

Problem Set 2: Written Part

Your Name Goes Here

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. This means it should be organized, and any Rmd files should knit to pdf without issue.

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.

Problem I: M&M's

Starting the week after next, we will be looking at Bayesian inference. The central idea in Bayesian statistics is that probability is used to express our state of knowledge about the value of an unknown parameter. The written part of this problem should take you 1 or 2 minutes, and is intended to introduce this way of thinking about probability.

Consider the question "What proportion of peanut M&M's are blue?" Note that the maker of M&M's has set the proportion of peanut M&M's that are blue, but they have not published this number. We will denote this proportion by θ . Although we don't know the value of θ , we can probably make some reasonable guesses about what it is based on our past experience with eating M&M's.

(1) What is your best guess at the proportion of M&M's that are blue? This should be a number between 0 and 1.

(2) Give an interval $[a, b]$ so that you think there is a 50% chance the proportion of M&M's that are blue is in that interval. In other words, you are picking numbers a and b between 0 and 1 so that you think $P(\theta \in [a, b]) = 0.5$. This is called a 50% credible interval. We'll return to this idea in class later.

(3) Give an interval $[c, d]$ so that you think there is a 90% chance that the proportion of M&M's that are blue is in that interval. In other words, you are picking numbers c and d between 0 and 1 so that you think $P(\theta \in [c, d]) = 0.9$. This is called a 90% credible interval. We'll return to this idea in class later.

Problem II: Method of Moments

(1) Suppose that $X_i \stackrel{\text{i.i.d.}}{\sim} \text{Geometric}(p)$ for $i = 1, \dots, n$. Find the method of moments estimator of p .

Note that there are multiple parameterizations of the Geometric distribution. Please use the parameterization as given on the "Common Probability Distributions" handout.

Problem III: Wind Speeds

This problem is adapted from an example in "Mathematical Statistics with Resampling and R" by Chihara and Hesterberg (2011), who write:

"[U]nderstanding the characteristics of wind speed is important. Engineers use wind speed information to determine suitable locations to build a wind turbine or to optimize the design of a turbine. Utility companies use this information to make predictions on energy availability during peak demand periods (say, during a heat wave) or to estimate yearly revenue.

The Weibull distribution is the most commonly used probability distribution used to model wind speed . . . The Weibull distribution has a density function with two parameters, the shape parameter $k > 0$ and the scale parameter $\lambda > 0$."

If $X \sim \text{Weibull}(k, \lambda)$, then it has pdf

$$f(x|k, \lambda) = \frac{kx^{k-1}}{\lambda^k} e^{-(x/\lambda)^k}$$

We will consider fitting a Weibull distribution to measurements of daily average wind speeds in meters per second at the site of a wind turbine in Minnesota over the course of 168 days from February 14 to August 1, 2010 (there were no data for July 2). Although these data are gathered over time, let's treat the measurements as independent (this is unrealistic but will make the problem more approachable).

We adopt the model $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \text{Weibull}(k, \lambda)$.

The likelihood function is

$$\begin{aligned} \mathcal{L}(k, \lambda|x_1, \dots, x_n) &= \prod_{i=1}^n \frac{kx_i^{k-1}}{\lambda^k} e^{-(x_i/\lambda)^k} \\ &= \frac{k^n}{\lambda^{kn}} \exp\left\{-\sum_{i=1}^n (x_i/\lambda)^k\right\} \prod_{i=1}^n x_i^{k-1} \end{aligned}$$

The log-likelihood is

$$\ell(k, \lambda|x_1, \dots, x_n) = n \log(k) - kn \log(\lambda) + (k-1) \sum_{i=1}^n \log(x_i) - \sum_{i=1}^n \left(\frac{x_i}{\lambda}\right)^k$$

The partial derivatives of the log-likelihood with respect to the unknown parameters are:

$$\frac{\partial}{\partial k} \ell(k, \lambda|x_1, \dots, x_n) = \frac{n}{k} - n \log(\lambda) + \sum_{i=1}^n \log(x_i) - \sum_{i=1}^n \left(\frac{x_i}{\lambda}\right)^k \log\left(\frac{x_i}{\lambda}\right) \quad (1)$$

and

$$\frac{\partial}{\partial \lambda} \ell(k, \lambda | x_1, \dots, x_n) = \frac{-kn}{\lambda} + \frac{k}{\lambda^{k+1}} \sum_{i=1}^n x_i^k \quad (2)$$

Setting (2) equal to 0 and solving for λ gives

$$\lambda = \left[\frac{1}{n} \sum_{i=1}^n x_i^k \right]^{1/k} \quad (3)$$

Substituting this into (1) and setting it equal to 0 gives

$$\frac{1}{k} + \frac{1}{n} \sum_{i=1}^n \log(x_i) - \frac{\sum_{i=1}^n x_i^k \log(x_i)}{\sum_{i=1}^n x_i^k} = 0$$

This equation cannot be analytically solved for k ; numerical optimization methods must be used to maximize the log-likelihood. Note that if we plug (3) back into the log-likelihood function, we obtain a function of just k to maximize:

$$\begin{aligned} \tilde{\ell}(k | x_1, \dots, x_n) &= n \log(k) - kn \log \left[\left\{ \frac{1}{n} \sum_{i=1}^n x_i^k \right\}^{1/k} \right] + (k-1) \sum_{i=1}^n \log(x_i) - \sum_{i=1}^n \left(\frac{x_i}{\left\{ \frac{1}{n} \sum_{i'=1}^n x_{i'}^k \right\}^{1/k}} \right)^k \\ &= n \log(k) - n \log \left[\frac{1}{n} \sum_{i=1}^n x_i^k \right] + (k-1) \sum_{i=1}^n \log(x_i) - \sum_{i=1}^n \frac{x_i^k}{\frac{1}{n} \sum_{i'=1}^n x_{i'}^k} \\ &= n \log(k) - n \log \left[\frac{1}{n} \sum_{i=1}^n x_i^k \right] + (k-1) \sum_{i=1}^n \log(x_i) - n \end{aligned}$$

This is not the log-likelihood function, but if we find the value of k that maximizes $\tilde{\ell}(k | x_1, \dots, x_n)$, we will have found the value of k that maximizes the log-likelihood.

The first and second derivatives of $\tilde{\ell}(k | x_1, \dots, x_n)$ are:

$$\frac{d}{dk} \tilde{\ell}(k | x_1, \dots, x_n) = \frac{n}{k} - n \frac{\sum_{i=1}^n x_i^k \log(x_i)}{\sum_{i=1}^n x_i^k} + \sum_{i=1}^n \log(x_i)$$

and

$$\frac{d^2}{dk^2} \tilde{\ell}(k | x_1, \dots, x_n) = \frac{-n}{k^2} - n \frac{\sum_{i=1}^n x_i^k \{\log(x_i)\}^2}{\sum_{i=1}^n x_i^k} - n \frac{\{\sum_{i=1}^n x_i^k \log(x_i)\}^2}{\{\sum_{i=1}^n x_i^k\}^2}$$

(1) Write down a complete statement of a Newton algorithm for finding a maximum likelihood estimate of k by maximizing the function $\tilde{\ell}$. Your algorithm will run until either a user-specified maximum number of iterations is reached or a user-specified tolerance is reached for the magnitude of the change in estimates of k on consecutive iterations. Include a specific statement of how the estimate of k will be updated in each iteration of the algorithm. (You do not need to do any new calculus for this – all the results you need are above.)

(2) Once the algorithm you specified in part a has finished running, how could you use the results to find the maximum likelihood estimate of λ ? (You can basically just write down one of the equations derived above.)

(3) Write a 1 or 2 paragraph explanation of what Newton's method is in the context of optimization. Please explain what the goal of the method is, how the method works, and how it relates to Taylor's theorem. You should include a supporting illustration (this can be drawn by hand).