

# Comparing estimators: mean squared error

Uniform  $\mathcal{U}(0, \theta)$ , Monte Carlo and the bootstrap

Hedibert Freitas Lopes

Insper

# Setup

Suppose  $x_1, \dots, x_n \mid \theta \stackrel{iid}{\sim} \mathcal{U}[0, \theta]$ ,  $\theta > 0$ .

We want to compare four estimators:

$$\begin{cases} \hat{\theta}_1 = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \\ \hat{\theta}_2 = x_{(n)} = \max(x_1, \dots, x_n) \end{cases} \quad \begin{cases} \hat{\theta}_3 = 2\hat{\theta}_1 \\ \hat{\theta}_4 = \left(\frac{n+1}{n}\right) \hat{\theta}_2 \end{cases}$$

where  $x_{(1)} \leq \dots \leq x_{(n)}$  are the order statistics.

## Means and variances

|                   | $\hat{\theta}_1$       | $\hat{\theta}_2$                  | $\hat{\theta}_3$      | $\hat{\theta}_4$          |
|-------------------|------------------------|-----------------------------------|-----------------------|---------------------------|
| $E(\hat{\theta})$ | $\frac{\theta}{2}$     | $\frac{n}{n+1} \theta$            | $\theta$              | $\theta$                  |
| $V(\hat{\theta})$ | $\frac{\theta^2}{12n}$ | $\frac{n \theta^2}{(n+1)^2(n+2)}$ | $\frac{\theta^2}{3n}$ | $\frac{\theta^2}{n(n+2)}$ |

$\hat{\theta}_3$  and  $\hat{\theta}_4$  are the **unbiased versions** of  $\hat{\theta}_1$  and  $\hat{\theta}_2$ , respectively.

Useful fact:  $x_{(n)}/\theta \sim \text{Beta}(n, 1)$ , i.e.  $p(x_{(n)} | \theta) = n x_{(n)}^{n-1} / \theta^n$ ,  $0 \leq x_{(n)} \leq \theta$ .

# Mean squared error

To compare the estimators we use the mean squared error

$$\begin{aligned}\text{MSE}(\hat{\theta}) &= E_{x|\theta} \left\{ (\hat{\theta} - \theta)^2 \right\} \\ &= \underbrace{V(\hat{\theta})}_{\text{variance}} + \underbrace{[E(\hat{\theta}) - \theta]^2}_{\text{bias}}.\end{aligned}$$

## Mean squared error

To compare the estimators we use the mean squared error

$$\begin{aligned}\text{MSE}(\hat{\theta}) &= E_{x|\theta} \left\{ (\hat{\theta} - \theta)^2 \right\} \\ &= \underbrace{V(\hat{\theta})}_{\text{variance}} + \underbrace{[E(\hat{\theta}) - \theta]^2}_{\text{bias}}.\end{aligned}$$

Obviously,  $\text{MSE}(\hat{\theta}_i) = V(\hat{\theta}_i)$  for  $i = 3, 4$ . One can show that

$$\text{MSE}(\hat{\theta}_1) = \left( \frac{1+3n}{12n} \right) \theta^2, \quad \text{MSE}(\hat{\theta}_2) = \left( \frac{2}{(n+1)(n+2)} \right) \theta^2.$$

# Ranking the four estimators

Comparing the expressions (all proportional to  $\theta^2$ ):

- ▶  $\text{MSE}(\hat{\theta}_4) < \text{MSE}(\hat{\theta}_2) \iff n + 1 < 2n \iff n > 1$
- ▶  $\text{MSE}(\hat{\theta}_2) < \text{MSE}(\hat{\theta}_3) \iff 6n < (n + 1)(n + 2) \iff (n - 1)(n - 2) > 0 \iff n > 2$
- ▶  $\text{MSE}(\hat{\theta}_3) < \text{MSE}(\hat{\theta}_1) \iff 4 < 1 + 3n \iff n > 1$

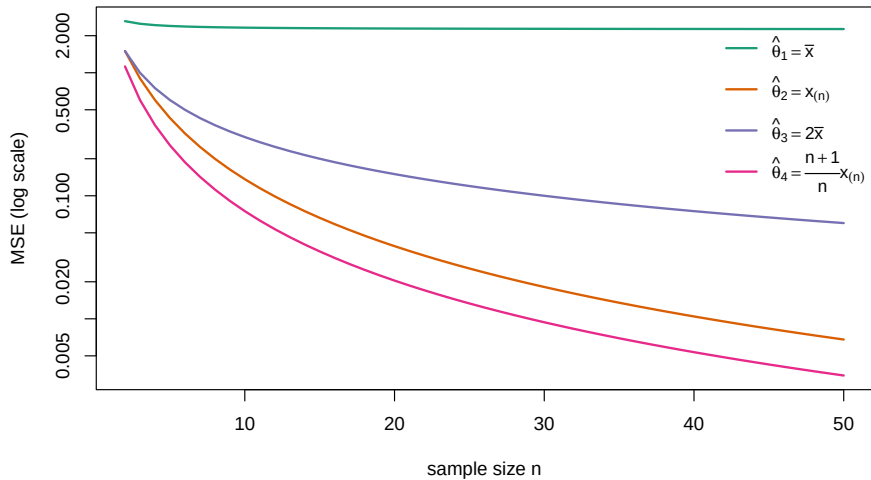
Therefore, for  $n \geq 3$ ,

$$\text{MSE}(\hat{\theta}_4) < \text{MSE}(\hat{\theta}_2) < \text{MSE}(\hat{\theta}_3) < \text{MSE}(\hat{\theta}_1)$$

(At  $n = 2$ ,  $\text{MSE}(\hat{\theta}_2) = \text{MSE}(\hat{\theta}_3) = \theta^2/6$ .)

Note the **rates**:  $\hat{\theta}_2, \hat{\theta}_4$  have MSE of order  $1/n^2$ ;  $\hat{\theta}_1, \hat{\theta}_3$  of order  $1/n$  (and  $\hat{\theta}_1$  does not even go to zero:  $\text{bias} = -\theta/2$ ).

## MSE as a function of $n$ ( $\theta = 3$ )



Is  $\hat{\theta}_1 = \bar{x}$  a “natural choice”?

$\hat{\theta}_1 = \bar{x}$  is a “natural choice”.

Is  $\hat{\theta}_1 = \bar{x}$  a “natural choice”?

$\hat{\theta}_1 = \bar{x}$  is a “natural choice”.

**Not quite!** All  $x_i$ 's are less than or equal to  $\theta$ , so  $\bar{x}_n$  is, most likely, even further away from  $\theta$ : it estimates  $E(x) = \theta/2$ , not  $\theta$ .

The fix  $\hat{\theta}_3 = 2\bar{x}$  removes the bias, but doubles the standard deviation.

A better route is to look at the likelihood.

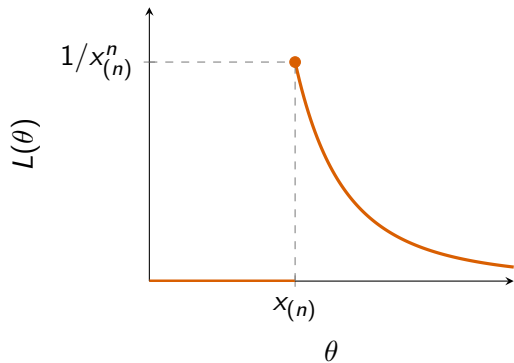
# Maximum likelihood estimator

$$\begin{aligned} p(x_1, \dots, x_n \mid \theta) &= \left(\frac{1}{\theta}\right)^n \prod_{i=1}^n \mathbb{I}(x_i \leq \theta) \\ &= \left(\frac{1}{\theta}\right)^n \mathbb{I}(x_{(n)} \leq \theta) \end{aligned}$$

where  $x_{(n)} = \max(x_1, \dots, x_n)$ . So,

$$L(\theta \mid x_1, \dots, x_n) = \frac{1}{\theta^n} \mathbb{I}(\theta \geq x_{(n)}),$$

$$\boxed{\hat{\theta}_{\text{MLE}} = x_{(n)} = \hat{\theta}_2}$$



# Monte Carlo exercise

Set  $\theta = 3$ ,  $n = 10$ ,  $N = 1000$ . For  $j = 1, \dots, N$ :

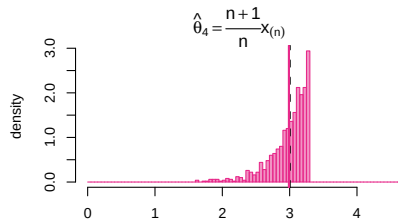
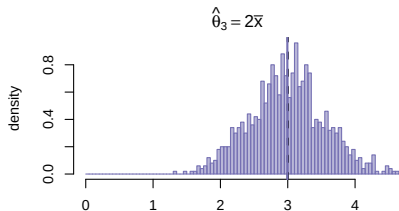
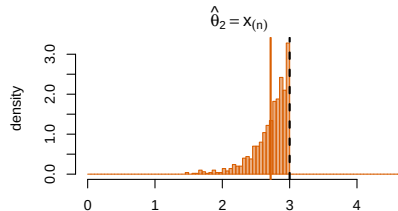
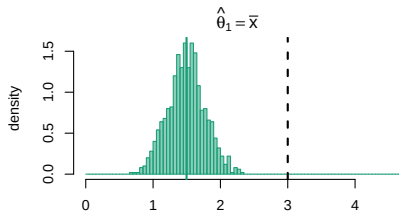
1. Draw  $x_1^{(j)}, \dots, x_n^{(j)} \stackrel{iid}{\sim} \mathcal{U}(0, \theta)$ .
2. Compute  $\hat{\theta}_1^{(j)}, \hat{\theta}_2^{(j)}, \hat{\theta}_3^{(j)}, \hat{\theta}_4^{(j)}$ .

Then  $\{\hat{\theta}_k^{(j)}\}_{j=1}^N$  approximates the sampling distribution of  $\hat{\theta}_k$ , and

$$\widehat{\text{MSE}}(\hat{\theta}_k) = \frac{1}{N} \sum_{j=1}^N \left( \hat{\theta}_k^{(j)} - \theta \right)^2.$$

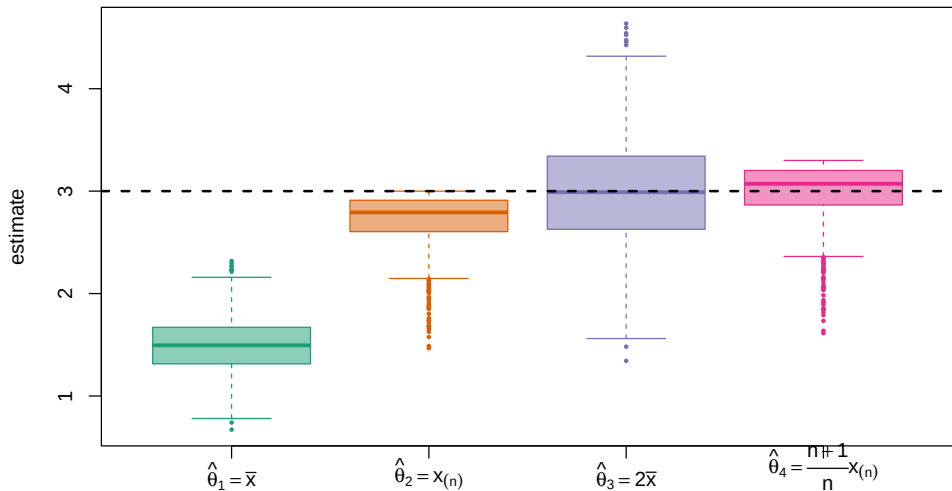
```
estimators = function(x){ n = length(x); c(mean(x), max(x), 2*mean(x), (n+1)/n*max(x)) }  
draws = matrix(runif(N*n, 0, theta), N, n)  
est = t(apply(draws, 1, estimators)) # N x 4
```

# MC sampling distributions: histograms



Dashed: true  $\theta = 3$ . Solid: MC mean of each estimator.

# MC sampling distributions: boxplots



## Confirming the theory ( $\theta = 3$ , $n = 10$ , $N = 1000$ )

|                                         | $E(\hat{\theta})$ |        | $SD(\hat{\theta})$ |        | $MSE(\hat{\theta})$ |        |
|-----------------------------------------|-------------------|--------|--------------------|--------|---------------------|--------|
|                                         | MC                | Theory | MC                 | Theory | MC                  | Theory |
| $\hat{\theta}_1 = \bar{x}$              | 1.497             | 1.500  | 0.278              | 0.274  | 2.336               | 2.325  |
| $\hat{\theta}_2 = x_{(n)}$              | 2.717             | 2.727  | 0.264              | 0.249  | 0.149               | 0.136  |
| $\hat{\theta}_3 = 2\bar{x}$             | 2.994             | 3.000  | 0.555              | 0.548  | 0.308               | 0.300  |
| $\hat{\theta}_4 = \frac{n+1}{n}x_{(n)}$ | 2.989             | 3.000  | 0.290              | 0.274  | 0.084               | 0.075  |

- ▶ MC and theoretical values agree up to MC error (with  $N = 1000$ , SDs are accurate to roughly  $\pm 2\%$ ).
- ▶  $SD(\hat{\theta}_3)$  is twice  $SD(\hat{\theta}_1)$ ;  $SD(\hat{\theta}_4)$  is  $(n+1)/n = 1.1$  times  $SD(\hat{\theta}_2)$ .
- ▶ Curiosity: at  $n = 10$ ,  $12n = n(n+2)$ , so  $SD(\hat{\theta}_1) = SD(\hat{\theta}_4)$  exactly. Same spread, but  $\hat{\theta}_1$  is centered at 1.5!
- ▶ The MSE ranking  $\hat{\theta}_4 < \hat{\theta}_2 < \hat{\theta}_3 < \hat{\theta}_1$  is reproduced.

## Bootstrap: one observed sample

In practice we observe **one** sample, not  $N$  of them. The (nonparametric) bootstrap replaces the unknown  $\mathcal{U}(0, \theta)$  by the empirical distribution  $\hat{F}_n$  of the data.

Given  $x_1, \dots, x_n$ , for  $b = 1, \dots, B$  ( $B = 1000$ ):

1. Draw  $x_1^{*b}, \dots, x_n^{*b}$  *with replacement* from  $\{x_1, \dots, x_n\}$ .
2. Compute  $\hat{\theta}_k^{*b}$ ,  $k = 1, \dots, 4$ .

Bootstrap standard error:  $SD^*(\hat{\theta}_k) = \text{sd}\{\hat{\theta}_k^{*1}, \dots, \hat{\theta}_k^{*B}\}$ .

Data: a single sample of 100 draws from  $\mathcal{U}(0, 3)$ ; we use the first 10, the first 50, and all 100.

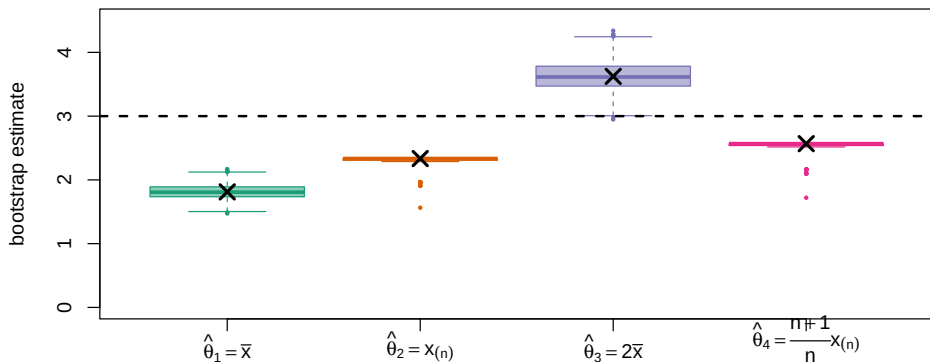
```
bt = t(replicate(B, estimators(sample(x, n, replace = TRUE))))  
apply(bt, 2, sd)
```

## Bootstrap standard errors vs theory

|           |                | $\hat{\theta}_1$ | $\hat{\theta}_2$ | $\hat{\theta}_3$ | $\hat{\theta}_4$ |
|-----------|----------------|------------------|------------------|------------------|------------------|
| $n = 10$  | estimate       | 1.812            | 2.335            | 3.624            | 2.568            |
|           | bootstrap SD   | 0.117            | 0.062            | 0.234            | 0.068            |
|           | theoretical SD | 0.274            | 0.249            | 0.548            | 0.274            |
| $n = 50$  | estimate       | 1.651            | 2.869            | 3.302            | 2.926            |
|           | bootstrap SD   | 0.113            | 0.041            | 0.226            | 0.042            |
|           | theoretical SD | 0.122            | 0.058            | 0.245            | 0.059            |
| $n = 100$ | estimate       | 1.617            | 2.999            | 3.234            | 3.029            |
|           | bootstrap SD   | 0.089            | 0.036            | 0.177            | 0.036            |
|           | theoretical SD | 0.087            | 0.029            | 0.173            | 0.030            |

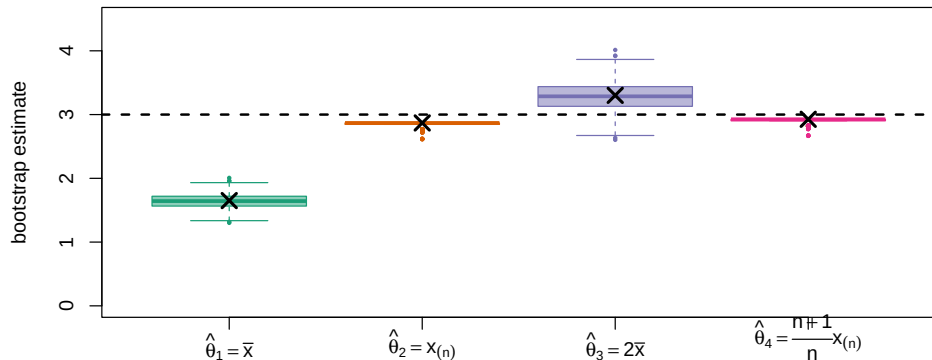
Theoretical SDs evaluated at the true  $\theta = 3$ .

# Bootstrap distributions, $n = 10$



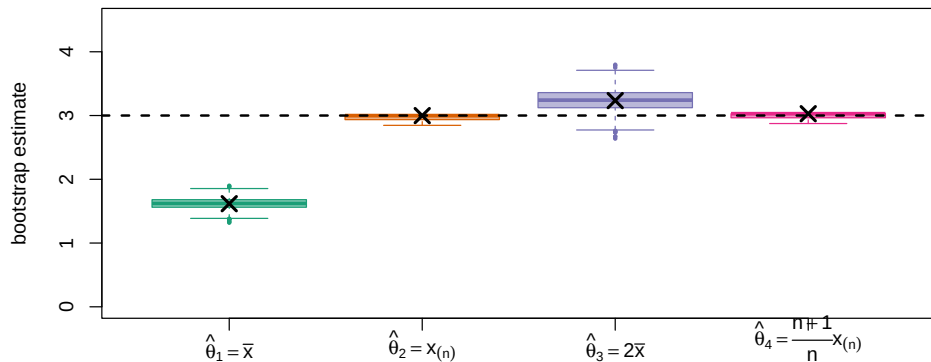
- ▶  $\times$ : estimate from the observed sample. Dashed:  $\theta = 3$ .
- ▶ These 10 draws happen to lie in  $[1.35, 2.33]$ : small spread  $\Rightarrow$  bootstrap SDs well below theory, and  $x_{(10)} = 2.33$  is far from 3.
- ▶ The bootstrap cannot recover what the sample does not contain.

# Bootstrap distributions, $n = 50$



- ▶ For  $\bar{x}$ -based estimators the bootstrap SD is now close to theory.
- ▶ For  $x_{(n)}$ -based estimators the bootstrap SD is still too small.

## Bootstrap distributions, $n = 100$



- $\bar{x}$ -based: bootstrap SD  $\approx$  theory.  $x_{(n)}$ -based: the boxes collapse onto  $x_{(n)}$ , a warning sign.

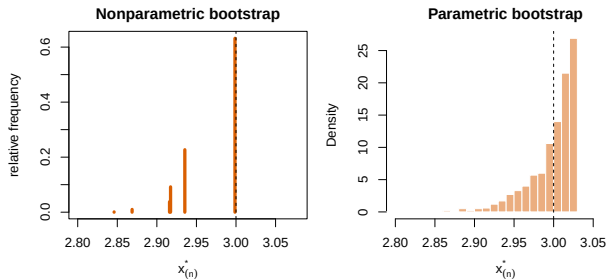
# Why the bootstrap struggles with $x_{(n)}$

A bootstrap sample reproduces the observed maximum whenever  $x_{(n)}$  is drawn at least once:

$$P^*\left(x_{(n)}^* = x_{(n)}\right) = 1 - \left(1 - \frac{1}{n}\right)^n \longrightarrow 1 - e^{-1} \approx 0.632.$$

Observed: 0.666 ( $n = 10$ ), 0.622 ( $n = 50$ ), 0.632 ( $n = 100$ ).

So  $x_{(n)}^*$  has a point mass of  $\approx 63\%$  at  $x_{(n)}$ , while the true sampling distribution,  $\theta \text{Beta}(n, 1)$ , is continuous: the nonparametric bootstrap is **inconsistent** for the maximum.



## A fix: parametric bootstrap

Use the model: draw  $x_1^*, \dots, x_n^* \stackrel{iid}{\sim} \mathcal{U}(0, \hat{\theta}_4)$  instead of resampling the data.

|           |                  | $\hat{\theta}_1$ | $\hat{\theta}_2$ | $\hat{\theta}_3$ | $\hat{\theta}_4$ |
|-----------|------------------|------------------|------------------|------------------|------------------|
| $n = 10$  | nonparametric SD | 0.117            | 0.062            | 0.234            | 0.068            |
|           | parametric SD    | 0.237            | 0.213            | 0.475            | 0.234            |
|           | theoretical SD   | 0.274            | 0.249            | 0.548            | 0.274            |
| $n = 50$  | nonparametric SD | 0.113            | 0.041            | 0.226            | 0.042            |
|           | parametric SD    | 0.119            | 0.054            | 0.238            | 0.055            |
|           | theoretical SD   | 0.122            | 0.058            | 0.245            | 0.059            |
| $n = 100$ | nonparametric SD | 0.089            | 0.036            | 0.177            | 0.036            |
|           | parametric SD    | 0.084            | 0.031            | 0.167            | 0.032            |
|           | theoretical SD   | 0.087            | 0.029            | 0.173            | 0.030            |

The parametric SD is the plug-in  $\text{SD}(\hat{\theta}_k; \hat{\theta})$ ; it tracks the theory as long as  $\hat{\theta}_4$  is close to  $\theta$ .

# Take-aways

- ▶  $\text{MSE} = \text{variance} + \text{bias}^2$ : unbiasedness alone is not the criterion.
- ▶  $\hat{\theta}_4 = \frac{n+1}{n}x_{(n)}$  wins for  $n \geq 3$ ; estimators based on  $x_{(n)}$  (sufficient statistic) converge at rate  $1/n$  in SD, those based on  $\bar{x}$  at rate  $1/\sqrt{n}$ .
- ▶ Monte Carlo with  $N = 1000$  replicates confirms the analytical means, SDs and MSEs.
- ▶ With one sample, the nonparametric bootstrap works well for  $\bar{x}$  but fails for  $x_{(n)}$ ; the parametric bootstrap repairs it.