The proposal is fully nonparametrical. The idea is to develope tools/approaches that can be used in all the cases in which we have no idea about the actual distribution of the data. However, we have some knowledge given by the data: we have a good approximation of the actual distribution: the empirical cumulative distribution function. We want a fully data driven approach relying only on the available sample. We want to learn about the distribution of the sample just by looking at the sample (more precisely, we're interested in finding the distribution of the test statistic just looking at the data).

— based on **LLN** (so generally we need very large sample size)

BOOTSTRAPPING (EFRON 1979)

THE PRIMARY TASK OF BOOTSTRAPPING IS ESTIMATING FROM A RANDOM SAMPLE THE DISTRIBUTION OF A STATISTIC (i.e, AN ESTIMATOR $\hat{\theta}$ OF AN UNKNOWN PARAMETER θ)

WHY? TO BUILD/COMPUTE IN MATHEMATICALLY COMPLICATED SETTINGS:

- STANDARD DEVIATION AND/OR BIAS OF $\hat{\theta}$
- CONFIDENCE INTERVALS FOR θ
- NULL HYPOTHESIS TESTS FOR θ

PLUG-IN PRINCIPLE

We suppose the data to come from an unknown distribution **F**. $\underline{X} = (X_1, .., X_n)$ is a random sample that we observe just once (:= realization).
<u>Inference</u>: we want to get some informations about some quantities related to **F** (mean, variance, percentiles, ..). We call the quantity of interest with θ. θ is a function of the unknown distribution F: $\theta(F)$.
<u>Estimation strategy</u>:
$\hat{\theta} = \hat{\theta}(x)$ ← it depends on the collected data.
$\hat{\theta}$ is modeled as a random variable depending on the random sample.
The realization of this random variable $\hat{\theta}$ is what we call our estimate of the parameter. The distribution is fixed and unknown, the random sample is considered as random (just one of the many samples that we can get).

UNKNOWN

$X = (X_1, X_2, \dots, X_m) \sim \text{iid } F$

$\theta = \theta(F)$ [FIXED]

$\hat{\theta} = \hat{\theta}(X)$ $\begin{bmatrix} \text{RANDOM} \\ \text{WRT } F \end{bmatrix}$

"REAL" WORLD

KNOWN
CONSIDERED AS FIXED AND DEPENDING ON x

$X^* = (X_1^*, X_2^*, \dots, X_m^*) \sim \text{iid } \hat{F}$

$\hat{\theta}(x) = \theta(\hat{F})$ [FIXED]

$\hat{\theta}^* = \hat{\theta}(X^*)$ $\begin{bmatrix} \text{RANDOM} \\ \text{WRT } \hat{F} \end{bmatrix}$

"BOOTSTRAP" WORLD

$\xleftarrow{\;m \to +\infty\;}$

The bootstrap world is our approximation of the real world. This approximation gets better and better as the sample size increases. The bootstrap world is easier to manage because the distribution is known (is the empirical cumulative distribution).

True distribution:
$\mathbf{F}(x) = \mathbb{P}(X_i \leq x)$
Empirical cumulative distr.:
$\hat{\mathbf{F}}(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{X_i \leq x\}}$

Everything works because:

└ standard statistic settings

LAW OF LARGE NUMBERS

$X_1, \dots, X_m \sim \text{iid } F$

$\hat{F}(x) \xrightarrow{a.s.} F(x) \quad \forall x \in \mathbb{R}$

$$F(x) = \mathbb{P}(X_i \leq x) \quad i = 1, \dots, m$$

$$\hat{F}(x) = \frac{1}{m} \sum_{i=1}^{m} \mathbb{1}_{\{X_i \leq x\}}$$

$$\boxed{\begin{array}{c} \hat{F} \xrightarrow{a.s.} F \\ \hat{\theta}(x) \xrightarrow{a.s.} \theta \\ \boxed{\hat{\theta}^* \xrightarrow{d} \hat{\theta}} \end{array}}$$

MOST IMPORTANT

Thanks to LLN we can say that the bootstrap world converges to the real world. So, if **n** is large enough working in one or the other settings is the same (with the large difference that in the bootstrap world $\hat{\mathbf{F}}$ is known, so we can do whatever we need).

The bootstrap estimator $\hat{\theta}$ is a function of the random sample drawn from the empirical distribution. What do we mean with this? it means that the bootstrap sample $\mathbf{X^*}$ is simply generated by picking randomly data among the real sample (with replacement). $X_1^* \sim \hat{\mathbf{F}}$ (where X_1^* is the first component of X^*) means that with probability $\frac{1}{N}$ we pick X_1, with probability $1/N$ we pick X_2, ... Same for $X_2^*, X_3^*, \dots, X_n^*$.
The bootstrap samples are simply generated by sampling independently with replacement from the original dataset.
For each bootstrap sample we have a different realization of the bootstrap estimator.

we underline continuous because in the case of discrete data we may have ties

Similarly to permutation tests, being the distribution uniform, we use **MC** estimation of the distr. Instead of going through all the possible n^n bootstrap samples X^*, we simply randomly pick some of these samples.

FOR CONTINUOUS RANDOM VARIABLES

NB THE BOOTSTRAP SAMPLE X^* IS UNIFORMELY DISTRIBUTED OVER THE n^n POSSIBLE VALUES OF THE BOOTSTRAP SAMPLE.

this is huge since we asked for large sample sizes (since bootstrap is based on LLN)

$\Rightarrow$ COMPUTING THE BOOTSTRAP DISTRIBUTION OF $\hat{\theta}^*$ IS THEORETICALLY VERY EASY BUT COMPUTATIONALLY PROHIBITIVE.

$\Rightarrow$ THE BOOTSTRAP DISTRIBUTION OF $\hat{\theta}^*$ IS USUALLY ESTIMATED VIA MC SAMPLING:

If, for example, $B = 1.000.000$ we generate 1.000.000 bootstrap samples that will generate 1.000.000 realizations of the bootstrap estimator ($\hat{\theta}^*$) and then we build the cumulative distribution.

$$\mathbb{P}_{\hat{F}}(\hat{\theta}^* \leq t) \approx \frac{1}{B}\sum_{b=1}^{B} \mathbb{1}\{\hat{\theta}(x_b^*) \leq t\}$$

random sample from the bootstrap distr.

WITH $X_b^* \sim iid\ (\hat{F})^n \quad b = 1, \dots, B$

(i.e. X_b^* IS OBTAINED BY RANDOMLY SAMPLING WITH REPLACEMENT FROM x)

EXAMPLE IN R: BOOTSTRAP ESTIMATION.R

BOOTSTRAP ESTIMATION OF MSE, VARIANCE AND BIAS OF $\hat{\theta}$

$$MSE_F(\hat{\theta}(x)) = \underbrace{E_F\left[\left(\hat{\theta}(x) - E_F(\hat{\theta}(x))\right)^2\right]}_{VAR_F(\hat{\theta}(x))} + \underbrace{\left[E_F(\hat{\theta}(x)) - \theta\right]^2}_{BIAS_F^2(\hat{\theta}(x))}$$

$MSE(\hat{\theta}) = \mathbb{E}[(\theta - \hat{\theta})^2]$

PLUG-IN PRINCIPLE

To enter the bootstrap world, we replace everything: $F \to \hat{F}$, $\theta \to \hat{\theta}(x)$, $\hat{\theta}(x) \to \hat{\theta}(x^*)$

1st level of approximation:

$$MSE_{\hat{F}}(\hat{\theta}(x^*)) = \underbrace{E_{\hat{F}}\left[\left(\hat{\theta}(x^*) - E_{\hat{F}}(\hat{\theta}(x^*))\right)^2\right]}_{VAR_{\hat{F}}(\hat{\theta}(x^*))} + \underbrace{\left[E_{\hat{F}}(\hat{\theta}(x^*)) - \hat{\theta}(x)\right]^2}_{BIAS_{\hat{F}}^2(\hat{\theta}(x^*))}$$

$\downarrow$ MC

2nd level of approximation:

$$\frac{1}{B}\sum_{b=1}^{B}\left[\left(\hat{\theta}(x_b^*) - \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}(x_b^*)\right)^2\right] + \left[\frac{1}{B}\sum_{b=1}^{B}\hat{\theta}(X_b^*) - \hat{\theta}(x)\right]^2$$

EXAMPLE IN R: BOOTSTRAP ESTIMATION.R

We have 2 approximations:

1st approx. : **PLUG-IN PRINCIPLE** : we approximate the real world with the bootstrap world (this approximation is related with **n**, which unfortunately is a fixed parameter)

2nd approx. : **MC APPROXIMATION** : empirical implementation in the bootstrap world via MC simulation. The quality of this approximation is given by a second parameter which is **B**. In the practice we make B as large as we can.

$\Longrightarrow$ **n** is the real constraint

Ways to make inference: point estimation, CI, hypothesis testing. Let's focus on CI.

Comment: permutational tests have hypothesis testing as the natural approach. In the bootstrap inference the approaches are more confidence intervals based. In permutational inference CIs come out hypothesis testing, in bootstrap inference it's the other way around: testing comes out of CIs (we can always build a test using CIs).

BOOTSTRAP CONFIDENCE INTERVALS → 3 approaches

CLASSICAL PARAMETRIC z-INTERVALS → CI for gaussian data with known variance

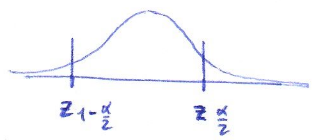
$$\mathbb{P}\left(z_{1-\alpha/2} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < z_{\alpha/2}\right) = 1-\alpha$$

$$\mathbb{P}\left(\underbrace{\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}_{\alpha/2 \text{ SUP QUANTILE OF } \bar{X}-\mu} < \mu < \underbrace{\bar{X} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}}_{\alpha/2 \text{ INF QUANTILE OF } \bar{X}-\mu}\right) = 1-\alpha$$

Important: in general the quantiles are $\neq$, here the distribution is simmetric ($z_{1-\alpha/2} = z_{\alpha/2}$) but in general it's not. What is important to know is that the left edge of the CI is droven by the right quantile, while the right edge of the CI is droven by the left quantile:

$$[\, \underbrace{.. z_{\frac{\alpha}{2}}}_{\text{right quantile}} ,\ \underbrace{.. z_{1-\frac{\alpha}{2}}}_{\text{left quantile}} \,]$$

NOTATIONS:



MORE IN GENERAL: (for non-simmetric cases)

$$\mathbb{P}\left(\hat{\theta} - (\hat{\theta}_{\alpha/2} - \theta) < \theta < \hat{\theta} - (\hat{\theta}_{1-\alpha/2} - \theta)\right) = 1-\alpha$$

---

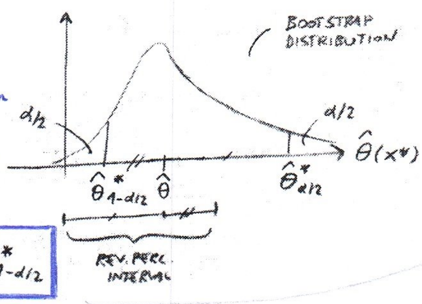
1. REVERSE PERCENTILE INTERVALS (BASIC BOOTSTRAP)

QUANTILES OF $(\hat{\theta}-\theta)$ ARE ESTIMATED VIA BOOTSTRAP (AND MC)

$$\hat{\theta}_{\alpha/2} - \theta \approx \hat{\theta}^*_{\alpha/2} - \hat{\theta}$$

$$\hat{\theta}_{1-\alpha/2} - \theta \approx \hat{\theta}^*_{1-\alpha/2} - \hat{\theta}$$

= we compute the bootstrap distribution of the estimator and then we compute the superior $\alpha/2$ quantile and the inferior $1-\alpha/2$ quantile



$$\boxed{2\hat{\theta} - \hat{\theta}^*_{\alpha/2} < \theta < 2\hat{\theta} - \hat{\theta}^*_{1-\alpha/2}}$$

Why like this? If the distr. has a heavy right tail (which means that we may have very large values) it makes sense to extend (make longer) the CI to the left.

Long-right-tail ⟹ CI elongated on the left

Long-left-tail ⟹ CI elongated on the right

NB THEY ARE ONLY ASYMPTOTICALLY EXACT

NB FOR SMALL SAMPLE SIZE THEY TYPICALLY HAVE A REAL COVERAGE MUCH SMALLER THAN THE NOMINAL ONE.

---

2. PERCENTILE INTERVALS (DO NOT USE THIS)

SIMPLER (BUT WRONG) EXCEPT FOR $\hat{\theta}$ WITH A SYMMETRIC DISTRIB.)

$$\boxed{\hat{\theta}^*_{1-\alpha/2} < \theta < \hat{\theta}^*_{\alpha/2}}$$

NB THEY SHOULD NOT BE CONSIDERED AS CI FOR θ BUT RATHER AS PREDICTION INTERVALS FOR $\hat{\theta}$!

NB THEY ARE SIMILAR TO RP-INTERVALS FOR SIMMETRIC we do not bootstrap the estimator, we bootstrap a pivotal quantity associated to the estimator

3. BOOTSTRAP T INTERVALS

THE IDEA IS VERY SIMILAR TO REVERSE PERCENTILE INTERVALS

CLASSICAL PARAMETRIC T INTERVALS:

$$\bar{X} - \underbrace{t_{\alpha/2}(m-1)}_{\substack{\frac{\alpha}{2}\text{ SUP. QUANTILE OF} \\ T=\frac{\bar{X}-\mu}{S/\sqrt{m}}}} \frac{S}{\sqrt{m}} < \mu < \bar{X} - \underbrace{t_{1-\alpha/2}(m-1)}_{\substack{\frac{\alpha}{2}\text{ INF. QUANTILE OF} \\ T=\frac{\bar{X}-\mu}{S/\sqrt{m}}}} \frac{S}{\sqrt{m}}$$

MORE IN GENERAL:

$$\hat{\theta} - \underbrace{\tilde{t}_{\alpha/2}}_{\substack{\frac{\alpha}{2}\text{ SUP. QUANT. OF} \\ T=\frac{\hat{\theta}-\theta}{\hat{\sigma}_{\hat{\theta}}}}} \hat{\sigma}_{\hat{\theta}} < \mu < \hat{\theta} - \underbrace{\tilde{t}_{1-\alpha/2}}_{\substack{\frac{\alpha}{2}\text{ INF. QUANT. OF} \\ T=\frac{\hat{\theta}-\theta}{\hat{\sigma}_{\hat{\theta}}}}} \hat{\sigma}_{\hat{\theta}}$$

$\hat{\theta}$ estimate

$\hat{\sigma}_{\hat{\theta}}$ estimate of the standard deviation of the estimator

$\tilde{t}_{\frac{\alpha}{2}}, \tilde{t}_{1-\frac{\alpha}{2}}$ quantiles of **T**

$\tilde{t}_{\alpha/2}$ AND $\tilde{t}_{1-\alpha/2}$ ARE ESTIMATED VIA BOOTSTRAP (AND MC)

$\tilde{t}_{\alpha/2} \approx t^*_{\alpha/2}$ $\qquad$ $t^*_{\alpha/2}$ IS $\alpha/2$ SUPERIOR BOOTSTRAP QUANTILE

$\tilde{t}_{1-\alpha/2} \approx t^*_{1-\alpha/2}$ $\qquad$ OF $\dfrac{\hat{\theta}^* - \hat{\theta}}{\hat{\sigma}^*_{\hat{\theta}}}$

$\hat{\theta}^*$ = estimator computed on the bootstrap sample

$\hat{\theta}$ = estimate associated to the original sample

$\hat{\sigma}^*_{\hat{\theta}}$ = estimate of the standard deviation of the estimator, computed on the bootstrap sample

$\Longrightarrow$ we need a way to estimate the standard deviation of the estimator

$$\boxed{\hat{\theta} - t^*_{\alpha/2}\, \hat{\sigma}_{\hat{\theta}} < \theta < \hat{\theta} - t^*_{1-\alpha/2}\, \hat{\sigma}_{\hat{\theta}}}$$

NB THEY ARE ASYMPTOTICALLY EXACT

NB THEIR COVERAGE IS USUALLY CLOSER TO $1-\alpha$ THAN RP INT.

reverse percentile

Summary

REVERSE PERCENTILE INTERVALS $\rangle$ BOOTSTRAPPING OF $\hat{\theta} - \theta$

BOOTSTRAP T-INTERVALS $\rangle$ BOOTSTRAPPING OF $\dfrac{\hat{\theta} - \theta}{\hat{\sigma}_{\hat{\theta}}}$

EXAMPLE IN R: BOOTSTRAPCI.R

BOOTSTRAP TESTS

BOOTSTRAP TESTS ARE TRIVIALLY BUILT FROM BOOTSTRAP C.I.
ALSO BOOTSTRAP P-VALUES ARE COMPUTED ACCORDINGLY.

IMPROVING BOOTSTRAP ESTIMATION

If we have some extra knowledge (than only the observed sample) we can use it to approximate better **F** (as $\hat{F}$)

NAIVE BOOTSTRAP

= resampling from the same sample (we're not using other informations but the ones coming from the original sample)

$$F(x) \approx \hat{F}(x) = \frac{1}{m}\sum_{i=1}^{m} \mathbb{1}_{\{x_i \le x\}}$$

(ALWAYS GOOD)

[MC: SAMPLE WITH REPLACEMENT FROM $x_1, \dots, x_m$]

SMOOTH BOOTSTRAP

→ we take the empirical distribution function and we perform a smoothing ⟹ **DENSITY ESTIMATION** starting from a set of points

$$F(x) \approx \int_{-\infty}^{x} \underbrace{\frac{1}{m}\sum_{i=1}^{m} K(x - x_i)}_{\text{A DENSITY CENTERED IN } x_i} \, dx$$

(e.g. $N(x_i, 1)$)

(GOOD WHEN F IS SMOOTH)

[MC: SAMPLE WITH REPLACEMENT FROM $x_1, \dots, x_m$ PLUS A NOISE WHICH IS K-DISTRIBUTED]

The idea of smooth bootstrap is to sample not from the distr. generated by the sample but from the **KERNEL DENSITY ESTIMATION** *

PARAMETRIC BOOTSTRAP

→ we know (suppose) the family of the distribution but we don't know the parameters

$$F_\theta(x) \approx F_{\hat{\theta}}(x)$$

(GOOD WHEN F IS KNOWN TO BELONG TO A PARAMETRIC FAMILY)

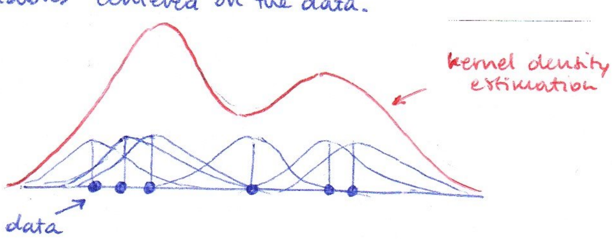
[MC: SAMPLING FROM $F_{\hat{\theta}}$]

EXAMPLE IN R: BOOTSTRAP METHOD.R

Idea: we estimate the unknown parameters from the sample and then we plug-in the distribution replacing the unknown parameters with the estimates

* Kernel density estimation

We have a set of points and we want to have an estimate of the probability density function.
We define the notion of kernel
(:= probability density function, usually a simmetric one (very often a gaussian)).
We estimate the density of these points by putting together a lot of gaussian random variables centered on the data.



The area under the blue curves is $1/n$, so that the area under the red is 1.

BOOTSTRAP REGRESSION

$$Y_i | (X_i = x_i) = \beta_0 + \beta_1 x_i + \varepsilon_i \qquad \varepsilon_i \sim \text{iid } \varepsilon$$

IDEALLY WE WOULD LIKE TO BOOTSTRAP THE ERRORS BUT ERRORS ARE NOT OBSERVABLE $\Rightarrow$ WE ESTIMATE THE ERRORS ε_i WITH THE RESIDUALS $\hat{\varepsilon}_i$

Scheme:

input $\longrightarrow$ $(y_1, x_1), (y_2, x_2), \dots, (y_n, x_n)$

- FIT THE REGRESSION MODEL + compute the residuals

$$y_i = \underbrace{\hat{\beta}_0 + \hat{\beta}_1 x_i}_{\hat{y}_i} + \hat{\varepsilon}_i \qquad i = 1, \dots, n$$

Note: here we're computing the residuals from the H_1 model ($\neq$ permutation regr.)

- Generate bootstrap samples by sampling with replacement from the residuals
  RESAMPLE FROM THE EMP. DISTRIBUTION OF RESIDUALS

$$y_i^* = \hat{\beta}_0 + \hat{\beta}_1 x + \hat{\varepsilon}_i^* \qquad \hat{\varepsilon}_i^* \sim \hat{F}_{\hat{\varepsilon}}$$

($\neq$ permutation settings, where we sample without replacement)

output $\longrightarrow$ $(y_1^*, x_1), (y_2^*, x), \dots, (y_n^*, x_n)$

$\Longrightarrow$ WE CAN THEN MAKE INFERENCE ON PARAMETERS:

- $\hat{\beta}_0^*$
- $\hat{\beta}_1^*$
- $\hat{\beta}_0^* + \hat{\beta}_1^* x_0$ (N.B. SYNCHRONIZED)

R EXAMPLE: BOOTSTRAP REGRESSION.R

TWO OR MORE INDEPENDENT SAMPLES

WE INDEPENDENTLY BOOTSTRAP THE DIFFERENT SAMPLES.

R EXAMPLE: BOOTSTRAP TWO SAMPLES.R