Nonparametric Bootstrap

Empirical cumulative distribution function (ecdf)

Def: For a batch of numbers $x_1, x_2, ..., x_n$ we define the ecdf as

$$F_n(x) = \frac{1}{n}(\#x \leq x). \tag{1}$$

Ex. Suppose we have the following batch of numbers: 1, 14, 10, 9, 11, 9. Draw the ecdf for this batch of numbers.

Introduction to the Nonparametric Bootstrap

The Nonparametric Bootstrap is useful for estimating the variability of a location parameter.

Note: nonparametric means distribution-free.

Suppose $x_1, x_2, ..., x_n$ are observations from an unknown distribution F, it is appropriate to investigate the variability and sampling distribution of a location estimate calculated from the sample.

Let $\hat{\theta} = \hat{\theta}(x_1, x_2, ..., x_n)$ be a location estimate.

Problems:

1. We do not know F and don't necessarily want to make a parametric assumption about F.
2. $\hat{\theta}$ may be complicated. (median, trimmed mean, etc.)

The nonparametric bootstrap proceeds as follows:

1. Use the ecdf $F_n(x)$ as an approximation to $F(x)$.
2. Take B samples of size n from $F_n(x)$, *with replacement*.
   (a) Estimate the sampling distribution of $\hat{\theta}$ by plotting a histogram of
   $$\hat{\theta}^*_1, \hat{\theta}^*_2, ..., \hat{\theta}^*_B \tag{2}$$
   (b) Estimate the variability by calculating the sd
   $$s_{\hat{\theta}} = \sqrt{\frac{\sum_{i=1}^{B}(\hat{\theta}^*_i - \bar{\hat{\theta}}^*)^2}{B-1}} \tag{3}$$

Example: Heat of Sublimation of Platinum. Find the sampling distribution of the mean, median, and 10% trimmed mean.

Reference:

Rice, *Mathematical Statistics, 3ed*, Duxbury, 2007.