Modeling Risk and Realities: Week 3

Senthil Veeraraghavan
Operations, Information and Decisions Department
The Wharton School

Week 3: Choosing Distributions that fit your Data

- Data and visualization: Graphical representation

- Choosing among the family of distributions: Discrete and continuous distributions.

- How good does a certain distribution fit? Hypothesis testing and goodness of fit.

Week 3: Choosing Distributions that fit your Data

- Data and visualization: Graphical representation

Session 1

- Choosing among the family of distributions: Discrete and continuous distributions.

- How good does a certain distribution fit? Hypothesis testing and goodness of fit.

Week 3: Choosing Distributions that fit your Data

- Data and visualization: Graphical representation

- Choosing among the family of distributions: Discrete and continuous distributions.

Session 2

- How good does a certain distribution fit? Hypothesis testing and goodness of fit.

Week 3: Choosing Distributions that fit your Data

- Data and visualization: Graphical representation

- Choosing among the family of distributions: Discrete and continuous distributions.

- How good does a certain distribution fit? Hypothesis testing and goodness of fit.

Session 3

Week 1: Modeling Business Decisions in Low-Uncertainty Settings

- Data and visualization: Graphical representation

Session 1

- Choosing among the family of distributions: Discrete and continuous distributions.

- How good does a certain distribution fit? Hypothesis testing and goodness of fit.

Fitting Distributions to Your Data

- Many future events in the world are inherently uncertain.
  - At what value will a firm be acquired?
  - How many units of a new smartphone will be sold on its introduction?
  - How much dividend will a firm pay to its shareholders next quarter?
  - Will a firm meet its growth target next year? If not, by how much will it fall short of its target?
- As long as the uncertainty in the world is stationary (i.e., the future is not fundamentally different from the past), we can use past data to model the future realities.

How to Model Uncertainty?

- Real world uncertainty can be modeled using **random variables**.
- Some real world uncertainties may include:
  - Profit values from an investment
  - Percentage growth in revenues.
  - Customer orders for a new product.
  - Service time (for e.g. time to process a loan application).
- Random variable is a numerical description of such outcomes.
- Typically, (but not always), a random variable is denoted by a capital letter such as X or Y.
- Random variables can be discrete or continuous.
- We can characterize random variables by their **probability distribution**.

How to Model Uncertainty? (contd…)

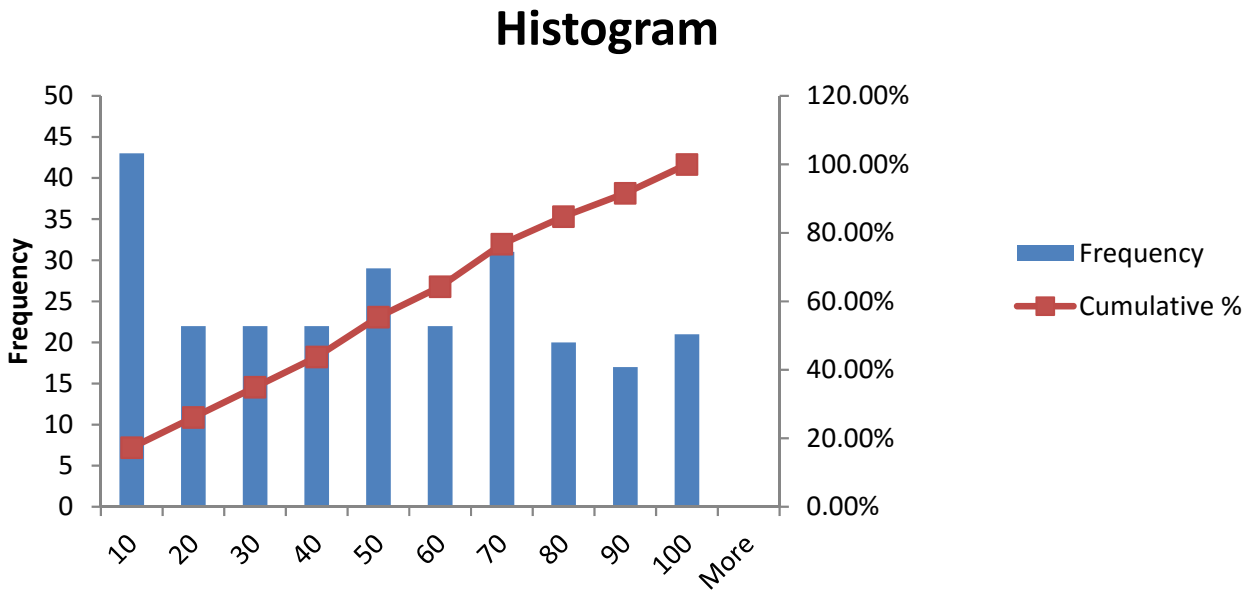
- Before we model random variables using distributions, we ask the following question.
- How important is it to model the distribution?
- I will make the case for choosing a correct distribution using two example datasets.
- You can find them on the course site as
  - Dataset1_Template.xlsx
  - Dataset2_Template.xlsx
- We generate histograms and descriptive parameters and save them as Dataset1_histogram.xlsx and Dataset2_histogram.xlsx

Understanding Uncertainty

- One way to understand and describe uncertainty is to visualize it.
- Generating a histogram is a good method to visualize data.
- A *histogram* is a graphical representation of the distribution of numerical data.
- For instance, a histogram helps to understand the shape and scale of the distribution of the random variable.
- As we go forward, we will examine a family of distributions and their parameters.

Dataset 1 Distribution

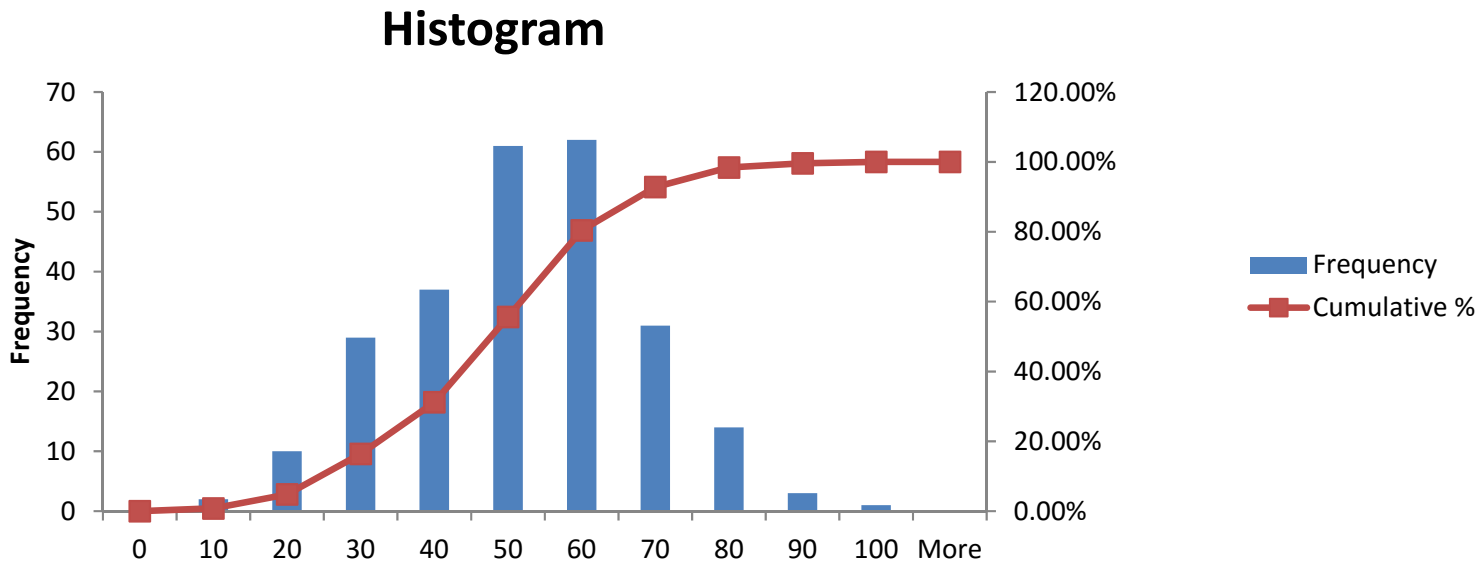
- Now we can examine the results of Dataset 1 with 250 data points.



- You can see in general low and high outcomes are more or less comparable in frequency to mid-level outcomes.

Dataset 2 Histogram

- Now we can examine the results of Dataset 1 with 250 data points.
  - However, the histogram looks quite different.



- You can see that low and high occurrences are less frequent than outcomes in the middle.
- Histogram is shaped like a “bell” curve.

How to Model Uncertainty?

- The datasets we saw are different – how can we model them differently?
- Real world uncertainty can be modeled using **random variables**.
- Typically, (but not always), a random variable is denoted by a capital letters such as X or Y.
- Random variables can be discrete or continuous.
- We can characterize random variables by their **probability distribution**.
- By visualization, one could argue that the two datasets have two different distributions.

Discrete and Continuous Random Variables

- Random variables can be discrete (i.e., countable).
  - Number of M&M candies in a box.
  - Number of shares of a firm that were sold in a day.
  - Number of investors who attended a meeting.
  - Number of cars that pass by a toll booth in an hour.

- Random variables can be continuous
  - The exact rainfall in a particular region during the monsoon season.
  - The heights of men and women living in a region.
  - Speeds of vehicles passing by a sign board.
  - Length of fabric thread yarned from a ton of cotton.

Describing a Random Variable

- A random variable is typically described by parameters. Some descriptors:
- **Mean**: represents the "average" value of the random variable over infinite number of repetitions.
- **Standard deviation**: measures the extent of "spread" of outcomes, or how "far" the random variable could be from the mean.
- **Variance**: square of standard deviation.
- **Median**: The value of the random variable at "mid-point" – the value of the random variable that lies in the middle of, or separates, the upper half from the lower half.
- **Mode**: the most likely value a random variable takes.

Distributions

- In the next session, we will look at some families of distributions that are often used to model realities.

- See you in Week 3 Session 2.