STAT 5014: Homework One

Bobby Soule

September 6th, 2017

Problem Two

Click **here** to see my GitHub page.

Problem Three

I completed the following lessons from the "R Programming E" course:

- Basic Building Blocks (1)
- Workspace and Files (2)
- Sequences of Numbers (3)
- Base Graphics (15)

Problem Four

Part A

From my undergraduate education and professional work experience, I am already well-acquainted with the R and Python programming languages; however, I am always looking for more practice to hone my programming skills. I have never used Latex or RMarkdown before and I am looking forward to learning how to create well-organized outputs with these tools.

Below are some of my specific learning objectives:

- Become comfortable with Latex/RMarkdown syntax
- Learn how to write Python code in R Notebooks
- Get experience working with ARC

Part B

Binomial(n, p) Distribution:

$$f(x|n,p) = \binom{n}{x} p^x (1-p)^{n-x} \tag{1}$$

Normal(μ, σ^2) Distribution:

$$f(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{2}$$

Exponential(β) Distribution:

$$f(x|\beta) = \frac{1}{\beta} e^{-\frac{x}{\beta}} \tag{3}$$

Problem Five

1. Keep track of how every result is produced, including details such as programs, parameters, and manual procedures
    - Can be tedious to keep track of every single detail
    - Unless details are precisely specified, it can be difficult to follow procedures
2. Do not use manual data manipulation steps; instead, use automated programs
    - Often times, people manually play with data before writing an automated script and may forget the manual manipulations they made
3. Archive the current versions of all external programs that are used to avoid errors that arise from differences in program versions
    - Any individual program may rely on numerous other programs, so it may be difficult to archive all programs and the dependencies of those programs
4. Maintain an archive of all program versions so that results can be reproduced using the original version
5. Record intermediate results in a standardized format so that errors can be traced to their source when they arise
    - With complex processes, there may be a plethora of intermediate results that are hard to follow
6. Take note of the seeds used for random number generators so that results can be reporduced exactly
7. Store the raw data that is used to create published plots
    - With very large datasets, it may take up too much memory to store every dataset for every plot
8. Provide the detailed values that underly any summaries so that layers of increasing detail can be inspected
9. Connect textual statements made in analyses to underlying results so that interpretations can be easily reevaluated
10. Make all data, scripts, and results publicly available

Problem Six

```
par(mfrow=c(1,2))
plot(rivers, main="River Length vs. Index", ylab="Length")
hist(rivers, main="Distribution of River Lengths", xlab="Length")
```