ISyE 6416: Computational Statistics
Spring 2017

Lecture 12: Bootstrap

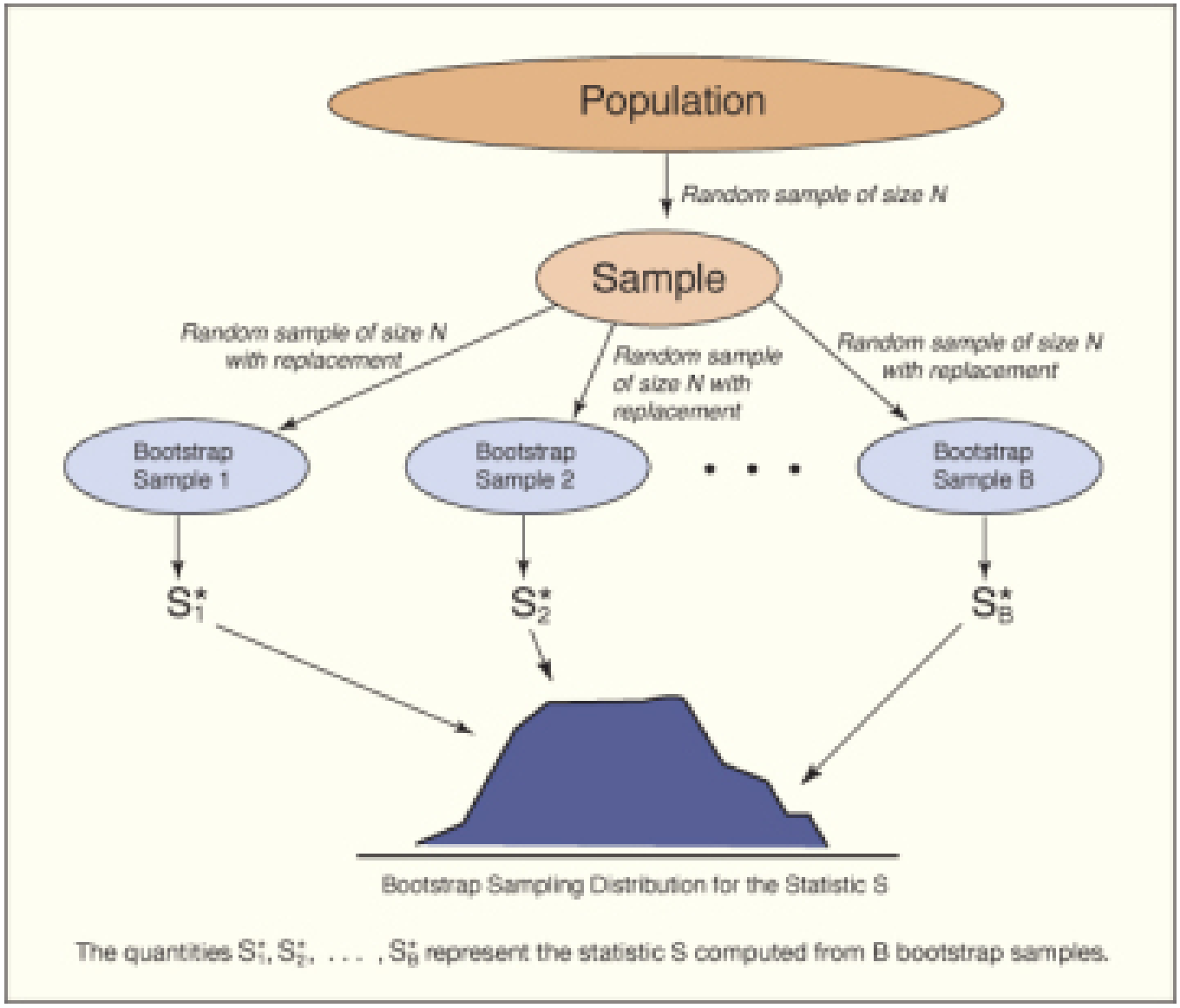
Prof. Yao Xie

H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Main idea

- in statistics, we learn about characteristics of the population by taking samples. As sample represents the population, analogous characteristics of the sample should give us information about the population characteristics
- bootstrapping learns about the sample characteristics by taking resamples and use the information to infer to the population
- resample: we retake samples from the original samples
- it provides an powerful too to calculate the standard error of an estimator, construct confidence intervals, and many other uses
- theory behind bootstrap is sophisticated, being based on Edge-worth expansions

Main idea

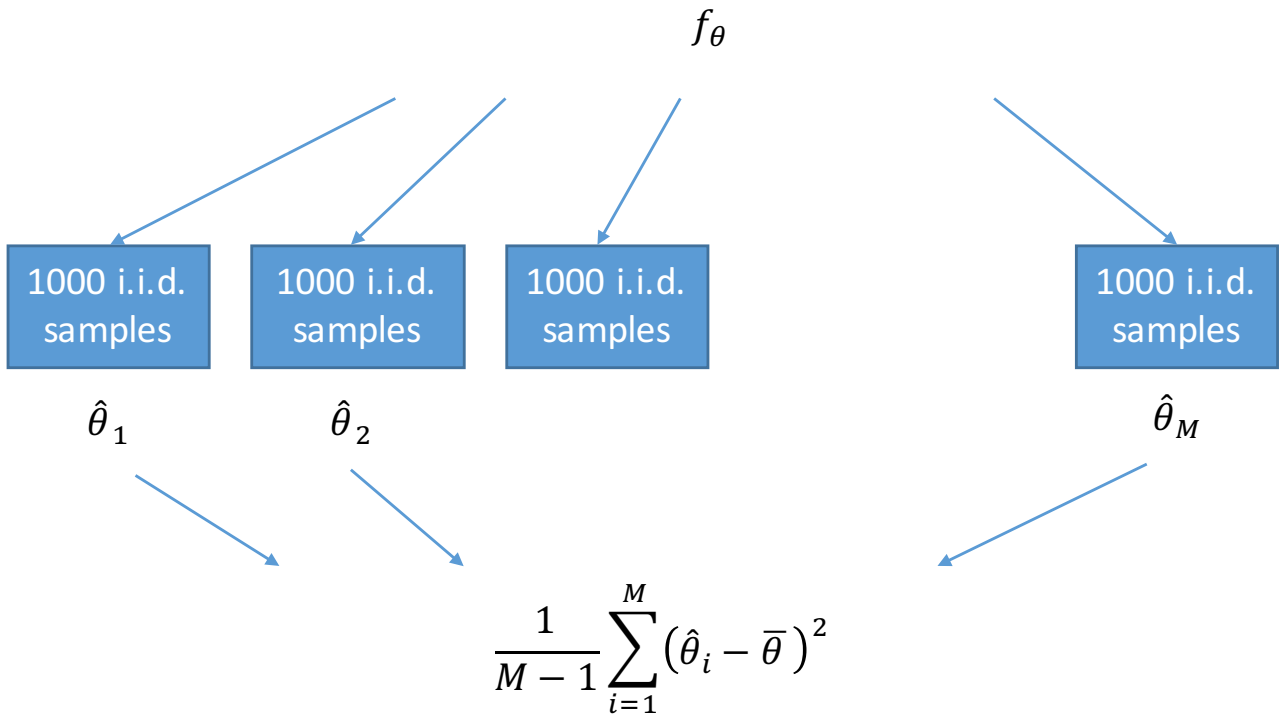

Population
Random sample of size N
Sample
Random sample of size N with replacement
Random sample of size N with replacement
Random sample of size N with replacement
Bootstrap Sample 1
Bootstrap Sample 2
Bootstrap Sample B
S_1^*
S_2^*
S_B^*
Bootstrap Sampling Distribution for the Statistic S
The quantities $S_1^*, S_2^*, \ldots, S_B^*$ represent the statistic S computed from B bootstrap samples.

Basic setup

- bootstrap was first men.oned by Bradley Efron in 1979.
- assume independent samples $z_1, \ldots, z_n \sim f_\theta$
- subscript θ emphasizes that it is the parameter of interest
- let $\hat{\theta}$ be an estimate of θ that we computed from the samples $z_1, \ldots, z_n$
- Example: suppose we are interested in knowing $\text{Var}(\hat{\theta})$

Example: estimating $\mathrm{Var}(\hat{\theta})$

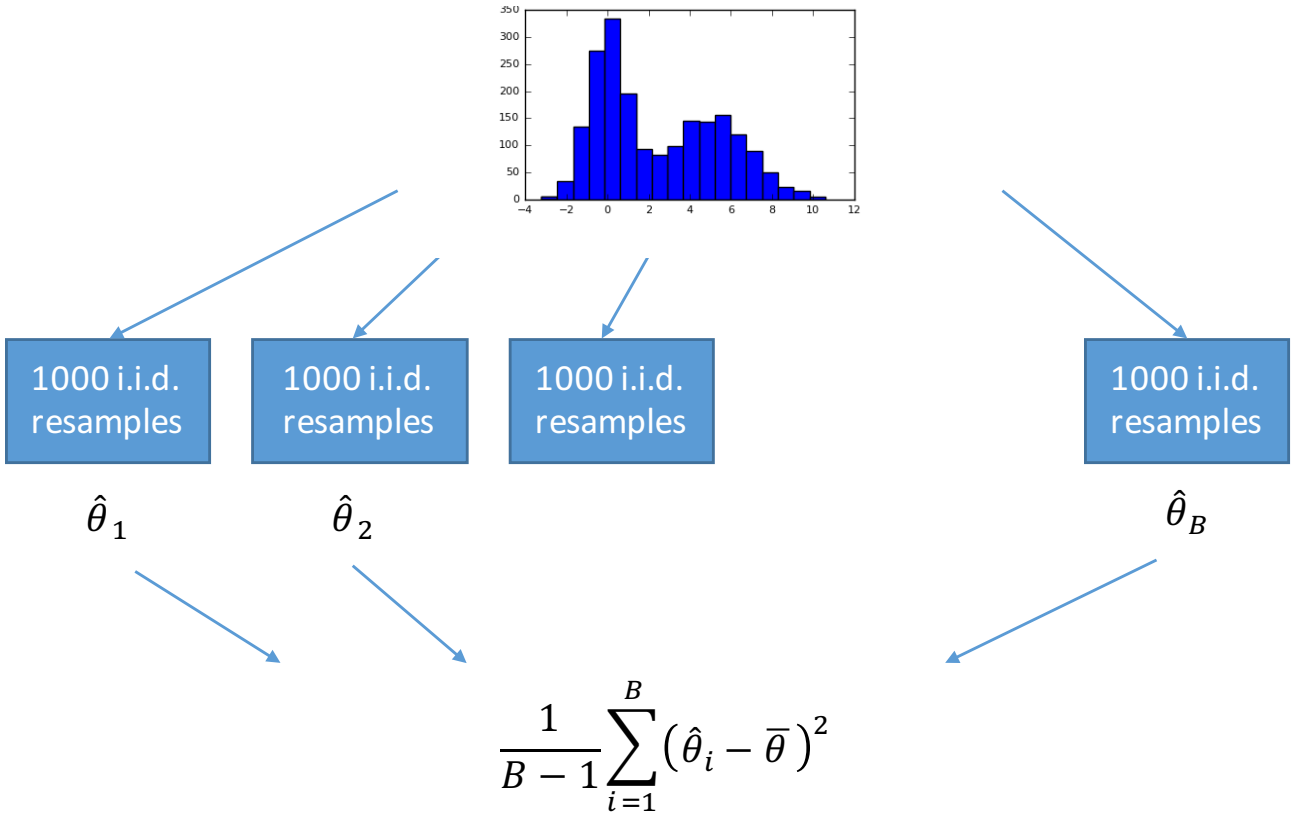
- If we know f_θ, we can generate new samples to recompute the statistic, and take the sample variance of these estimators



- However, we do not know f_θ

Example: estimating $\mathrm{Var}(\hat{\theta})$

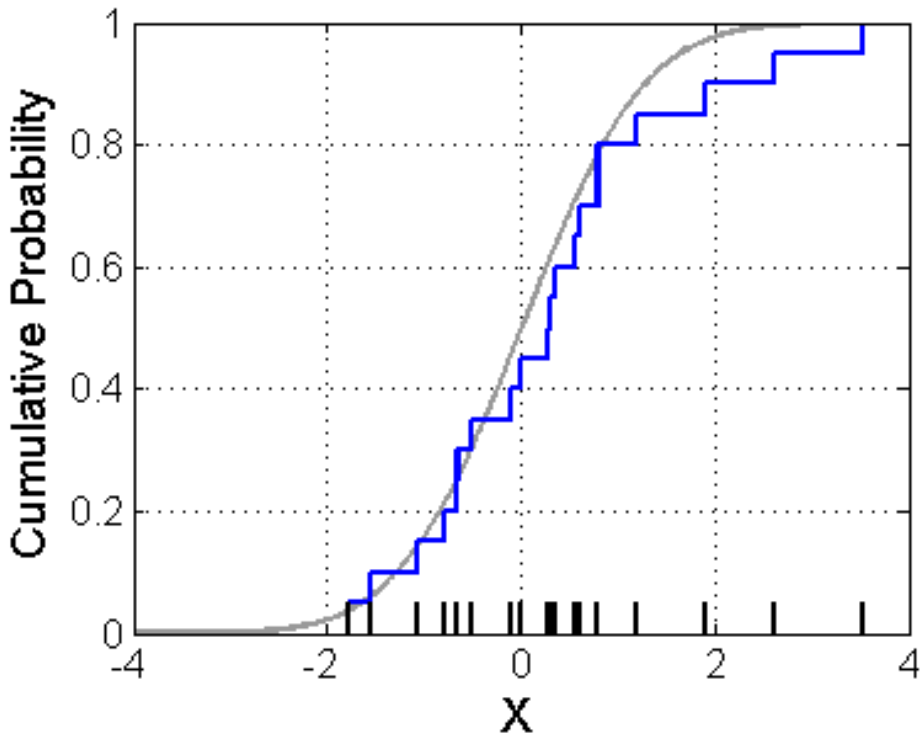
- Idea: use the observed samples $z_1, \ldots, z_n$ to generate n "new" samples, as if they come from f_θ



- precisely, the new samples are i.i.d. $\tilde{z}_j \sim \mathrm{Unif}\{z_1, \ldots, z_n\}$ sample with replacement

Interpretation

- we can think $\tilde{z}_1, \ldots, \tilde{z}_n$ as coming from an *empirical distribution function* of the original samples $z_1, \ldots, z_n$, which is a discrete distribution with probability mass $1/n$ at each of the values $z_1, \ldots, z_n$



Resampling

- Suppose we roll an 8-sided die 10 times and get the following data, written in increasing order:

$$1, 1, 2, 3, 3, 3, 3, 4, 7, 7.$$

- Imagine writing these values on 10 slips of paper, putting them in a hat and drawing one at random.
- The probability of drawing a 3 is 4/10 and the probability of drawing a 4 is 1/10.
- Empirical distribution

value x	1	2	3	4	7
$p(x)$	2/10	1/10	4/10	1/10	2/10

- To resample is to draw a number j from the uniform distribution on $\{1, 2, ..., 10\}$ and take x_j as our resampled value.
- If we want a resampled data set of size 5, then we roll the 10-sided die 5 times and choose the corresponding elements from the list of data.
- if the 5 rolls are

$$5, 3, 6, 6, 1$$

- the resample is

$$3, 2, 3, 3, 1.$$

- same data point can appear multiple times when we resample.

Source: J. Orloff and J. Bloom.

A running example

- Suppose we have two random variables $X, Y \in \mathbb{R}$ representing yields of two financial assets
- we will invest a fraction θ of our money in X, and $1-\theta$ in Y
- total yield $\theta X + (1-\theta)Y$
- we want to minimize the variance (risk) our our decision $\mathsf{Var}(\theta X + (1-\theta)Y)$
- one can show that the minimizer is

$$\theta = \frac{\sigma_Y^2 - \sigma_{XY}}{\sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}}$$

$\sigma_X^2 = \mathsf{Var}(X)$, $\sigma_Y^2 = \mathsf{Var}(Y)$, $\sigma_{XY} = \mathsf{Cov}(X, Y)$

Chapter 5 of Elements of Statistical Learning.

- Given n samples of past measurements (x_1, y_1), ..., (x_n, y_n) for the yields, we can compute the estimates for $\hat{\sigma}_X^2$, $\hat{\sigma}_Y^2$, and $\hat{\sigma}_{XY}$
- sample solution

$$\hat{\theta} = \frac{\hat{\sigma}_Y^2 - \hat{\sigma}_{XY}}{\hat{\sigma}_X^2 + \hat{\sigma}_Y^2 - 2\hat{\sigma}_{XY}}$$

- we may use bootstrap to estimate the variance of $\hat{\theta}$
- even if we have the parametric distribution from for joint distribution of (X, Y), computing the variance of $\hat{\theta}$ is still hard because the estimates are used in both the numerator and the denominator

bootstrap procedure illustration

- let $z_1 = (x_1, y_1)$, ..., $z_n = (x_n, y_n)$
- pick a large number B, say $B = 1000$, and repeat for $b = 1, \ldots, B$
  - draw a bootstrap sampling $z_1^{(b)}, \ldots, z_n^{(b)}$ as described earlier
  - compute the value of the statistics on these resamples $\hat{\theta}^{(b)}$
- To estimate the standard error of $\hat{\theta}$, we use

$$\mathsf{SE}(\hat{\theta}) \approx \left[\frac{1}{B} \sum_{b=1}^{B} \left(\hat{\theta}^{(b)} - \frac{1}{B} \sum_{r=1}^{B} \hat{\theta}^{(r)} \right)^2 \right]^{1/2}$$

Example: estimating correlation

Efron (1982) analyzes data on law school admission, with the object being to examine the correlation between LSAT score and the first year GPA. For each of 15 law schools, we have the pairs

(576, 3.93)	(635, 3.30)	(558, 2.81)	(578, 3.03)	(666, 3.44)
(580, 3.07)	(555, 3.00)	(661, 3.43)	(651, 3.36)	(605, 3.13)
(653, 3.12)	(575, 2.74)	(545, 2.76)	(572, 2.88)	(594, 2.96)

- estimate for correlation

$$\hat{\rho} = \frac{\sqrt{\hat{\sigma}_{XY}}}{\hat{\sigma}_X \hat{\sigma}_Y}$$

- use $B = 1000$ resamples bootstrap to estimate the variance of the estimator

Usage

- Can be easily used to estimate the mean

$$\frac{1}{B}\sum_{i=1}^{B}\hat{\theta}_i$$

  and standard error (square-root of variance) of the estimator $\hat{\theta}$
- Bootstrap confidence interval is a very useful technique, and it requires slightly more work

Bootstrap confidence interval

- Confidence interval for θ computed over $z_1, \ldots, z_n$

$$\mathbb{P}(L \leq \theta \leq U) = 1 - \alpha.$$

- idea: use the bootstrap statistics $\hat{\theta}_1, \ldots, \hat{\theta}_B$, denote these as samples from a random variable $\tilde{\theta}$
- find the $\alpha/2$ and $1 - \alpha/2$ quantiles of boostrap statistics $\hat{\theta}_1, \ldots, \hat{\theta}_B$, denoted as $q_{\alpha/2}$ and $q_{1-\alpha/2}$
- approximate the distribution $\hat{\theta} - \theta$ by $\tilde{\theta} - \hat{\theta}$
- Use $[2\hat{\theta} - q_{1-\alpha/2}, 2\hat{\theta} - q_{\alpha/2}]$ as an approximate $(1 - \alpha)$ confidence interval for θ
  Proof:

Parametric bootstrap

- Previous approach is called *empirical bootstrap*, without making any assumption about the data distribution
- For *parametric bootstrap*, we make use of assumed parametric distribution form to generate the bootstrap samples
- Example: Suppose the data $x_1, \ldots, x_{300}$ is drawn from an exp(λ) distribution. Estimate λ and give a 95% parametric bootstrap confidence interval for λ