MA 677 Midterm Project

Due: 2024 March 28

MSSP's self-consciously generalist approach to statistics education attracts students with a variety of academic backgrounds who pursue an even wider variety of career paths after graduation. With its emphasis on "conceptual foundations", I have never been comfortable assigning MA677 grades based solely on exams comprised of problems to be solved. Projects offer students more meaningful opportunities to demonstrate their ability to tackle complex problems. Project that can provide a basis for grading are difficult to write. that may require time.

This midterm project assignment follows a whirlwind review of probability that emphasizes topics useful for statisticians. The project is organized as five problems each based on a different topic. You must respond to a total of three of the five listed topics, including the required problem on order statistics.

I hope that you will consider all five of the problems and present the ones that you found most interesting. Curiosity is the leavening of education. So, be curious and rise to the occasion!

Order statistics - *Required*

This problem is required because it requires you to combine mathematical and simulation approaches as we have discussed in bootcamp, MA615 and MA677. Your response to this topic should include deriving the distribution of order statistics and simulations that validate your analysis. Using the approach I demonstrated in class, derive the distribution of order statistics for the uniform, exponential and normal distributions. Use five order statistics – the minimum, maximum and points in between.. Discuss your results. Do you recognize the distribution of the order statistics from samples of the uniform distribution?

Markov chains

You probably remember that we discussed Markov chains in boot camp. Markov chains have been used to model a wide variety of applications. Using an application of your choice, explain how Markov chains work and produce an example in Python or R. Here are some examples with references:

(which can be presentation slides)

Psychology:
Affects affect affects: A Markov Chain

Google page ranks:
Markov chains and Google's PageRank algorithm

Google Page Rank and Markov Chains

Manufacturing:
Application of Markov chains in manufacturing: A review

Linguistics:
Markov Models Applications in Natural Language Processing: A Survey

A SURVEY OF MARKOV CHAIN MODELS IN LINGUISTICS APPLICATIONS

Exponential distribution

Start with the PDF of the exponential distribution. Derive the CDF and the MGF. Verify the mean and variance with the MGF. Also, find the third, fourth, and fifth moments about the mean. What do these moments mean? Is the exponential distribution related to other distributions? Consider, for example, the distribution of the sum of exponential variates. Use the MGF to find justify your conclusion.

Variance stabilizing transformations

We discussed an approach to determining a variance stabilizing transformation for the Poisson distribution using the delta method. Could that approach be applied to other distributions where the variance of the distribution is a function of, say, the mean? We also talked about transformation when we discussed functions of random variables. The last paper in the list below uses this approach, but not exclusively. In the second to the last paper,Bradley Efron focuses on how transformations affect the distributions encountered during modeling, citing a 1915 paper by Fisher. The papers below are intended to be a comprehensive list.

4.6: Data Transformations

distribution theory

Transforming Density Functions

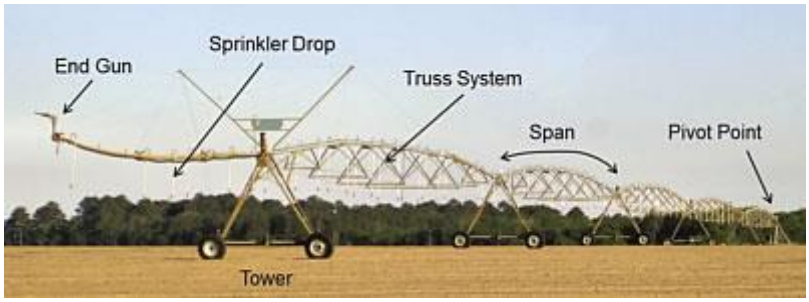
change of variable

Transformation Theory: How Normal is a Family of Distributions?

Variance stabilizing transformations of Poisson, binomial and negative binomial distributions

MA677 Midterm Project **Irrigation -- Those green circles**

Center Pivot Irrigation Systems are throughout the world and in about half the sprinkler irrigation systems in the U.S. Even if you haven't seen one up close, you have probably seen them from the air due to the distinctive visual pattern they create.



In simple terms, the system consists of a large arm, like the one pictured above, comprised of spans and towers. The arm is connected to a pivot point which provides power, water, and chemicals. The towers have powered wheels that propel the arm slowly around the pivot as water and chemicals are applied to the crops at controlled rates. Power required for each tower is about 1HP.

In the U.S., the large rotating arm is typically 1320 feet long (1/4 of a mile (440 yards) to the outer wheels. As shown in the picture above, there is usually an end-gun that extends about 100 feet beyond the outer wheels. Typical full rotation around the pivot takes slightly less than a day.

Center pivot systems are operated with a minimum amount of supervision, monitored with just a few simple statistics. One statistic is the rotation speed of the system. Although the speed is not directly observed, full rotation time for the system provides a way to calculate average rotation speed.

The file rot35.txt contains the rotation times for an irrigation system.

Calculate a 90% confidence interval for the speed of the rotating arm at the outer wheels.

Prepare a short explanation of your analysis for an audience that includes both data scientists and farmers.