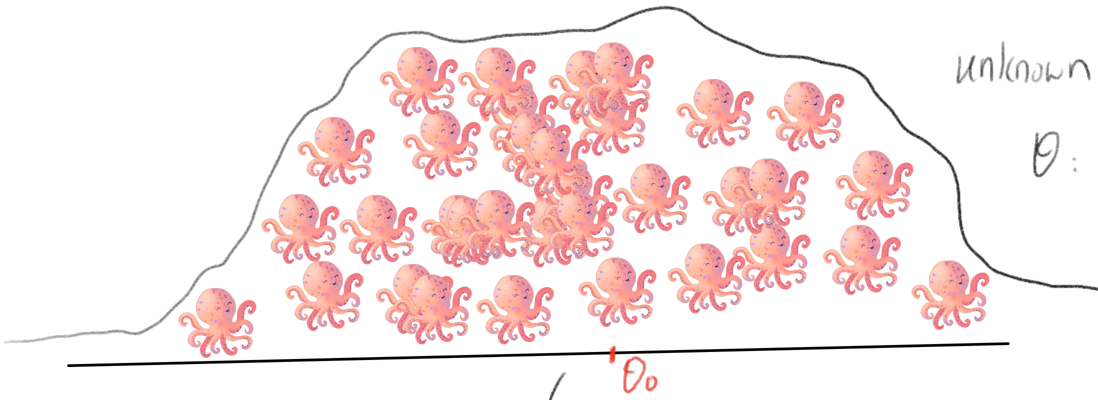
STA 104 Applied Nonparametric Statistics

Chapter 6: Bootstrap

Xiner Zhou

Department of Statistics, University of California, Davis

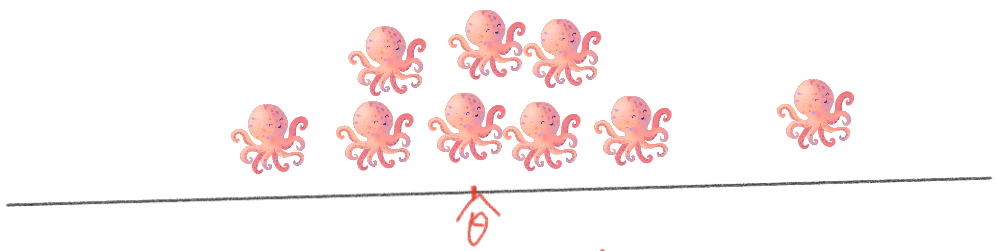
Table of contents



unknown population

θ: typical value

observe one sample



how good is your estimate $\hat{\theta}$ for the truth θ_0?

① typically how far away your estimate $\hat{\theta}$ from θ_0?

$SE(\hat{\theta})$

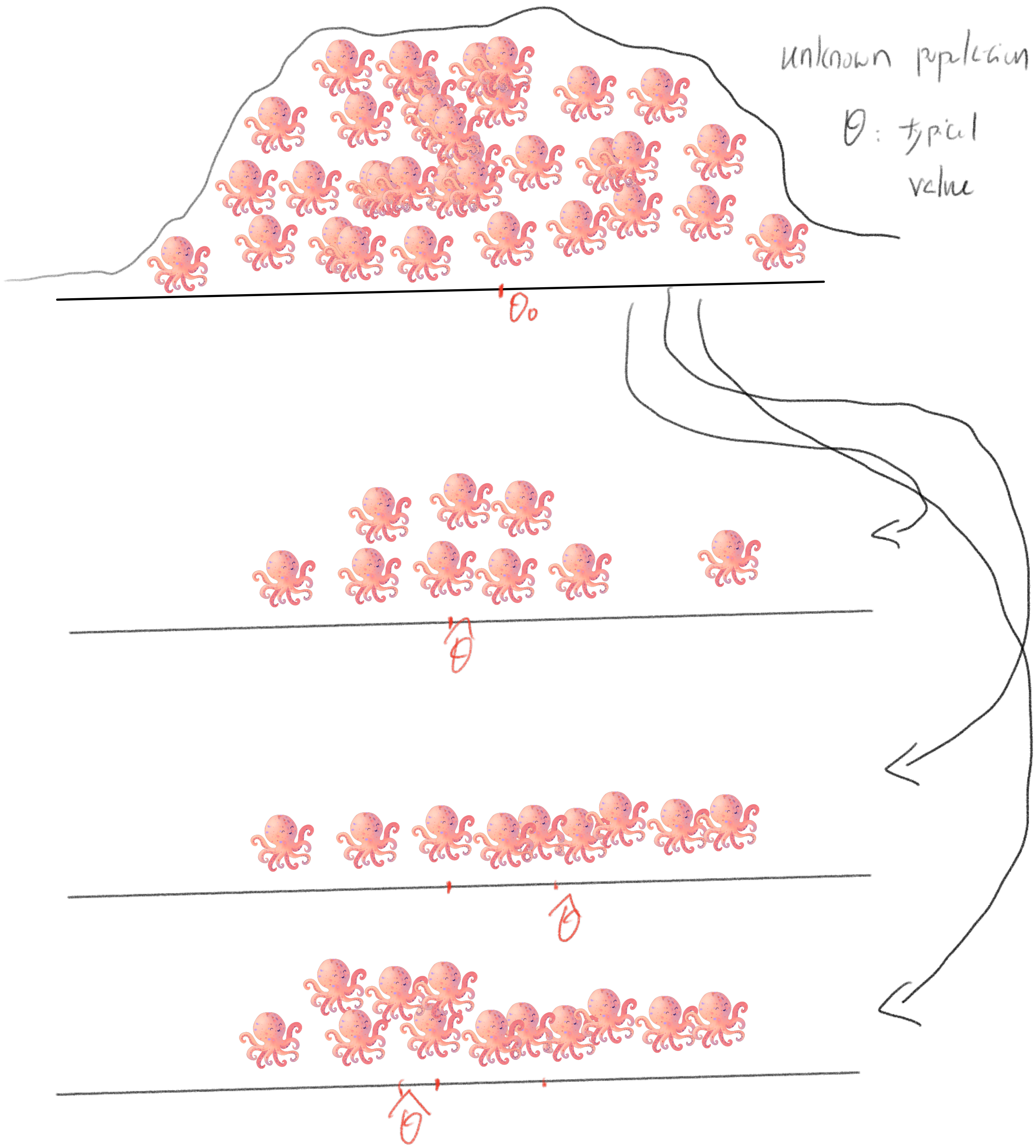
② Confidence interval: give an interval that contains the truth θ_0 almost surely / with very high probability?

two measures to assess the "quality" of your estimate

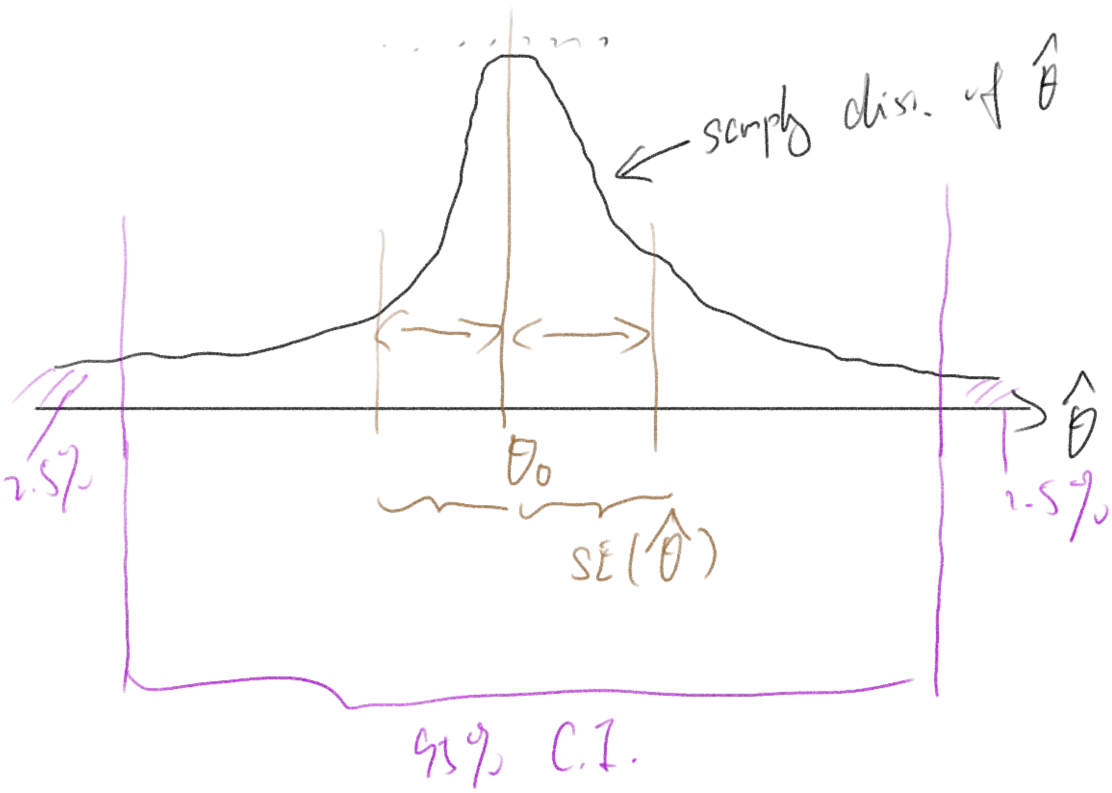
Frequentist's way = Repeated Sampling idea:

if we could repeatedly draw random samples from the underlying population, and estimate $\hat{\theta}$ each time.

⇒ "sampling distribution" of $\hat{\theta}$ describes its behavior!



$\Rightarrow$ repeatedly 1,000,000 times



Problem :

We can't repeated sampling !

Example : One - sample mean μ

① $X_1 \cdots X_n \sim N(\mu, \sigma^2)$

$\hat{\theta} : \bar{X}_n$

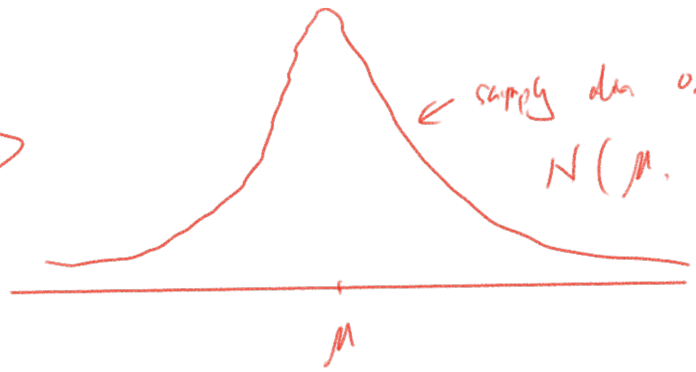
$SE(\hat{\theta}) = \sqrt{\frac{\sigma^2}{n}}$

CI for θ : $\bar{X}_n \pm 1.96 \sqrt{\frac{\sigma^2}{n}} \leftarrow s^2$

$t_{1-\alpha}$ if σ^2 unknown

$\boxed{\text{CLT : Central limit theorem}}$

$\Rightarrow$



$\leftarrow$ sampling dsn of $\bar{X} = \hat{\theta}$

$N(\mu, \frac{\sigma^2}{n})$

A central element of frequentist inference is uncertainty quantification through the standard error or confidence interval.

- no theoretical result
- large sample approximations
  - in many cases (such as with the sample median), requires knowledge about the underlying distribution unknown in real data situations.
- Direct standard error formulas exist for various forms of averaging (sample mean, linear regression), but for hardly anything else.

$\Rightarrow$ modern computer-intensive, nonformulaic, statistical method for estimating quantities like standard error and confidence interval

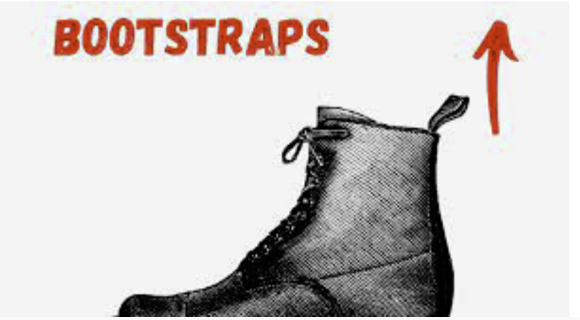


Figure 1: get oneself out of some situation using existing resources

Figure 2: The word "bootstrap" comes from an old story about a hero - Baron Munchausen - who is riding around on his horse in a forest and suddenly gets stuck in a swamp. He screams for help but there is no one around who hears his voice! Luckily our hero does not give up and gets a great idea: "what if I just pull myself out of this swamp?". He grabs the straps of his boots and pulls himself loose. Fantastic - he just invented bootstrapping. Physics-defying stories aside, bootstrapping has become a common term for something seemingly impossible or counterintuitive.

Why Bootstrap?

- modern computer power
- automates a wide variety of inferential calculations, including standard errors, confidence interval.
- sparing statisticians the exhaustion of tedious routine calculations
- opened the door for more complicated estimation algorithms, so that their accuracy would be easily assessed.

Bootstrap for Assessing the Quality of Estimators: Variance and Standard Error

Motivation

permutation

Bootstrap

Intuition: Resampling from your data to approximate resampling from a population.

- The standard error of an estimate $\hat{\theta} = s(\boldsymbol{x})$ is, ideally, the standard deviation we would observe by repeatedly sampling new versions of $\boldsymbol{x}$ from F.
- This is impossible since F is unknown.
- Instead, the bootstrap substitutes an estimate $\hat{F}$ for F and then estimates by direct simulation, a feasible tactic only since the advent of electronic computation.

Motivation

$\hat{\theta}$ is obtained in two steps: first $\boldsymbol{x}$ is generated by iid sampling from probability distribution F, and then $\hat{\theta}$ is calculated from $\boldsymbol{x}$ according to algorithm $s(\cdot)$,

$$F \xrightarrow{\text{iid}} \boldsymbol{x} \xrightarrow{s} \hat{\theta}.$$

We don't know F, but we can estimate it by the empirical probability distribution $\hat{F}$ that puts probability $1/n$ on each point x_i.

Bootstrap replications $\hat{\theta}^*$ are obtained

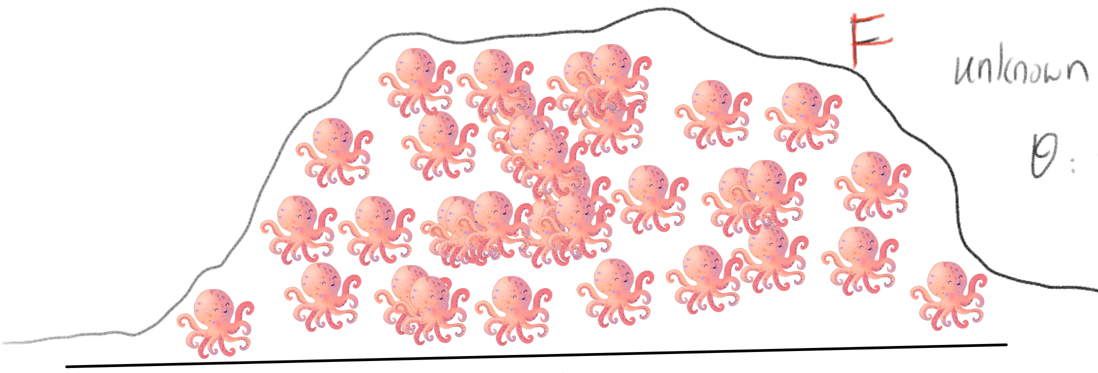
$$\hat{F} \xrightarrow{\text{iid}} x^* \xrightarrow{s} \hat{\theta}^*.$$

In the real world we only get to see the single value $\hat{\theta}$, but the bootstrap world is more generous: we can generate as many bootstrap replications $\hat{\theta}^{*b}$ as we want, or have time for, and directly estimate their variability.

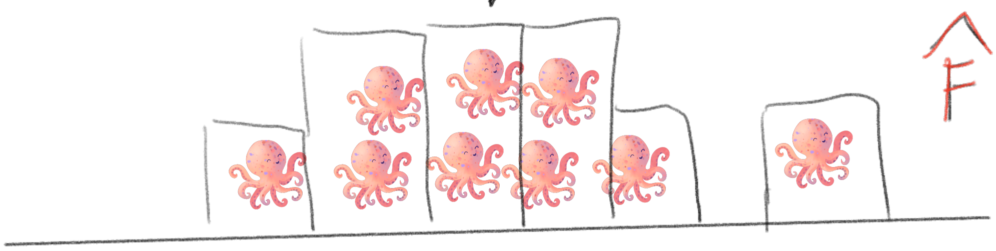
F

unknown population

θ: typical value



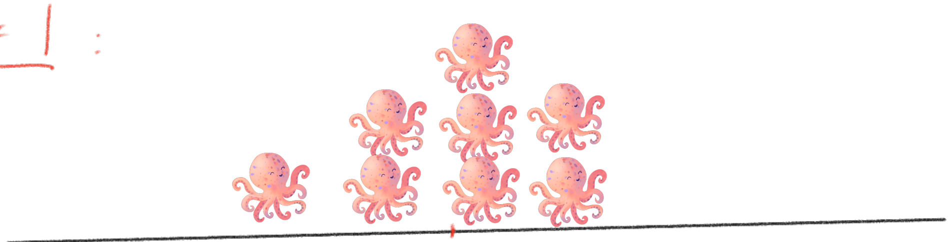
observe one sample



$\hat{F}$: empirical dsn of F

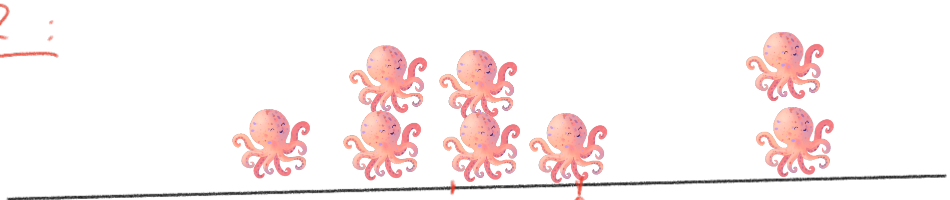
Bootstrap:

b=1:



$\hat{\theta}_1$

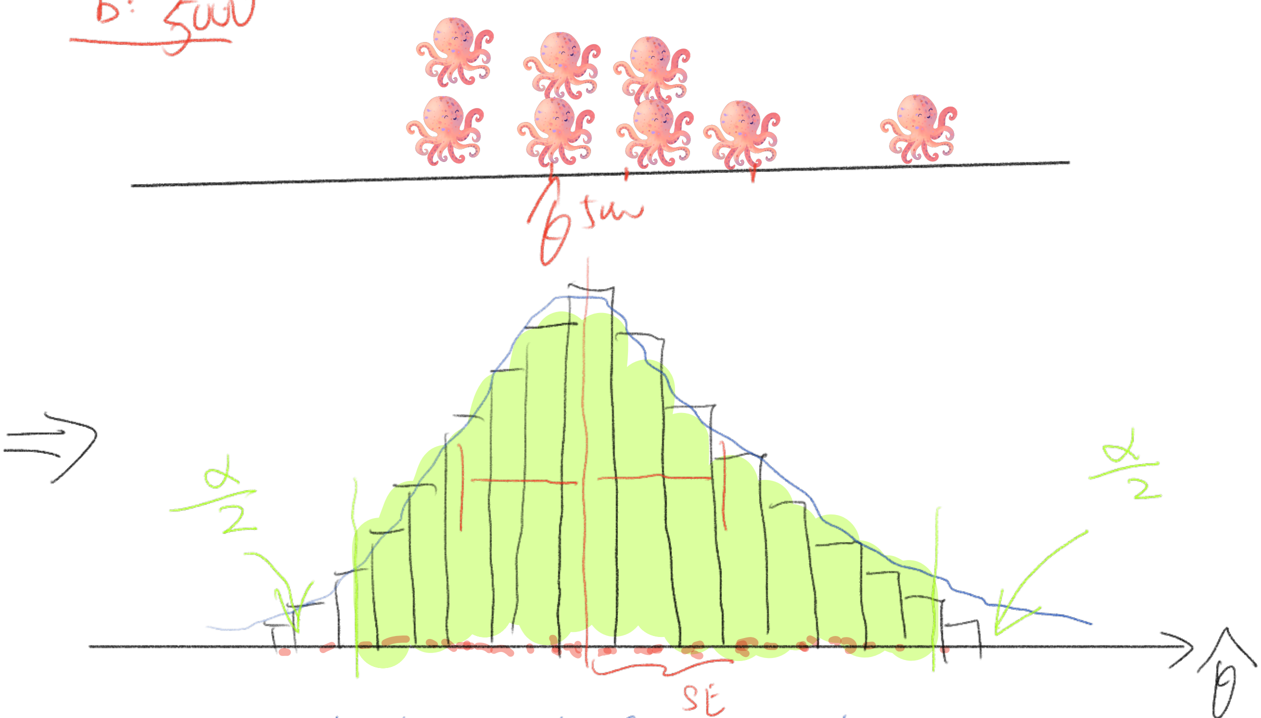
b=2:



$\hat{\theta}_2$

repeate for B=500 times:

b: 3000



distribution of Bootstrap replications

$\approx$ sampling distribution of $\hat{\theta}$

idea of Bootstrap CI:

Motivation

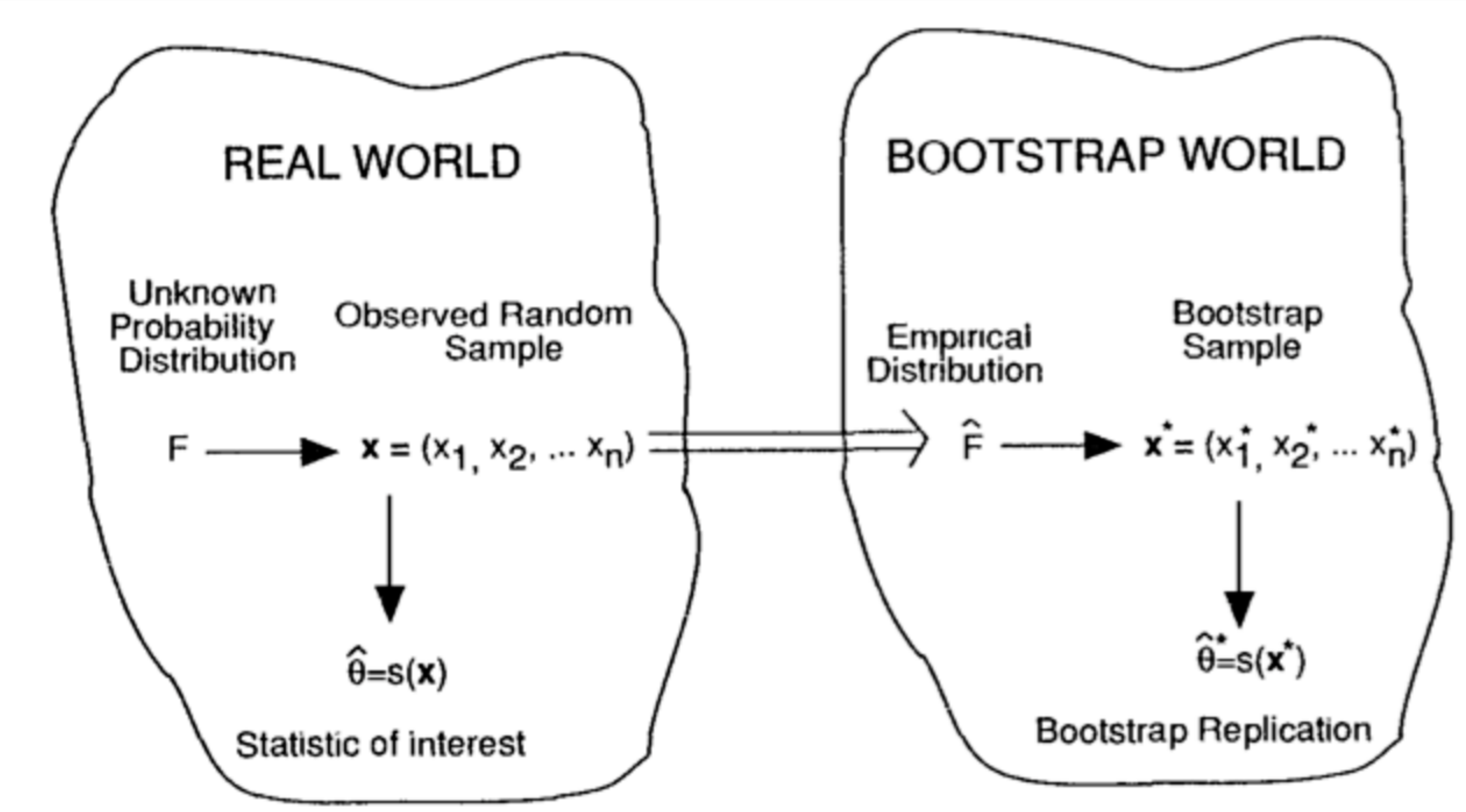


Figure 3: An Introduction to the Bootstrap (Efron Tibshirani, 1993).

Procedure

Bootstrap sample

The bootstrap estimate of standard error for a statistic $\hat{\theta} = s(\boldsymbol{x})$ computed from a data set $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ begins with the notion of a bootstrap sample

$$\boldsymbol{x}^* = (x_1^*, x_2^*, \ldots, x_n^*),$$

sample in original sample

where each x_i^* is drawn randomly with equal probability and with replacement from $\{x_1, x_2, \ldots, x_n\}$.

Each bootstrap sample provides a bootstrap replication of the statistic of interest,

$$\hat{\theta}^* = s(\boldsymbol{x}^*).$$

function (data)

Procedure

Bootstrap for SE

- Some large number B of bootstrap samples are independently drawn (say $B = 1000$). The corresponding bootstrap replications are calculated, say

$$\hat{\theta}^{*b} = s\left(\boldsymbol{x}^{*b}\right) \qquad \text{for } b = 1, 2, \ldots, B.$$

- The resulting bootstrap estimate of standard error for $\hat{\theta}$ is the empirical standard deviation of the $\hat{\theta}^{*b}$ values,

$$\widehat{\mathrm{se}}_{\mathrm{boot}} = \left[\underbrace{\sum_{b=1}^{B}\left(\hat{\theta}^{*b} - \hat{\theta}^{*\cdot}\right)^2 / (B-1)}_{\text{definition of SD}}\right]^{1/2} \approx SE(\hat{\theta}), \qquad \text{with } \boxed{\hat{\theta}^{*\cdot} = \sum_{b=1}^{B} \hat{\theta}^{*b} / B.}$$

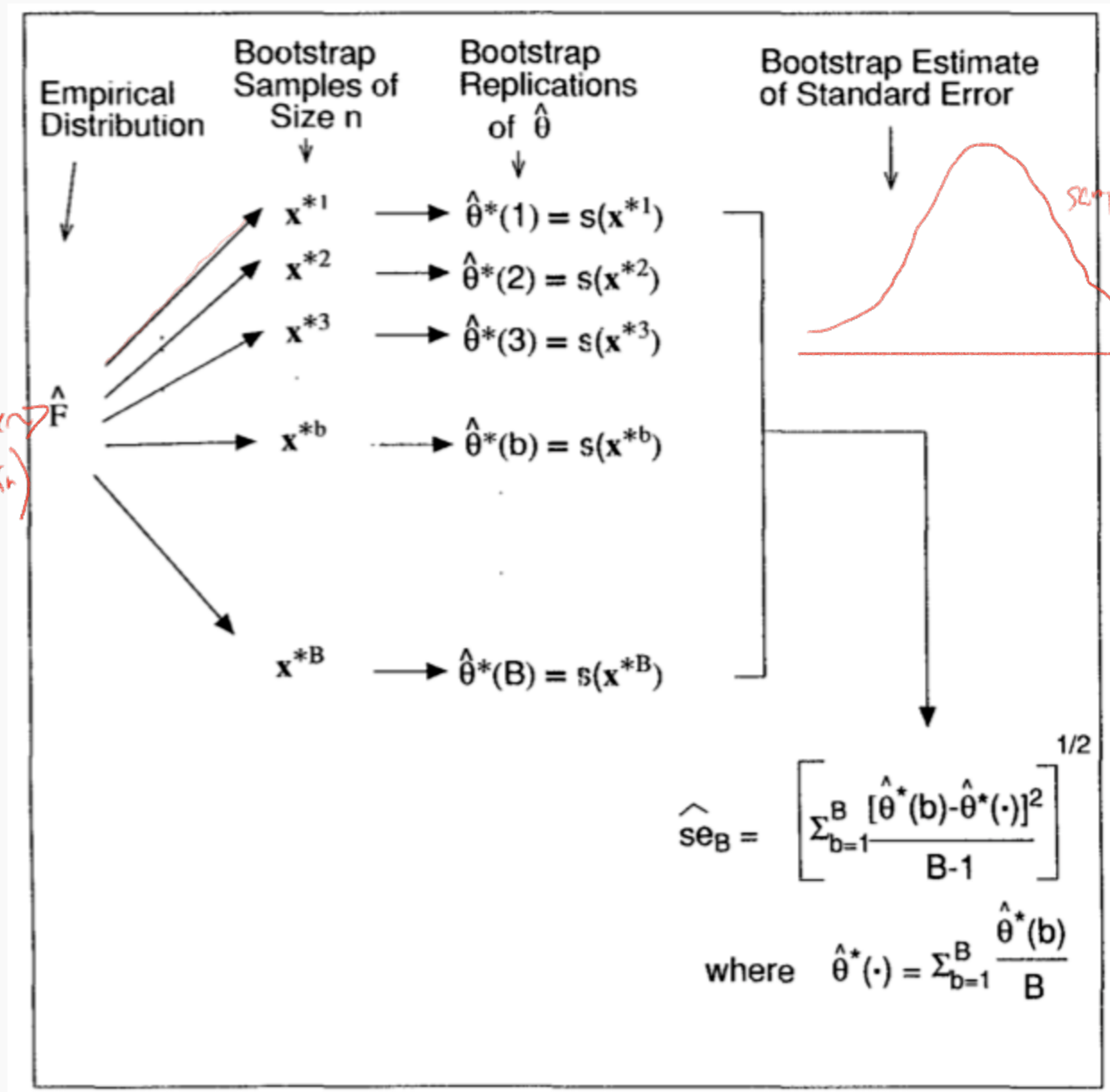
□ data : $X_1 \cdots X_n$

$$SD = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}}$$

$$\sqrt{\frac{\sum_{b=1}^{B}\left(\hat{\theta}^{*b} - \hat{\theta}^{*\cdot}\right)^2}{B-1}}$$

Procedure

F : unknown ~> sample ~> $\hat{F}$
$(X_1 \ldots X_n)$



sampling dist of $\hat{\theta}$