The Bootstrap Method

Statistical Intervals.

$P \pm CV \cdot SE$

P - statistic $(\bar{x}, s^2, \hat{\beta})$

CV follows a distribution
based the point estimate

$x_1, \ldots, x_n \sim N(\mu, \sigma^2)$

$\bar{x} \sim N(\mu, \sigma^2/n)$

$(1-\alpha)100\%$ CI for μ:

$\bar{x} \pm z_{\alpha/2}\, \sigma/\sqrt{n}$ $\quad \sigma$ is known

$z \sim N(0,1)$

σ is unknown

$\bar{x} \pm t_{n-1}\, s/\sqrt{n}$

$(1-\alpha)100\%$ CI of σ^2

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{(n-1)}$$

$$\left(\frac{(n-1)s^2}{\chi^2_{(n-1,\, 1-\alpha/2)}} \,,\, \frac{(n-1)s^2}{\chi^2_{(n-1,\, \alpha/2)}} \right)$$

Bigger Number

Smaller number

$$P(\underline{\chi^2_{(n-1,\,1-\alpha/2)}} < X) = 1-\alpha/2$$

Of the 3 intervals

What was on thing required? CV (Distribution)

How do we construct intervals when the distribution of your point estimate is unknown?

We utilize The Bootstrap Method

The Bootstrap Method is an approach that constructs the sampling distribution of a statistical estimator.

The main idea.

$\mathbb{X} = \{X_1, \ldots, X_n\} \sim F(\theta)$

The data $X_1, \ldots, X_n$ is generated from $F(\theta)$, $\boxed{\text{if}}$ n is large enough, all the information about $F(\theta)$ is contained in the data $\mathbb{X}$

Therefore, the $\hat{F}_n$ will be an excellent estimator of $F(\theta)$

$$\lim_{n \to \infty} \hat{F}_n \to F(x; \theta)$$

The empirical estimator of $F(\theta)$

$$\hat{F}_n(x) = \begin{cases} 0 & x < x_{(1)} \\ \frac{i}{n} & x_{(i)} \le x \le x_{(i+1)} \\ 1 & x > x_{(n)} \end{cases}$$

$x_{(1)}$ — minimum value

$x_{(n)}$ — max

$X_{(i)}$: ith order statistic

Glivenk - Cantelli Theorem.

$$\hat{F}_n \longrightarrow F(\theta)$$

as $n \to \infty$

If n is sufficiently large then $\hat{F}_n$ has the same information as $F(\theta)$.

$$\underbrace{X_1, \ldots\ldots, X_n}_{\mathbb{X}} \overset{iid}{\sim} F(\theta)$$

If we collect sample, with replacement from $\mathbb{X}$

It is like sampling from $F(\theta)$

Each new sample from $\mathbb{X}$, with replacement, is being sampled by $\hat{F}_n$.

The Bootstrap Method

Given a large sample of $\mathbb{X}$
Data is iid sample size n

1. We are going to create a new sample called $\mathbb{X}_B^{(1)}$ from $\mathbb{X}$.
   1. We draw a datapoint from $\mathbb{X}$, record it, and put it back in $\mathbb{X}$
   2. Repeat 1. until $\mathbb{X}_B^{(1)}$ is

the same size a $\mathbb{X}$ (n)

(This is known as a Bootstrap sample)

② Using $\mathbb{X}_B^{(1)}$ compute youre desired statistic $T_B^{(1)}$

③. Repeat steps ① and ② until you have B sample and statistics.

1. What is B?
2. $B = n$
3. As much as possible

④. $T^* = (T_1^{(1)}, T_2^{(2)}, \ldots, T_B^{(B)})$

$$\mathbb{X} = \{X_1, \ldots, X_n\} \sim F(\theta)$$

$$T(\mathbb{X}) \sim G(\tau)$$

$$T^* \sim \hat{G}_n$$

$$\hat{G} \to G(\tau)$$

T^* is the sampling Distribution of T!

To construct $(1-\alpha)100\%$ CI of θ

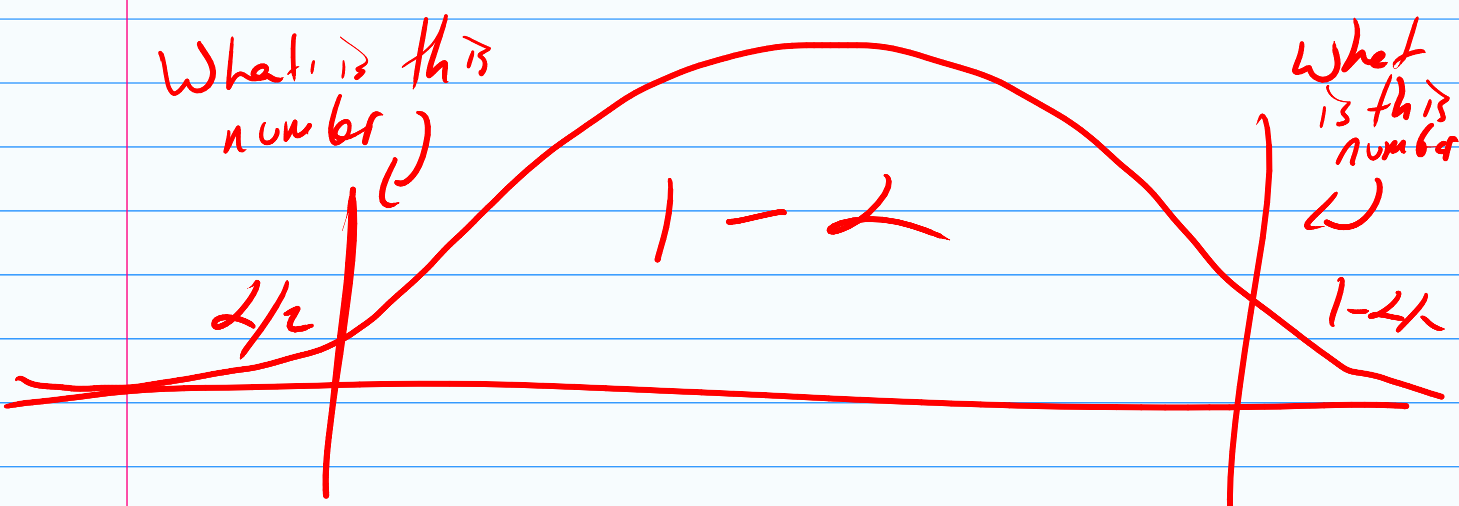
Obtain the $\alpha/2$th and $(1-\alpha/2)$th percentile from T^*

This will be your $(1-\alpha)100\%$ CI of θ

$$(T^*_{\alpha/2}, T^*_{1-\alpha/2})$$

Why set percentiles

We plot the density plot T^*



If you want the point estimate, use the original data.