

Bootstrap Methods for Uncertainty Analysis

**(Sects. 1.15.3, 1.15.4, and 4.5 in text;
Hesterberg, 2011, reading)**

- **Part 1:** Selected Review of Estimation Theory
- **Part 2:** Bootstrap Methods for Determining Uncertainty Bounds in Estimates
- **Part 3:** Case Study: System Reliability Estimation and Confidence Regions from Subsystem and Full System Tests

PART 1

Selected Review of Estimation Theory

(Sects. 1.15.3 and 1.15.4 in text plus other sources)

Note: In order to allow for more general discussion of bootstrap in Part 2, the presentation here covers aspects of estimation theory in more general way than the textbook. The discussion in the textbook is a special case of results here.

Selected Review of Estimation Theory

(Sects. 1.15.3 & 1.15.4 in text plus any source in basic estimation theory)

- Fundamental role of data analysis is to **extract information from physical or simulated data**
- Parameter estimation for models is central to process of extracting information
- Uncertainty bounds on estimates essential in making inference
- All known bounds require knowledge of the moments (e.g., mean and variance) of estimator or knowledge of the finite-sample or asymptotic distribution of estimator
- Chebyshev (or other) inequalities use moments to provide conservative (too wide) uncertainty bounds
- Exact distribution known for certain simple estimators (linear regression estimates, sample means and variances from i.i.d. Gaussian data)
- Asymptotic distribution via asymptotic normality for sample averages or maximum likelihood estimates (MLEs)

3

Basic Formulation

- We have considered at least two forms for generic estimators:
 - Performance measures $\hat{\ell}$ from simulation runs
 - Estimate $\hat{\theta}$ for parameters θ in models (e.g., MLEs)
- Note that θ is a general vector of parameters
 - Special case of θ is scalar $\theta = \ell$ (unknown true mean of simulation output)
- Have available asymptotic theory that governs probability distribution for estimators (central limit theory for sample averages and asymptotic normality/Fisher info. matrix for MLEs)
- *Note:* In part 1 here, we use θ as a generic parameter vector to be estimated ($\theta = \ell$ or something else in simulation, such as parameters *inside* of simulation)
- *Note:* Data can be anything, including simulation data or physical system data

4

Problem Setting

- Consider classical problem of estimating parameter vector θ from n data vectors $\mathbf{Z}^{(n)} \equiv \{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n\}$
- Suppose have probability density or mass function (PDF or PMF) associated with data
- Parameters θ appear in PDF/PMF and affect distribution
 - Example: $\mathbf{z}_i \sim N(\text{mean}(\theta), \text{covariance}(\theta))$ for all i
- Let $L(\theta|\mathbf{Z}^{(n)})$ represent **likelihood function**, i.e., $L(\cdot)$ is PDF or PMF viewed as function of θ conditioned on data
- MLE found by maximizing L with respect to θ
- **Overall issue:** Generally impossible to know probability distribution of $\hat{\theta} \Rightarrow$ impossible to specify exact P -values or uncertainty bounds
 - Asymptotic normality and bootstrap—discussed below—are two leading ways to cope with issue

5

Information Matrix

- Recall likelihood function $L(\theta|\mathbf{Z}^{(n)})$
- Information matrix defined as
$$\mathbf{F}_n(\theta) = E\left(\frac{\partial \log L}{\partial \theta} \frac{\partial \log L}{\partial \theta^T}\right)$$
where expectation is w.r.t. $\mathbf{Z}^{(n)}$
- If Hessian matrix exists, equivalent form based on Hessian matrix:
$$\mathbf{F}_n(\theta) = -E\left(\frac{\partial^2 \log L}{\partial \theta \partial \theta^T}\right)$$
- $\mathbf{F}_n(\theta)$ is positive semidefinite of dimension $p \times p$ ($p = \dim(\theta)$)

6

Two Famous Results

- Connection of $F_n(\theta)$ and uncertainty in estimate $\hat{\theta}_n$ is rigorously specified via following results (θ^* = true value of θ):

1. Asymptotic normality:

$$\sqrt{n}(\hat{\theta}_n - \theta^*) \xrightarrow{\text{dist}} N(\mathbf{0}, \bar{F}^{-1})$$

where $\bar{F} \equiv \lim_{n \rightarrow \infty} F_n(\theta^*)/n$

2. Cramér-Rao inequality:

$$\text{cov}(\hat{\theta}_n) \geq F_n(\theta^*)^{-1} \text{ for all } n \text{ (unbiased } \hat{\theta}_n)$$

Above two results indicate: greater variability of $\hat{\theta}_n \Rightarrow$ “smaller” $F_n(\theta)$ (and vice versa)

7

Supplement: Interpretation of Information Matrix

- Known that

$$E\left[\frac{\partial \log L(\theta | \mathbf{Z}^{(n)})}{\partial \theta}\right] = \mathbf{0}$$

- Hence, $F_n(\theta)$ is covariance matrix of score vector $\partial \log L / \partial \theta$
- MLE usually found by solving $\partial \log L / \partial \theta = \mathbf{0}$
- High variance of $\partial \log L / \partial \theta$ is “good”: implies score vector sensitive to data
 - Implies data contain much information about θ
- Roughly speaking, $F_n(\theta)$ is related to precision of estimator of θ
 - Low precision (high variance) of estimator \Leftrightarrow “small” $F_n(\theta)$
 - High precision (low variance) of estimator \Leftrightarrow “large” $F_n(\theta)$

8

Confidence Intervals Using Fisher Information

- Confidence intervals are useful mechanism for characterizing uncertainty in estimate
 - Random interval that covers “truth” with specified probability
- Confidence intervals for large n based on normal distribution
- Given Fisher information \bar{F} (or approximation) for scalar θ , level α interval based on normal distribution approximation is

$$P\left(\hat{\theta}_n - z_{1-\alpha/2}\sqrt{\frac{\bar{F}^{-1}}{n}} \leq \theta^* \leq \hat{\theta}_n + z_{1-\alpha/2}\sqrt{\frac{\bar{F}^{-1}}{n}}\right) \approx 1 - \alpha$$

($z_{1-\alpha/2}$ is $1 - \alpha/2$ percentile point for normal distribution)

- Typical values for α are 0.90, 0.95, or 0.99, yielding 90%, 95%, or 99% confidence intervals:

$$\left(\hat{\theta}_n - z_{1-\alpha/2}\sqrt{\frac{\bar{F}^{-1}}{n}}, \hat{\theta}_n + z_{1-\alpha/2}\sqrt{\frac{\bar{F}^{-1}}{n}}\right)$$

- Special case of above for $\theta = \ell$ is interval

$$\left(\hat{\ell} - z_{1-\alpha/2}\sqrt{\frac{\sigma^2}{n}}, \hat{\ell} + z_{1-\alpha/2}\sqrt{\frac{\sigma^2}{n}}\right)$$

9

PART 2

Bootstrap Methods for Determining Uncertainty Bounds in Estimates

(Sect. 4.5 in text; Hesterberg, 2011)

Note: The presentation here covers bootstrap in a more general—and broadly applicable—way than the textbook. The discussion in the textbook is a special case of the methods here.

10

Motivation

- Two general methods for constructing uncertainty (confidence) bounds for estimate
 - (i) large-sample approach based on asymptotic normal distribution (central limit theorem)
 - (ii) finite-sample approach based on Monte Carlo (bootstrap) sampling.
- Prime rationale for bootstrap is that asymptotic results may not be useful in small samples
 - E.g., in reliability problem (earlier slides), asymptotic approach is problematic when true reliabilities very close to unity
- Discussion below is summary of bootstrap method
- Bootstrap idea applies in very general estimation problems (e.g., MLE)
 - Textbook focuses on **specific instance** of estimating ℓ using repeated values of H ; determine distribution of $\hat{\ell}$
 - In fact, textbook is special case of general θ : scalar $\theta = \ell$ (unknown true mean of simulation output)

11

Bootstrap: Class Connection

- Two ways in which bootstrap connects to class:
 1. Method is based on Monte Carlo sampling from distribution constructed from set of sample data
 2. “Real” data (the sampling data) may come from simulation output
 - E.g., using simulation runs to estimate performance measure $\theta = \ell$ (Sects. 4.1–4.4 in textbook):

$$\ell \equiv E[H(\mathbf{X})] = \int H(\mathbf{x})f(\mathbf{x})d\mathbf{x}$$

12

Bootstrap: General Algorithm and Principles

(1) Collect n "real" data from system of interest.

(2) Use Monte Carlo methods to sample from above real data (directly or indirectly) to form bootstrap sample of n "fake" data.

(3) Use bootstrap sample in step (2) to form bootstrap estimate of whatever quantity is relevant (e.g., mean value or general parameter vector).

(4) Repeat steps (2) and (3) many times; distribution of bootstrap estimates in step (3) is nearly same as unknown distribution of estimate based on real data.

- Comments:

- Approximate distribution from step (4) can be used for confidence bounds, variance and bias estimation (as in textbook), or other characterizations of uncertainty in estimate from set of real data
- Detailed algorithm ("standard" or "parametric" form) below

13

Bootstrap: Overall Principles

- Bootstrap not used to get better estimates
- Mean of bootstrap distribution close to mean of original observed values; no additional information relative to unknown true value of mean
- Bootstrap is useful in understanding spread of unknown (true sampling) distribution
- Implications of above is that bootstrap is useful in things like
 - Confidence intervals
 - Quantile calculation
 - Bias
 - Etc.
- Not useful in improving estimate per se
- In small samples, bootstrap distribution tends to be too narrow; probabilities for intervals off by factor of $O(1/\sqrt{n})$
- Improvements possible with advanced methods (e.g., BC_a method, which has error $O(1/n)$) (Hesterberg, 2011)

14

Standard Bootstrap vs. Parametric Bootstrap

- Bootstrap method is based on resampling from original data, directly or indirectly (“standard” or “parametric”)
- Suppose have n values of \mathbf{X}_i from real system or from simulation
- Special case: Interested in estimating $\ell = E[H(\mathbf{X})]$
- Suppose each \mathbf{X}_i is sample from distribution function F
- Two forms of bootstrap:
 - **Standard (nonparametric) bootstrap** forms fake samples $H^* = H(\mathbf{X}^*)$ from empirical distribution function (EDF) formed from real data
 - **Parametric bootstrap** samples from known pdf based on parameter estimates formed from real data (e.g., real data used to estimate reliabilities for subsystems in full system; then sample from “fake” system with estimated subsystem reliabilities as truth)
- Standard bootstrap method based on using bootstrap sample \mathbf{X}^* as a representation of “typical” data from F

15

Sampling for Standard Bootstrap

- Standard (nonparametric) bootstrap is based on sampling with replacement from original data $\{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n\}$
 - Sampling probability is $1/n$ for each \mathbf{X}_i
- Original observed data set $\{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n\}$ (could be real or simulated data in general)
- EDF is discrete and assigns probability $1/n$ to each \mathbf{X}_i
- Consider *scalar* data $\{X_1, X_2, \dots, X_n\}$ that are *ordered* from smallest to largest. Then,

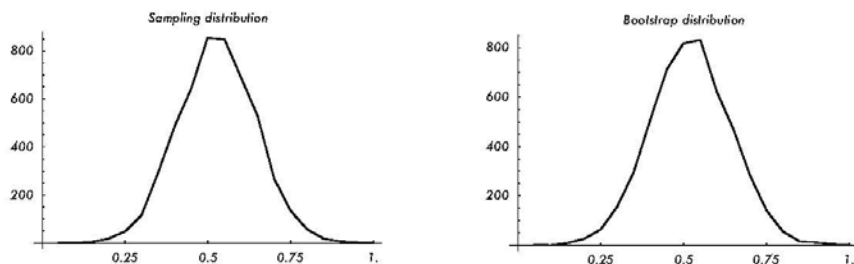
$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I_{\{x \geq X_i\}}$$

- From Chapter 2, to generate a random variable X from EDF of ordered (scalar) data:
 - (1) Generate U according to a Uniform (0, 1) distribution
 - (2) Find smallest integer $k \geq 0$ such that $U \leq F(x_k)$; then $X = x_k$
- Repeat above process n times to generate bootstrap sample of same size as true data set: $\{X_1^*, X_2^*, \dots, X_n^*\}$

16

Bootstrap Sampling Distribution: Example

- Consider distribution of sample mean of 25 $U(0,1)$ random variables (each X_i represents one sample mean)
- By CLT, know that asymptotic distribution is normal
- Plots below show sampling distribution of 5000 replicates of sample means (each of 25 points) and **one** bootstrap sample, $\{X_1^*, X_2^*, \dots, X_n^*\}$, representing $n = 5000$ sample means from sampling distribution (Varian, 2005)



- Above distributions are similar: resampling from single sample provides reasonable way to determine what would happen if actually drew many separate samples

17

Standard (Nonparametric) Bootstrap Method for Computing Confidence Intervals for $\hat{\ell}_n$

- Suppose we wish to put a confidence interval on standard sample mean estimate of ℓ (discussed in textbook).
Special case of general model: $\theta = \ell$
- **Step 0:** Run simulation n independent times to collect n values of $H(\mathbf{X}_i)$
- **Step 1:** Form EDF from the n values of $H(\mathbf{X}_i)$
- **Step 2:** Generate a bootstrap sample of values H (sample of size n) from the EDF in Step 1.
- **Step 3:** Calculate a bootstrap estimate, say $\hat{\ell}_n^*$, from the bootstrap sample in Step 2.
- **Step 3:** Repeat Steps 2 and 3 a large number of times (perhaps 1000 or 5000) and rank order resulting values $\hat{\ell}_n^*$; one- or two-sided confidence intervals are available by determining appropriate quantiles from ranked sample of these bootstrap estimates.

18

Parametric Bootstrap Method for Computing Confidence Intervals (General Models and θ)

- **Parametric bootstrap:** Consider pdf with unknown parameters θ ; use real data to estimate θ by whatever means is appropriate (e.g., MLE)
- **Step 0:** Treat above estimate of θ as true value of θ .
- **Step 1:** Generate by Monte Carlo a set of bootstrap data of same collective sample size n as real data using assumed pdf associated with value of θ from Step 0.
- **Step 2:** Calculate new estimate of θ from bootstrap data in Step 1.
- **Step 3:** Repeat Steps 1 and 2 large number of times (perhaps 1000 or 5000). For scalar θ , one- or two-sided confidence intervals are available by determining appropriate quantiles from ranked sample of bootstrap estimates in Step 2. (Multivariate θ requires other approaches.)

19

PART 3

Case Study in Bootstrap: System Reliability Estimation and Confidence Regions from Subsystem and Full System Tests

Spall (2012)*

*Setting here is special case of general problem of estimation with binary subsystems (e.g., reliability, networks, Internet tomography, etc.). See Spall (2014) for general setting and full theoretical justification.

20

Introduction: General Motivation

- Reliability determination is essential for assessment of large-scale systems
- Often difficult or infeasible to directly evaluate reliability of complex systems through large number of full system tests alone; e.g.:
 - Full system is very costly (or dangerous) to operate
 - Full system test requires destruction of system itself
- Nevertheless, sometimes there are **at least a few tests** of full system available; obviously desirable to include such information in overall reliability assessment
- Development of general method for estimating full-system reliability is long-standing problem, especially with limited full-system tests
- **Note:** General method is currently being used at JHU for modeling transportation systems (w/ former student Xilei Zhao and with Morgan State Univ.) (Zhao and Spall, 2016)

21

Introduction: General Motivation (cont'd)

- Consider **combination** of information from tests on subsystems (and/or components or other aspects of system) **and** tests on full system
- Subsystems tests may be much less expensive and/or more feasible to obtain than full system tests
- Goal of project:

Formation of “proper” maximum likelihood estimate (MLE) for system reliability—including uncertainty bounds—from combination of full system and subsystem tests
- MLE is “gold standard” for most parameter estimation problems
 - Desirable theoretical, numerical, and statistical properties

22

Uncertainty (Confidence) Bounds on Estimates

- Key part of approach is calculation of uncertainty (confidence) bounds on estimates
- Two methods for computing confidence regions:
 1. Asymptotic theory (asymptotic normality with Fisher information matrix)
 2. Monte Carlo (bootstrap)-based method
- Additional comments below

23

MLE Formulation and Notation

- System may have *dependent* or *independent subsystems* in *operational* mode; assume independent samples are available for *test* mode, as given below
- Let Y represent number of successes in n independent and identically distributed (i.i.d.) random trials for **full system**
- Let X_j represent number of successes in n_j i.i.d. random trials for **subsystem j** (p subsystems)
 - i.i.d. trials required only in test mode (subsystems may interact in full system operations); subsystems have different distributions
- Let ρ and ρ_j represent system and subsystem reliabilities (success probabilities); vector $\theta \equiv [\rho_1, \dots, \rho_p]^T$ contains unknown parameters to be estimated
 - Full system ρ uniquely determined from the ρ_j according to $\rho = h(\theta)$ for known function h
- Full system ρ is the performance measure ℓ here

24

Performance Metric for Estimation: Likelihood and Log-Likelihood Functions

- From independence of *test data* (key assumption!), likelihood function is:

$$\underbrace{\binom{n}{Y} \rho^Y (1-\rho)^{(n-Y)}}_{\text{system}} \underbrace{\binom{n_1}{X_1} \rho_1^{X_1} (1-\rho_1)^{(n_1-X_1)} \dots \binom{n_p}{X_p} \rho_p^{X_p} (1-\rho_p)^{(n_p-X_p)}}_{p \text{ subsystems}}$$

- Above leads to log-likelihood $\mathcal{L}(\theta)$ (to within a constant):

$$\mathcal{L}(\theta) = Y \log \rho + (n - Y) \log(1 - \rho) + \sum_{j=1}^p [X_j \log \rho_j + (n_j - X_j) \log(1 - \rho_j)]$$

- Generic performance metric $\mathcal{L}(\theta)$ applies **regardless** of configuration of system (series, parallel, operationally dependent, or other)
 - Constraints in MLE optimization process (e.g., $\rho = h(\theta)$) will reflect configuration

25

Example MLE Formulation: Series Systems

- Assume feasible values $\Theta \equiv \{0 < \rho_j < 1 \text{ for all } j\}$
- Suppose system has independent subsystems in series in operational mode (as example)
- MLE in such **series system** is found according to:

$$\hat{\theta} \equiv \arg \max_{\theta \in \Theta} \mathcal{L}(\theta)$$

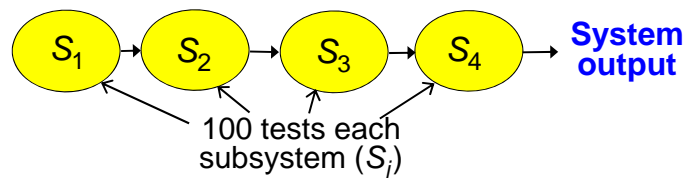
$$\text{subject to } \rho = h(\theta) = \prod_{j=1}^p \rho_j$$

- Series case is of wide practical interest, but other forms often needed
 - Constraint is altered to handle other types of systems
 - Next slide provides gives example of series system for use in numerical study

26

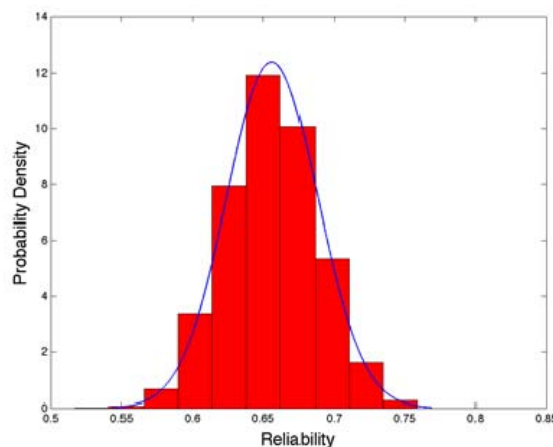
Numerical Example: Comparison of Asymptotic Normality and Bootstrap in Series System

- Consider reliability example with four independent subsystems in series
- Assume $n_1 = n_2 = n_3 = n_4 = 100$ subsystem samples; also have $n = 100$ full system samples
- Compare asymptotic density function (from asymptotic normality and Fisher info. matrix for θ) with bootstrap-based histogram
- Other studies (including parallel systems) show very similar results



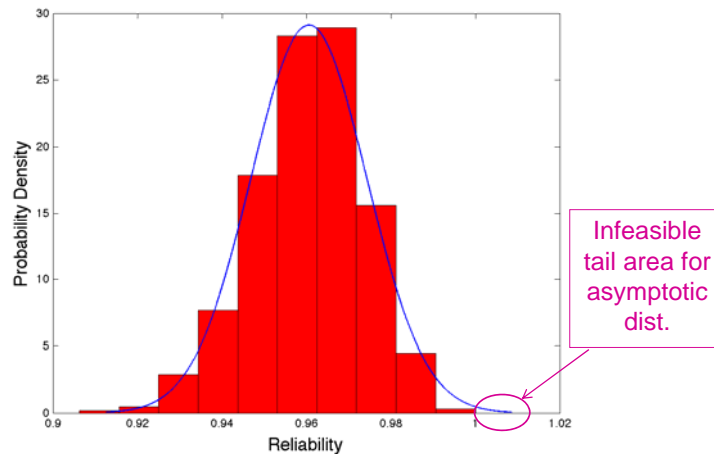
27

Comparison of Asymptotic Density and Bootstrap Histogram: Common Sample Sizes of 100, True $\rho = \ell = 0.656$ (low reliability system)



Good agreement between asymptotic distribution (using Fisher information) and bootstrap-based histogram.

Comparison of Asymptotic Density and Bootstrap Histogram: Common Sample Sizes of 100, True $\rho = \ell = 0.961$ (high reliability system)



Reasonable agreement between asymptotic distribution and bootstrap-based histogram; however, asymptotic dist. has infeasible tail area > 1 .

References

- Hesterberg, T. (2011), "Bootstrap," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 3, pp. 497–526.
- Spall, J. C. (2012), "Asymptotic Normality and Uncertainty Bounds for Reliability Estimates from Subsystem and Full System Tests," *Proceedings of the American Control Conference*, 27–29 June 2012, Montréal, Canada, pp. 56–61 (paper WeA02.3).
- Spall, J. C. (2014), "Identification for Systems with Binary Subsystems," *IEEE Transactions on Automatic Control*, vol. 59(1), pp. 3–17. [dx.doi.org/10.1109/TAC.2013.2275664](https://doi.org/10.1109/TAC.2013.2275664)
- Varian, H. (2005), "Bootstrap Tutorial," *Mathematica Journal*, vol.9(4).
- Zhao, X. and Spall, J. C. (2016), "Estimating Travel Time in Urban Traffic by Modeling Transportation Network Systems with Binary Subsystems," *Proceedings of the American Control Conference*, Boston, MA, 6–8 July 2016, pp. 803–808. [http://dx.doi.org/10.1109/ACC.2016.7525012](https://doi.org/10.1109/ACC.2016.7525012)