Nuisance Parameters and Confidence Intervals

Bradley Efron

brad@stat.stanford.edu

Department of Statistics
Stanford University

The Problem of Nuisance Parameters

- Last year's PHYSTAT talk:

Bootstrap Confidence Intervals

- This year's talk:

How to Deal with Nuisance Parameters
(and Bootstrap Confidence Intervals!)

One-parameter Models

(no nuisance parameters)

- θ is parameter of interest (usually $\theta \in \mathcal{R}^1$)
- **Observe** data X with density $X \sim f_\theta(x)$
- Confidence interval $\mathcal{C}$ for θ eliminates those values of θ for which

$$\Pr_\theta\{\text{observed } X\}$$

  is too small in some sense.
- If "too small" means < 0.05 then $\mathcal{C}$ is 0.95 interval.

Nuisance Parameters

- Density $\quad X \sim f_{\theta,\nu}(x)$ with
  - θ the parameter of interest
  - ν the vector of nuisance parameters
- Now $\Pr_{\theta,\nu}\{\text{observed } X\}$ depends on ν as well as θ

Tactics for Getting Rid of ν

- Integrate out ν (Bayes)
- Uninformative priors (Jeffries)
- Profile likelihood ($\max_\nu f_{\theta,\nu}(x)$)
- Pivotal methods (Neyman construction)
- Standard intervals ($\hat{\theta} \pm 1.96\widehat{\mathrm{se}}$)
- Bootstrap intervals (all of above)

Pivotal Methods

(Student's t)

- $x_1, x_2, \ldots, x_n \overset{\text{ind}}{\sim} \mathcal{N}(\theta, \sigma^2)$ [σ a nuisance parameter]
- $\bar{x} = \sum_1^n x_i/n$ and $\widehat{\text{se}} = \left[\sum_1^n (x_i - \bar{x})^2/n(n-1)\right]^{1/2}$
- Pivotal quantity

$$T_\theta = \frac{\bar{x} - \theta}{\widehat{\text{se}}} \sim t_{n-1}$$

  depends on θ but not σ^2

- *Confidence interval* (0.95 level)

$$\mathcal{C} = \bar{x} \pm t_{n-1}^{(.975)} \widehat{\text{se}}$$

Standard Intervals

- **Extended central limit theorem** $\hat{\theta} \dot{\sim} \mathcal{N}(\theta, \mathrm{se}^2)$
  - $\mathcal{N}$ Normal
  - θ unbiased
  - se constant
- 0.95 standard interval $\hat{\theta} \pm 1.96\widehat{\mathrm{se}}$
- $\widehat{\mathrm{se}}$ from Taylor series, Fisher information, bootstrap
- *Actual non-coverage probability* $0.05 \pm O\left(1/\sqrt{n}\right)$

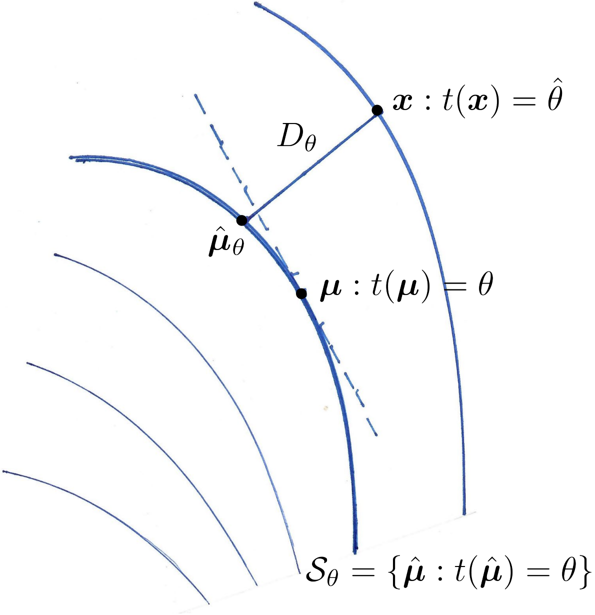
A Convenient Class of Parametric Problems

- **Observe** $\boldsymbol{x} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{I})$ where

$$\left[x_i \overset{\text{ind}}{\sim} \mathcal{N}(\mu_i, 1), \;\; i = 1, 2, \dots, p\right]$$

- *Confidence interval wanted* for real-valued parameter $\theta = t(\boldsymbol{\mu})$
- MLE $\quad \hat{\boldsymbol{\mu}} = \boldsymbol{x}, \quad \hat{\theta} = t(\boldsymbol{x})$
- Example $\quad \theta = t(\boldsymbol{\mu}) = \|\boldsymbol{\mu}\|$
- Nuisance $\nu =$ angular coordinates of $\boldsymbol{\mu}/\|\boldsymbol{\mu}\|$

$$x \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{I}),\ \theta = t(\boldsymbol{\mu}),\ \hat{\theta} = t(\boldsymbol{x})$$



Signed Distance D_θ

- $\boldsymbol{x} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{I})$ • $\theta = t(\boldsymbol{\mu})$
- Level surface $\mathcal{S}_\theta = \{\tilde{\boldsymbol{\mu}} : t(\tilde{\boldsymbol{\mu}}) = \theta\}$
- MLE $\hat{\boldsymbol{\mu}} = \boldsymbol{x}$ • $\hat{\theta} = t(\boldsymbol{x})$
- $\hat{\boldsymbol{\mu}}_\theta$ is closest point to $\boldsymbol{x}$ on $\mathcal{S}_\theta$
- **Signed distance** $D_\theta = \text{sign}(\hat{\theta} - \theta) \cdot \|\boldsymbol{x} - \hat{\boldsymbol{\mu}}_\theta\|$
- Is $D_\theta \dot{\sim} \mathcal{N}(0, 1)$? • Curvature of $\mathcal{S}_\theta$?

More Accurate Distribution of D_θ

- Local coordinate system for $\mathcal{S}_\theta$ around $\boldsymbol{\mu}$:

$$u = v^\top \boldsymbol{d_\mu} v \qquad \text{(locally quadratic)}$$

when $\boldsymbol{d_\mu}$ is $(p-1) \times (p-1)$ matrix

Lemma

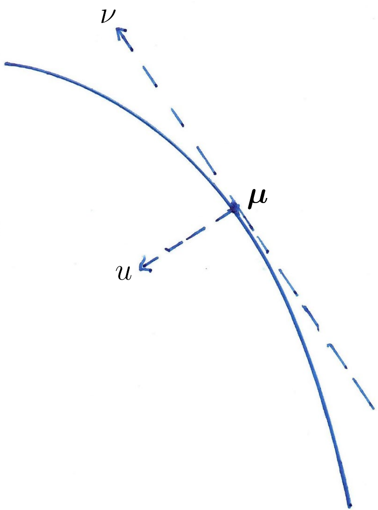
To a third order of accuracy,

$$D_\theta \dot{\sim} \mathcal{N}\left[\mathrm{tr}(\boldsymbol{d_\mu}), \left(1 - \mathrm{tr}(\boldsymbol{d}_{\boldsymbol{\mu}}^2)\right)^2\right]$$

where $\mathrm{tr}(\boldsymbol{d_\mu})$ *is the first-order correction to mean* D *and* $\mathrm{tr}(\boldsymbol{d}_{\boldsymbol{\mu}}^2)$ *is the second-order correction to variance 1.*

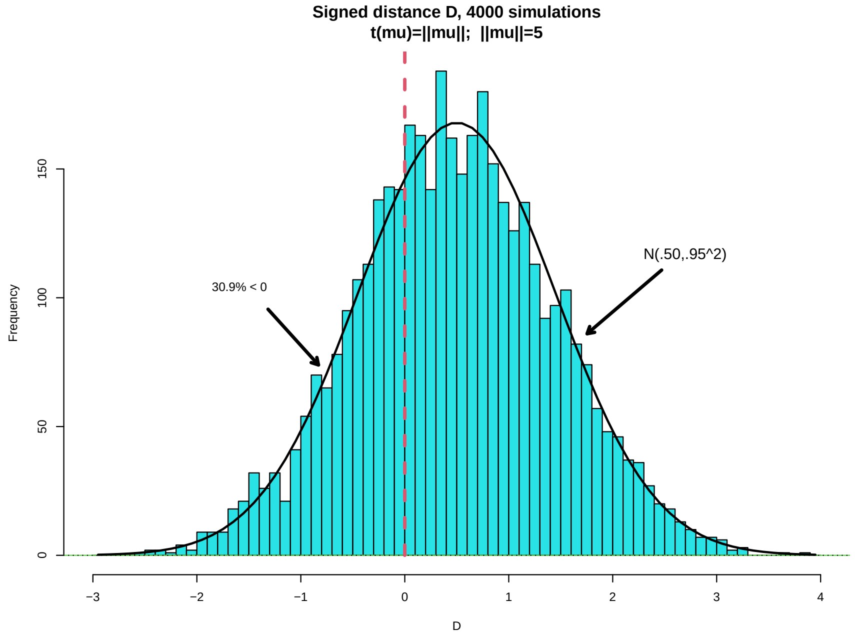
- Bias $\mathrm{tr}(\boldsymbol{d_\mu})$ can be big!

Local coordinates (u, v) for $\mathcal{S}_\theta$ near $\boldsymbol{\mu}$



Radius Example

- $\boldsymbol{x} \sim \mathcal{N}_6(\boldsymbol{\mu}, \boldsymbol{I})$
- $\theta = t(\boldsymbol{\mu}) = \|\boldsymbol{\mu}\| = 5$
- Surfaces $\mathcal{S}_\theta$ are 6-dimensional concentric spheres
- *Lemma* $D_\theta \sim \mathcal{N}(0.50, 0.95^2) \qquad (\theta = 5)$
- Simulated 4000 replications of D_θ
- Only 30.9% of 4000 replications $\leq \theta$
  (Not a coincidence: $\Phi^{-1}(0.309) = 0.50$)


Signed distance D, 4000 simulations
t(mu)=||mu||; ||mu||=5
30.9% < 0
N(.50,.95^2)
Frequency
0
50
100
150
−3
−2
−1
0
1
2
3
4
D

High-order Pivotal Quantity T_θ

- Pivot

$$T_\theta = \frac{D_\theta - \operatorname{tr}(\boldsymbol{d}_\theta)}{1 - \operatorname{tr}(\boldsymbol{d}_\theta^2)} \,\dot{\sim}\, \mathcal{N}(0,1)$$

- Nuisance parameters gone!
- Neyman construction 0.95 confidence intervals

$$\mathcal{C} = \{\theta : |T_\theta| \leq 1.96\}$$

- Using D_θ as pivot gives standard intervals

Bootstrap Intervals

(parametric)

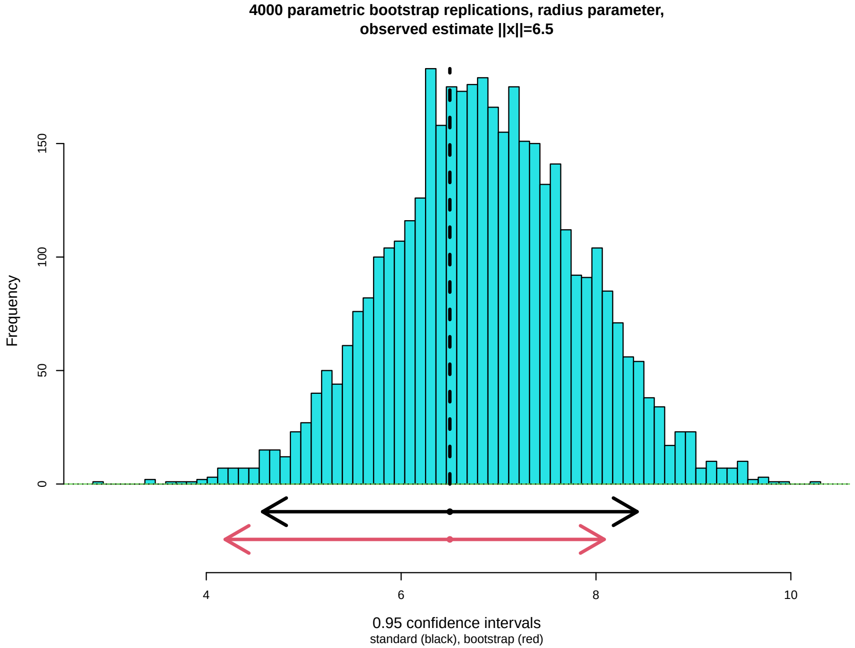
- **Real world**

$$\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{I}) \longrightarrow \boldsymbol{x} = \hat{\boldsymbol{\mu}} \longrightarrow \hat{\theta} = t(\boldsymbol{x}) = \|\boldsymbol{x}\|$$

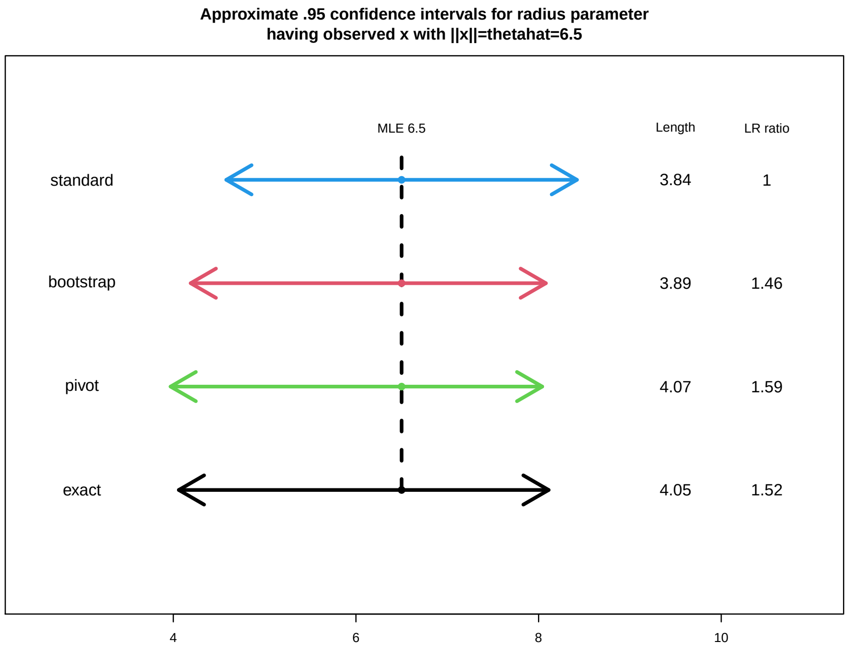
- *Bootstrap world*

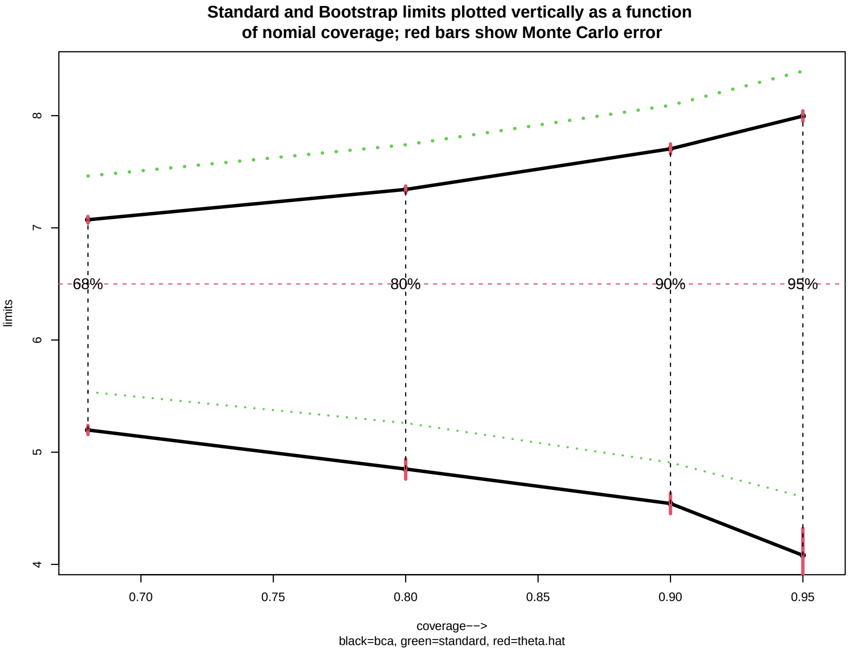
$$\mathcal{N}_p(\hat{\boldsymbol{\mu}}, \boldsymbol{I}) \longrightarrow \boldsymbol{x}^* \longrightarrow \hat{\theta}^* = t(\boldsymbol{x}^*) \qquad \text{(4000 times)}$$

- `bcapar` $\left(\hat{\theta}, \underset{4000}{\hat{\boldsymbol{\theta}}^*}, \underset{4000\times 5}{\boldsymbol{x}^*}\right)$ (from CRAN `bcaboot`)
- Exponential family models: $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, logistic regression, Dirichlet, . . .
- Efron and Narasimhan (2020)



4000 parametric bootstrap replications, radius parameter,
observed estimate ||x||=6.5
Frequency
0
50
100
150
4
6
8
10
0.95 confidence intervals
standard (black), bootstrap (red)


Approximate .95 confidence intervals for radius parameter
having observed x with ||x||=thetahat=6.5
MLE 6.5
Length
LR ratio
standard
3.84
1
bootstrap
3.89
1.46
pivot
4.07
1.59
exact
4.05
1.52
4
6
8
10


Standard and Bootstrap limits plotted vertically as a function
of nomial coverage; red bars show Monte Carlo error
limits
4
5
6
7
8
68%
80%
90%
95%
0.70
0.75
0.80
0.85
0.90
0.95
coverage−−>
black=bca, green=standard, red=theta.hat

Accuracy of Approximate Confidence Intervals

- **Observe** $x_1, x_2, \ldots, x_n \overset{\text{ind}}{\sim} f_{\theta,\nu}(x)$
- *Want 0.95 confidence limits* for $\hat{\theta} = t(x_1, \ldots, x_n)$

Method	Actual non-coverage	
Standard	$.05 + O(1/rn)$	"1st-order accurate"
Bootstrap	$.05 + O(1/n)$	"2nd-order accurate"
Pivot	$.05 + O(1/n^{3/2})$	"3rd-order accurate"

Supernova Data

- Data matrix $\underset{39\times 25}{\boldsymbol{X}}$
  - 39 Type Ia supernovas, each measured at 25 frequencies
  - fluxes from 350 to 850 nanometers
- Statistic of interest $\hat{\theta} = t(\boldsymbol{X})$
  $= 2\text{nd eigenvalue of } \boldsymbol{X}^\top \boldsymbol{X} = 5.29$
- 2nd eigenvector contrasts fluxes 1 and 2 with others

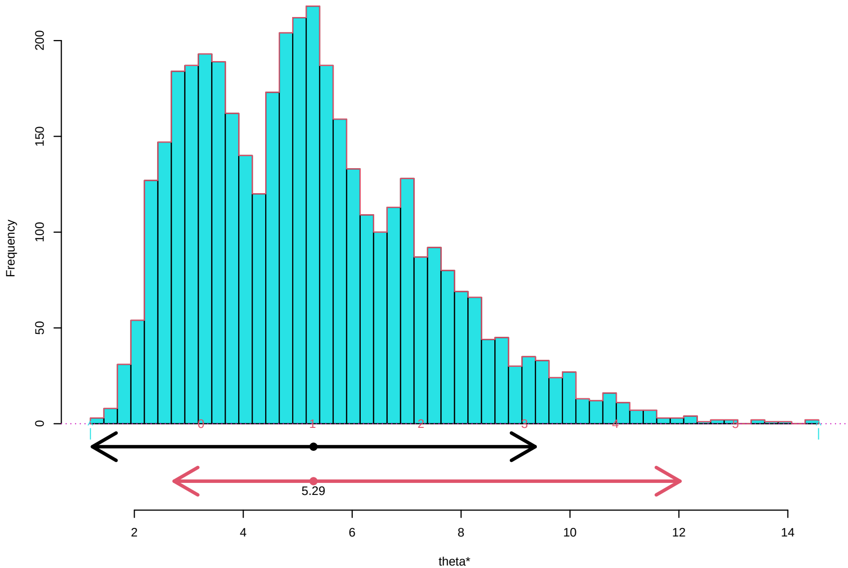
Nonparametric Bootstrap Confidence Limits

- Nonparametric bootstrap replication

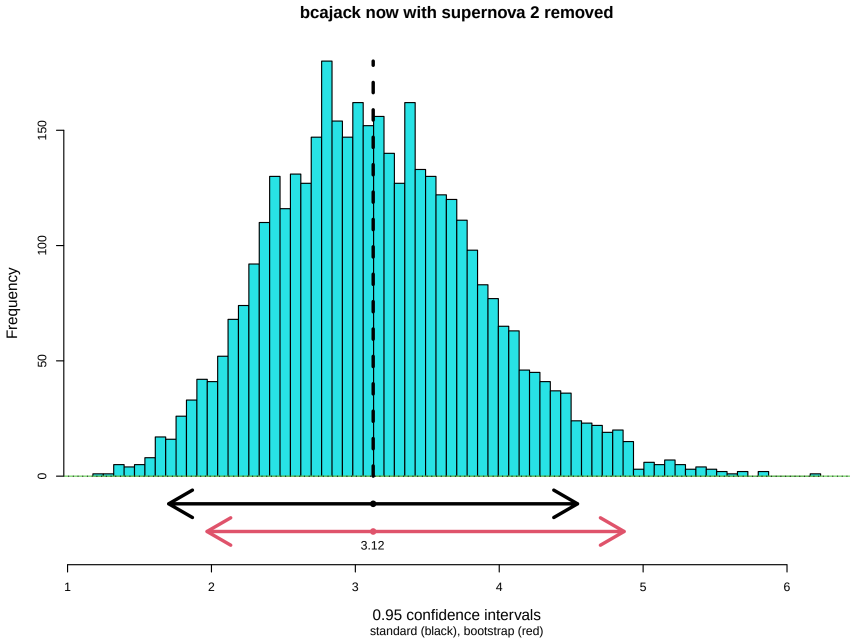
$$\boldsymbol{X}^* = \text{ sample 39 rows of } \boldsymbol{X}\text{, with replacement}$$

- `bcajack (X,4000,func)`
- $\texttt{func}(\boldsymbol{x}) = \text{2nd eigenvalue of } \boldsymbol{x}^\top \boldsymbol{x}$

4000 nonparametric boot reps, 2nd eigenvalue;
95% approx confidence intervals: standard(black), boot(red)



red numbers: boot repetitions of supernova 2

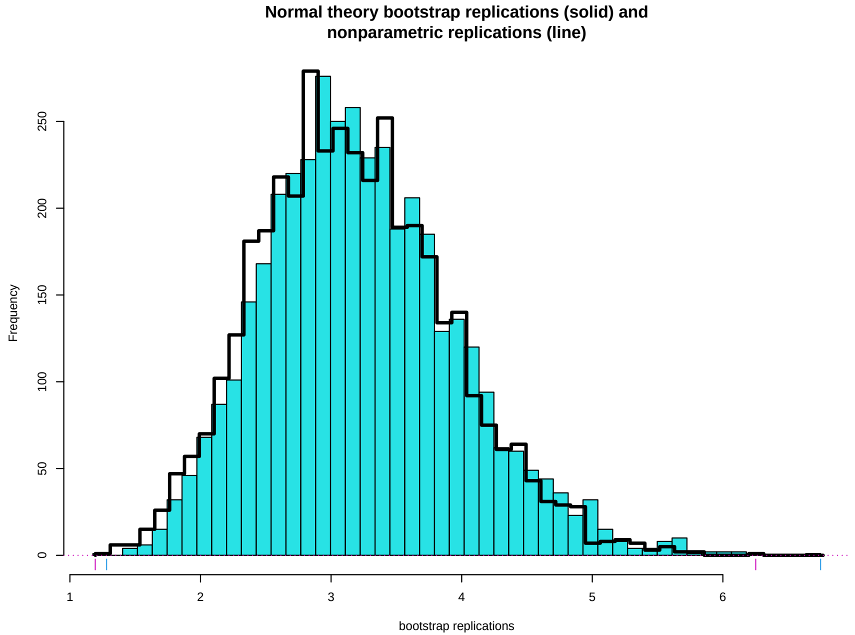

bcajack now with supernova 2 removed
Frequency
0
50
100
150
3.12
1
2
3
4
5
6
0.95 confidence intervals
standard (black), bootstrap (red)

Parametric Analysis

- **Assume** rows $\boldsymbol{x}$ of $\underset{38\times 25}{\boldsymbol{X}}$ are multivariate normal

$$\boldsymbol{x}_i \overset{\text{ind}}{\sim} \mathcal{N}_{25}\left(\underset{25}{\boldsymbol{\mu}}, \underset{25\times 25}{\boldsymbol{\Sigma}}\right), \qquad i = 1, 2, \ldots, 38$$

- 350 parameters!
- Resample $x_i^* \overset{\text{ind}}{\sim} \mathcal{N}_{25}(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}), \qquad i = 1, 2, \ldots, 38$


Normal theory bootstrap replications (solid) and
nonparametric replications (line)
Frequency
0
50
100
150
200
250
1
2
3
4
5
6
bootstrap replications

References

Efron, B. (1985). Bootstrap confidence intervals for a class of parametric problems. *Biometrika* 72: 45–58.

Efron, B. and Narasimhan, B. (2020). The automatic construction of bootstrap confidence intervals. *J. Comput. Graph. Stat.* 29: 608–619.