STA 243 Assignment 1

Chen Zihao 915490404

(Chen Zihao:100%) 1. The Cauchy.....

(a)

the likelihood function is that

$$L(\theta|x) = f_n(x|\theta) = \Pi_{i=1}^{n} \frac{1}{\pi[1 + (x - \theta)^2]}$$

then we can get:

1) the log likelihood function is that

$$l(\theta) = -nlog\pi - \sum_{i=1}^{n} \log[1 + (\theta - x_i)^2]$$

2) take the first derivative and we can get

$$l'(\theta) = -\sum_{i=1}^{n} \frac{2(\theta - x_i)}{1 + (\theta - x_i)^2}$$

3) take the second derivative and we can get

$$l''(\theta) = -2\sum_{i=1}^{n} \frac{1 + (\theta - x_i)^2 - 2(\theta - x_i)(\theta - x_i)}{[1 + (\theta - x_i)^2]^2} = -2\sum_{i=1}^{n} \frac{1 - (\theta - x_i)^2}{[1 + (\theta - x_i)^2]^2}$$

(b) Show that the Fisher information is $I(\theta) = \frac{n}{2}$

$$\begin{aligned}
I(\theta) = -E(l''(\theta)) &= 2nE(\frac{1 - (\theta - x_1)^2}{[1 + (\theta - x_1)^2]^2}) \\
&= \frac{2n}{\pi} \int_{-\infty}^{\infty} \frac{1}{[1 + (x - \theta)^2]^2} - \frac{2(x - \theta)^2}{[1 + (x - \theta)^2]^3} dx \\
&= \frac{2n}{\pi} \int_{-\infty}^{\infty} \frac{1}{[1 + x^2]^2} - \frac{2x^2}{[1 + x^2]^3} dx \\
&= \frac{2n}{\pi} \int_{-\infty}^{\infty} \frac{1}{[1 + x^2]^2} - \frac{2}{[1 + x^2]^2} + \frac{2}{[1 + x^2]^3} dx \\
&= \frac{2n}{\pi} \int_{-\infty}^{\infty} \frac{-1}{[1 + x^2]^2} + \frac{2}{[1 + x^2]^3} dx
\end{aligned}$$

let

$$F_k = \int_{-\infty}^{\infty} \frac{1}{[1 + x^2]^k} dx$$

we can get

$$
\begin{aligned}
F_k &= \int_{-\infty}^{\infty} \frac{1}{[1+x^2]^k} dx \\
&= \int_{-\infty}^{\infty} \frac{1+x^2}{[1+x^2]^{k+1}} dx \\
&= F_{k+1} + \int_{-\infty}^{\infty} \frac{2kx}{[1+x^2]^{k+1}} \frac{x}{2k} dx \\
&= F_{k+1} + \frac{1}{2k} \int_{-\infty}^{\infty} \frac{1}{[1+x^2]^k} dx = F_{k+1} + \frac{1}{2k} F_k
\end{aligned}
$$

By an integration by parts.

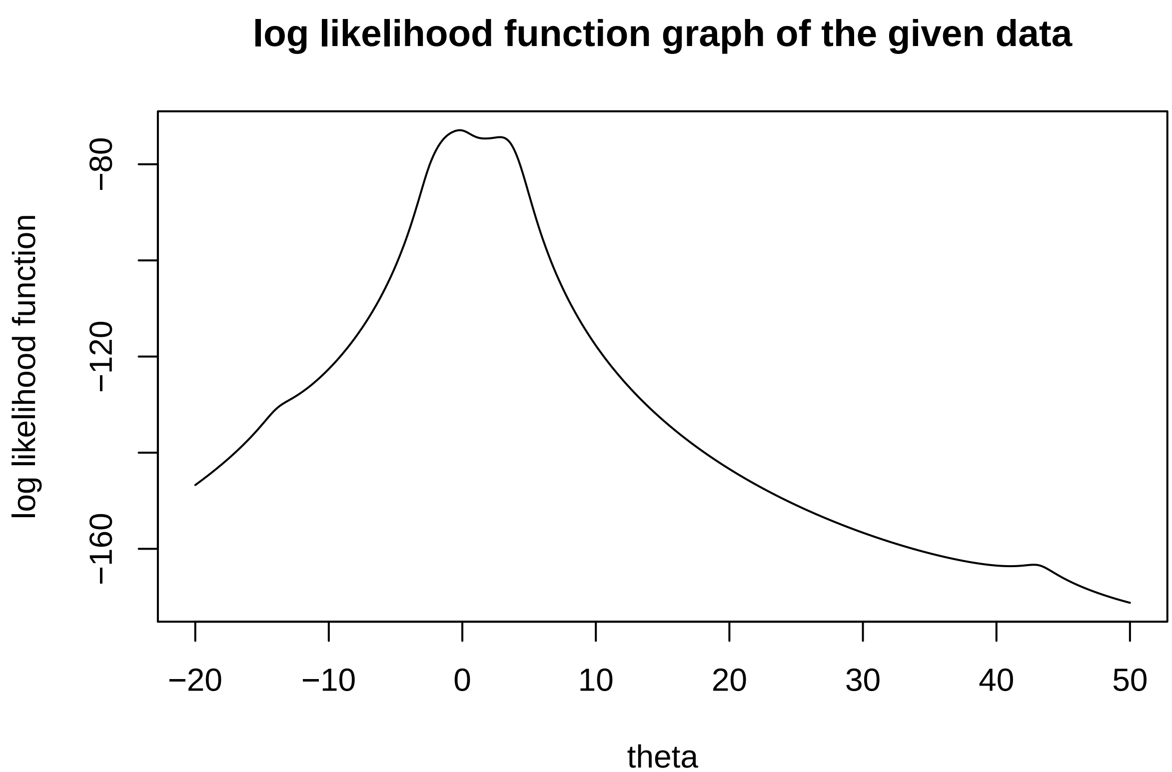
Hence we get

$F_1 = \pi$ and $F_{k+1} = \frac{2k-1}{2k} F_k, k > 1$

so that $F_1 = \pi, F_2 = \pi/2, F_3 = 3\pi/8$

$$
I(\theta) = \frac{2n}{\pi}[-F_2 + 2F_3] = \frac{n}{2}
$$

(c) Use the following data, graph the log likelihood function. . . .



It shows that the maximum is near 0,there is a local maximum and local minimum at little bit larger than 0, there seems a station point near 40.

(d) Find the MLE for θ using the Newton-Raphson method. . . .

with the given initial point, some points did not converge. The following table shows the results:

```
##                   [,1]        [,2]        [,3]       [,4]       [,5] [,6]
## start              -11  -1.0000000   0.0000000   1.400000   4.100000  4.8
## theta value         NA  -0.1922865  -0.1922866   1.713587   2.817473   NA
## function value      NA -72.9158196 -72.9158196 -74.642016 -74.360461   NA
##                        [,7] [,8]       [,9]
## start               7.00000    8   38.00000
## theta value        41.04085   NA   42.79538
## function value -163.60772   NA -163.31289
```

(e) First use Fisher scoring to find the MLE for . . . .

First use Fisher scoring to find the MLE for θ

the new initial point becomes

```
## [1] -0.2367353 -0.1939695 -0.1900633 -0.1906264  2.8203144  2.8208646
## [7]  2.8207965  2.8206372 36.2247409
```

Then refine my estimate using Newton-Raphson.

```
##                           [,1]        [,2]        [,3]        [,4]
## start              -11.0000000  -1.0000000   0.0000000   1.4000000
## new initialpoint    -0.2367353  -0.1939695  -0.1900633  -0.1906264
## theta value         -0.1922865  -0.1922866  -0.1922866  -0.1922866
## function value     -72.9158196 -72.9158196 -72.9158196 -72.9158196
##                          [,5]        [,6]       [,7]       [,8]      [,9]
## start                4.100000    4.800000   7.000000   8.000000 38.00000
## new initialpoint     2.820314    2.820865   2.820796   2.820637 36.22474
## theta value          2.817472    2.817472   2.817472   2.817472       NA
## function value     -74.360461 -74.360461 -74.360461 -74.360461       NA
```

Compare with the previous one, Most of them is now stable although some of them are still trapped in the local maximum/minimum point. Those can not converage in (d) can converage now.

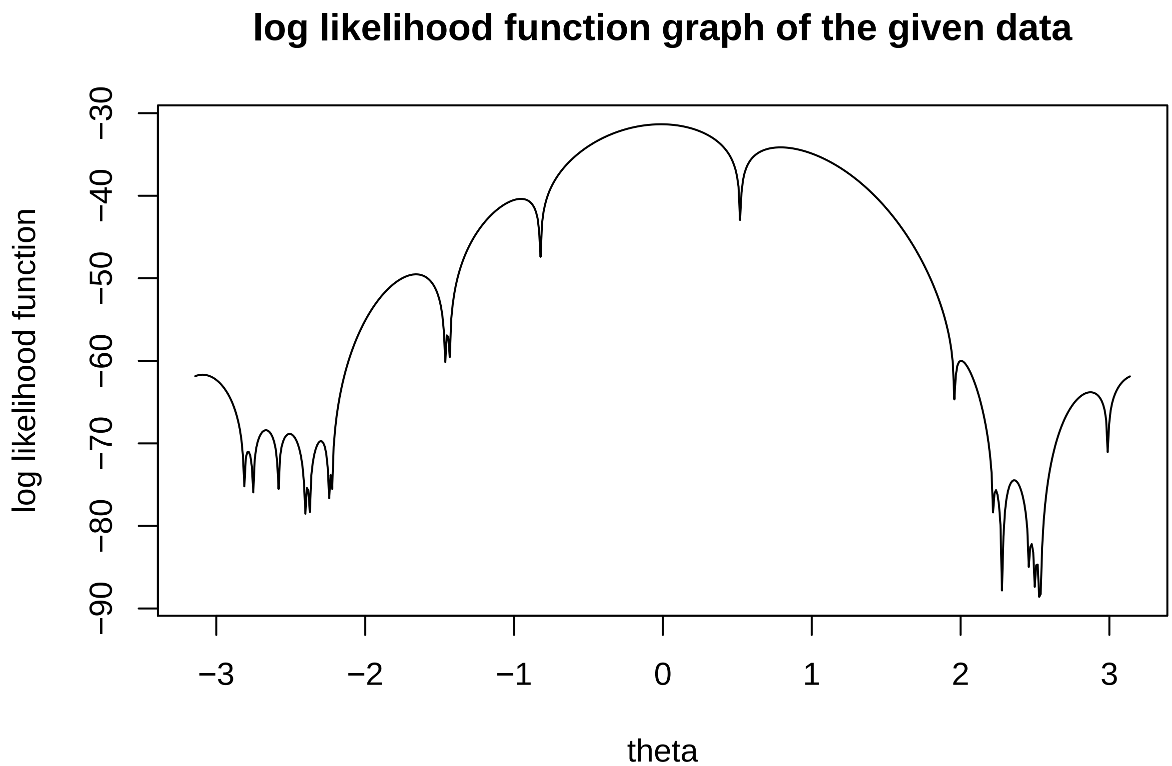
(Chen Zihao:100%) 2.Consider the following . . . ..

(a) Graph the log likelihood function.

$$L(\theta|x) = \frac{\prod_{i=1}^{n}[[1-\cos(x_i-\theta)]}{2^n\pi^n}$$

$$l(\theta) = \log L(\theta|x) = \sum_{i=1}^{n}\log(1-\cos(x_i-\theta)) - n\log(2\pi)$$

The graph is like below:



(b) Find the method of moments ....

$$
\begin{aligned}
E(x) &= \int_0^{2\pi} \frac{x - x\cos(x-\theta)}{2\pi} dx \\
&= \frac{1}{2\pi}\left(\int_0^{2\pi} x dx - \int_0^{2\pi} x\cos(x-\theta) dx\right) \\
&= \frac{1}{2\pi}\left(2\pi^2 - \int_{-\theta}^{2\pi-\theta} (x+\theta)\cos(x) dx\right) \\
&= \pi - \frac{1}{2\pi}\left(\theta\int_{-\theta}^{2\pi-\theta} \cos(x) dx + \int_{-\theta}^{2\pi-\theta} x\cos(x) dx\right) \\
&= \pi - \frac{1}{2\pi}\left(\int_{-\theta}^{2\pi-\theta} x\cos(x) dx\right) \\
&= \pi + \sin(\theta)
\end{aligned}
$$

so that we get $\hat{\theta}_{moment} = \arcsin(\bar{x} - \pi)$

plug the given data in we can get MOM=

```
## [1] 0.05844061
```

(c) Find the MLE for θ using Newton-Raphson with MOM.

$$l(\theta) = \sum_{i=1}^{n} \log(1 - \cos(x_i - \theta)) - n\log(2\pi)$$
$$l'(\theta) = \sum_{i=1}^{n} \frac{-\sin(x_i - \theta)}{1 - \cos(x_i - \theta)}$$
$$l''(\theta) = \sum_{i=1}^{n} \frac{-1}{1 - \cos(x_i - \theta)}$$

After Newton-Rapson we can get

```
##                    [,1]
## start        0.05844061
## theta_hat   -0.01197200
## func_value -31.34291259
```

(d) What solutions do you find when you start at -2.7 and 2.7

```
##                 [,1]       [,2]
## start       -2.70000   2.700000
## theta_hat   -2.66670   2.873095
## func_value -68.39919 -63.805650
```

It will converge to the local stationary point but not the global maximum point (as shown in the graph).

(e) Repeat the above using 200 equally-spaced...

Here is a table about the the $\hat{\theta}$ and its frequency.

```
##                  theta Freq
## 1    -3.09309172991194   11
## 2     -2.7861667516046    2
## 3    -2.66669992610095    5
## 4     -2.5076132262462    6
## 5    -2.38820049182006    1
## 6     -2.2972562196393    4
## 7      -2.232167292072    1
## 8    -1.65828322990256   24
## 9    -1.44747876505045    1
## 10   -0.95333632773287   16
## 11  -0.953336327732869    3
## 12 -0.0119720022874401   20
## 13   -0.01197200228744   12
## 14 -0.0119720022874399    8
## 15 -0.0119720022874398    2
## 16   0.790601310409971    2
## 17   0.790601310409972   44
## 18    2.00364488877485    8
## 19    2.23621938723281    2
## 20    2.36071817373184    6
## 21    2.47537362875389    1
## 22    2.51359317779379    1
```

```
## 23    2.87309451424508   15
## 24    3.19009357726764    5
```

(Chen Zihao:100%) 3.In chemical kinetics the Michaelis-Menten model…

(a) A quick way for finding rough estimates…

Using lm function in R, we can easily get the β_0 and β_1 with least squares

```
## [1] "beta0 is 0.00510718164158581 beta1 is 0.000247220961099642"
```

so that we can get $\hat{\theta}_1 = 1/\beta_0$ and $\hat{\theta}_2 = \beta_1/\beta_0$ as follows:

```
## [1] "theta1_hat is 195.80270884775 theta2_hat is 0.0484065338672541"
```

(b)

$$
\begin{aligned}
g(\theta_1, \theta_2) &= \sum_{i=1}^{n}(y_i - \frac{\theta_1 x_i}{x_i + \theta_2})^2 \\
\frac{\partial g}{\partial \theta_1} &= 2\sum_{i=1}^{n}(y_i - \frac{\theta_1 x_i}{x_i + \theta_2})(-\frac{x_i}{x_i + \theta_2}) \\
\frac{\partial g}{\partial \theta_2} &= 2\sum_{i=1}^{n}(y_i - \frac{\theta_1 x_i}{x_i + \theta_2})(\frac{\theta_1 x_i}{(x_i + \theta_2)^2}) \\
\frac{\partial^2 g}{(\partial \theta_1)^2} &= 2\sum_{i=1}^{n}(\frac{x_i}{x_i + \theta_2})^2 \\
\frac{\partial^2 g}{\partial \theta_1 \partial \theta_2} &= 2\sum_{i=1}^{n}(\frac{y_i x_i}{(x_i + \theta_2)^2} - \frac{2\theta_1 x_i^2}{(x_i + \theta_2)^3}) \\
\frac{\partial^2 g}{(\partial \theta_2)^2} &= 2\sum_{i=1}^{n}(\frac{-2y_i\theta_1 x_i}{(x_i + \theta_2)^3} + \frac{3\theta_1^2 x_i^2}{(x_i + \theta_2)^4})
\end{aligned}
$$

and then we can get the first derivative vector and the hessian matrix and do a Newton-Raphson algorithm. The code is in the code file, here is the results.

```
##        theta1     theta2
## [1,] 212.6837 0.06412128
```

(c) Repeat (b) with the steepest descent algorithm.

change the update rule to

$$X_{t+1} = X_t - \alpha_t g'(X_t)$$

if $g(X_{t+1}) > g(X_t)$ then half α

I firstly use $\alpha = 1$, but it seems the magnitude of θ_1 and θ_2 is so different so that the algorithm will not converage to the ideal point. I try to set the $\alpha = diag(\theta_0)$ which will solve the magnitude problem, and the result is shown as below.

```
##        theta1     theta2
## [1,] 212.6504 0.06409098
```

(d) Repeat (b) with the Gauss-Newton algorithm.

Consider

$$g(\theta) = -\sum_{i=1}^{n}(y_i - f_i(\theta))^2$$

where

$$\begin{aligned} f_i(\theta) &= \frac{\theta_1 x_i}{x_i + \theta_2} \\ f_i'(\theta) &= [\frac{x_i}{x_i + \theta_2}, -\frac{\theta_1 x_i}{(x_i + \theta_2)^2}]^T \end{aligned}$$

then we get

$$\begin{aligned} A = A(\theta) &= [f_1'(\theta)^T, ..., f_n'(\theta)^T]^T \\ Z = Z(\theta) &= [y_1 - f_1(\theta), ..., y_n - f_n(\theta)]^T \end{aligned}$$

The updating formula is

$$\theta_{t+1} = \theta_t + (A_t^T A_t)^{-1} A_t^T Z_t$$

here is the results

```
##         theta1     theta2
## [1,] 212.6837 0.06412128
```