Bootstrapping.

Bootstrap in iid case

Suppose $X_t \underset{iid}{\sim} \;?\;(\mu, \sigma^2)$

distribution unknown.

Use $\bar{X}$ as $\hat{\mu}$.

Q: what is the $V(\hat{\mu})$?

① Theoretical consideration

Assume $X_t \sim N(\mu, \sigma^2)$

then $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ Exact for all n.

$$V(\hat{\mu}) = V(\bar{X}) = \frac{\sigma^2}{n}$$

→ In many cases, you can only derive asymptotic $V(\hat{\mu})$.

$$V(\hat{\mu}) \xrightarrow[\text{as } n \to \infty]{} \text{formula}$$

(2) Monte Carlo Simulation

Assume $X_t \sim N(\mu_0, \sigma_0^2)$ for some chosen value of μ_0, σ_0.

In PC, generate ~~many~~ $\{X_1, \dots, X_n\} \to \overline{X}$

repeat many times. get $\underbrace{\text{many } \overline{X}}$

look at Distribution of this.

to estimate $V(\overline{X})$.

→ Must assume μ_0, σ_0 as well as N.

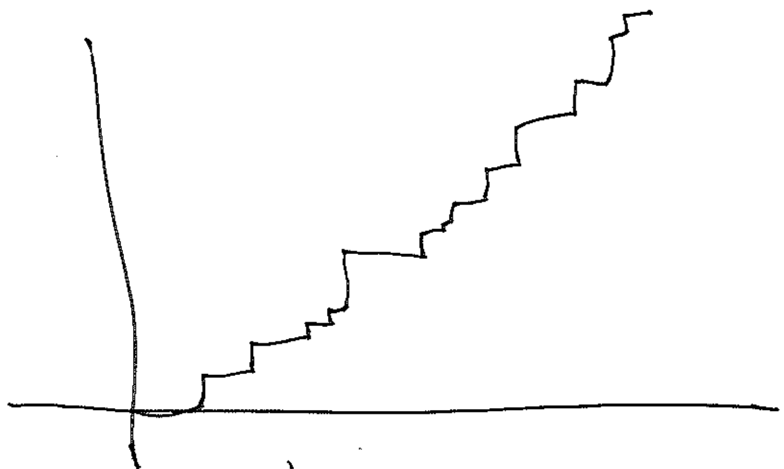
→ Can obtain result for small n.

③ Bootstrapping

<u>Do not assume</u> $X_t \sim N(\mu_0, \sigma_0^2)$.

Use the original data $\{x_1, \ldots, x_n\}$ to draw

Emprical Distribution Function $\hat{F}(x) = \sum_{i=1}^{n} \frac{1}{n} \mathbb{I}\{x_i < x\}$



Generate new set $\{X_1^*, \ldots, X_n^*\}$

from EDF.

Generate new set $\{x_1^*, \ldots, x_n^*\}$ from EDF made from original data $\{x_1, \ldots, x_n\}$

$$\parallel$$

Sample with replacement from $\{x_1, \ldots, x_n\}$

R code

```
Sample ( 1:10 , 20 , replace = TRUE )
```

Original data $\{x_1, \dots, x_n\}$

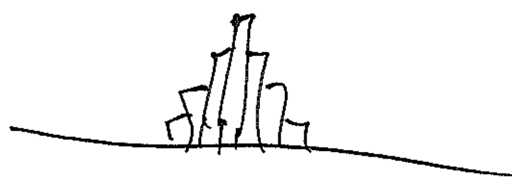
$$\left.\begin{array}{l} \hookrightarrow \{x_1^*, \dots, x_n^*\} \longrightarrow \bar{X}^* \\ \text{another } \{x_1^*, \dots, x_n^*\} \longrightarrow \text{another } \bar{X}^* \end{array}\right\} \text{repeat many times.}$$

$$\Downarrow$$

many $\bar{X}^*$

$\downarrow$

look at distribution of $\bar{X}^*$



$V(\bar{X}^*) \approx V(\bar{X})$.

- Did not assume anything about distribution of X_t.

- Completely data driven.

Bootstrapping in ARMA.

ARMA(1,1)

$$Y_t - \phi_1 Y_{t-1} = e_t - \theta_1 e_{t-1}$$

unknown.

$$e_t \sim_{iid} (0, \sigma^2)$$

Data $\{Y_1, \dots, Y_n\}$.

1) Fit ARMA(1,1) $\rightarrow$ get $\hat{\phi}_1$ $\hat{\theta}_1$.

Say, $\hat{\phi}_1 = .521$

$\hat{\theta}_1 = .297$.

2) Get residuals from the estimation

$$\hat{e}_t = \sum_{j=0}^{n-t} \hat{\psi}_j Y_{t+j} \qquad \text{(invertible representation)}$$

$\hat{\psi}_j$ calculated by $\hat{\phi}_1$ and $\hat{\theta}_1$.

3) Bootstrap residuals, $\{\hat{e}_1, \dots, \hat{e}_n\}$ and generate ARMA

$$Y_t^* - \underbrace{\hat{\phi}_1}_{.521} Y_{t-1}^* = e_t^* - \underbrace{\hat{\theta}_1}_{.297} e_{t-1}^*$$

bootstrapped errors.

4) For $\{Y_1^*, \ldots, Y_n^*\}$ pretend that you don't know $\hat{\phi}_1 = .521$ and $\hat{\theta}_1 = .297$, and estimate

$\hat{\phi}_1^*$ and $\hat{\theta}_1^*$.

5) repeat (3) and (4). original set of residuals $\{\hat{e}_1, \ldots, \hat{e}_n\}$ does not change, Neither do $\hat{\phi}_1 = .521$ and $\hat{\theta}_1 = .297$ in step (3).

In step (4), $\hat{\phi}_1^*$ and $\hat{\theta}_1^*$ are different each time.

Bootstrap

many $\hat{\phi}_1^*$ and $\hat{\theta}_1^*$

around

$\hat{\phi}_1 = .521$ and $\hat{\theta}_1 = .297$

$\approx$

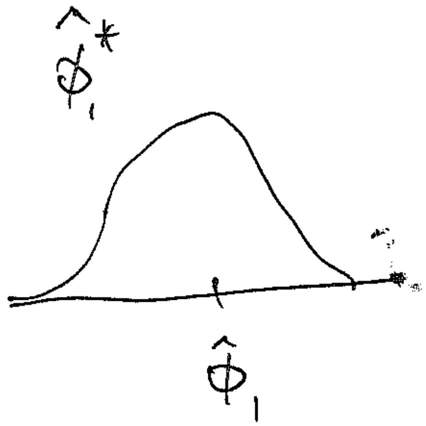
Monte Carlo

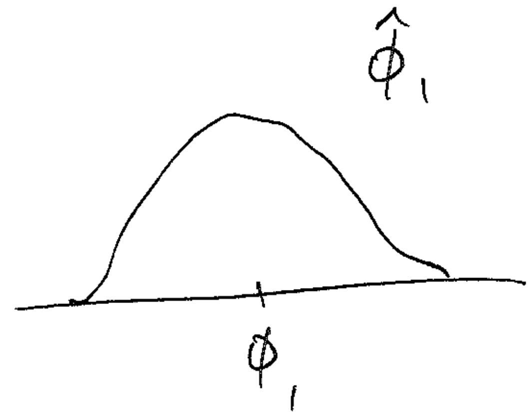
many $\hat{\phi}_1$ and $\hat{\theta}_1$

around

ϕ_1 and θ_1

(true values)





use $V(\hat{\phi}_1^*)$ as estimate of $V(\hat{\phi})$.

Order Selection of ARMA(p, q)

Property of ACF and PACF

	AR(p)	MA(q)	ARMA(p, q)
ACF	Tails off	Cuts off at q	tails off
PACF	Cuts off at p.	Tails off	tails off.

Akaike Information Criteria AIC

$$AIC = -2\log(\text{maximum likelihood}) + 2(p+q+1)$$

$p+q$ if there's no μ.

Best model = lowest AIC.

AICC

$$\text{AICC} = -2\log(\text{maximum likelihood}) + \frac{2(k+1)(k+2)}{n-k-2}$$

$$k = \begin{cases} p+q+1 & \text{with } \mu \\ p+q & \text{w/o } \mu \end{cases}$$

BIC Bayesian Information Criteria

$$BIC = -2 \log(\text{maximum likelihood}) + k \log(n)$$

$$k = \begin{cases} p+q+1 & w \ \mu \\ p+q & w/o \ \mu \end{cases}$$