Homework 1 Stat 550

Dr. Scott

August 22, 2023

Instructions: We will discuss generally in class on Thursday; due Tuesday. You can work in groups, but turn in your own solutions.

Where to look: Some files will be on Canvas, some will be on my own web site: http://www.stat.rice.edu/~scottdw/stat550/. The baseball data are there in the HW/hw1/ folder.

Where to turn in: Canvas or physical OK.

- Using any techniques you wish, compare the Sosa and McGuire baseball home run (HR) data. These data are the (estimated) distances traveled by each HR and were recorded during the 1998 baseball season, breaking Roger Maris's long-standing 1961 record of 61 home runs. Sosa played for the Chicago Cubs; McGuire for the St Louis Cardinals, and Maris for the New York Yankees. BTW, Babe Ruth (NYY) hit 60 HR's in 1927.
- 2. Show how to normalize a frequency histogram so its area is 1.
- 3. For a probability histogram with equally spaced bins

 $\{(kh, (k+1)h], -\infty < k < \infty\},\$

find a formula for log-likelihood given data $\{x_1, x_2, \ldots, x_n\}$. What is the MLE for the bin width h?

4. Future Research Problem: For this problem, I generated two vectors of length 1,000, one each from two different $N(\mu_{\ell}, \sigma_{\ell}^2)$ densities. To make life interesting, the data were truncated to the intervals (a_{ℓ}, b_{ℓ}) . The 4 parameters for each vector, $\{\mu_{\ell}, \sigma_{\ell}, a_{\ell}, a_{\ell}\}$, are all unknown. The primary goal is to estimate the normal parameters.

You can find these data in the HW/hw1/ directory in the URL above. The R command source("dump data.R") will put the two vectors, x1 and x2 in your working directory.

- (a) Try a graphical approach.
- (b) Try an optimization-based approach.