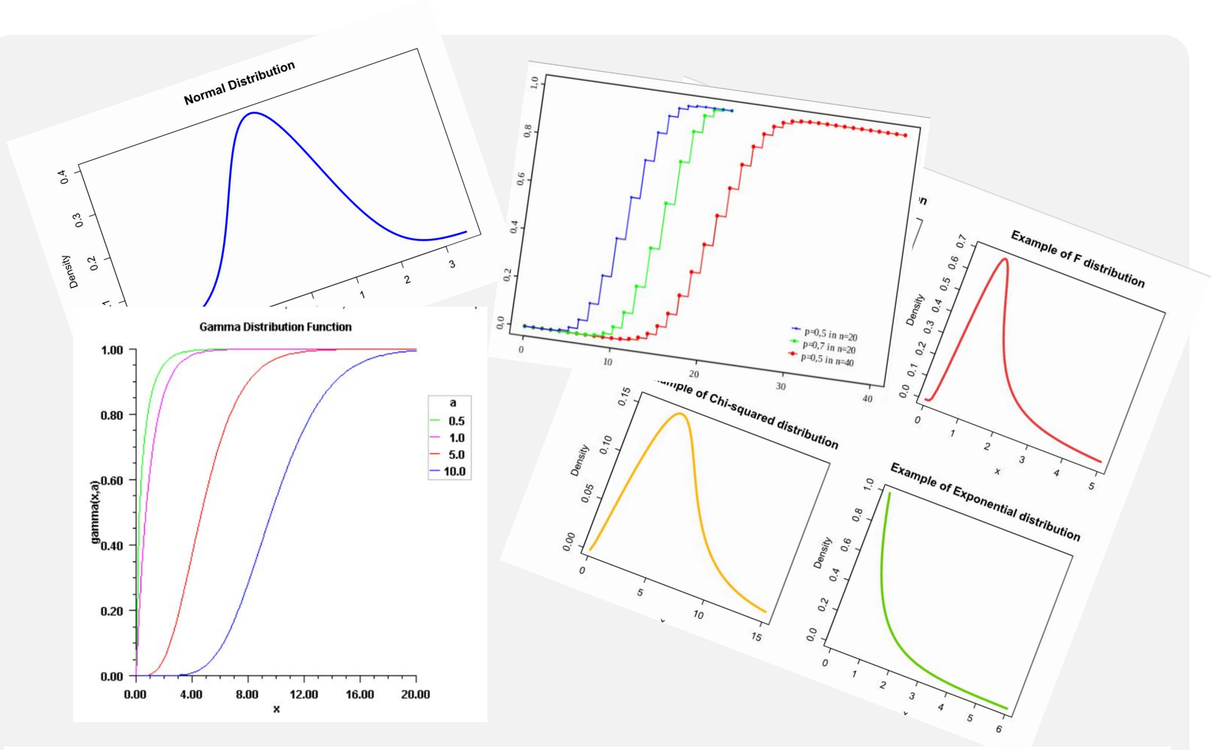
An Ultimate Guide to Statistical Distributions: An Interactive Visualization Tool for Statistics Students

Bohan Song, Dongyang Wang, Peihong Zhang, Wenjin Zhang

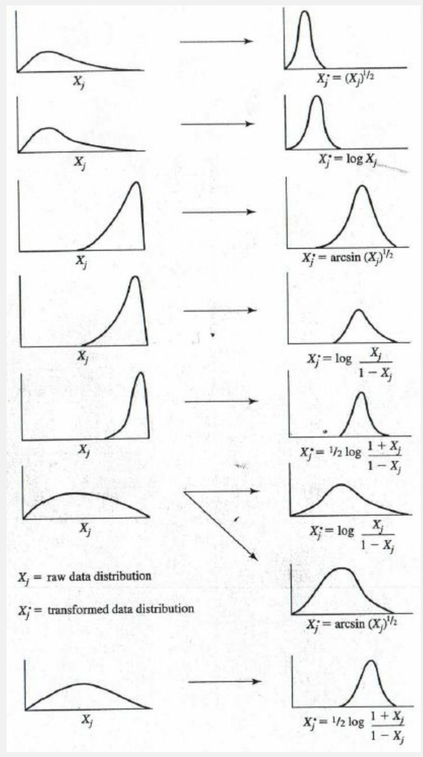
UW CSE 512: Data Visualization – Spring '22

Motivation



Familiarity with different distributions is essential for students in statistics, which narrows down to understanding the main properties of a distribution. Not only should one comprehend PMF/PDF, CDF, MGF, mean, variance in different parameters, but also be able to operate transformation between two related distributions. Although there are dispersed online resources regarding this, there is no summative tool for comparison across different distributions and very few interactive visualization tools to clarify their relationships.

Design Goals



M.S. Statistical Theory Exam Syllabus

Background

This exam is a three-hour exam on statistical theory. It is assumed that all candidates will have a background corresponding to Statistics 512 and 513. The exam will typically consist of 6 questions on the following topics:

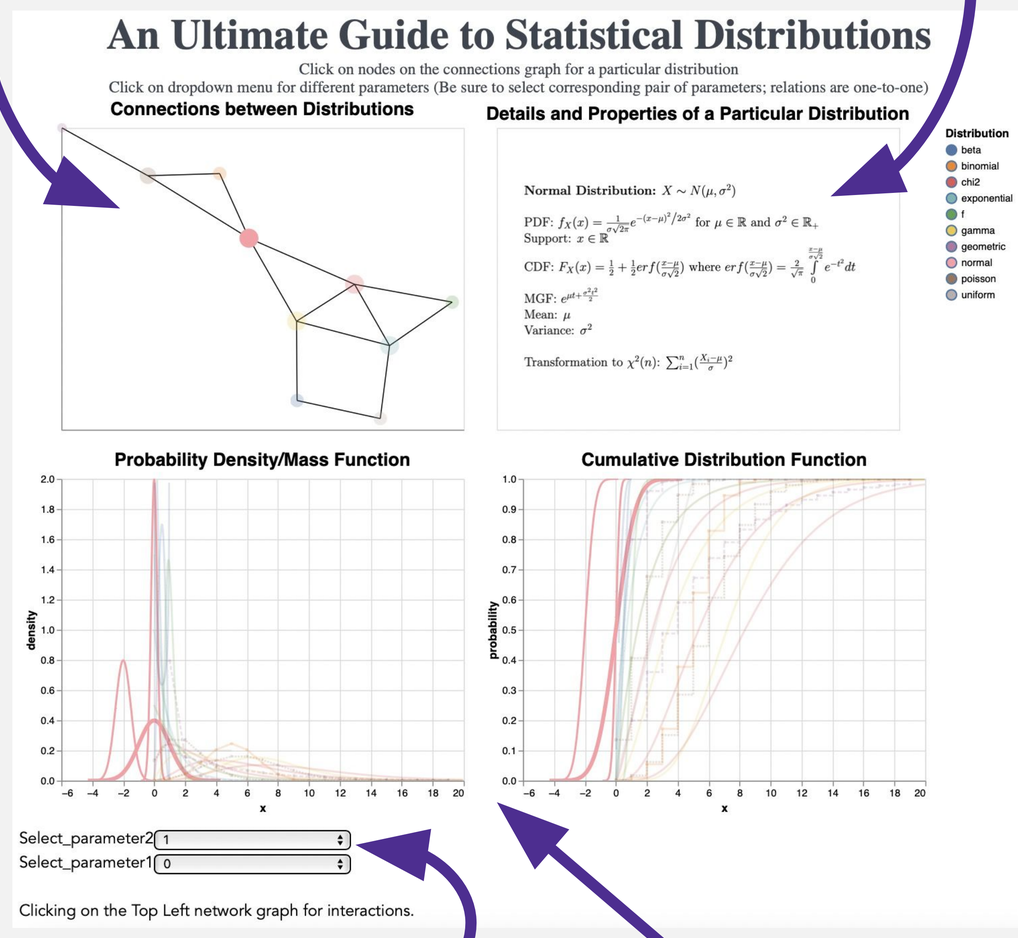
- Basic Probability theory
- Multivariable Models
- Sampling results and Asymptotic Theory
- Estimation
- Testing and Confidence Intervals
- Bavesian Methods

Therefore, our goal is to create a visualization tool to help statistics students understand how each of the commonly used distributions can be visualized and how they relate to each other. Ideally, this tool can also be used to prepare MS theory exams for MS statistics students.

Approach

Network Graph: click on the nodes

Details and properties



Dropdown menu for parameter selection

Distributions plots

1. Three Elements
   a. Distributions Plots: PMF/PDF & CDF
   b. Network Graph: Connections b/w distributions
   c. Text Information: Properties of distributions
2. Interaction
3. Synthesis of the elements & interactions

Evaluation/Feedback

Effectiveness:
The network graph gives the viewers a direct impression that these distributions are can mutually transform;
The visualization gives a good snapshot on commonly used distributions;
Expressiveness:
The adoption of network combining line chart makes the plot neat and fancy;
The selection using click and drop down menu gives a good interaction

The subtitle serving as instructions could be more eye-catching; (SOLVED)
Hierarchical drop down menu will make parameter selection more user-friendly;
The network graph cannot show discrete and continuous distribution differently;
The legend whose points and circles with different colors is a little confusing.

Department of STATISTICS

References:
1. https://www.statisticssolutions.com/transforming-data-for-normality/
2. https://en.wikipedia.org/wiki/Normal_distribution
3. https://en.wikipedia.org/wiki/Chi-squared_distribution
4. https://commons.wikimedia.org/wiki/File:Binomial_distribution_cdf_sl.svg