MATH 559 - ASSIGNMENT 5
FINAL PROJECT

Please submit your solutions by 11.59 pm (EST) on December 5th by uploading a single pdf to myCourses.

This project considers robust regression using Bayesian methods. The linear model

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i \qquad i = 1, \ldots, n$$

when $\epsilon_1, \ldots, \epsilon_n$ are independent and identically but **not Normally distributed** may provide a way of introducing robustness into the estimation of the regression parameters (β_0, β_1). This makes inference about the parameters robust to **outliers**.

One robustifying assumption is to assume $Cauchy \equiv Student(1)$ errors; data generated under this assumption are depicted below.

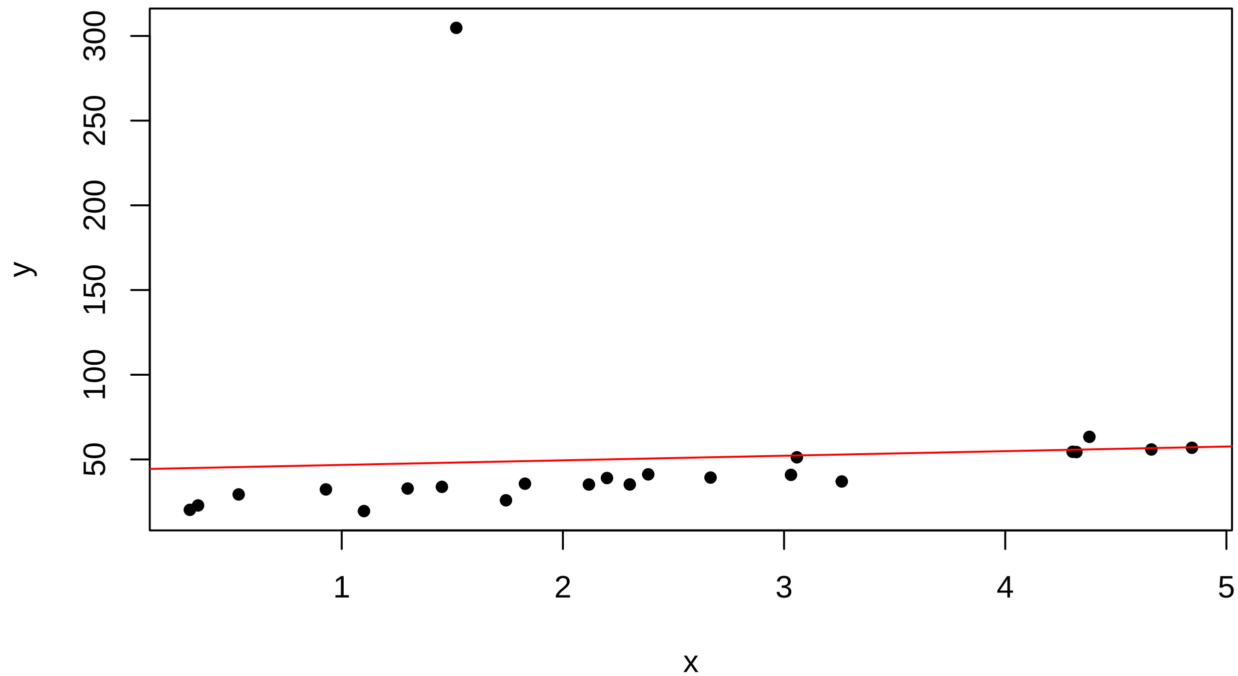
```
n<-23
set.seed(81642)
be0<-20
be1<-8
sig<-4
x<-sort(runif(n,0,5))
y<-be0+be1*x+rt(n,1)*sig
cbind(x,y)

+               x         y
+  [1,] 0.3132980  20.23565
+  [2,] 0.3502758  22.83761
+  [3,] 0.5340795  29.33431
+  [4,] 0.9285782  32.30061
+  [5,] 1.1006034  19.50876
+  [6,] 1.2979742  32.81472
+  [7,] 1.4529997  33.81671
+  [8,] 1.5180004 304.76622
+  [9,] 1.7426959  25.88330
+ [10,] 1.8285557  35.66872
+ [11,] 2.1175934  35.18131
+ [12,] 2.1994464  39.01334
+ [13,] 2.3025292  35.23135
+ [14,] 2.3859049  41.18901
+ [15,] 2.6675895  39.28907
+ [16,] 3.0313951  40.89708
+ [17,] 3.0580300  51.24407
+ [18,] 3.2609863  36.99516
+ [19,] 4.3048254  54.50957
+ [20,] 4.3220233  54.34492
+ [21,] 4.3806682  63.33452
+ [22,] 4.6609183  55.88011
+ [23,] 4.8442350  56.88866

summary(lm(y~x))$coef

+              Estimate Std. Error   t value   Pr(>|t|)
+ (Intercept) 44.045757  24.079428 1.8291862 0.08160871
+ x            2.712521   8.773261 0.3091805 0.76023093
```

```r
par(mar=c(4,4,2,2))
plot(x,y,pch=19,cex=0.75)
abline(coef(lm(y~x)),col='red')
```



Problem: For the data generated above, construct and implement a Bayesian analysis (using a suitably chosen prior) for the parameters $(\beta_0, \beta_1, \sigma^2)$ in the Cauchy regression model. Recall that the Cauchy assumption means that the density for the observed data takes the form

$$f_{Y|X}(y|x;\beta_0,\beta_1,\sigma^2) = \frac{1}{\pi}\frac{1}{\sigma}\left\{1+\left(\frac{y-\beta_0-\beta_1 x}{\sigma}\right)^2\right\}^{-1} \qquad y \in \mathbb{R}$$

where β_0, β_1 are real-valued parameters, and $\sigma^2 > 0$, and where the x values may be treated as fixed constants.

As well as mathematical derivations that define the form of the posterior distribution, you should include in your solution annotated computational code and demonstrate that it is producing Bayesian inference. You may re-use code from the course if it is annotated appropriately.

20 MARKS