Bootstrap

M. Cichrová, K. Mečiarová, M. Vronková

May 25, 2022

Bootstrap

- Suppose we observe iid random vectors $\mathbf{X}_1, \ldots, \mathbf{X}_n$ from some distribution F_X
- Let $\theta_X = \theta(F_X)$ be a quantity of interest, we are interested in the distribution of a random vector $\mathbf{R}_n = g\left(\hat{\theta}_n, \theta_X\right)$, e.g. $\mathbf{R_n} = \sqrt{n}\left(\hat{\theta}_n - \theta_X\right)$, where $\hat{\theta}_n$ is an estimator of θ_X
- The algorithm is as follows:
  - estimate F_X by the empirical distribution function $\hat{F}_n$
  - Choose B large enough
  - For each $b \in \{1, \ldots, B\}$ generate datasets $\mathbb{X}_b^* = (\mathbf{X}_{1,b}^*, \ldots, \mathbf{X}_{n,b}^*)^T$ from the distribution $\hat{F}_n$
  - let $\mathbf{R}_{n,b}^* = g\left(\hat{\theta}_{n,b}^*, \hat{\theta}_n\right)$, e.g. $\mathbf{R}_{n,b}^* = \sqrt{n}\left(\hat{\theta}_{n,b}^* - \hat{\theta}_n\right)$
  - The distribution function $H_n(x)$ of $\mathbf{R_n}$ is now estimated as $\hat{H}_{n,b}^* = \frac{1}{B}\sum_{b=1}^{B} \mathbb{1}\{\mathbf{R}_{n,b}^* \leq x\}$

Example

- data: `https://msekce.karlin.mff.cuni.cz/~omelka/Soubory/nmst434/data/Hosi0.RData`
- Perform a bootstrap test of the null hypothesis that the weight of a baby at 12 months (weight12) follows a lognormal distribution with some unknown values of the parameters μ and σ^2
- 9999 replicates on a single processor

```
   user  system elapsed
  29.75    0.16   29.94
```

- 9999 replicates on 7 processors

```
   user  system elapsed
   3.30    0.28    7.58
```

- The time needed decreased almost 4 times, the memory needed should stay roughly the same