STA 248 - Tutorial 2 Independent Work

DUE: Feb. 12 @ 10 PM EST + 1 hr grace period

Show all your thought processes and steps, using appropriate notation, paying careful attention to the notation differences for estimators and parameters. Since this is independent work, try to answer each question on your own. Should you absolutely need additional clarification, any questions should be **addressed via email** and not posted on the weekly discussion boards.

This week's tutorial will involve practice and exploration of different estimators of the same parameter. You will be expected to simulate the sampling distributions of these estimators, which you have learned from transformations of random variables in your probability course, by using observations from a Uniform(0, 1) distribution. We will review this on Tuesday, but a refresher guide can be found on February 2nd's R Markdown file. If you are rusty with this, you should plan to do some personal review prior to starting this exercise. For all work in R, you are expected to do all the work in an R Markdown file, knitted to pdf.

Problem 1 [10 points]. This problem can be done with pen and paper. The purpose of this problem is for you to demonstrate that you can implement the two methods of deriving estimators covered: maximum likelihood estimation and method of moment estimation.

a) (6 points) Suppose a random sample $X_1, X_2, ..., X_n$ is drawn from a distribution believed to have the following probability density function:

$$f(x|\alpha,\beta) = \begin{cases} \alpha\beta^{\alpha}x^{-(\alpha+1)}, & x \geq \beta, \text{ and } \alpha,\beta > 0 \\ 0, & otherwise \end{cases}$$

Find the first moment of X supposing that the parameter α is known. Use it to find the method of moment estimator for unknown parameter β. You may find it easier to use the notation $m_1, m_2, ..., m_k$ to denote the first sample moment, second sample moment, ..., k$^{\text{th}}$ sample moment, respectively. Is the estimator you derived unbiased?

b) (4 points) Suppose a random sample $Y_1, Y_2, ..., Y_m$ represent random observations a uniformly distributed quantity that takes on values on the interval $[\alpha, \beta]$, where the endpoints are both unknown. Derive the maximum likelihood estimator that can be used to estimate α and β. Explain your process.

Hint 1: Do not just follow the steps of MLE, you should keep the goal of MLE in mind when finding the estimator here. Consider what would happen if you end up with an estimator $\hat{\alpha}$ that yields a point estimate larger than some of your data.

Hint 2: If Hint #1 doesn't trigger any 'AHA!' moments, consider reading the example found on p. 318 of MIPS, starting from 'Estimating the upper endpoint of a uniform distribution'. Either way, this problem relies mostly on skills you have learned in Calculus when it comes to finding maxima/minima of a function!

Problem 2 [R - 5 points]. This portion should be completed in R Markdown. Referring to the method reviewed on February 2nd's R Markdown file on simulating data from a particular distribution, apply the method to simulate random samples from **part (a) of problem 1**. With each random sample, compute the point estimate using the estimator you derived above.

To sample from the above distributions, we need to have a completely specified model. Thus:

- For $f(x|\alpha, \beta)$, use $\alpha = 2$, $\beta = 3$
- Use the simulation procedure below to see how well your $\hat{\beta}$ estimates $\beta = 3$. You may find it helpful to know the cumulative distribution function $F(x|\alpha, \beta) = 1 - \left(\frac{\beta}{x}\right)^{\alpha}$.

If you collect enough point estimates, you can recover the sampling distribution of the estimator. This is the same idea behind plotting a collected data set as a histogram – it represents the distribution from which the data are drawn from. In this case, you will be plotting your simulated point estimates as a histogram to see the distribution of your estimator.

1. Generate sampled data from the distribution $f(x|\alpha, \beta)$ in part (a) above.
2. From your sampled data, calculate $\hat{\beta}$. Save each estimate into a vector/data frame/tibble.
3. Repeat steps (1) and (2) many times with the same sample size. In order to get a more complete picture of the distribution of your estimator, you should plot with as many point estimates as you can. Start with at least 1000 point estimates of $\hat{\beta}$.
4. Plot a properly labeled histogram of your point estimates, and briefly comment on the distribution of the estimator (e.g. is it symmetric? distribution shape? approximate centre/spread?)

Do the above for sample size of 10, 50, 100, and 1000. You will have four histograms corresponding to each sample size. Include appropriate titles and labels to communicate which graph represents which sampling distribution. Use the `grid.arrange()` in the `gridExtra` package to display all four histograms in a 2x2 grid. If you have not previously installed the `gridExtra` package, you will need to do so first by running `install.packages('gridExtra')` in the console.

You might be wondering: What's the point if we already have these parameter values? The focus of this exercise has now shifted from '*find the missing parameter*' to '*assess how the estimator performs by examining its behaviour when distributions are known completely*'.