Homework 2: Written 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.

Problem I: Method of Moments

(1) Suppose that $X_i \overset{\text{i.i.d.}}{\sim}$ 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.

- The first theoretical moment for X is $\mu_1 = E(X) = \frac{1-p}{p}$.
- The first sample moment is $\frac{1}{n}\sum_{i=1}^n X_i = \bar{X}$.

We set the first sample moment equal to the first theoretical moment and solve for the unknown parameter, p:

$$
\begin{aligned}
\bar{X} &= \frac{1-p}{p} \\
\Rightarrow p\bar{X} &= 1-p \\
\Rightarrow p\bar{X} + p &= 1 \\
\Rightarrow p(\bar{X}+1) &= 1 \\
\Rightarrow \hat{p}^{MOM} &= \frac{1}{\bar{X}+1}
\end{aligned}
$$

Note: We only need one equation (and one sample moment) because we only have one unknown parameter we are trying to estimate.

Problem II: 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 engery avaiabilty 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 $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \text{Weibull}(k, \lambda)$.

The likelihood function is

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

The log-likelihood is

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

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

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

and

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

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

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

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

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

(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.) Note: you are not being asked to actually program the algorithm.

1. Let α be the set number of maximum iterations and let ϵ be the set tolerance. Choose an initial estimate k_0 for k (this is often chosen using method of moments to get a good starting value).
2. Then, update the estimate of k to k_1, using the following:

$$k_1 = k_0 - \frac{\frac{\partial}{\partial k}\ell(k_0|x_1, ..., x_n)}{\frac{\partial^2}{\partial k^2}\ell(k_0|x_1, ..., x_n)}$$

to calculate the updated estimate of k (which maximizes $\ell(k_0|x_1, ..., x_n)$.

3. Repeat this process of updating the estimate of k for k_i, $i = 2, 3, ...$, continuing until we reach the maximum number of iterations, α, or until $|k_i - k_{i-1}| < \epsilon$, the set tolerance.

(2) Once the algorithm you specified in part (1) 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.) Suppose that k_m is the estimate we arrived at through our algorithm as the value of k that maximizes the second order Taylor series approximation to the log-likelihood. In this case, it also maximizes the log-likelihood, because the second order approximation will well-approximate the log-likelihood function at this point. Then, we can substitute the value from our algorithm into the solve equation for λ:

$$\hat{\lambda} = \left[\frac{1}{n}\sum_{i=1}^{n} x_i^{k_m}\right]^{1/k_m}$$

(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). The goal of Newton's method is to estimate the parameter value that maximizes the log-likelihood function by maximizing the second-order Taylor series approximation of the log-likelihood function centered around successive guesses of the unknown parameter value. We can write the second order Taylor series approximation expanded around θ_i:

$$P_2(\theta) = \ell(\theta_i|x_1, ..., x_n) + \frac{d}{d\theta}\ell(\theta_i|x_1, ..., x_n)(\theta - \theta_i) + \frac{d^2}{d\theta^2}\ell(\theta_i|x_1, ..., x_n)(\theta - \theta_i)^2.$$

We need to maximize $P_2(\theta)$. Under the condition that $\frac{d^2}{d\theta^2}\ell(\theta_i|x_1, ..., x_n)(\theta - \theta_i)^2 < 0$, we maximize a Taylor series by setting its first derivative equal to 0. Therefore

$$\theta_i = \theta_{i-1} - \frac{\frac{\partial}{\partial k}\ell(\theta_{i-1}|x_1, ..., x_n)}{\frac{\partial^2}{\partial k^2}\ell(\theta_{i-1}|x_1, ..., x_n)}$$

This gives us an updated estimate of θ, which maximizes the Taylor series approximation around θ_{i-1}. We repeat this process until we reach a maximum, within some tolerance ϵ ($|\theta_m - \theta_{n-1}| < \epsilon$) or reach the maximum number of iterations, in which case we have failed to find a MLE.