STAT 5361 - Homework 2

Kristen Sandberg[*]

07 February 2018

Contents

Problem 1

The Cauchy$(\theta, 1)$ distribution has probability density

$$p(x;\theta) = \frac{1}{\pi[1+(x-\theta)^2]},\ x \in \mathbb{R},\ \theta \in \mathbb{R}. \tag{1}$$

0.0.0.1 Problem 1(a)

Let $x_1, \ldots, x_n$ be an i.i.d. sample and $l(\theta)$ the log-likelihood of θ based on the sample. Show that

$$l(\theta) = -n \ln \pi - \sum_{i=1}^{n} \ln\left[1+(\theta - x_i)^2\right], \tag{2}$$

$$l'(\theta) = -2\sum_{i=1}^{n} \frac{\theta - x_i}{1+(\theta-x_i)^2}, \tag{3}$$

$$l''(\theta) = -2\sum_{i=1}^{n} \frac{1-(\theta-x_i)^2}{[1+(\theta-x_i)^2]^2}, \tag{4}$$

$$I(\theta) = n\int \frac{\{p'(x)\}^2}{p(x)}dx = \frac{4n}{\pi}\int_{-\infty}^{\infty} \frac{x^2dx}{(1+x^2)^3} = \frac{n}{2}. \tag{5}$$

0.0.0.2 Solution 1(a)

Let $x_1, \ldots, x_n$ be an i.i.d. sample and $l(\theta)$ the log-likelihood of θ based on the sample. Observe,

[*]kristen.sandberg@uconn.edu; M.S. in Applied Financial Mathematics, Department of Mathematics, University of Connecticut.

$$
\begin{aligned}
l(\theta) &= \ln L(\theta) && (6)\\
&= \ln\left(\prod_{i=1}^{n}\left(\frac{1}{\pi[1+(x_i-\theta)^2]}\right)\right) && (7)\\
&= \ln\left(\prod_{i=1}^{n}\left(\pi[1+(x_i-\theta)^2]\right)^{-1}\right) && (8)\\
&= \sum_{i=1}^{n}\ln\left(\pi[1+(x_i-\theta)^2]\right)^{-1} && (9)\\
&= \sum_{i=1}^{n}\left[\ln(1)-\ln\left(\pi[1+(x_i-\theta)^2]\right)\right] && (10)\\
&= \sum_{i=1}^{n}\left[0-\ln\left(\pi[1+(x_i-\theta)^2]\right)\right] && (11)\\
&= \sum_{i=1}^{n}\left[(-\ln(\pi))-\ln\left(1+(x_i-\theta)^2\right)\right] && (12)\\
&= -n\ln\pi-\sum_{i=1}^{n}\ln[1+(\theta-x_i)^2]. && (13)
\end{aligned}
$$

Next, we have

$$
\begin{aligned}
l'(\theta) &= 0-\sum_{i=1}^{n}\left(\frac{1}{1+(\theta-x_i)^2}\right)(2(\theta-x_i)(1)) && (14)\\
&= -2\sum_{i=1}^{n}\frac{\theta-x_i}{1+(\theta-x_i)^2}. && (15)
\end{aligned}
$$

Then for the second derivative we have,

$$
\begin{aligned}
l''(\theta) &= -2\sum_{i=1}^{n}\frac{[1+(\theta-x_i)^2](1)-(\theta-x_i)[2(\theta-x_i)(1)]}{[1+(\theta-x_i)^2]^2} && (16)\\
&= -2\sum_{i=1}^{n}\frac{1+(\theta-x_i)^2-2(\theta-x_i)^2}{[1+(\theta-x_i)^2]^2} && (17)\\
&= -2\sum_{i=1}^{n}\frac{1-(\theta-x_i)^2}{[1+(\theta-x_i)^2]^2}. && (18)
\end{aligned}
$$

Lastly,

$$
\begin{aligned}
I(\theta) &= \mathbb{E}\left[\left(l'(\theta)\right)^2\right] && (19)\\
&= \mathbb{E}\left[\left(-2\sum_{i=1}^{n}\frac{\theta - x_i}{1+(\theta-x_i)^2}\right)^2\right] && (20)\\
&= 4\mathbb{E}\left[\sum_{i=1}^{n}\frac{(\theta-x_i)^2}{[1+(\theta-x_i)^2]^2}\right] && (21)\\
&= \frac{4n}{\pi}\int \frac{(\theta-x)^2}{[1+(\theta-x)^2]^2}\frac{1}{1+(\theta-x)^2}dx && (22)\\
&= \frac{4n}{\pi}\int \frac{(\theta-x)^2}{[1+(\theta-x)^2]^3}dx && (23)\\
&= \frac{4n\pi}{\pi^2}\int \frac{(\theta-x)^2}{[1+(\theta-x)^2]^4}[1+(\theta-x)^2]dx && (24)\\
&= \frac{4n}{\pi^2}\int \frac{(\theta-x)^2}{[1+(\theta-x)^2]^4}\left(\pi[1+(\theta-x)^2]\right)dx && (25)\\
&= n\int \frac{\{p'(x)\}^2}{p(x)}dx && (26)
\end{aligned}
$$

This means, when $u = \theta - x$ and $du = dx$, we have

$$
\begin{aligned}
I(\theta) &= n\int \frac{\{p'(x)\}^2}{p(x)}dx && (27)\\
&= \frac{4n}{\pi^2}\int_{-\infty}^{\infty}\frac{u^2}{[1+u^2]^4}\left(\frac{1}{\pi[1+u^2]}\right)^{-1}du && (28)\\
&= n\int_{-\infty}^{\infty}\frac{4\pi u^2(1+u^2)}{\pi^2[1+u^2]^4}du && (29)\\
&= \frac{4n}{\pi}\int_{-\infty}^{\infty}\frac{u^2du}{(1+u^2)^3}. && (30)
\end{aligned}
$$

Thus, if we substitute $u = \tan(\theta)$ and $du = sec^2(\theta)$, we have

$$
\begin{aligned}
I(\theta) &= \frac{4n}{\pi}\int_{-\infty}^{\infty}\frac{u^2du}{(1+u^2)^3} && (31)\\
&= \frac{4n}{\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\frac{\tan^2\theta}{(1+\tan^2\theta)^3}\sec^2\theta d\theta && (32)\\
&= \frac{4n}{\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\frac{\tan^2\theta}{(\sec^2\theta)^3}\sec^2\theta d\theta && (33)\\
&= \frac{4n}{\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\frac{\tan^2\theta}{(\sec^2\theta)^2}d\theta && (34)\\
&= \frac{4n}{\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\frac{\sin^2\theta}{\cos^2\theta}(\cos^2\theta)^2d\theta && (35)\\
&= \frac{4n}{\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\sin^2\theta\cos^2\theta d\theta && (36)\\
&= \frac{4n}{\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}(1-\cos^2\theta)\cos^2\theta d\theta && (37)\\
&= \frac{4n}{\pi}\left[\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\cos^2\theta d\theta - \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\cos^4\theta d\theta\right] && (38)\\
&= \frac{4n}{\pi}\left[\frac{1}{2}\theta+\frac{1}{4}\sin 2\theta-\frac{1}{4}\cos^3\theta\sin\theta\right]\Bigg|_{-\frac{\pi}{2}}^{\frac{\pi}{2}} - \frac{3}{4}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\cos^2\theta d\theta && (39)\\
&= \frac{4n}{\pi}\left[\frac{1}{2}\theta+\frac{1}{4}\sin 2\theta-\frac{1}{4}\cos^3\theta\sin\theta\right]\Bigg|_{-\frac{\pi}{2}}^{\frac{\pi}{2}} - \frac{3}{4}\left(\frac{1}{2}\theta+\frac{1}{4}\sin 2\theta\right)\Bigg|_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\cos^2\theta && (40)\\
&= \frac{4n}{\pi}\left[\frac{1}{4}\left(\frac{1}{2}\theta+\frac{1}{4}\sin 2\theta\right)-\frac{1}{4}\cos^3\theta\sin\theta\right]\Bigg|_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\cos^2\theta && (41)\\
&= \frac{4n}{\pi}\left[\frac{1}{4}\left(\frac{1}{2}\left(\frac{\pi}{2}-\frac{-\pi}{2}\right)\right]+\left[\frac{1}{4}\sin 2\frac{\pi}{2}\right]-\left[\frac{1}{4}\sin 2\frac{-\pi}{2}\right]\right)-\frac{1}{4}\left[\cos^3\frac{\pi}{2}\sin\frac{\pi}{2}-\cos^3\frac{-\pi}{2}\sin\frac{-\pi}{2}\right] && (42)\\
&= \frac{4n}{\pi}\left(\frac{\pi}{8}+0\right) && (43)\\
&= \frac{n}{2}. && (44)
\end{aligned}
$$

0.0.0.3 Problem 1(b)

Suppose that the observed sample is

```
x_q1 <- c(1.77, -0.23, 2.76, 3.80, 3.47, 56.75, -1.34, 4.24, -2.44, 3.29, 3.71,
          -2.40, 4.53, -0.07, -1.05, -13.87, -2.53, -1.75)
```

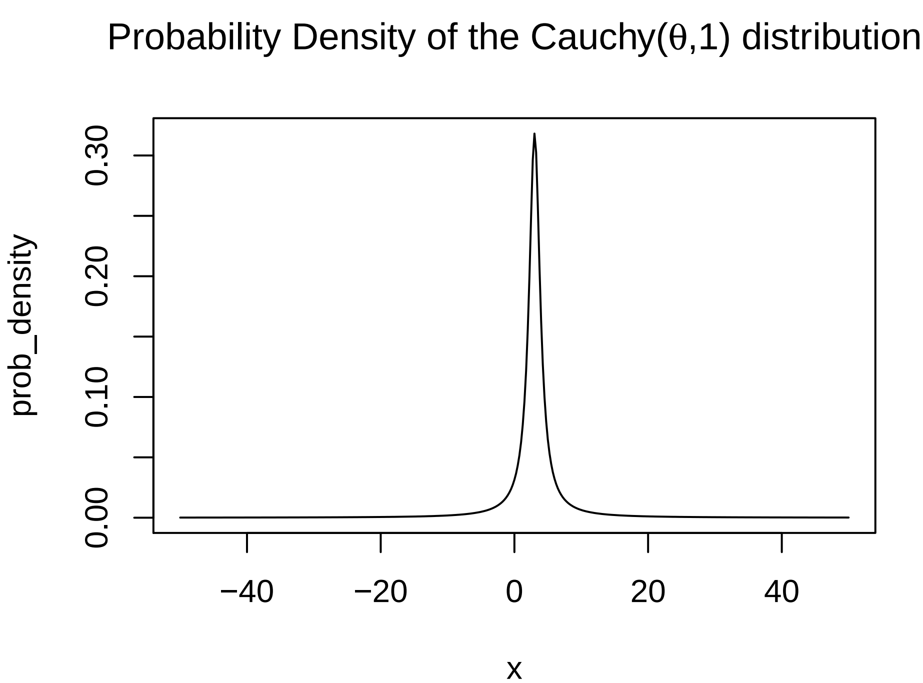
Graph the log-likelihood function. Find the MLE for θ using the Newton-Raphson method. Try all the following starting points: $-11, -1, 0, 1.5, 4, 4.7, 7, 8$, and 38. Compare your results. Is the sample mean a good starting point?

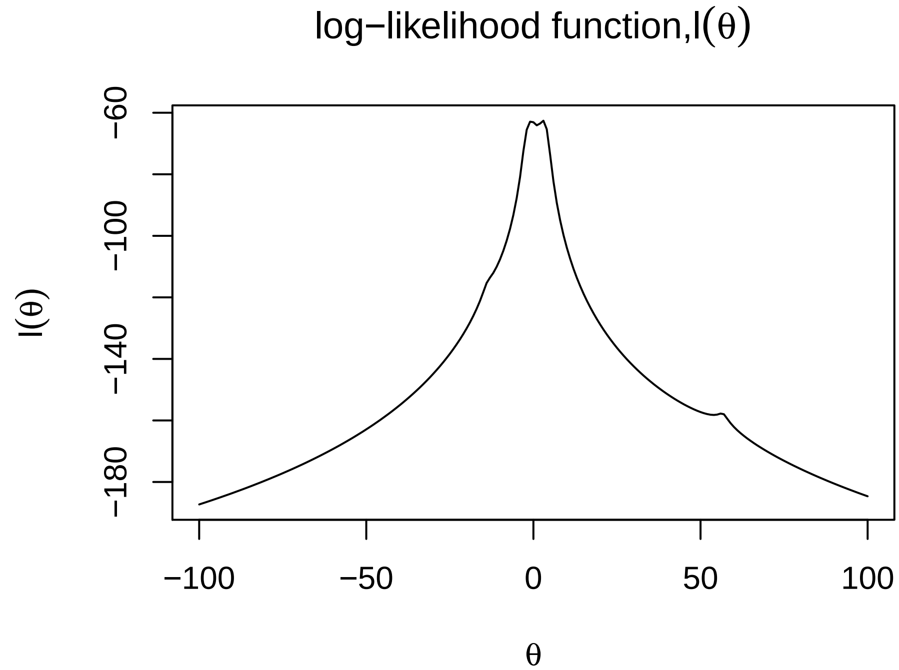
Table 1: The MLE for θ using the Newton-Raphson method.

Starting Point	Converges to
-11.0000	-3.660385e+30
-1.0000	-5.915000e-01
0.0000	-5.915000e-01
1.5000	1.092700e+00
4.0000	3.021300e+00
4.7000	-5.915000e-01
7.0000	1.527426e+30
8.0000	8.084521e+29
38.0000	1.135250e+31
3.2578	3.021300e+00

0.0.0.4 Solution 1(b)

The probability densiy function and log-likelihood function are plotted below for the observed sample. I built a function that finds the MLE for θ using the Newton-Raphson method. The results are listed in the table below. Most of the starting points converged to two maxima, one of which is the global maximum. The sample mean was a good starting point because it did lead us to the MLE.





0.0.0.5 Problem 1(c)

Apply fixed-point iterations using $G(x) = \alpha l'(\theta) + \theta$, with scaling choices of $\alpha \in \{1, 0.64, 0.25\}$ and initial value -1. Try the same starting points as above.

0.0.0.6 Solution 1(c)

When we apply the fixed-point iteration method with $G(\theta) = \alpha l'(\theta) + \theta$, with scaling choices of $\alpha \in \{1, 0.64, 0.25\}$, we find the following results in Table 2.

0.0.0.7 Problem 1(d)

First use Fisher scoring to find the MLE for θ, then refine the estimate by running Newton-Raphson method. Try the same starting points as above.

0.0.0.8 Solution 1(d)

We found that the starting points converged to two maxima, one of which was the global maximum, or MLE.

Table 2: The results of the fixed-point iterations using $G(x) = \alpha l'(\theta) + \theta$ where $\alpha \in \{1, 0.64, 0.25\}$.

Starting Point	Alpha	Attracted to Theta
-1	1.00	0.1035
-1	0.64	-0.5915
-1	0.25	-0.5915
-11	1.00	-0.5915
-11	0.64	-0.5915
-11	0.25	-0.5915
0	1.00	-1.1063
0	0.64	-0.5915
0	0.25	-0.5915

Starting Point	Alpha	Attracted to Theta
1.5	1.00	0.1035
1.5	0.64	3.2398
1.5	0.25	3.0213
4.0	1.00	-1.1063
4.0	0.64	-0.5915
4.0	0.25	3.0213
4.7	1.00	-1.1714
4.7	0.64	-0.5915
4.7	0.25	3.0213

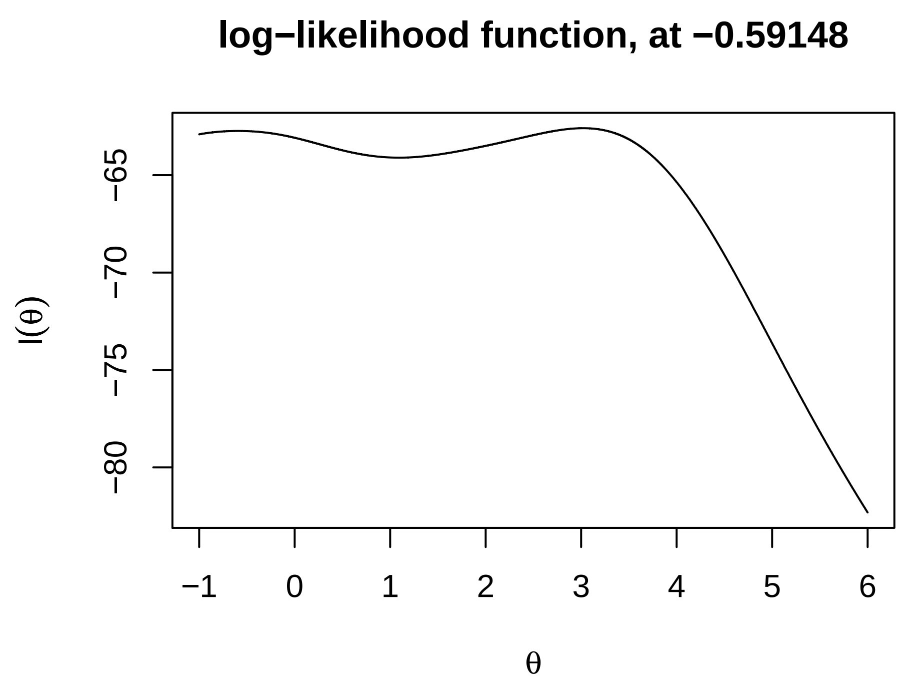
Starting Point	Alpha	Attracted to Theta
7	1.00	-1.1714
7	0.64	2.5915
7	0.25	3.0213
8	1.00	0.2417
8	0.64	-0.5915
8	0.25	3.0213
38	1.00	0.2417
38	0.64	2.5915
38	0.25	3.0213

Table 3: The results of using Fisher scoring to find the MLE from different starting points are shown in the table.

Starting Point	Converges to
-11.00000	-0.59148
-1.00000	-0.59148
0.00000	-0.59147
1.50000	3.02134
4.00000	3.02135
4.70000	3.02135
7.00000	3.02135
8.00000	3.02135
38.00000	3.02135
3.25778	3.02135

Table 4: The result of using the Newton-Raphson method to refine our results from the Fisher scoring method is shown in the table.

Starting Point from Fisher Scoring	Converges to using Newton-Raphson
-0.59147684369266007032	-0.59147352051799439998



Using Fisher scoring, we found the MLE to be θ = -0.59147684369266007032. Once we refined this estimate using the Newton-Raphson method, we found the MLE to be θ = -0.59147352051799439998.

0.0.0.9 Problem 1(e)

Comment on the results from different methods (speed, stability, etc.).

0.0.0.10 Solution 1(e)

For the Newton-Raphson method in part 1b, some of the starting points diverged, so the iterations had to be cut off to prevent the code from running too long or not working, but that was only true for $\theta \in \{-11, 7, 8, 38\}$. The other starting points took five or six iterations. In part 1c, the fixed-point iteration method was slower than the Newton-Raphson method. The speed of the fixed-point method depended on both the starting point and the value of α. Convergence was faster when $\alpha = 0.25$, which is not a surprise because it was the smallest of the scaling choices. Finally, when I used the Fisher scoring method, each of the stating points took between 1000 and 1500 iterations, however, that may be due to the tolerance I have set. The positive aspect of the Fisher scoring method is all of the starting points converged to just two maxima, one of which is the MLE while the other is a local maxima.

Problem 2

Consider the probability density with parameter θ

$$p(x;\theta) = \frac{1-\cos(x-\theta)}{2\pi},\ 0 \le x \le 2\pi,\ \theta \in (-\pi, \pi). \tag{45}$$

A random sample from the distribution is

```
x_q2 <- c(3.91, 4.85, 2.28, 4.06, 3.70, 4.04, 5.46, 3.53, 2.28, 1.96,
       2.53, 3.88, 2.22, 3.47, 4.82, 2.46, 2.99, 2.54, 0.52)
```

0.0.0.11 Problem 2(a)

What is the log-likelihood function of θ based on the sample? Graph the function between $-\pi$ and π.

0.0.0.12 Solution 2(a)

The log-likelihood function of θ based on the given sample can be computed as follows:

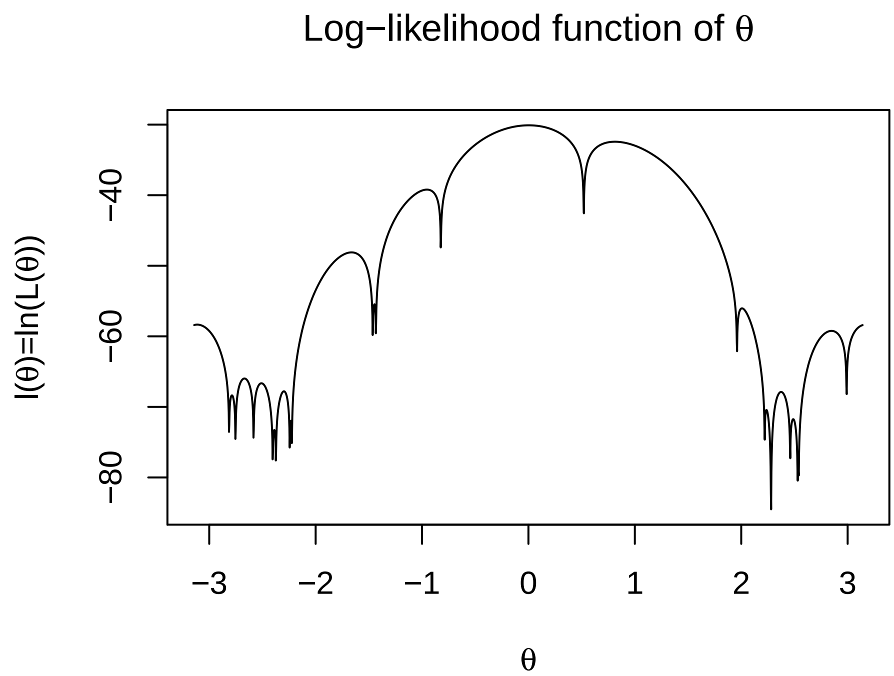


Figure 1: Log-likelihood function

$$
\begin{aligned}
l(\theta) &= \ln\left[\prod_{i=1}^{n}\left(\frac{1-\cos(x-\theta)}{2\pi}\right)\right] && (46)\\
&= \sum_{i=1}^{n}\left[\ln\left(\frac{1-\cos(x-\theta)}{2\pi}\right)\right] && (47)\\
&= \sum_{i=1}^{n}\left[\ln\left(\frac{1}{2\pi}\right)+\ln(1-\cos(x-\theta))\right] && (48)\\
&= n\ln\left(\frac{1}{2\pi}\right)+\sum_{i=1}^{n}\left[\ln(1-\cos(x-\theta))\right] && (49)\\
&= n\ln(1)-n\ln(2\pi)+\sum_{i=1}^{n}\left[\ln(1-\cos(x-\theta))\right] && (50)\\
&= -n\ln(2\pi)+\sum_{i=1}^{n}\left[\ln(1-\cos(x-\theta))\right]. && (51)
\end{aligned}
$$

The graph of the log-likelihood function of θ is plotted from $[-\pi, \pi]$ in Figure 1.

0.0.0.13 Problem 2(b)

Find the method-of-moments estimator of θ. That is, the estimator $\hat{\theta}_{\text{moment}}$ is value of θ with

$\mathbb{E}[X|\theta] = \bar{x}$ where x is the sample mean. This means you have to first find the expression for $\mathbb{E}[X|\theta]$.

0.0.0.14 Solution 2(b)

To find the method-of-moments estimator of θ, we will first find the expression for $\mathbb{E}[X|\theta]$, as shown below:

$$\begin{aligned}
\mathbb{E}[X|\theta] &= \int_0^{2\pi} xp(x;\theta)dx && (52)\\
&= \frac{1}{2\pi}\int_0^{2\pi} x(1-\cos(x-\theta))dx && (53)\\
&= \frac{1}{2\pi}\left[\int_0^{2\pi} xdx - \int_0^{2\pi} x\cos(x-\theta)dx\right] && (54)\\
& && (55)
\end{aligned}$$

Using integration by parts, we find that

$$\begin{aligned}
\mathbb{E}[X|\theta] &= \frac{1}{2\pi}\left[\frac{x^2}{2}\bigg|_0^{2\pi} - x\sin(x-\theta)\bigg|_0^{2\pi} + \int_0^{2\pi}\sin(x-\theta)dx\right] && (56)\\
&= \left(\frac{4\pi^2}{4\pi} - 0\right) - \frac{1}{2\pi}\left[2\pi\sin(2\pi-\theta) - 0\right] - \frac{1}{2\pi}\left[\cos(x-\theta)\right]\bigg|_0^{2\pi} && (57)\\
&= \pi - \sin(2\pi-\theta) - \frac{1}{2\pi}\left[\cos(2\pi-\theta) - \cos(0-\theta)\right] && (58)\\
&= \pi - \left[\sin(2\pi)\cos(-\theta) + \cos(2\pi)\sin(-\theta)\right] && (59)\\
&- \frac{1}{2\pi}\left[\cos(2\pi)\cos(-\theta) - \sin(2\pi)\sin(-\theta)\right] + \frac{1}{2\pi}\left[\cos(0)\cos(-\theta) - \sin(0)\sin(-\theta)\right] && (60)\\
&= \pi - \sin(-\theta) - \frac{1}{2\pi}\cos(-\theta) + \frac{1}{2\pi}\cos(-\theta) && (61)\\
&= \pi - \sin(-\theta) && (62)\\
&= \pi - \sin(\theta). && (63)
\end{aligned}$$

Now we set this equal to the sample mean, $\bar{x}$, and we have

$$\mathbb{E}[X|\theta] = \pi - \sin(\theta) = \bar{x}, \tag{64}$$

so we subtract π from both sides and we have

$$-\sin(\theta) = \bar{x} - \pi. \tag{65}$$

Finally we multiply by -1 and take the inverse sine function of both sides and we find that

$$\theta = \arcsin(\pi - \bar{x}). \tag{66}$$

Thus, the method-of-moments estimator of θ, denoted $\hat{\theta}_{\text{moment}} = \arcsin(\pi - \bar{x})$.

Given the observations, the sample mean of the given random sample is $\bar{x} = 3.2368421$. The method-of-moments estimator of θ is $\hat{\theta}_{\text{moment}}$ = -0.0953941

0.0.0.15 Problem 2(c)

Find the MLE for θ using the Newton-Raphson method with $\theta_0 = \hat{\theta}_{\text{moment}}$.

0.0.0.16 Solution 2(c)

When we start with $\theta_0 = \hat{\theta}_{\text{moment}}$ = -0.0953941 we find that the MLE for θ using the Newton-Raphson method is 0.0031182.

0.0.0.17 Problem 2(d)

What solutions do you find when you start at $\theta_0 = -2.7$ and $\theta_0 = 2.7$?

0.0.0.18 Solution 2(d)

When we start with $\theta_0 = -2.7$ we find that the MLE for θ using the Newton-Raphson method is -2.6688575 and when $\theta_0 = 2.7$ the MLE is 2.8484153.

0.0.0.19 Problem 2(e)

Repeat the above using 200 equally spaced starting values between $-\pi$ and π. Partition the values into sets of attraction, That is, divide the set of starting values into separate groups, with each group corresponding to a separate unique outcome of the optimization.

0.0.0.20 Solution 2(e)

Using 200 equally spaced starting values between $-\pi$ and π, I found the following unique outcomes and divided the set into groups that are attracted to the same outcome.

Table 5: The unique outcomes of the optimization using 200 equally spaced starting values between $-\pi$ and π. The table lists the lower and upper bounds of values for θ that are attracted to each unique outcome.

Attracted to Theta	Lowest Theta in Set	Highest Theta in Set
-3.1125	-3.1416	-2.8259
-2.7866	-2.7943	-2.7627
-2.6689	-2.7311	-2.6048
-2.5094	-2.5733	-2.4154
-2.3883	-2.3838	-2.3838
-2.2979	-2.3522	-2.2575
-2.2322	-2.2260	-2.2260
-1.6627	-2.1944	-1.4682
-1.4475	-1.4366	-1.4366
-0.9544	-1.4050	-0.8367
0.0031	-0.8051	0.4894
0.8126	0.5210	1.9418
2.0072	1.9734	2.1944
2.2370	2.2260	2.2575
2.3747	2.2891	2.4470
2.4884	2.4785	2.5101
2.8484	2.5417	2.9837
3.1707	3.0153	3.1416

Problem 3

The counts of a floor beetle at various time points (in days) are given in a dataset.

```
beetles <- data.frame(
  days = c(0, 8, 28, 41, 63, 69, 97, 117, 135, 154),
  beetles = c(2, 47, 192, 256, 768, 896, 1120, 896, 1184, 1024))
```

A simple model for population growth is the logistic model given by

$$\frac{\mathrm{d}N}{\mathrm{d}t} = rN\left(1 - \frac{N}{K}\right), \tag{67}$$

where N is the population size, t is time, r is an unknown growth rate parameter, and K is an unknown parameter that represents the population carrying capacity of the environment. The solution to the differential equation is given by

$$N_t = f(t) = \frac{KN_0}{N_0 + (K - N_0)\exp(-rt)}, \tag{68}$$

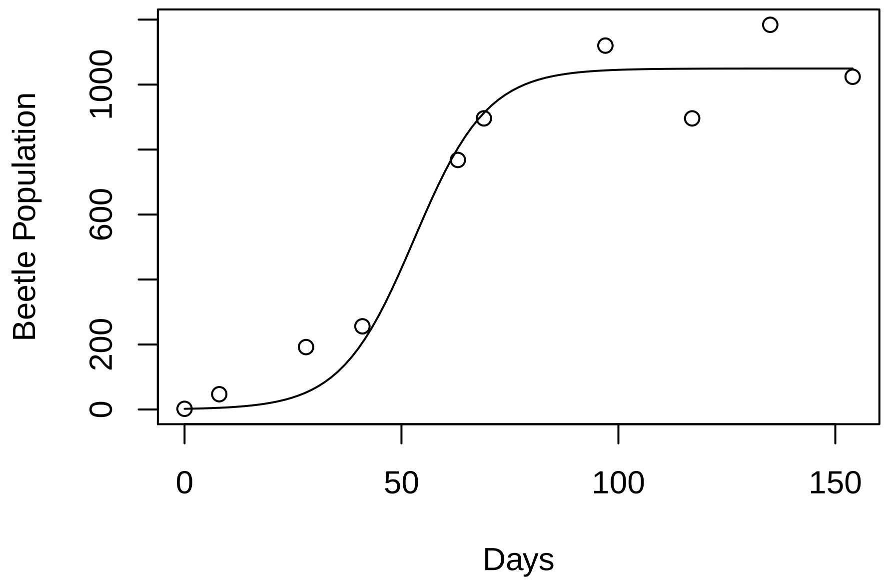
where N_t denotes the population size at time t.

0.0.0.21 Problem 3(a)

Fit the population growth model to the beetles data using the Gauss-Newton approach, to minimize the sum of squared errors between model predictions and observed counts.

0.0.0.22 Solution 3(a)

The following figure is the result of fitting the population growth model to the beetles data using the Gauss-Newton approach.



To use the Gauss-Newton approach, we need to compute the following partial derivatives:

$$
\begin{aligned}
\frac{\partial f(t)}{\partial K} &= \frac{[N_0 + (K - N_0)\exp(-rt)](N_0) - [KN_0][\exp(-rt)]}{[N_0 + (K - N_0)\exp(-rt)]^2} && (69)\\
&= \frac{N_0^2 + KN_0\exp(-rt) - N_0^2\exp(-rt) - KN_0\exp(-rt)}{[N_0 + (K - N_0)\exp(-rt)]^2} && (70)\\
&= \frac{N_0^2(1 - \exp(-rt))}{[N_0 + (K - N_0)\exp(-rt)]^2} && (71)
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{\partial f(t)}{\partial r} &= \frac{[N_0 + (K - N_0)\exp(-rt)](0) - [KN_0][-t(K - N_0)\exp(-rt)]}{[N_0 + (K - N_0)\exp(-rt)]^2} && (72)\\
&= \frac{tKN_0(K - N_0)\exp(-rt)}{[N_0 + (K - N_0)\exp(-rt)]^2} && (73)
\end{aligned}
$$

0.0.0.23 Problem 3(b)

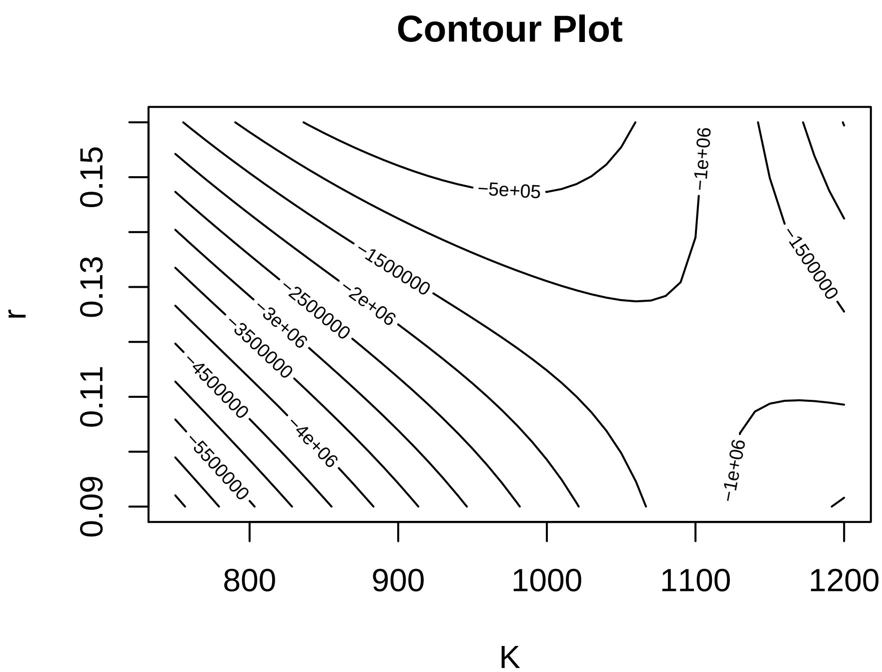
Show the contour plot of the sum of squared errors.

Table 6: The maximum likelihood estimators for K and r are shown in the following table.

K	r
1049.407	0.11827

0.0.0.24 Solution 3(b)

The following figure is the contour plot of the sum of squared errors that resulted from using the Gauss-Newton approach. The MLE appears to be near $K = 1100$ and $r = 0.11$.



0.0.0.25 Problem 3(c)

In many population modeling application, an assumption of lognormality is adopted. That is , we assume that $\log N_t$ are independent and normally distributed with mean $\log f(t)$ and variance σ^2. Find the maximum likelihood estimators of r, K, σ^2 using any suitable method of your choice. Estimate the variance your parameter estimates.

0.0.0.26 Solution 3(c)

I compute the maximum likelihood estimators for K and r using the Gauss-Newton approach.

I found the MLEs for $K = 1049.4072443$ and $r = 0.1182684$. The variance is $\sigma^2 = 234001.4654$. See solution 3a for the details of the Gauss-Newton approach.

Acknowledgment

I would like to thank Surya Eada for his collaboration on problem 1a of this assignment prior to when the groups were split up.

Reference