Bootstrap Methods Short Course

By Michael Chernick

To the International Biometrics Society German Region

November 4 – November 5, 2010

Bootstrap Topics Day 1

- Introduction to Bootstrap
- Wide Variety of Applications
- Estimating Bias
- Error Rate Estimation in Discriminant Analysis
- Confidence regions and hypothesis tests

Bootstrap Topics Day 2

- **Examples of bootstrap applications: (1) P-value adjustment - consulting example, (2) Confidence Interval for Process Capability C_{pk}, (3) Bioequivalence - Efron's Patch Data example**
- **Examples where bootstrap is not consistent: (1) infinite variance case for a population mean, (2) extreme values, (3) survey sampling**
- **Available Software**
- **Efficient Algorithms in SAS**
- **Examples with Software Solutions**

Introduction

- The bootstrap is a general method for doing statistical analysis without making strong parametric assumptions.
- Efron’s nonparametric bootstrap, resamples the original data.
- It was originally designed to estimate bias and standard errors for statistical estimates much like the jackknife.

Introduction (continued)

The bootstrap is similar to earlier techniques which are also called re-sampling methods:

(1) jackknife,

(2) cross-validation,

(3) delta method,

(4) permutation methods, and

(5) subsampling.

- It is called bootstrap because Efron viewed it as an analysis tool based solely on the data.
- The data are the bootstraps and the statistician does inference by “picking himself up by his own bootstraps” as was attributed to the fictional Baron von Münchausen.

Introduction (continued)

The technique was extended, modified and refined to handle a wide variety of problems including:

(1) confidence intervals and hypothesis tests,
(2) linear and nonlinear regression,
(3) time series analysis,
(4) complex survey sampling data,
(5) missing and censored data,
(6) spatial data analysis,
(7) point processes, and
(8) model subset selection.

Introduction (continued)

- The bootstrap has now seen applications in many disciplines including:

(1) process capability, (2) reliability, (3) communications,

(4) image and signal processing, (5) auditing, (6) meteorology,

(7) sports medicine, (8) chemistry, (9) ornithology, (10) physics,

(11) political science, (12) ecology, (13) evolution, (14) genetics,

(15) behavioral sciences, (16) biology and medicine, (17) psychology,

(18) geology, (19) astronomy, and (20) economics.

Introduction (continued)

- Definition of Efron's nonparametric bootstrap.
- Given a sample of n independent identically distributed (i.i.d.) observations $X_1, X_2, \ldots, X_n$ from a distribution F and a parameter θ of the distribution F with a real valued estimator

  $\theta(X_1, X_2, \ldots, X_n)$, the bootstrap estimates the accuracy of the estimator by replacing F with F_n, the empirical distribution, where F_n places probability mass 1/n at each observation X_i.

Introduction (continued)

- Let $X_1^*, X_2^*, \ldots, X_n^*$ be a bootstrap sample, that is a sample of size n taken with replacement from F_n .
- The bootstrap, estimates the variance of

  $\theta(X_1, X_2, \ldots, X_n)$ by computing or approximating the variance of $\theta^* = \theta(X_1^*, X_2^*, \ldots, X_n^*)$.

Introduction (continued)

- The bias, the median or any other property of $\theta(X_1, X_2, \ldots, X_n)$ is estimated by applying the bootstrap version of that quantity.
- Sometimes, the bootstrap estimate can be obtained analytically, as with the case of the standard deviation for a median or mean.
- However, in most cases the bootstrap estimate is approximated by Monte Carlo.

Introduction (continued)

- The k θ^*s, provide a sample distribution of θs to obtain standard deviations, bias or any other property of the distribution for θ.
- Note that there is variability among bootstrap samples because we sample with replacement.
- So, in a particular bootstrap sample, some observations will appear two or more times and others not at all.
- For estimating standard deviations or biases, k is recommended to be at least 100, although some suggest that even larger values may be required.
- For confidence interval estimation or hypothesis testing at least 1000 bootstrap replications are recommended.

Introduction (continued)

- Statistical Functionals - A functional is a mapping that takes functions into real numbers.
- Parameters of a distribution can usually be expressed as functionals of the population distribution.
- Often the standard estimate of a parameter is the same functional applied to the empirical distribution.
- Such functionals whose arguments are random quantities are called statistical functionals.

Introduction (continued)

- Statistical Functionals and the bootstrap.
- A parameter θ is a functional T(F) where T denotes the functional and F is a population distribution.
- An estimator of θ is $\theta_h = T(F_n)$ where F_n is the empirical distribution function.
- Many statistical problems involve properties of the distribution of $\theta - \theta_h$, its mean (bias of θ_h), variance, median etc.

Introduction (continued)

- Bootstrap idea: Cannot determine the distribution of $\theta - \theta_h$ but through the bootstrap we can determine, or approximate through Monte Carlo, the distribution of $\theta_h - \theta^*$, where $\theta^* = T(F_n^*)$ and F_n^* is the empirical distribution for a bootstrap sample $X_1^*, X_2^*,\ldots,X_n^*$ (θ^* is a bootstrap estimate of θ).
- Based on k bootstrap samples the Monte Carlo approximation to the distribution of $\theta_h - \theta^*$ is used to estimate bias, variance etc. for θ_h .
- In bootstrapping θ_h substitutes for θ and θ^* substitutes for θ_h . Called the bootstrap principle.

Introduction (continued)

- Basic Theory: Mathematical results show that bootstrap estimates are consistent in particular cases.
- Mathematical tools including Edgeworth and Cornish-Fisher expansions are used to demonstrate rates of convergence for bootstrap estimates satisfying certain smoothness criteria.
- Basic Idea: Empirical distributions behave in large samples like population distributions. Glivenko-Cantelli Theorem tells us this. Shorack (2000) also discusses this.
- The smoothness condition is needed to transfer consistency to functionals of F_n, such as the estimate of the parameter θ.

Wide Variety of Applications

- Initially the bootstrap was used as an alternative to the jackknife to provide estimates of standard deviations and biases in complex estimation problems but for independent identically distributed observations (i.i.d.).
- However Efron and others recognized that through the power of fast computing the Monte Carlo approximation could be extended to many different statistical problems (not just i.i.d. situations).

Wide Variety of Applications (continued)

- The bootstrap and other computer-intensive procedures such as permutation methods are attractive because they free the researcher from restrictive parametric assumptions and oversimplified models.
- Often data are skewed, multimodal or have outlying values due to heavy-tailed distributions.
- Regression models do not need to be linearized and the outcome variable does not even need to be expressed as a closed form function of input variables (a computer algorithm will do).

Wide Variety of Applications (continued)

- The bootstrap has been applied in survival and reliability analyses where data are right censored.
- It is used for subset selection in regression (linear and non-linear) and logistic regression.
- It is used to estimate error rates for discriminant functions.

Wide Variety of Applications (continued)

- It can estimate process capability indices for non-Gaussian data.
- It is used to adjust p-values in a variety of multiple comparison situations.
- It can be extended to problems involving dependent data including multivariate, spatial and time series data and in sampling from finite populations.

Wide Variety of Applications (continued)

- It also has been applied to problems involving missing data.
- In many cases, the theory justifying the use of bootstrap (e.g. consistency theorems) has been extended to these non i.i.d. settings.
- In other cases, the bootstrap has been modified to "make it work." The general case of confidence interval estimation is a notable example.

Wide Variety of Applications (continued)

- For some problems, the bootstrap seems to work without theoretical justification, but with support from simulation studies.
- A primary example has been the estimation of error rates for discriminant functions in small samples where a bootstrap variant called the .632 estimator has been shown to be superior to Lachenbruch and Mickey's popular leave-one-out estimator in a variety of situations with small training sample sizes.

Wide Variety of Applications (continued)

- The bootstrap has been called “computer-intensive” because in many of the applications where it is valuable, the Monte Carlo approximation is required. Sometimes a variant called bootstrap iteration is even more helpful but it requires much more Monte Carlo sample generation.

- However, there are cases where the bootstrap estimate can be obtained analytically (e.g. standard error of a sample mean, standard deviation of a sample median and censored matched pairs test for equality of distributions). So sometimes no Monte Carlo approximation is needed and calculations are not intensive at all!

Estimating Bias

How to do it by bootstrapping.

- Let E(X) denote the expected or mean value of a random variable X. For an estimator θ_h of a parameter θ, $\theta_h - \theta$ represents our X.
- The bias of θ_h is usually defined as

  $b = E(\theta_h - \theta)$.

Estimating Bias (continued)

- A bootstrap estimate for the bias b of θ_h is given by b* = $E(\theta^* - \theta_h)$.
- As is common with bootstrapping, a bootstrap estimate of θ, θ^* takes the place of θ_h and θ_h takes the place of θ.
- The Monte Carlo approximation for the bootstrap estimate is $B_{MONTE} = \Sigma_i (\theta^*_i - \theta_h)/k$ where θ^*_i is the i^{th} bootstrap sample estimate of θ for $1 \leq i \leq k$.

Estimating Bias (continued)

- Generally, the purpose of estimating bias is to improve a biased estimator by subtracting an estimate of its bias.
- Efron took a poor estimate, the re-substitution estimate which has a large bias, estimated its bias by bootstrap methods and subtracted the estimated bias from the re-substitution estimate.
- This produced a better estimate of the error rate than the best known method at that time (1983).

Estimating Bias (continued)

Examples:

(1) error rate estimation in discriminant analysis (ref. Chernick (1999) pp. 50-82 or (2007) pp. 28-44) .

(2) ratio estimates (Efron's patch data example { refs. Chernick (1999) pp. 86 – 89 or (2007) pp.44-46 and Efron and Tibshirani (1993) pp. 126 –133 }).

Estimating Bias - Error Rate Estimation in Discriminant Analysis

- Two class discrimination problem: There are two classes of objects along with the variables that differ between classes.
- The objective is to construct a rule to classify objects based on the values of the variables.
- Example: Targets - Reentry vehicles with warheads. Decoys - Balloons made to look like reentry vehicles.

Estimating Bias - Error Rate Estimation in Discriminant Analysis (continued)

- Problem: Given a training set of feature vectors associated with known groups, targets and decoys construct a classification or discrimination rule to identify unknown objects based on values of the features.
- Given a priori information on the ratio of decoys to targets and known densities (or in practice estimated densities based on the training data), Bayes theorem provides the optimal rule that minimizes the expected classification error rates.

Estimating Bias - Error Rate Estimation in Discriminant Analysis (continued)

- This optimal rule is called the Bayes rule.
- Results: (1) If the feature vectors are assumed to have a multivariate Gaussian distribution with the same covariance matrix for each class, the Bayes rule is linear. (2) If the feature vectors are Gaussian with different covariances by class, the Bayes rule is quadratic.

Estimating Bias - Error Rate Estimation in Discriminant Analysis (continued)

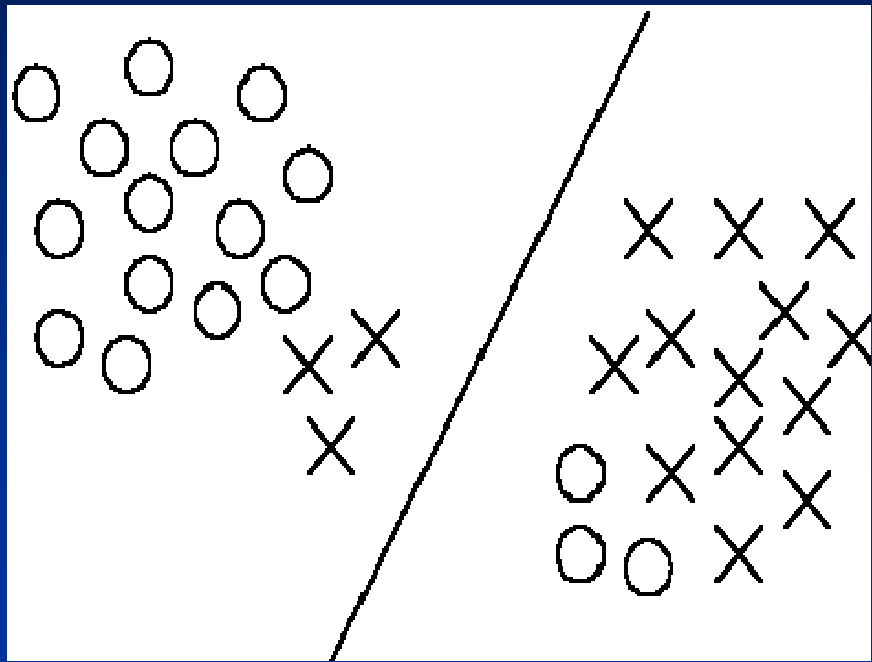
- For more details on the discrimination problem see Duda and Hart (1972) or Chernick (1999, 2007) Sections 2.1.2 and 2.1.3.
- In practice, a form for the discriminant rule is chosen and parameters are estimated from training data.
- Fisher's linear discriminant rule is one such example.
- Kernel discriminant rules and quadratic rules are other choices.

Estimating Bias - Error Rate Estimation in Discriminant Analysis (continued)

- Given a discriminant function, we wish to characterize its performance based on its error rates.
- The re-substitution method, estimates the error rate by classifying the training vectors and counting the proportion misclassified.
- It is very biased in small samples because the rule is fit to the same data that it is tested on.

Estimating Bias - Error Rate Estimation in Discriminant Analysis (continued)

- The picture on the right shows discriminant function boundaries for a two-dimensional two class problem. The Xs to the left of the linear boundary and the Os to the right of the boundary are classification errors



Estimating Bias - Error Rate Estimation in Discriminant Analysis (continued)

- The leave-one-out method provides an error rate estimator first given by Lachenbruch and Mickey (1968).
- It constructs the classification rule with one training vector left out and classifies only the one left out.
- This is done in turn for each training vector left out. The error rate is the percentage of cases with the left out case misclassified.
- This method takes advantage of all the training data and is unbiased. It use to be very popular.

Estimating Bias - Error Rate Estimation in Discriminant Analysis (continued)

- However, Efron (1983) showed that the leave-one-out method could be improved on by bootstrapping.
- The basic bootstrap approach takes the re-substitution estimate, gets a bootstrap estimate of its bias and subtracts it from the re-substitution estimate to get a good estimate of the error rate.

Estimating Bias - Error Rate Estimation in Discriminant Analysis (continued)

- Efron (1983) introduces the bootstrap, double bootstrap, e_0 and the .632 estimator (all variations on the bootstrap) and compares them to the leave-one-out estimate using a variety of small sample simulations with Gaussian features.
- The .632 estimator is a weighted average of re-substitution and e_0, giving weight 0.632 to e_0 and weight 0.368 to re-substitution.
- Other authors including Chernick, Murthy and Nealy (1985, 1986, 1988a, and 1988b), Chatterjee and Chatterjee (1983) and Jain, Dubes and Chen (1987) have done similar small sample Monte Carlo comparisons for Gaussian and non-Gaussian feature distributions with linear and quadratic classifiers.

Estimating Bias - Error Rate Estimation in Discriminant Analysis (continued)

- All these papers show advantages to bootstrap type estimates particularly the .632 estimator.
- Efron and Tibshirani (1997) have a modification called 632+ which appears to be even better.
- The following simulation results appear in Chernick (1999) and are the summary tables from Chernick, Murthy and Nealy (1988b) and Chernick, Murthy and Nealy (1986).
- The 1986 paper dealt only with Gaussian populations.
- Results from all the simulations were summarized and the various estimators were listed based on how many times they ranked first, second or third.

Estimating Bias - Error Rate Estimation in Discriminant Analysis (continued)

- In addition to bootstrap (BOOT), .632 (632), e_0, resubstitution (APP) and leave-one-out (U) two other bootstrap variations, MC and convex bootstrap (CONV) were considered.
- See Chernick, Murthy and Nealy (1985) for definitions.
- Efron (1983) also considered a randomized bootstrap which was not consider by other researchers in later Monte Carlo studies.

Estimating Bias - Error Rate Estimation in Discriminant Analysis (continued)

- Chernick, Murthy and Nealy (1988b) compared the same estimators for a class of Pearson VII multivariate distributions.
- This family was chosen since the tail behavior of the distribution could be controlled by a single parameter m and contours of constant probability density are elliptic as is also the case for Gaussian densities.
- The family has second moments for m > 2.5 when p, the number of features in the feature vector is 2 and it has a first moment only when m > 1.5 and p = 2.
- For the results we present only the case p = 2 is considered.

Table for Comparison of Error Rate Estimators - Pearson VII Case from Chernick, Murthy and Nealy (1988b) with ranks based on root mean square error of estimated error rate

Rank	632	MC	e_0	BOOT	CONV	U	APP	Total
	– M = 1.3							
First	0	0	2	0	10	0	0	12
Second	3	0	0	9	0	0	0	12
Third	0	9	0	1	2	0	0	12
Total	3	9	2	10	12	0	0	36
	– M = 1.5							
First	6	1	8	5	12	0	1	33
Second	8	4	0	14	7	0	0	33
Third	3	15	2	4	8	0	1	33
Total	17	20	10	23	27	0	2	99

Table for Comparison of Error Rate Estimators - Pearson VII Case from Chernick, Murthy and Nealy (1988b) with ranks based on root mean square error of estimated error rate

M = 2.0

Rank	632	MC	e_0	BOOT	CONV	U	APP	Total
First	18	1	3	0	1	0	7	30
Second	10	4	4	2	5	2	3	30
Third	1	9	3	8	5	0	3	30
Total	29	14	10	10	11	2	13	90

– M = 2.5

Rank	632	MC	e_0	BOOT	CONV	U	APP	Total
First	21	0	8	1	0	0	3	33
Second	10	3	4	5	4	2	5	33
Third	1	13	1	6	10	0	2	33
Total	32	16	13	12	14	2	10	99

Table for Comparison of Error Rate Estimators - Pearson VII Case from Chernick, Murthy and Nealy (1988b) with ranks based on root mean square error of estimated error rate

Rank	632	MC	e_0	BOOT	CONV	U	APP	Total
	– M = 3.0							
First	21	0	6	0	0	0	3	30
Second	9	3	5	3	2	2	6	30
Third	0	8	1	8	11	1	1	30
Total	30	11	12	11	13	3	10	90

Summary of Table on Pearson VII cases

- When m < 1.7 the convex bootstrap appears to be the best with the ordinary bootstrap second.
- The e_0 estimate is also very competitive for low values of m.
- When m is 2.0 or higher the .632 estimate takes over as the clear winner.

Summary of Table on Pearson VII cases (continued)

- Generally the .632 estimate appropriately weight the e_0 and re-substitution estimates adjusting for their opposite biases.
- However for low m the pessimistic bias of e_0 disappears and .632 no longer provides the best weighted average of these two estimators.
- Refer to Chernick, Murthy and Nealy (1988b) for specifics on individual cases.

Table for Comparison of Error Rate Estimators -Gaussian Cases
from Chernick, Murthy and Nealy (1986) with ranks based on root mean square error of estimated error rate

Rank	632	MC	e_0	BOOT	CONV	U	APP	Total
First	72	1	29	6	0	0	1	109
Second	21	13	27	23	11	1	13	109
Third	7	20	8	25	37	7	5	109
Total	100	34	64	54	48	8	19	327

Summary of Table on Gaussian cases

- The .632 estimator appropriately weights the e_0 and re-substitution estimates to nearly cancel their opposite biases.
- This make it the clear winner in most of the cases (72 out of 109).
- It finishes in the top three in 100 out of the 109 cases.
- See Chernick, Murthy and Nealy (1985, 1986) for a discussion of the individual cases.

References on estimating bias

(1) Chatterjee, S., and Chatterjee, S., (1983) Estimation of misclassification probabilities by bootstrap methods*. Commun. Statist. Simulation and Computation* 12, 645-656.

(2) Chernick, M.R. (1999). *Bootstrap Methods: A Practitioner's Guide.* Wiley, New York.

(3) Chernick, M.R. (2007). *Bootstrap Methods: A Guide for Practitioners and Researchers, 2nd Edition.* Wiley, New York.

(4) Chernick, M.R., Murthy, V.K., and Nealy, C.D. (1985). Application of bootstrap and other resampling techniques: evaluation of classifier performance. *Pattern Recogn. Lett.* 3, 167-178.

(5) Chernick, M.R., Murthy, V.K., and Nealy, C.D. (1986). Correction note to application bootstrap and other resampling techniques: evaluation of classifier performance. *Pattern Recogn. Lett.* 4, 133-142.

References on estimating bias (continued)

(6) Chernick, M.R., Murthy, V.K., and Nealy, C.D. (1988a). Estimation of error rate for linear discriminant functions by resampling: non-Gaussian populations. *Comput. Math. Applic.* 15, 29-37.

(7) Chernick, M.R., Murthy, V.K., and Nealy, C.D. (1988b). Resampling-type error rate estimation for linear discriminant functions: Pearson VII distributions. *Comput. Math. Applic.* 15, 897-902.

(8) Duda, R.O. and Hart, P.E. (1973). *Pattern Recognition and Scene Analysis.* Wiley, New York.

(9) Efron, B. and Tibshirani, R. (1993). *An Introduction to the Bootstrap.* Chapman & Hall, New York

References on estimating bias (continued)

(10) Jain, A.K., Dubes, R.C., and Chen, C. (1987). Bootstrap techniques for error rate estimation. *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-9,** 628-633.

(11) Lachenbruch, P.A., and Mickey, M.R. (1968). Estimation of error rates in discriminant analysis. *Technometrics* **10,** 1-11.

Confidence regions and hypothesis tests

- The simplest way to generate approximate nonparametric confidence intervals by the bootstrap is by taking the appropriate percentiles of the bootstrap samples, i. e. from the k Monte Carlo replications of bootstrap samples.
- For example a two-sided approximate 95% confidence interval for a parameter θ would be obtained as the interval from the 2.5 to the 97.5 percentile of the distribution of bootstrap samples.
- This method is called Efron's percentile method.

Confidence regions and hypothesis tests (continued)

- The percentile method and other bootstrap variations may require 1000 or more bootstrap replications to be very useful.
- The percentile method only works under special conditions.
- Bias correction and other adjustments are sometimes needed to make the bootstrap “accurate” and “correct” when the sample size n is small or moderate.

Confidence regions and hypothesis tests (continued)

- Confidence intervals are **accurate** or **nearly exact** when the stated confidence level for the intervals is approximately the long run probability that the random interval contains the "true" value of the parameter.
- Accurate confidence intervals are said to be **correct** if they are approximately the shortest length confidence intervals possible for the given confidence level.

Confidence regions and hypothesis tests (continued)

- The BC_a method, the iterated bootstrap (or double bootstrap) and the bootstrap t method are methods for constructing bootstrap confidence intervals that are closer to being exact (accurate) and correct than the percentile method in many circumstances.
- See Chernick (2007) pp. 57-65 for details on these methods.

Confidence regions and hypothesis tests (continued)

- Motivation for percentile method:
- Hartigan’s typical value theorem was the motivation for the percentile method.
- It applies to random sub-sampling and only for a particular class of estimates called M-estimates.
- Also he assumed that the population distribution had a symmetric density.

Confidence regions and hypothesis tests (continued)

- Under those conditions he showed that a particular confidence set was exact.
- Efron saw the following analogies:

(1) Hartigan's confidence set was very similar to the percentile method confidence set.

(2) Bootstrap sampling is similar to random sub-sampling.

Confidence regions and hypothesis tests (continued)

- Efron wanted to drop the symmetry and M-estimator requirement (perhaps they were not necessary conditions).
- However the percentile method and all other bootstrap methods do **<u>not</u>** give exact confidence intervals.
- Whenever the class of alternative hypotheses is very large as is the general “nonparametric” setting for the bootstrap there can be no uniformly most powerful test and consequently no correct and exact confidence interval for the hypothesized parameter (Bahadur and Savage [1956])

Confidence regions and hypothesis tests (continued)

- Hartigan's result motivated Efron to propose the percentile bootstrap method for confidence intervals.
- Efron found situations where the percentile method works well (namely when there exists a monotone transformation $\phi = g(\theta)$ such that $\phi_h = g(\theta_h)$ has an approximate Gaussian distribution).

Confidence regions and hypothesis tests (continued)

- In more general situations modifications have been found which work better. See Chernick (2007) pp. 57-64.
- Hartigan's theorem is discussed in Chernick (2007) pp. 55-57 and Efron (1982) pp 69-73.

Confidence regions and hypothesis tests (continued)

- Hall and Martin have shown the rate at which various bootstrap estimates approach their advertised confidence levels as the size n of the original sample increases.
- They use Edgeworth and Cornish-Fisher expansions to prove these results.
- See Hall (1992) Chapter 3 or Chernick (2007) Section 3.1 for more discussion of this.
- See Ewens and Grant (2001) Chapter 12 for another nice treatment and comparison with permutation tests.

Four Methods for Setting Approximate Confidence Intervals for a Real-Valued Parameter θ

Method	Abbrev-iation	α-Level Endpoint	Correct if
1. Standard Normal Approximation	$\theta_S[\alpha]$	$\theta_h + \sigma_h z^{(\alpha)}$	$\theta_h \approx N(\theta, \sigma^2)$ with σ constant
2. Percentile	$\theta_P[\alpha]$	$G_h^{-1}(\alpha)$	There exists a monotone transformation such that $\phi_h = g(\theta_h)$ where $\phi = g(\theta)$ and $\phi_h \approx N(\phi, \tau^2)$ and τ is constant
3. Bias-corrected	$\theta_{BC}[\alpha]$	$G_h^{-1}(\phi\{2z_0 + z^{(\alpha)}\})$	There exists a monotone transformation such that $\phi_h = g(\theta_h)$ where $\phi = g(\theta)$ and $\phi_h \approx N(\phi - z_0\tau, \tau^2)$ and τ and z_0 are constant
4. BC_a	$\theta_{BCa}[\alpha]$	$G_h^{-1}(\phi\{z_0 + [z_0 + z^{(\alpha)}]/[1-a(z_0 + z^{(\alpha)}]\})$	There exists a monotone transformation such that $\phi_h = g(\theta_h)$ where $\phi = g(\theta)$ and $\phi_h \approx N(\phi - z_0\tau_0, \tau_0^2)$ where $\tau_0 = 1 + a\phi$ and z_0 and a are constant.

More About Confidence regions

- Results on previous slides depend on the following assumptions

  (1) asymptotic results apply

  (2) distributions are not heavy-tailed or highly skewed
- For small to moderate sample sizes these properties may not apply as (1) and (2) may not be satisfied

More About Confidence regions (continued)

- Example: Estimating the variance of a population
- Chernick and LaBudde (2010) have shown for heavy-tailed distributions with finite second moments that BCa is not always the most accurate bootstrap estimate unless the sample size is very large
- This also has implications on the comparison of two variances by the bootstrap

Results from Chernick and LaBudde (2010)

- Can we expect small sample behavior of the bootstrap to be similar to large sample behavior?
- For estimating variance Chernick and LaBudde compared various population distributions and bootstrap confidence interval methods to see how they perform in terms of coverage.

Results from Chernick and LaBudde (2010)

The key parameters of the simulations are:
nSize: The sample size of the originating data which is to be bootstrapped.
nReal: The number of bootstrap resamples used to estimate the bootstrap resampling distribution (The number of possible unique resamples is always no more than $nSize^{nSize}$).
nRepl: The number of Monte Carlo replications of the entire experiment, based on generating new samples of size nSize from the underlying assumed distribution, in order to estimate coverage accuracy and other errors of the bootstrap methodology.

Results from Chernick and LaBudde (2010)

- In the article we reported on the following bootstrap methods:
- Normal-t: A parametric Student-t confidence interval, with center point the sample variance and the standard error of variance estimated from that of the resampling distribution. This differs from a parametric normal bootstrap in that the percentile of the t distribution with *n*-1 degrees of freedom is used instead of the standard normal percentile. In Hesterberg's bootstrap chapter (Hesterberg et al. [2003]) it is referred to as the bootstrap-t but that confuses it with the bootstrap percentile t presented earlier in the chapter.
- EP: Efron percentile interval, with endpoints the plug-in quantiles of the resampling distribution.

Results from Chernick and LaBudde (2010)

- BC: The Efron bias-corrected interval. Simulations have shown that the BC method is, as expected, virtually identical in estimates to the BCa interval with the acceleration a = 0 (i.e., adjustment for median bias only).
- BCa: The Efron bias-corrected-and-accelerated interval, with median bias correction and skew correction via a jackknife estimate of the (biased) coefficient of skewness from the original sample.
- ABC: The Efron-DiCiccio approximate bootstrap confidence interval.

Results from Chernick and LaBudde (2010)

- Distributions Simulated
  - Gamma(2,3)
  - Uniform(0,1)
  - Student’s t with 5 degrees of freedom
  - Normal(0,1)
  - Lognormal(0,1)

Uniform (0, 1) Distribution: Results for various confidence intervals:

(1) Results: nSize=25, nReal=1000.

Confidence Level	Normal-t	EP
50%	49.5%	49%
60%	60.4%	57.3%
70%	67.2%	68%
80%	78.4%	77.9%
90%	86.7%	86.8%
95%	92.7%	92%
99%	97.4%	96.5%

Uniform (0, 1) Distribution: Results for 90% confidence intervals:

(2) Results: Sample Size (nSize), Bootstrap samples (nReal) and the number of Monte Carlo samples generated (nRepl) are varied and the Asymptotic Confidence Level is 90%

Sample Size (nSize)	nRepl	nReal	Normal-t	EP	ABC	BCa
10	64,000	16,000	86.42%	84.31%	81.65%	83.35%
20	64,000	16,000	88.89%	88.11%	88.35%	88.28%
25	64,000	16,000	89.21%	88.66%	88.15%	87.95%
30	64,000	16,000	89.41%	88.98%	88.98%	88.53%
40	64,000	16,000	89.69%	89.36%	88.30%	88.58%
50	64,000	16,000	90.17%	89.86%	89.95%	90.40%
100	64,000	16,000	90.11%	89.97%	Not done	Not done

ln[Normal(0,1)] Distribution:

Results for 60% coverage.

nSize	nRepl	nReal	Normal-t	EP	BC	ABC	BCa
10	64,000	16,000	22.68%	28.61%	26.52%	20.84%	25.78%
25	64,000	16,000	31.45%	35.87%	35.32%	30.08%	35.95%
50	16,000	16,000	37.02%	40.28%	39.38%	35.59%	40.24%
100	16,000	16,000	41.51%	43.76%	43.19%	40.13%	43.55%
250	16,000	16,000	45.21%	46.80%	46.42%	44.42%	46.68%
1000	16,000	16,000	50.74%	51.59%	48.98%	49.94%	49.28%
2000	16,000	16,000	52.85%	53.64%	52.24%	*	52.13%

ln[Normal(0,1)] Distribution:

Results for 80% coverage.

nSize	nRepl	nReal	Normal-t	EP	BC	ABC	BCa
10	64,000	16,000	35.04%	33.76%	35.01%	30.51%	36.74%
25	64,000	16,000	44.84%	43.74%	46.70%	43.90%	48.71%
50	16,000	16,000	51.14%	50.51%	53.11%	51.49%	55.34%
100	16,000	16,000	56.48%	56.19%	58.61%	57.38%	60.54%
250	16,000	16,000	62.26%	62.14%	63.29%	63.06%	64.81%
1000	16,000	16,000	69.31%	69.03%	69.35%	*	69.80%
2000	16,000	16,000	71.80%	71.40%	71.28%	*	71.58%

ln[Normal(0,1)] Distribution:

Results for 90% coverage.

nSize	nRepl	nReal	Normal-t	EP	BC	ABC	BCa
10	64,000	16,000	39.98%	37.12%	39.38%	37.11%	41.03%
25	64,000	16,000	50.32%	50.11%	52.52%	53.03%	56.13%
50	16,000	16,000	56.93%	57.39%	60.43%	62.04%	64.63%
100	16,000	16,000	62.93%	63.93%	66.71%	68.50%	70.27%
250	16,000	16,000	69.35%	70.56%	72.74%	74.41%	75.33%
1000	16,000	16,000	77.63%	78.40%	79.59%	*	80.81%
2000	16,000	16,000	80.38%	80.85%	81.83%	*	82.36%

ln[Normal(0,1)] Distribution:

Results for 99% coverage.

nSize	nRepl	nReal	Normal-t	EP	BC	ABC	BCa
10	64,000	16,000	49.51%	42.84%	44.92%	34.14%	41.03%
25	64,000	16,000	60.05%	59.00%	61.68%	65.43%	67.29%
50	16,000	16,000	67.64%	68.90%	71.58%	77.44%	78.06%
100	16,000	16,000	74.11%	76.45%	78.99%	84.69%	84.82%
250	16,000	16,000	80.93%	83.71%	85.47%	90.16%	89.82%
1000	16,000	16,000	88.49%	90.74%	91.77%	*	94.20%
2000	16,000	16,000	91.31%	93.13%	93.83%	*	95.49%

Hypothesis tests

- Since there is a 1-1 correspondence between hypothesis tests and confidence intervals, a hypothesis test about a parameter θ can be constructed based on a bootstrap confidence interval for θ.
- See Chernick (1999 or 2007) Section 3.2.
- Examples of hypothesis tests can be found in Section 3.3 of Chernick (1999 or 2007).
- Advice on which method to use is also given in Carpenter and Bithell (2000). But for cases like those studied by Chernick and LaBudde (2008) <u>do not</u> follow this advice.

References on confidence intervals and hypothesis tests

(1) Chernick, M.R. (1999). *Bootstrap Methods: A Practitioner's Guide.* Wiley, New York.

(2) Chernick, M.R. (2007). *Bootstrap Methods: A Guide for Practitioners and Researchers, 2nd Edition.* Wiley, New York.

(3) Hall, P. (1992). *The Bootstrap and Edgeworth Expansion.* Springer-Verlag, New York.

(4) Efron, B. (1982) *The Jackknife, the Bootstrap and Other Resampling Plans.* Society for Industrial and Applied Mathematics CBMS-NSF Regional Conference Series **38,** Philadelphia.

References on confidence intervals and hypothesis tests (continued)

(5) Carpenter, J. and Bithell, J. (2000). Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. *Statistics in Medicine* **19,** 1141-1164.

(6) Bahadur, R.R. and Savage, L.J. (1956). The nonexistence of certain statistical procedures in nonparametric problems. *Annals of Mathematical Statistics* **27,** 1115-1122.

(7) Ewens, W.J. and Grant, G.R. (2001). *Statistical Methods in Bioinformatics An Introduction* Springer-Verlag, New York.

(8) Chernick, M. R. and LaBudde, R. (2010). More Qualms About Bootstrap Confidence Intervals. *Am. J. Math. Manag. Sci.* **29,** 437-456.

End of Day 1