Statistical Computing Homework 3

110024516 邱繼賢

Problem 1.

Define loss function

$$g(\beta) \ = \ -l(\beta) \ + \ (\text{constant}) \ = \ -\sum_{i=1}^{n} y_i \left(\beta_0 + \beta_1 x_i\right) \ + \ \sum_{i=1}^{n} e^{\beta_0+\beta_1 x_i}$$

and gradient function

$$g'(\beta) \ = \ \begin{bmatrix} \frac{\partial}{\partial \beta_0} g \\ \frac{\partial}{\partial \beta_1} g \end{bmatrix} \ = \ \begin{bmatrix} -\sum\limits_{i=1}^{n} y_i + \sum\limits_{i=1}^{n} e^{\beta_0+\beta_1 x_i} \\ -\sum\limits_{i=1}^{n} x_i y_i + \sum\limits_{i=1}^{n} x_i e^{\beta_0+\beta_1 x_i} \end{bmatrix}$$

(a) Newton method

Algorithm :

- Set an initial $\beta^{(0)} \ = \ (3,4)$

- Iteratively approximate the solution by :

$$\beta^{(t+1)} \ = \ \beta^{(t)} \ - \ \left[H\left(\beta^{(t)}\right)\right]^{-1} g'\left(\beta^{(t)}\right) \quad , \quad t = 0,1,2,...,349$$

  where

$$H(\beta) \ = \ \left[\frac{\partial^2}{\partial \beta_0 \partial \beta_1} \ g(\beta)\right] \ = \ \begin{bmatrix} \sum\limits_{i=1}^{n} e^{\beta_0+\beta_1 x_i} & \sum\limits_{i=1}^{n} x_i e^{\beta_0+\beta_1 x_i} \\ \sum\limits_{i=1}^{n} x_i e^{\beta_0+\beta_1 x_i} & \sum\limits_{i=1}^{n} x_i^2 e^{\beta_0+\beta_1 x_i} \end{bmatrix}$$

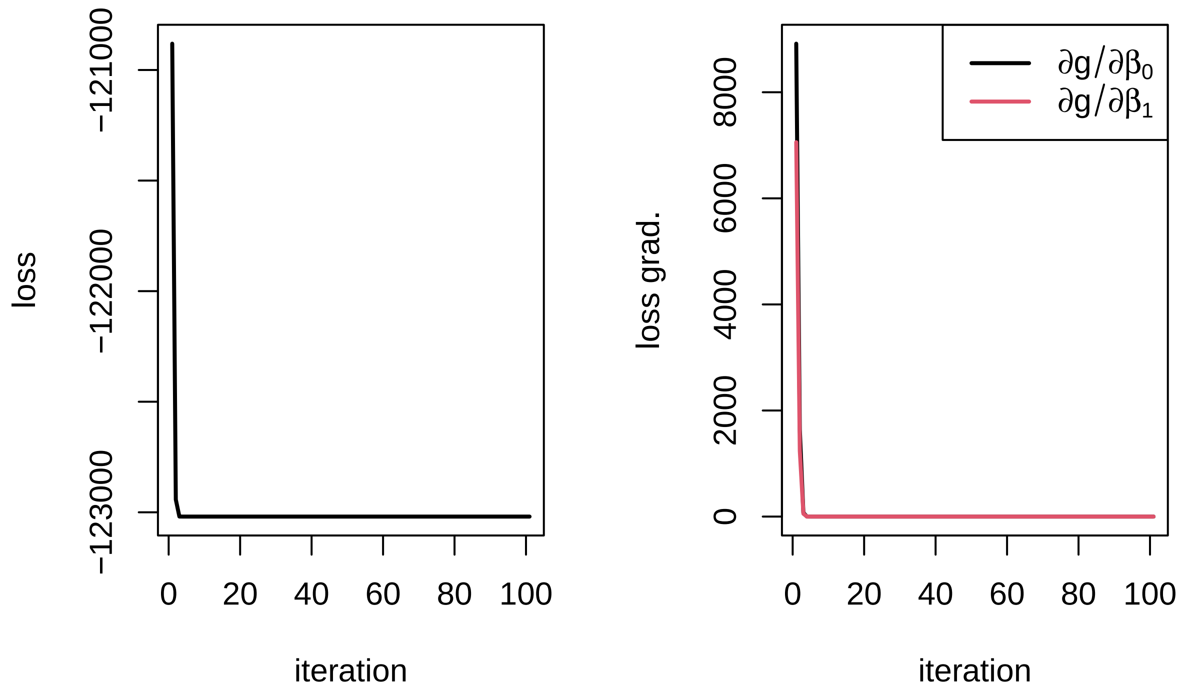
- Check convergence : $||\beta^{(t)} \ - \ \beta^{(t-1)}|| \ \to \ 0$

```
diff(newton$beta)[n.iter,]
```

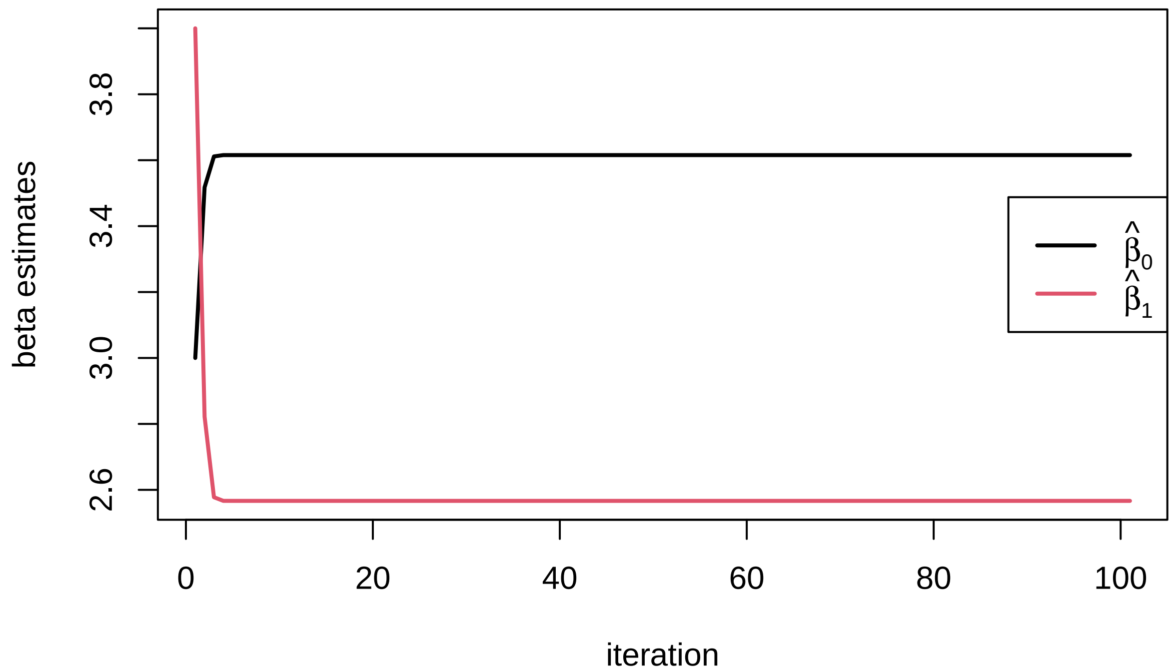
```
## [1] 0 0
```

Then, we can show the result of MLE and their standard error (square root of the diagonal terms of H^{-1})

$\hat{\beta}_0$	$s.e.(\hat{\beta}_0)$	$\hat{\beta}_1$	$s.e.(\hat{\beta}_1)$
3.615397	0.01918	2.566539	0.0311078



We can see that *loss function* is decreasing as iterating, and converges to a small value after about the 10^{th} iteration. The both two *gradient functions* converge to zero after about the 10^{th} iteration.



We can see that (β_0 , β_1) converge to two stable values after about the 10^{th} iteration.

Let's try another method by using funcition *optim()* in order to check whether the results are similar

$\hat{\beta}_0$	$\hat{\beta}_1$	$s.e.(\hat{\beta}_0)$	$s.e.(\hat{\beta}_1)$
3.615395	0.01918	2.566541	0.0311078

We can see that the two results are approximately the same.

(b) Gradient descent

Algorithm :

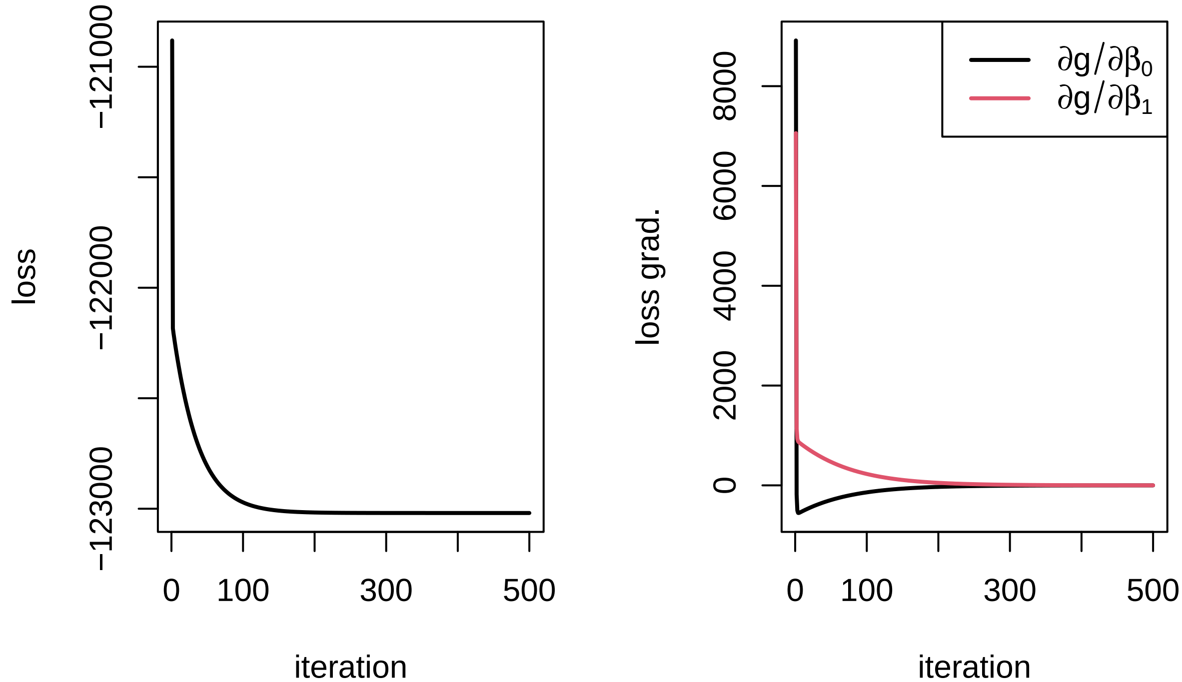
- Set an initial $\beta^{(0)} \ = \ (3, 4)$

- Iteratively approximate the solution by :

$$\beta^{(t+1)} \ = \ \beta^{(t)} \ - \ \alpha_t \ g'\left(\beta^{(t)}\right) \quad , \quad t = 0, 1, 2, ..., 498$$

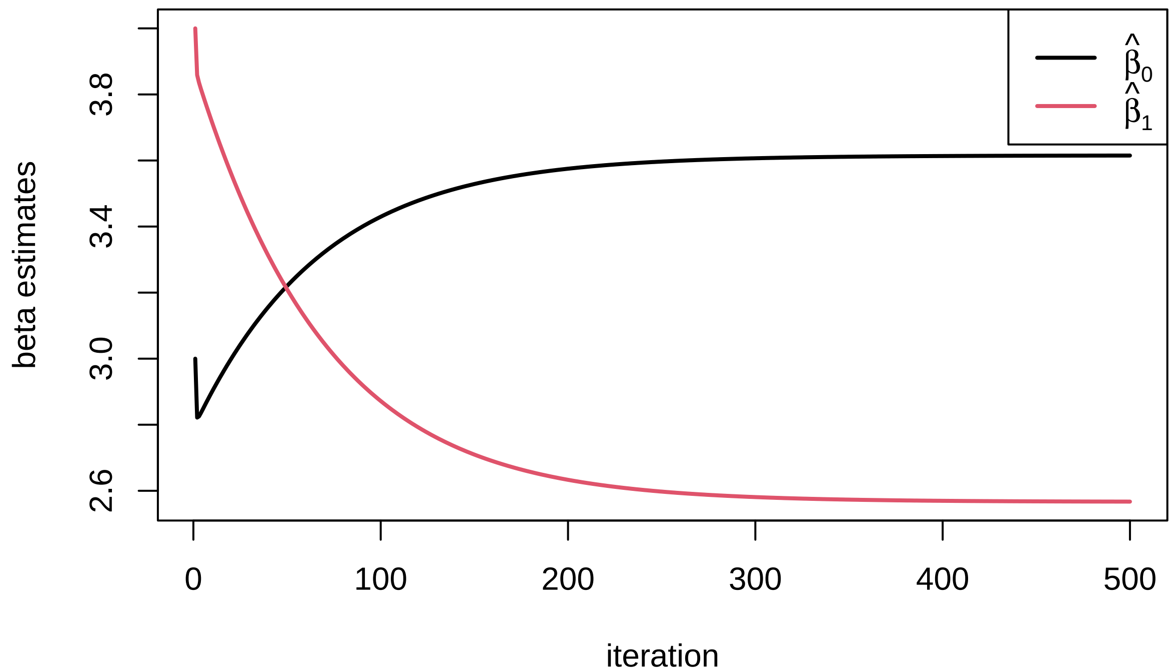
  where $\alpha_t \ = \ 0.00002$

Let's check the convergence of *loss function* and *gradient function*



We can see that *loss function* is decreasing as iterating, and converges to a small value after about the 200^{th} iteration. The both two *gradient functions* converge to zero after about the 300^{th} iteration.

And then check the convergence of β



We can see that $(\beta_0\ ,\ \beta_1)$ converge to two stable values after about the 400^{th} iteration.
Compute the difference between $\left(\beta_0^{(498)}\ ,\ \beta_1^{(498)}\right)$ and $\left(\beta_0^{(499)}\ ,\ \beta_1^{(499)}\right)$

```
diff(beta.hat)[n.iter-1,]
```

```
## [1]  6.212819e-06 -1.029887e-05
```

They are both really closed to zeros, so we can say that β is already converge.

$\hat{\beta}_0$	$\hat{\beta}_1$
3.614996	2.567204

Problem 2.

Compute log-likelihood function

$$\begin{aligned} l(\alpha\ ,\ \beta\ ;\ x\ ,\ y) &= -\frac{n}{2}log(2\pi)\ -\ \frac{1}{2}\sum_{i=1}^{n} log\left(\sigma_i^2\right)\ -\ \frac{1}{2}\sum_{i=1}^{n}\frac{\left(y_i-\mu_i\right)^2}{\sigma_i^2} \\ &= -\frac{n}{2}log(2\pi)\ -\ \frac{1}{2}\sum_{i=1}^{n}\left(\alpha_0+\alpha_1 x_i\right)\ -\ \frac{1}{2}\sum_{i=1}^{n}\left(y_i-\beta_0-\beta_1 x_i\right)^2\ e^{-\alpha_0-\alpha_1 x_i} \end{aligned}$$

Define *loss function*

$$g(\alpha\ ,\ \beta)\ =\ -2l(\alpha\ ,\ \beta)\ +\ (\text{constant})\ =\ \sum_{i=1}^{n}(\alpha_0+\alpha_1 x_i)\ +\ \sum_{i=1}^{n}(y_i-\beta_0-\beta_1 x_i)^2\ e^{-\alpha_0-\alpha_1 x_i}$$

Algorithm :

(1) Set an initial $\alpha^{(0)}\ =\ \left(\alpha_0^{(0)}\ ,\ \alpha_1^{(0)}\right)\ =\ (1,2)$

(2) Iteratively compute

$$\beta^{(t)}\ =\ arg\ min_{\beta}\ g\left(\alpha^{(t)}\ ,\ \beta\right)$$
$$\alpha^{(t+1)}\ =\ arg\ min_{\alpha}\ g\left(\alpha\ ,\ \beta^{(t)}\right)$$

for $t=1,2,...,99$

(3) Checking whether

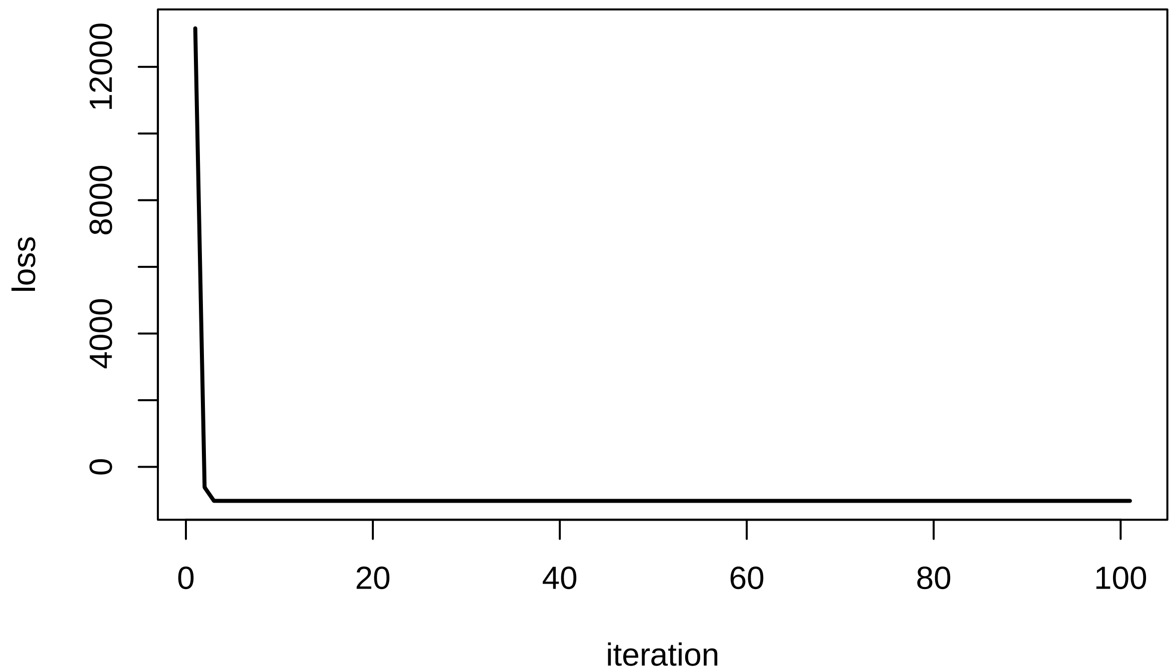
$$||\alpha^{(t+1)}\ -\ \alpha^{(t)}||\ \to\ 0$$
$$||\beta^{(t+1)}\ -\ \beta^{(t)}||\ \to\ 0$$

```
diff(cd_method$alpha)[n.iter,]
```

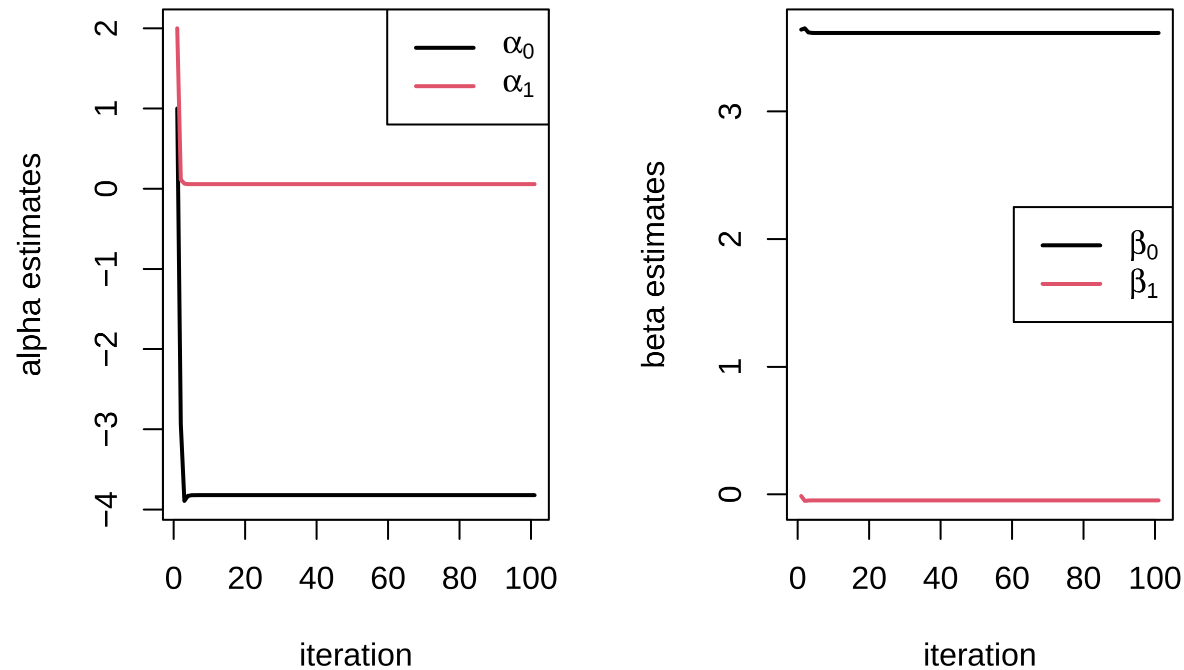
```
## [1] 0 0
```

```
diff(cd_method$beta)[n.iter,]
```

```
## [1] 1.776357e-15 2.983724e-16
```



We can see that *loss function* is decreasing as iterating, and converges to a small value after about the 10^{th} iteration.



We can see that both $(\alpha_0\ ,\ \alpha_1)$ and $(\beta_0\ ,\ \beta_1)$ converge to two stable values after about the 20^{th} iteration.

The MLE of $(\alpha\ ,\ \beta)$ are show as below

$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$
-3.821524	0.0572641	3.614922	-0.0475182

Problem 3.

Define

$$L_p(\beta\ ,\ \delta\ ,\ \lambda)\ =\ \frac{1}{2}||y-\beta||^2\ +\ \tau\ |\delta|\ +\ \lambda'(D\beta\ -\ \delta)\ +\ \frac{\rho}{2}||D\beta\ -\ \delta||^2$$

<u>*Algorithm*</u> :

(1) $\beta^{(t+1)}\ =\ \arg\ min\ L_p\left(\beta\ ,\ \delta^{(t)}\ ,\ \lambda^{(t)}\right)$

$$\Rightarrow\ \beta\ \leftarrow\ (I+\rho D'D)^{-1}\left[y+\rho D'(\delta-\lambda/\rho)\right]$$

(2) $\delta^{(t+1)}\ =\ \arg\ min\ L_p\left(\beta^{(t+1)}\ ,\ \delta\ ,\ \lambda^{(t)}\right)$

$$\Rightarrow\ \delta\ \leftarrow\ S_{\frac{\tau}{\rho}}\left(D\beta\ +\ \lambda/\rho\right)$$

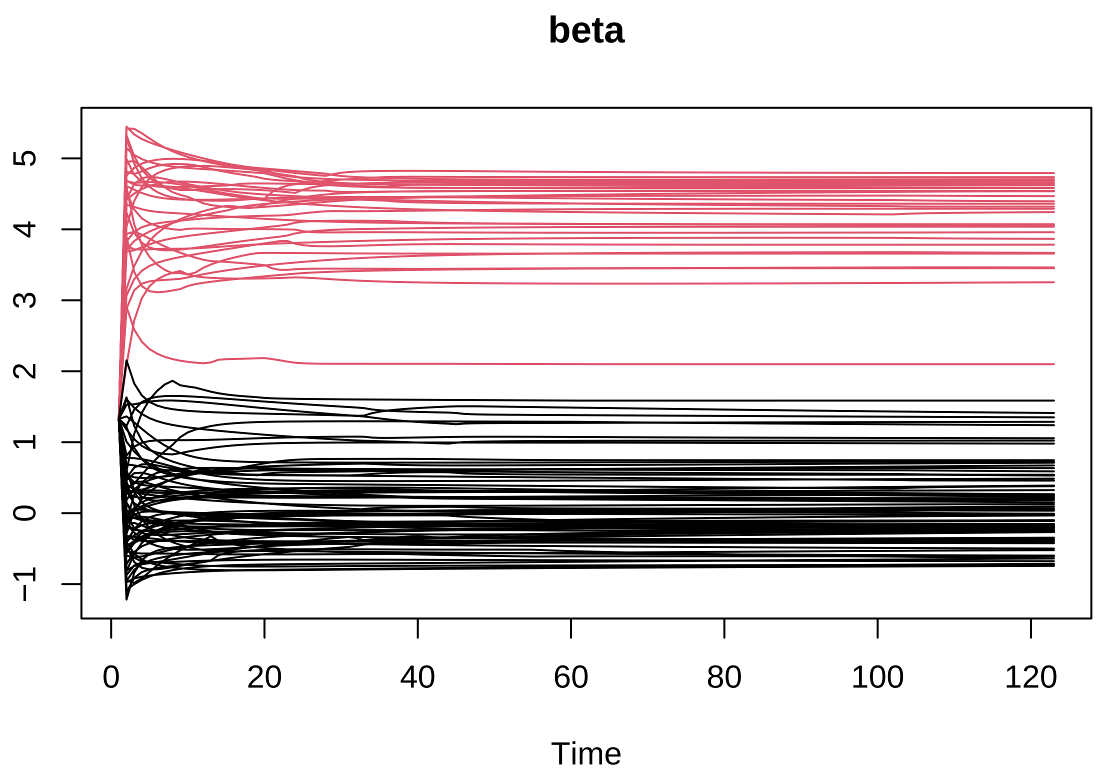
(3) $\lambda^{(t+1)} \leftarrow \lambda^{(t)} + \rho(D\beta - \delta)$

Take $\tau = 1$, $\rho = 0.1$ for example :

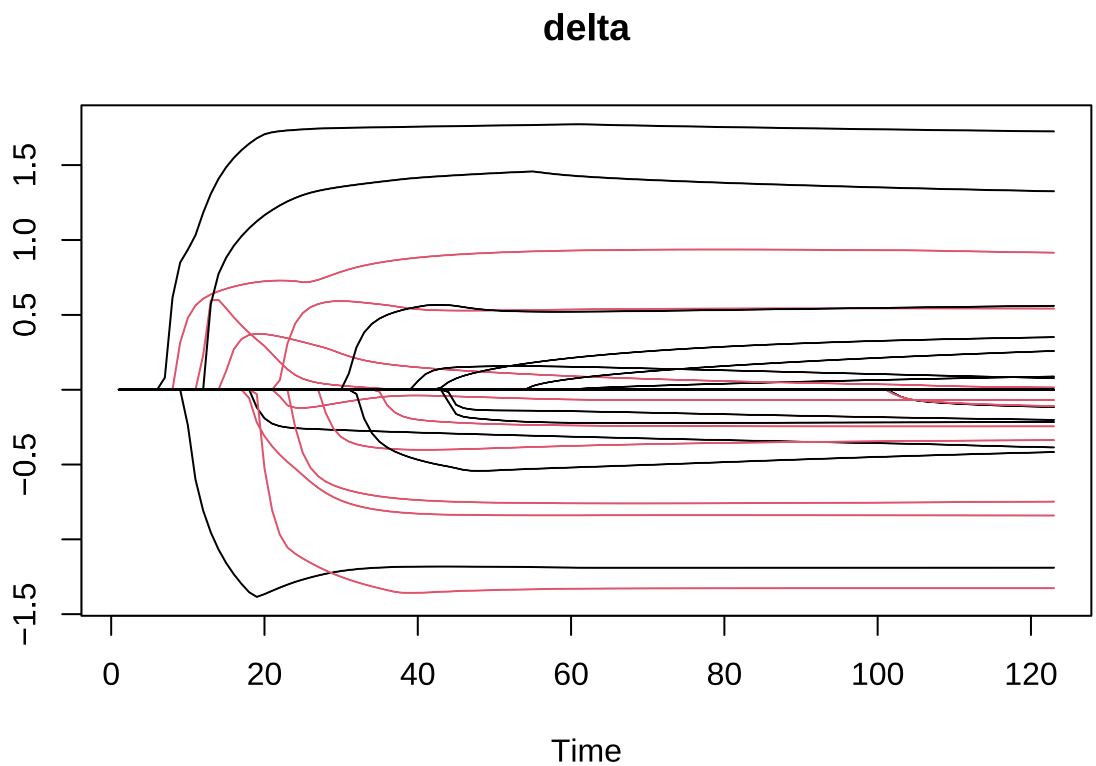
```
tmp = admm.3(data.3$yt, tau=1, rho=0.1, eps.conv = 0.001, iter.max = 5000)
tmp$count
```

```
## [1] 122
```

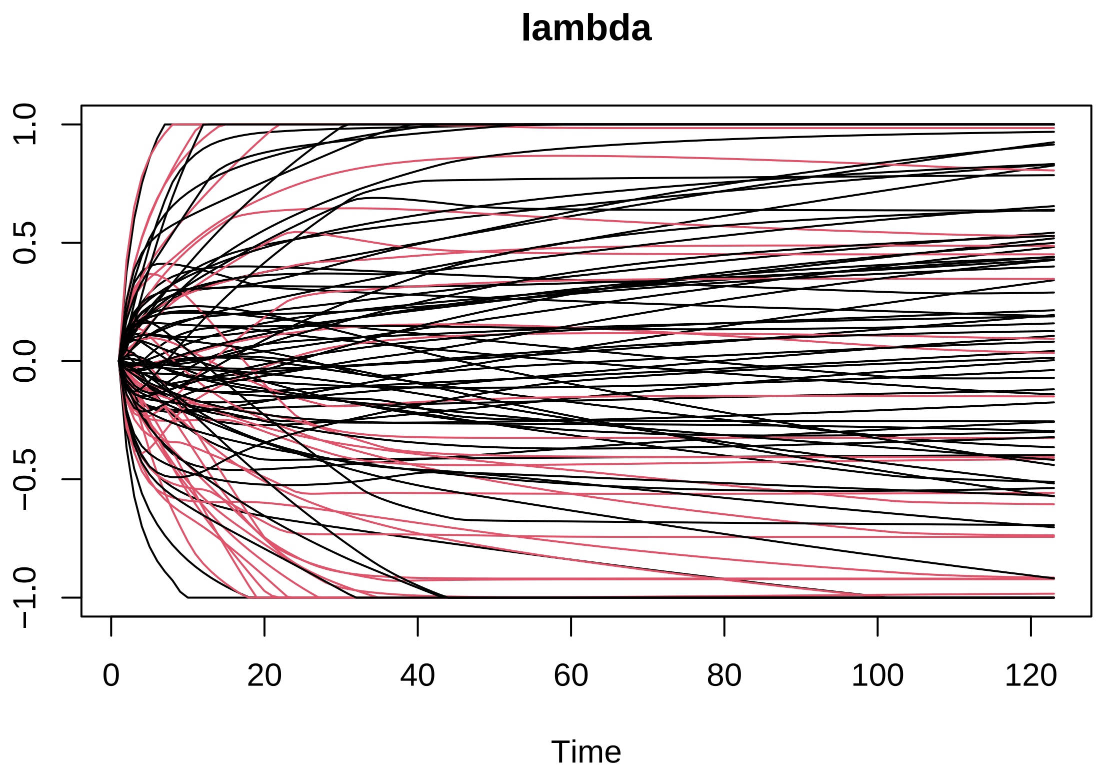
Iterate for 122 times to comverge



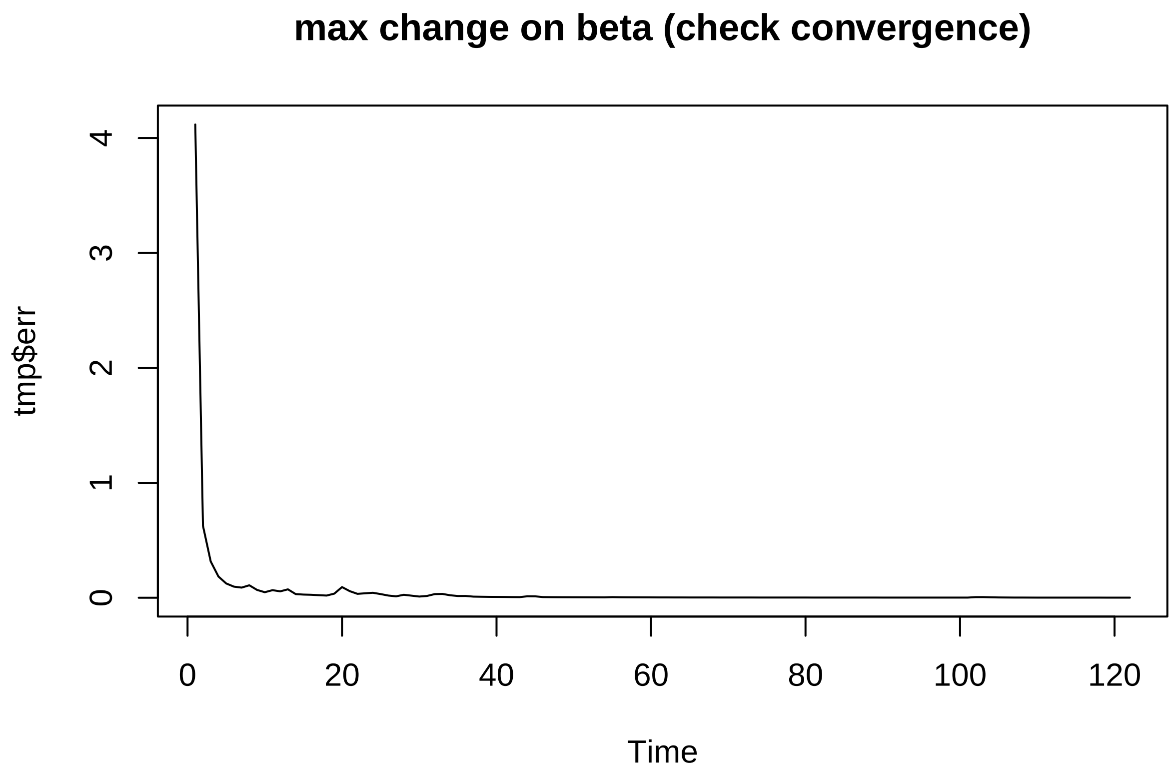
β converges to stable values after about 60th iteration.



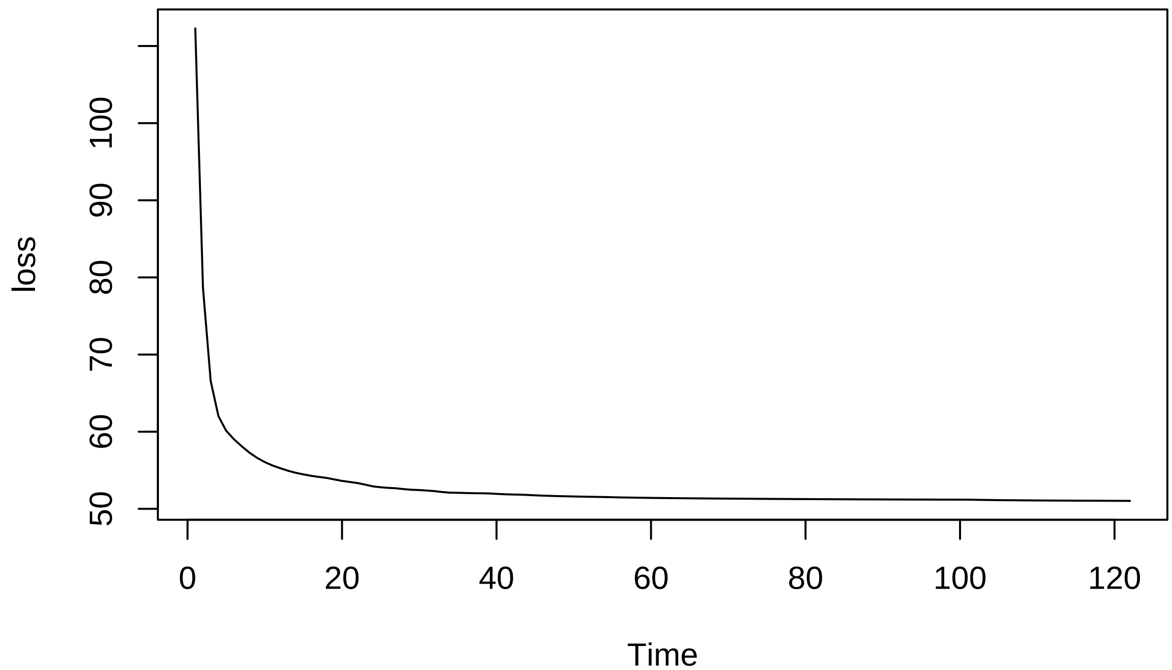
δ converges to stables values after about 110th iterations.



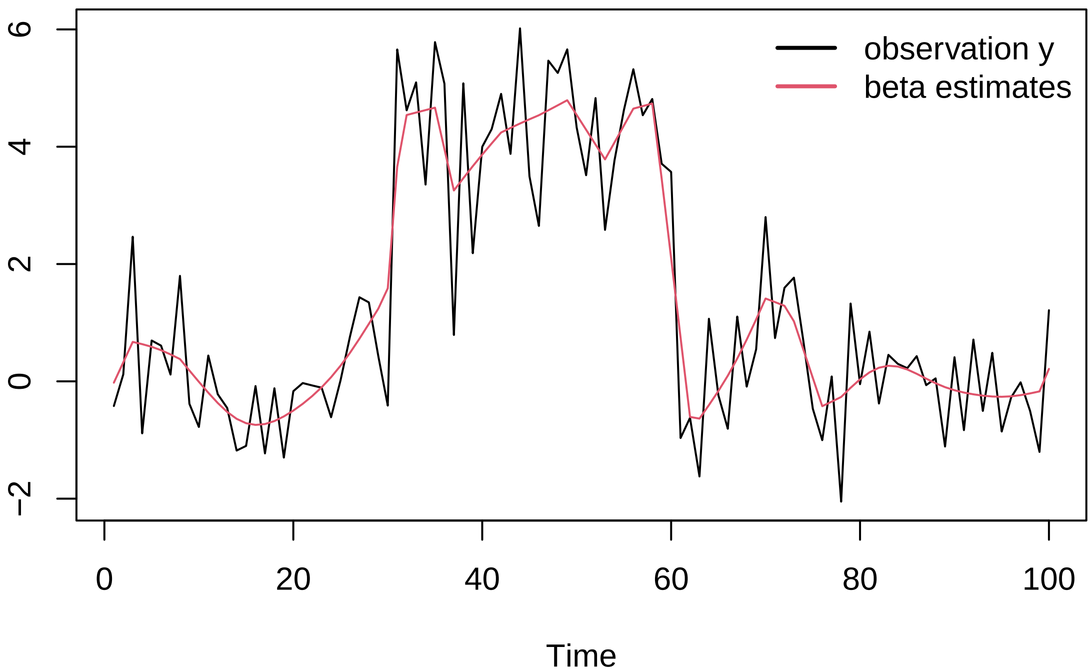
λ converges to stable values after about 100th iteration.



Error converge to 0 after about 60th iteration.



Loss function discreases to a stable value after about 80th iteration.



Our estimated β fit the observations well.