STAT 460/560- Assignments

Week IV:

1. Let $X_1, X_2, ..., X_n$ be an i.i.d. random variables from Weibull distribution with fixed scale parameter, whose density function is given by

   $$f(x;\theta) = \theta x^{\theta-1}\exp(-x^\theta), \quad x > 0, \theta > 0.$$

   (You may want to first go over the Example 6.1 in the Lecture Notes.)

   Suppose we observe the sample data

   ```
   0.6944788, 0.3285051, 0.7165376, 0.8865894, 1.0858084,
   0.4040884, 1.0538935, 1.2487677, 1.1523552, 0.9977360,
   0.7251880, 1.0716697, 1.0382114, 1.1535934, 0.9175693,
   0.5537849, 0.9701821, 0.5486354, 1.0168818, 0.5193687
   ```

   (a) For this sample, numerically find an upper bound θ_1 and a lower bound θ_2, so that the maximum point of the likelihood function is within the interval $[\theta_1, \theta_2]$.

   (b) Use a bisection algorithm (as discussed in section 6.2 in Lecture Notes) to numerically find the MLE of θ for this observed sample.

   You need to attach your code, preferably in R.

2. Let $X_1, X_2, ..., X_n$ be an i.i.d. random variables from Weibull distribution with fixed scale parameter.

   (a) Work out analytically the updating formulas for the parameter θ in the Newton-Raphson method (as discussed in section 6.2 in Lecture Notes).

   In other words, how do you obtain the value $\theta^{(k+1)}$ in the $(k+1)$-th iterative step based on the value $\theta^{(k)}$.

   (b) For the same observed sample as in Problem 1, numerically find the MLE of the parameter θ using the Newton-Raphson algorithm. Start from the initial value be $\theta^{(0)} = 1$ and report the first 5 values of the iteration.

You need to attach the code you use to your solution.

3. Let $X_1, X_2, \ldots, X_n$ be a set of i.i.d. random sample from $N(\theta, 1)$, and let $\bar{X}_n$ be the sample mean. Suppose θ^* is the true value of the mean parameter θ.

   Let
$$\tilde{\theta}_n = \begin{cases} 0 & \text{if } |\bar{X}_n| \le n^{-1/4}\,; \\ \bar{X}_n & \text{otherwise}\ . \end{cases}$$

   Remark: The problem has been simplified lately.

   (a) For $\theta^* = n^{-1/4}$ which changes with n, show that
$$P(\tilde{\theta}_n = 0) \to 0.5$$
   as $n \to \infty$.

   (b) Under the same condition as in (a), show that the MSE of $\tilde{\theta}_n$
$$n\mathbb{E}\{(\tilde{\theta}_n - \theta^*)^2\} \to \infty.$$
   Hint: develop an inequality based on result (a).

   (c) Use computer to generate data of size $n = 1600$ from $N(\theta^* = n^{-1/4}, 1)$, and compute the values of $\hat{\theta} = \bar{X}_n$ and $\tilde{\theta}_n$. Repeat it $N = 1000$ times so that you have N many pairs of these values. Compare their simulated total MSE:
$$\sum_{k=1}^{N} (\hat{\theta} - \theta^*)^2; \quad \sum_{k=1}^{N} (\tilde{\theta} - \theta^*)^2.$$

4. Let $X_1, X_2, ..., X_{2n+1}$ be an i.i.d. random sample from the Cauchy distribution with location parameter θ, whose density function is given by
$$f(x;\theta) = \frac{1}{\pi\left[1 + (x-\theta)^2\right]}.$$

The sample median is given by the order statistic $x_{(n+1)}$.

(a) Show that the sample median satisfies

$$P(n^{1/4}|x_{(n+1)} - \theta| \geq \epsilon) \to 0$$

for any $\epsilon > 0$ as $n \to \infty$.

Remark: this is challenging for inexperienced. It will be treated as a bonus question. Proving it by directly quoting an existing result will not qualify.

(b) Derive the explicit expression of the Newton-Raphson iteration for Cauchy distribution.

(c) Simulation $N = 1000$ times with $2n + 1 = 201$, $\theta = 0$ and obtain total MSEs in the same way as the last example. Clearly present your results. (That is, do not just write down two numbers without telling us what they are).

(d) Plot the histogram of the 1000 $X_{(n+1)}$. Do the same for the one-step Newtwon-Raphson estimator.

(e) Do these histogram support our asymptotic results on MLE and on median?

5. Derive the EM-iteration formulas for data from two component Binomial mixture model:

$$f(x; G) = \binom{m}{x} \{\pi\theta_1^x(1-\theta_1)^{m-x} + (1-\pi)\theta_2^x(1-\theta_2)^{m-x}\}$$

under the setting of n i.i.d. observations with $m \geq 3$. (Sizes n and m are not truly relevant in these formulas)

Be sure to have E-step and M-step clearly presented together with the corresponding Q function.