STAT 600 - HW 2

Kevin Korsurat

1. Your co-worker collected an iid sample, $X_1, \ldots, X_{25} \sim$ Cauchy(θ,1), so $f(x) = \frac{1}{\pi(1+(x-\theta)^2)}$ The data are:

```
-8.86 -6.82 -4.03 -2.84  0.14  0.19  0.24  0.27  0.49  0.62  0.76 1.09
 1.18  1.32  1.36  1.58  1.58  1.78  2.13  2.15  2.36  4.05  4.11 4.12
 6.83
```

   (a) Graph $l'(\theta)$.

   (b) Write functions in Rcpp to perform:

      i. Bisection
      ii. Newton-Raphson
      iii. Fisher Scoring
      iv. Secant Method

      Use each method to estimate $\hat{\theta}$.

   (c) Make a table of the results including the columns: method, $\hat{\theta}$, and number of iterations to converge.

   (d) Describe the convergence criteria that you used.

   (e) What is your best estimate of θ? What is the standard error of your estimate?

   (f) Discuss how you chose to initialize each method. Describe how the results may be sensitive to initialization. Provide empirical results if necessary.

   (g) Your co-worker the data is somewhat absent minded. He just found 25 more observations, listed below. Using these data together with the data listed above, what is your best estimate of θ? What is the standard error of your estimate?

```
-8.34  -1.73  -0.40  -0.24   0.60   0.94   1.05   1.06   1.45  1.50
 1.54   1.72   1.74   1.88   2.04   2.16   2.39   3.01   3.01  3.08
 4.66   4.99   6.01   7.06  25.45
```

2. Each step of the secant method requires only one new function evaluation, while each step of the Newton Raphson method requires two function evaluations (l' and l''). Since function evaluations constitute the principal computational burden in these algorithms, 2 steps in the secant method is comparable to 1 step in the Newton Raphson method.

   Prove that the convergence rate for 2 steps of the secant method is better than the convergence rate for 1 step of the Newton Raphson method.

3. The table below shows the binary outcome of individuals participating in a study of pancreatic cancer. Males and females were classified into categories based on the amount of coffee they consumed. There were r_i cancer cases out of a total of n_i participants.

   Assume a logit model where $Y_i \sim$Bernoulli($p_i = \frac{\exp(\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2})}{1+\exp(\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2})}$). Here x_1 reflects the amount of coffee consumed and x_2 reflects the gender.

   (a) Derive the log likelihood for this problem.

cups per day	gender	r_i	n_i	x_{i1}
0	male	9	41	0
1-2	male	94	213	2
3-4	male	53	127	4
5+	male	60	142	5
0	female	11	67	0
1-2	female	59	211	2
3-4	female	53	133	4
5+	female	28	76	5

(b) Use your Newton-Raphson algorithm from the previous problem to estimate the parameters and make a table of your results.

(c) Draw a plot showing coffee consumption versus probability of cancer for males and females. Show the estimated lines on the plot. Interpret the results and plot for a non-statistician.

(d) (Bonus) Are the parameter estimates significantly different from 0?