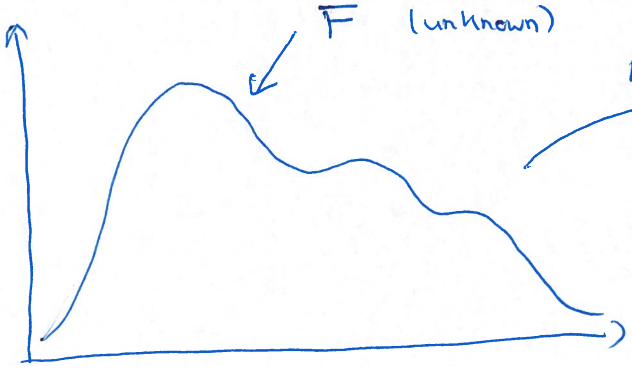
05.04.2016

Bootstrapping - rough idea

F (unknown)



$\xrightarrow{N}$ $x_1 = 2$, $x_2 = 5$, $x_3 = 7$, $\vdots$, x_N $\longrightarrow \hat{\theta} = s(\vec{x})$

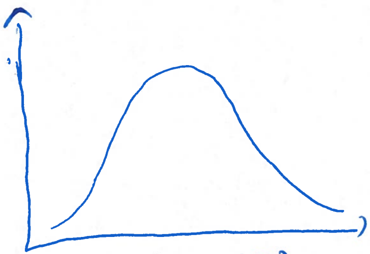
$$x^{1*} = \underbrace{(x_4, x_1, x_3, x_1, \ldots, x_N)}_{N}$$

$$\hookrightarrow s(\vec{x}^{1*}) = \hat{\theta}^{1*}$$

repeat a lot of times:

$$\begin{array}{ccccc} s(\vec{x}^{1*}) & s(\vec{x}^{2*}) & & & s(\vec{x}^{B*}) \\ \uparrow & \uparrow & & & \uparrow \\ \vec{x}^{1*} & \vec{x}^{2*} & & & \vec{x}^{B*} \\ \nwarrow & \uparrow & \cdots & & \nearrow \\ & & \vec{x} = x_1, x_2, \ldots x_N & & \end{array}$$

$\longleftarrow$ sampling with replacement



$s(\vec{x}^*)$

$$\sigma_B^2 = \frac{\sum_{i=1}^{B} \left(s(x^{i*} - \overline{s(x^*)})\right)^2}{B-1}$$

$$\lim_{N\to\infty} \left(1 - \frac{1}{N}\right)^N = e^{-1} \approx 0.367 \Rightarrow 0.633$$

to be included in the bootstrap sample

$$\Theta = E(X)$$

$$\hat{\Theta} = s(\vec{x}) = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$$

$$\hat{\Theta}^* = s(\vec{x}^*) = \frac{1}{n}\sum_{i=1}^{n} x_i^* = \bar{x}^*$$

Ideal bootstrap estimate (using the empirical distribution function):

$$SD_{\hat{F}}(\hat{\Theta}^*) = SD_{\hat{F}}\left(\frac{1}{n}\sum_{i=1}^{n} x_i^*\right)$$

$$= \sqrt{Var_{\hat{F}}\left(\frac{1}{n}\sum_{i=1}^{n} x_i^*\right)}$$

$$= \sqrt{\frac{1}{n^2}\sum_{i=1}^{n} Var_{\hat{F}}(x_i^*)} = \sqrt{\frac{1}{n} Var_{\hat{F}}(x_i^*)}$$

$$Var_{\hat{F}}(x_i^*) = \sum_{i=1}^{n} (x_i - \underbrace{\bar{x}}_{E_{\hat{F}}(x_i^*)})^2 \cdot \underbrace{\frac{1}{n}}_{p(x_i)}$$

Each bootstrap draw can take n possible values $(x_1, \dots, x_n)$ and each with prob $\frac{1}{n}$

$$\Rightarrow \quad SD_{\hat{F}}(\hat{\Theta}^*) = \sqrt{\frac{1}{n^2}\sum_{i=1}^{n}(x_i - \bar{x})^2} = \sqrt{\frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}}$$

Note: Here, we could compute the ideal bootstrap estimate analytically. However, in most cases we cannot do this

Note: Except for the factor $\frac{1}{n}$ instead of $\frac{1}{n-1}$ this is the usual estimate for the standard error.