Homework 2: R Part

STAT 343: Mathematical Statistics

SOLUTIONS

Details

How to Write Up

The written part of this assignment can be either typeset using latex or hand written.

Grading

5% of your grade on this assignment is for turning in something legible and organized.

An additional 15% of your grade is for completion. A quick pass will be made to ensure that you've made a reasonable attempt at all problems.

Across both the written part and the R part, in the range of 1 to 3 problems will be graded more carefully for correctness. In grading these problems, an emphasis will be placed on full explanations of your thought process. You don't need to write more than a few sentences for any given problem, but you should write complete sentences! Understanding and explaining the reasons behind what you are doing is at least as important as solving the problems correctly.

Solutions to all problems will be provided.

Collaboration

You are allowed to work with others on this assignment, but you must complete and submit your own write up. You should not copy large blocks of code or written text from another student.

Sources

You may refer to our text, Wikipedia, and other online sources. All sources you refer to must be cited in the space I have provided at the end of this problem set.

```
library(readr)
turbine <- read_csv("https://sites.google.com/site/chiharahesterberg/data2/Turbine.csv")
```

```
## Rows: 168 Columns: 4
```

```
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): Date2010
## dbl (3): AveKW, AveSpeed, Production
```

```
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Problem I. Using your results from Problem II of the written part, program the Newton Raphson algorithm to find the maximum likelihood estimate for the parameter k for the Weibull model for wind speed, using the turbine data included above. You will need to calculate the values of the first and second derivatives at each subsequent parameter guess as part of your algorithm.

```r
first_dk <- function(n,k,x){
  n/k-n*sum(x^k*log(x))/sum(x^k)+sum(log(x))
}

second_dk <- function(n,k,x){
  -n/k^2-n*sum(x^k*log(x)^2)/sum(x^k)-n*(sum(x^k*log(x)))^2/(sum(x^k))^2
}

max_itr <- 1000
tol <- 1e-8
i <- 0
k_old <- 3.1
eps <- k_old
n <- nrow(turbine)
while(eps > tol & i < max_itr){
  k_new <- k_old-first_dk(n=n,k=k_old,x=turbine$AveSpeed)/second_dk(n=n,k=k_old,x=turbine$AveSpeed)
  eps <- abs(k_new-k_old)
  k_old <- k_new
  i <- i+1
}
print(i)
```

```
## [1] 661
```

```r
print(eps)
```

```
## [1] 9.952741e-09
```

```r
print(k_old)
```

```
## [1] 3.16932
```

The maximum likelihood estimate for k is $\hat{k}^{MLE} = 3.169$.

Note that if you had been asked to find the maximum likelihood estimate for λ based on these data, you could calculate it as follows:

```r
lambda_mle <- function(n,k,x){
  (1/n*sum(x^k))^(1/n)
}

lambda_mle(n=nrow(turbine), k=k_old, x=turbine$AveSpeed)
```

```
## [1] 1.03916
```