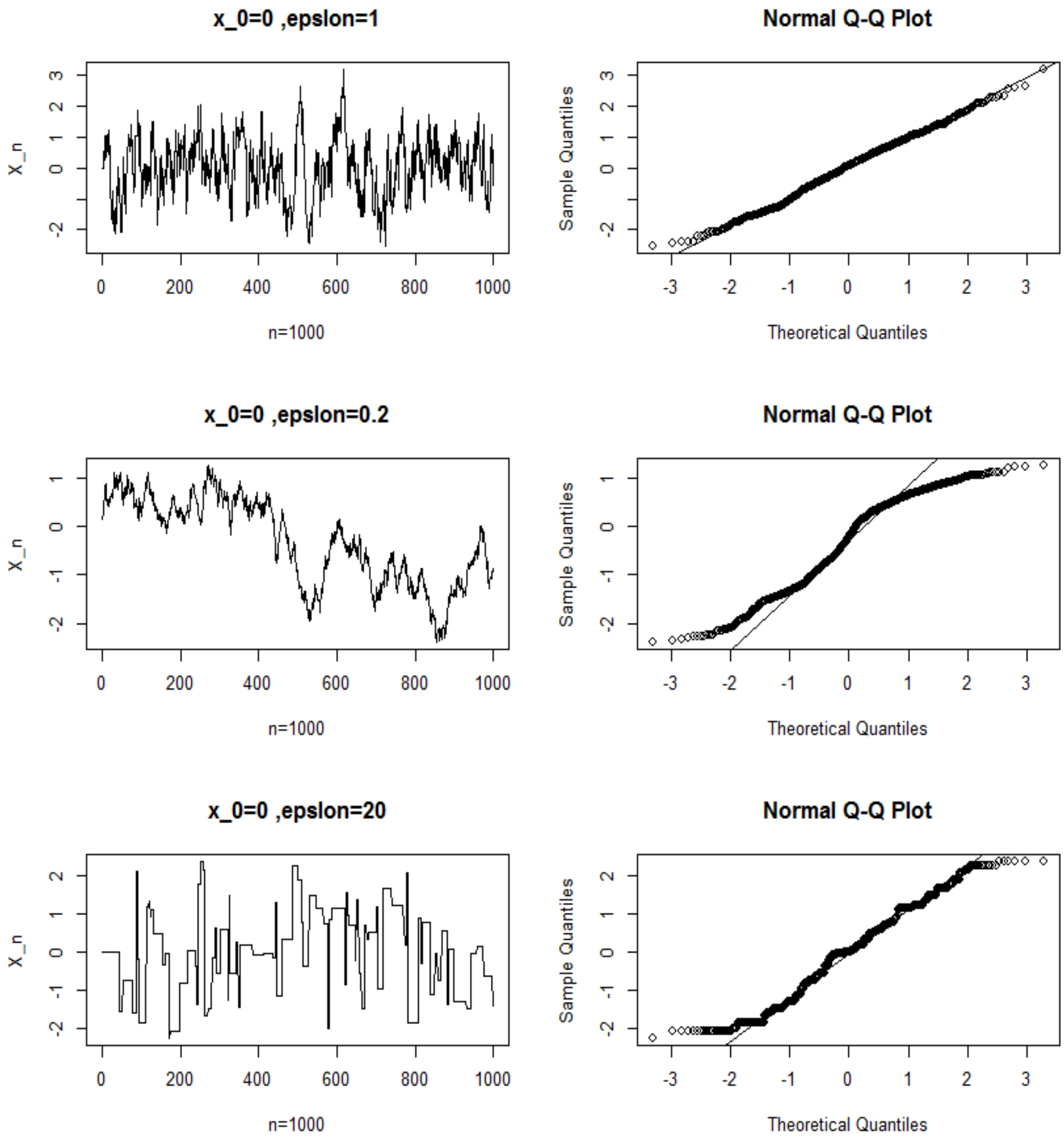
Case 1:

$x^{(0)} = 0$, $n = 1000$, $\varepsilon_1 = 1$, $\varepsilon_2 = 0.2$, $\varepsilon_3 = 20$.

我們發現不同的 ε 將造成生成的 $x^{(1)}, x^{(2)}, \ldots, x^{(1000)}$不一定是$N(0,1)$。 (如下的右圖)

首先，把 $g(y|x^{(m)})$ 想成是給定現在狀態 $x^{(m)}$轉移到狀態 y 的機率，而 $\alpha(x^{(m)}, y)$ 則是接受 "狀態 $x^{(m)}$轉移到狀態 y " 的機率。

因為 ε 會影響著 $y \sim g(\cdot|x^{(m)})$的可能值，所以如果 ε 大，則轉移前後的差異也相對的大；如果 ε 小，轉移前後的差異也相對的小。(如下的左圖)

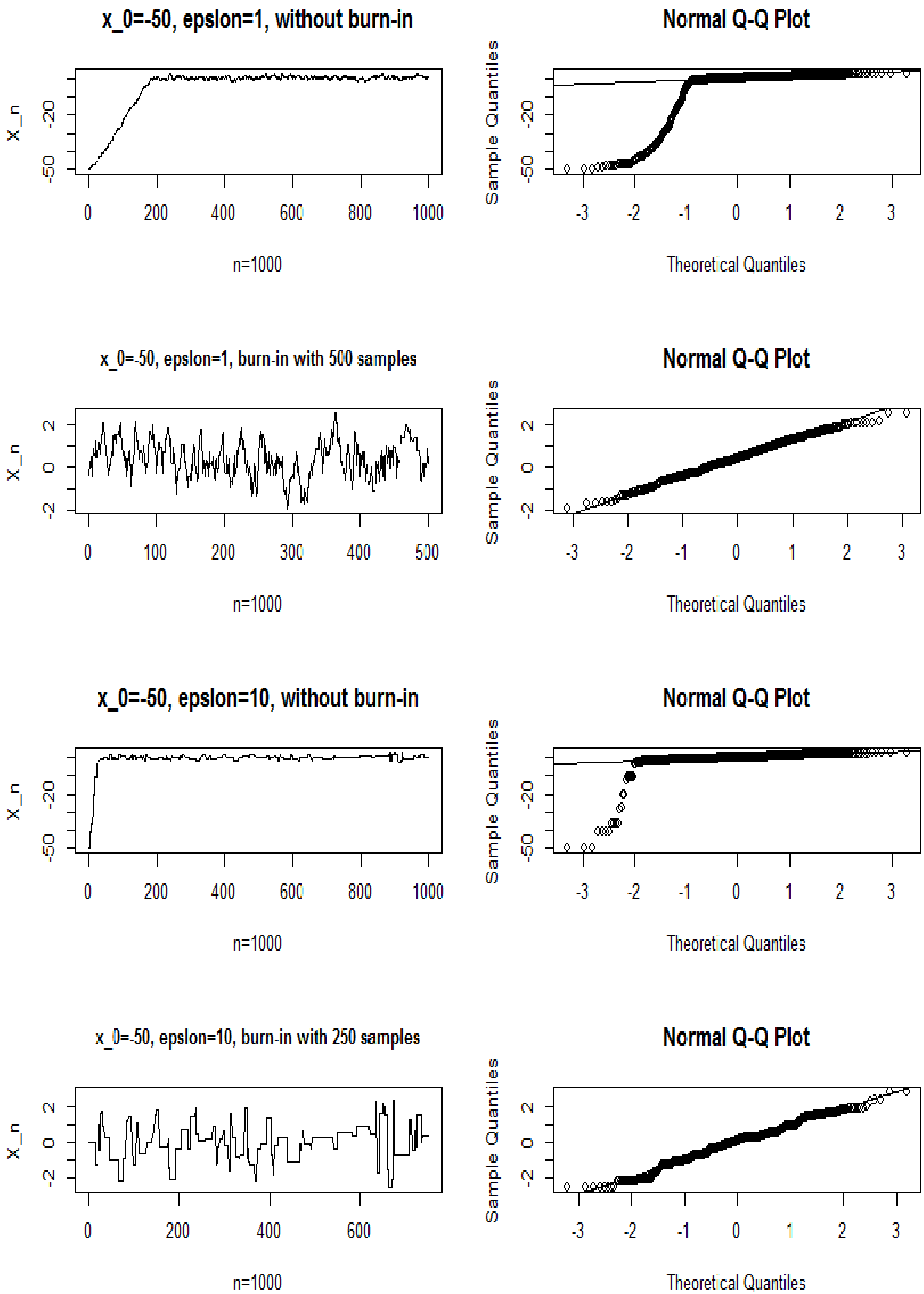


Case 2:

$x^{(0)} = -50$, $n = 1000$, $\varepsilon_1 = 1$, $\varepsilon_2 = 10$.

首先，起始狀態 $x^{(0)} = -50$，如果 $\varepsilon_1 = 1$ (上例中不算太大也不算太小)，則發現它需要一些步數才定進入服從 N(0,1) 的狀態。也就是說，如果我們起始點選的不夠好，此演算法需要所謂的暖機時間，才能夠穩定的生成我們想要的分佈。(如下圖上半部)

如果取 $\varepsilon_2 = 10$ ，雖然能夠減少暖機時間，但是，相對的會遇到 case 1 中 ε 大，轉移前後的差異也大，使得生成出來的不服從 N(0,1)。


x_0=-50, epslon=1, without burn-in
X_n
n=1000
Normal Q-Q Plot
Sample Quantiles
Theoretical Quantiles
x_0=-50, epslon=1, burn-in with 500 samples
X_n
n=1000
Normal Q-Q Plot
Sample Quantiles
Theoretical Quantiles
x_0=-50, epslon=10, without burn-in
X_n
n=1000
Normal Q-Q Plot
Sample Quantiles
Theoretical Quantiles
x_0=-50, epslon=10, burn-in with 250 samples
X_n
n=1000
Normal Q-Q Plot
Sample Quantiles
Theoretical Quantiles

(ii)

To estimate $\mathrm{E}(\exp(\mathrm{Z}^{16}))$.

Note:

By Ergodic theorem,

$\{\mathrm{X_k}\}_{k=1}$ be an irreducible and aperiodic with stationary distribution $\pi(\cdot)$

Then,

$$\frac{1}{\mathrm{n}}\sum_{i=1}^{n}\exp(X_i^{16}) \overset{P}{\to} E_\pi(\exp(Z^{16}))$$

<u>Simulation:</u>

By (i), with $\mathrm{x}^{(0)} = 0$, $\mathrm{n} = 1000$, $\varepsilon = 1$

Since $\frac{1}{\mathrm{n}}\sum_{i=1}^{n}\exp(X_i^{16}) = \infty$,

In this case, we can not estimate $\mathrm{E}(\exp(\mathrm{Z}^{16}))$.