Statistics

Author:
Ming Li

Department:
Mathematics

December 24, 2016

1 Introduction

The probability theory mainly research random variables such as x. x could be some of the values with a measure call probability:

$$P(x = k), \quad P(x \in A), \quad \text{where} \quad A \quad \text{is a Boreal set}$$

The general study of statistics is to find out the distribution of x, hopefully, the probability distribution function or cumulative distribution function. Moreover, for multivariate variables, the task could be to find out the joint distribution($P(x_1, x_2, \ldots)$) or conditional distribution($P(x|y)$). Typically, the following steps:

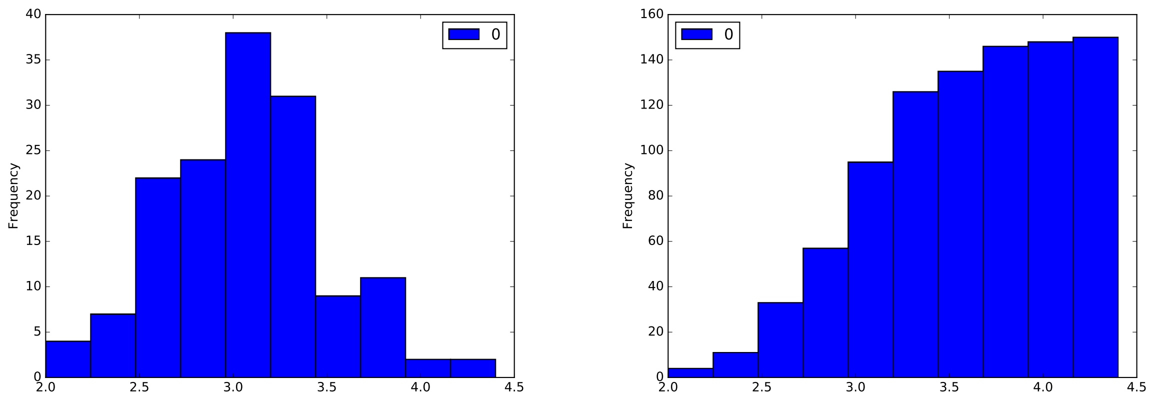
- Guess what should be the right distribution
- Estimate the parameters
- Prove your suppose(prove the distribution assumption or the parameters estimation is right)

So, this document is going to review the basic points of statistics follow the step: guess→estimate→prove.

2 Guess

2.1 Data perspective

Normally, the histogram chart can help us for a general perspective of the data probability distribution (data.hist(grid=...)) or cumulative distribution (data.plot.hist(cumulative=True)). And the estimation of probability distribution



(a) probability distribution

(b) cumulative distribution

Figure 1: Histogram

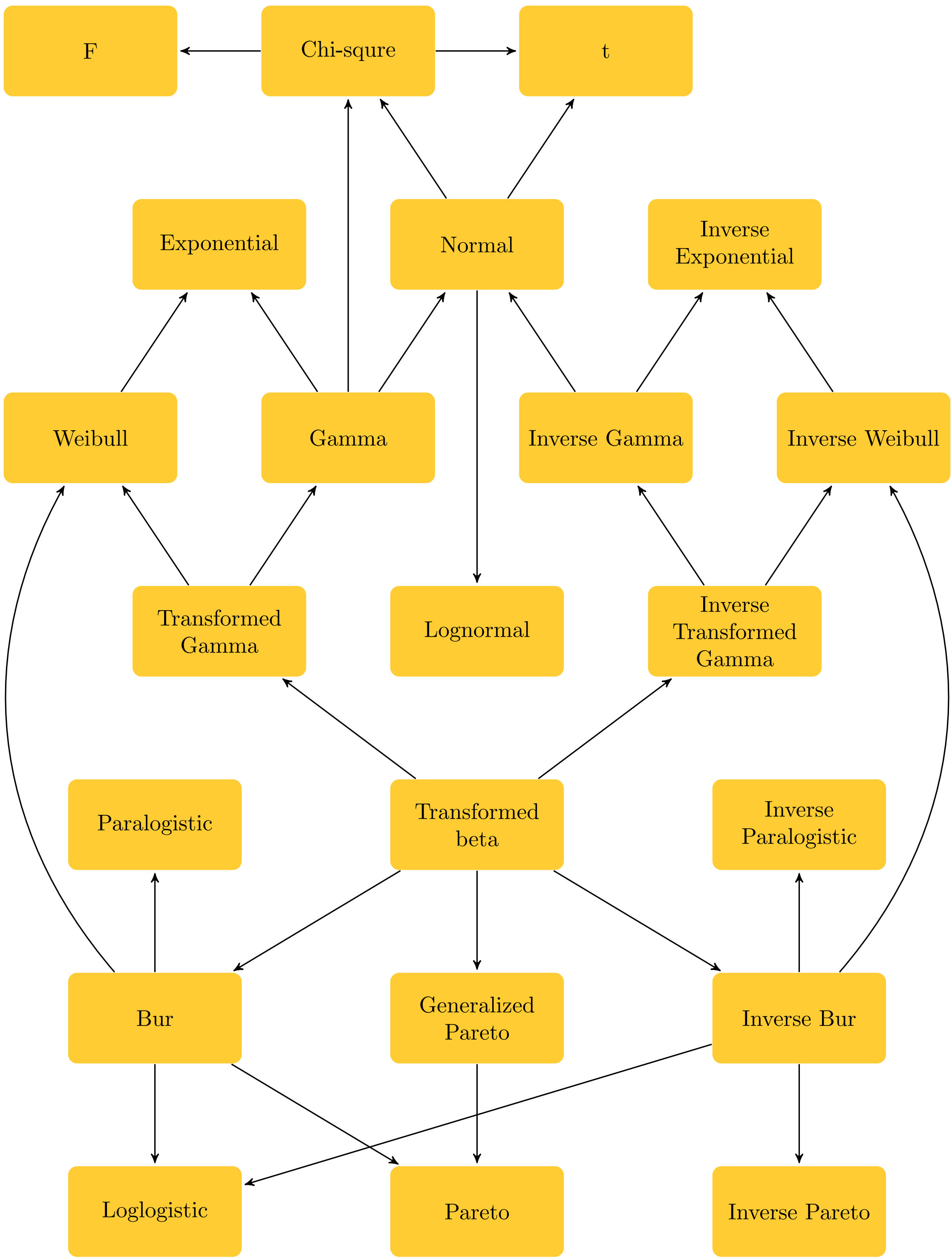
$$P(s_k < x \leq s_{k+1}) \approx \sum I(s_k < x_i \leq s_{k+1})/N$$

or cumulative distribution

$$P(x \leq s_{k+1}) \approx \sum I(x_i \leq s_{k+1})/N$$

2.2 Type of distributions

Three types of continuous distribution families are commonly used: Beta, Gamma and Normal. The following chart is the families relation.



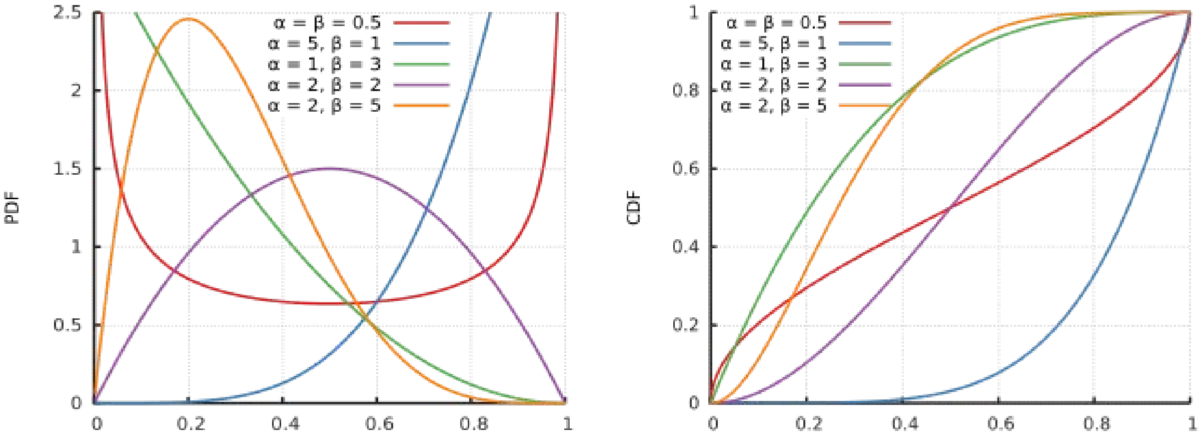
2.2.1 Beta distribution family

The PDF and CDF of Beta distribution (the commonly used version) are:

$$f(x;\alpha,\beta)=\frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)},\qquad F(x;\alpha,\beta)=\frac{B(x;\alpha,\beta)}{B(\alpha,\beta)}$$

where

$$B(\alpha, \beta) = \int_0^1 u^{\alpha}(1-u)^{\beta-1}, \qquad B(x;\alpha,\beta) = \int_0^x u^{\alpha}(1-u)^{\beta-1}$$



(a) PDF of Beta

(b) CDF of Beta

Figure 2: Beta distribution

3 Estimate

4 Prove