Assignment: Point Esimation, Confidence Interval, Bootstraps

Due March 16 11:59pm

Problem 1

Consider a parametric family of distributions with parameter $\alpha > 0$ with the following form:

$$f_x(x|\alpha) = \begin{cases} C(\alpha)x^{\alpha-1} & 0 < x < 1 \\ 0 & o/w \end{cases}$$

a) Calculate $C(\alpha)$ so that this is a proper density.
b) Calculate the expectation and variance of X.
c) Calculate the likelihood and log-likelihood of α, given data $X_1, ..., X_n$.
d) Calculate the MLE and method of moments estimators for α. (Food for thought, ungraded: what is the connection between the two functions of the data you get?)

Problem 2

Load the data from "normal_samples.csv." Using the formulas from class, calculate the MLEs for μ and σ. Calculate a two-sided 90% confidence interval for μ.

Problem 3

An overdispersed Poisson distribution can be thought of as when your sampling population consists of two distinct sub-populations, each of which follows its own Poisson distribution. This model has parameters $\lambda_1 > 0$ and $\lambda_2 > 0$ (the Poisson parameters of each of the subpopulations) and a third parameter $p$which is the proportion of the sampling population which is distributed according to the first Poisson distribution.

An example of this: suppose we are looking at the number of accidents on a given stretch of freeway in a day. On weekdays, we average 3 accidents, and on weekends, we average 1.2 - and if we look at just the weekdays and weekends seperately, the clearly have Poisson distributions. Then the overall distribution of daily accidents is an overdispersed Poisson which $\lambda_1 = 3$, $\lambda_2 = 1.2$ and $p = \frac{5}{7}$.

a) Given λ_1, λ_2, p, write down the pmf of an overdispersed Poisson. Use

$$P(X = i) = P(X = i|\lambda_1)P(\lambda_1) + P(X = i|\lambda_2)P(\lambda_2)$$

b) Given $X_1, ..., X_n$ write down the likelihood and log-likelihood.

c) Implement a function that calculates the log-likelihood of λ_1, λ_2, and p

d) Write a function that uses the function from part c and scipy.minimize to find the MLEs given data. You can hard-code the initial guess parameters to be $p = 0.5$, $\lambda_1 = 1$, $\lambda_2 = 2$.

e) Write a function that calculates bootstrap confidence intervals for each of the parameters (two-sided CIs, given level α).

e) Use the data from "accidents.csv" to calculate MLEs and 95% two-sided boostrap confidence intervals.