



Contents

[move to sidebar](#)[hide](#)

- (Top)
- 1History
- 2Approach
- 3Discussion
 - 3.1Advantages
 - 3.2Disadvantages
 - 3.3Recommendations
- 4Types of bootstrap scheme
 - 4.1Case resampling
 - 4.1.1Estimating the distribution of sample mean
 - 4.1.2Regression
 - 4.2Bayesian bootstrap
 - 4.3Smooth bootstrap
 - 4.4Parametric bootstrap
 - 4.5Resampling residuals
 - 4.6Gaussian process regression bootstrap
 - 4.7Wild bootstrap
 - 4.8Block bootstrap
 - 4.8.1Time series: Simple block bootstrap
 - 4.8.2Time series: Moving block bootstrap
 - 4.8.3Time series: Maximum entropy bootstrap
 - 4.8.4Cluster data: block bootstrap
- 5Methods for improving computational efficiency
 - 5.1Poisson bootstrap

- 5.2 Bag of Little Bootstraps
- 6 Choice of statistic
- 7 Deriving confidence intervals from the bootstrap distribution
 - 7.1 Bias, asymmetry, and confidence intervals
 - 7.2 Methods for bootstrap confidence intervals
- 8 Bootstrap hypothesis testing
- 9 Example applications
 - 9.1 Smoothed bootstrap
- 10 Relation to other approaches to inference
 - 10.1 Relationship to other resampling methods
 - 10.2 U-statistics
- 11 See also
- 12 References
- 13 Further reading
- 14 External links
 - 14.1 Software

Bootstrapping (statistics)

Bootstrapping is any test or metric that uses random sampling with replacement (e.g. mimicking the sampling process), and falls under the broader class of resampling methods. Bootstrapping assigns measures of accuracy (bias, variance, confidence intervals, prediction error, etc.) to sample estimates.^{[1][2]} This technique allows estimation of the sampling distribution of almost any statistic using random sampling methods.^{[3][4]}

Bootstrapping estimates the properties of an estimand (such as its variance) by measuring those properties when sampling from an approximating distribution. One standard choice for an approximating distribution is the empirical distribution function of the observed data. In the case where a set of observations can be assumed to be from an independent and identically distributed population, this can be implemented by constructing a number of resamples with replacement, of the observed data set (and of equal size to the observed data set).

It may also be used for constructing hypothesis tests. It is often used as an alternative to statistical inference based on the assumption of a parametric model when that assumption is in doubt, or where parametric inference is impossible or requires complicated formulas for the calculation of standard errors.

History

The bootstrap was published by Bradley Efron in "Bootstrap methods: another look at the jackknife" (1979),^{[5][6][7]} inspired by earlier work on the jackknife.^{[8][9][10]} Improved estimates of the variance were developed later.^{[11][12]} A Bayesian extension was developed in 1981.^[13] The bias-corrected and accelerated (BCa) bootstrap was developed by Efron in 1987,^[14] and the ABC procedure in 1992.^[15]

Approach

The basic idea of bootstrapping is that inference about a population from sample data (sample → population) can be modeled by *resampling* the sample data and performing inference about a sample from resampled data (resampled → sample). As the population is unknown, the true error in a sample statistic against its population value is unknown. In bootstrap-resamples, the 'population' is in fact the sample, and this is known; hence the quality of inference of the 'true' sample from resampled data (resampled → sample) is measurable.

More formally, the bootstrap works by treating inference of the true probability distribution J , given the original data, as being analogous to an inference of the empirical distribution \hat{J} , given the resampled data. The accuracy of inferences regarding \hat{J} using the resampled data can be assessed because we know \hat{J} . If \hat{J} is a reasonable approximation to J , then the quality of inference on J can in turn be inferred.

As an example, assume we are interested in the average (or mean) height of people worldwide. We cannot measure all the people in the global population, so instead, we sample only a tiny part of it, and measure that. Assume the sample is of size N ; that is, we measure the heights of N individuals. From that single sample, only one estimate of the mean can be obtained. In order to reason about the population, we need some sense of the variability of the mean that we have computed. The simplest bootstrap method involves taking the original data set of heights, and, using a computer, sampling from it to form a new sample (called a 'resample' or bootstrap sample) that is also of size N . The bootstrap sample is taken from the original by using sampling with replacement (e.g. we might 'resample' 5 times from [1,2,3,4,5] and get [2,5,4,4,1]), so, assuming N is sufficiently large, for all practical purposes there is virtually zero probability that it will be identical to the original "real" sample. This process is repeated a large number of times (typically 1,000 or 10,000 times), and for each of these bootstrap samples, we compute its mean (each of these is called a "bootstrap estimate"). We now can create a histogram of bootstrap means. This histogram provides an estimate of the shape of the distribution of the sample mean from which we can answer questions about how much the mean varies across samples. (The method here, described for the mean, can be applied to almost any other statistic or estimator.)

Discussion

Advantages

A great advantage of bootstrap is its simplicity. It is a straightforward way to derive estimates of standard errors and confidence intervals for complex estimators of the distribution, such as percentile points, proportions, odds ratio, and correlation coefficients. However, despite its simplicity, bootstrapping can be applied to complex sampling designs (e.g. for population divided into s strata with n_s observations per strata, bootstrapping can be applied for each strata).^[16] Bootstrap is also an appropriate way to control and check the stability of the results. Although for most problems it is impossible to know the true confidence interval, bootstrap is asymptotically more accurate than the standard intervals obtained using sample variance and assumptions of normality.^[17] Bootstrapping is also a convenient method that avoids the cost of repeating the experiment to get other groups of sample data.

Disadvantages

Bootstrapping depends heavily on the estimator used and, though simple, ignorant use of bootstrapping will not always yield asymptotically valid results and can lead to inconsistency.^[18] Although bootstrapping is (under some conditions) asymptotically consistent, it does not provide general finite-sample guarantees. The result may depend on the representative sample. The apparent simplicity may conceal the fact that important assumptions are being made when undertaking the bootstrap analysis (e.g. independence of samples or large enough of a sample size) where these would be more formally stated in other approaches. Also, bootstrapping can be time-consuming and there are not many available software for bootstrapping as it is difficult to automate using traditional statistical computer packages.^[16]

Recommendations

Scholars have recommended more bootstrap samples as available computing power has increased. If the results may have substantial real-world consequences, then one should use as many samples as is reasonable, given available computing power and time. Increasing the number of samples cannot increase the amount of information in the original data; it can only reduce the effects of random sampling errors which can arise from a bootstrap procedure itself. Moreover, there is evidence that numbers of samples greater than 100 lead to negligible improvements in the estimation of standard errors.^[19] In fact, according to the original developer of the bootstrapping method, even setting the number of samples at 50 is likely to lead to fairly good standard error estimates.^[20]

Ad r et al. recommend the bootstrap procedure for the following situations:^[21]

- When the theoretical distribution of a statistic of interest is complicated or unknown. Since the bootstrapping procedure is distribution-independent it provides an indirect method to assess the properties of the distribution underlying the sample and the parameters of interest that are derived from this distribution.
- When the sample size is insufficient for straightforward statistical inference. If the underlying distribution is well-known, bootstrapping provides a way to account for the distortions caused by the specific sample that may not be fully representative of the population.
- When power calculations have to be performed, and a small pilot sample is available. Most power and sample size calculations are heavily dependent on the standard deviation of the statistic of interest. If the estimate used is incorrect, the required sample size will also be

wrong. One method to get an impression of the variation of the statistic is to use a small pilot sample and perform bootstrapping on it to get impression of the variance.

However, Athreya has shown^[22] that if one performs a naive bootstrap on the sample mean when the underlying population lacks a finite variance (for example, a power law distribution), then the bootstrap distribution will not converge to the same limit as the sample mean. As a result, confidence intervals on the basis of a Monte Carlo simulation of the bootstrap could be misleading. Athreya states that "Unless one is reasonably sure that the underlying distribution is not heavy tailed, one should hesitate to use the naive bootstrap".

Types of bootstrap scheme

In univariate problems, it is usually acceptable to resample the individual observations with replacement ("case resampling" below) unlike subsampling, in which resampling is without replacement and is valid under much weaker conditions compared to the bootstrap. In small samples, a parametric bootstrap approach might be preferred. For other problems, a *smooth bootstrap* will likely be preferred.

For regression problems, various other alternatives are available.^[1]

Case resampling

The bootstrap is generally useful for estimating the distribution of a statistic (e.g. mean, variance) without using normality assumptions (as required, e.g., for a z-statistic or a t-statistic). In particular, the bootstrap is useful when there is no analytical form or an asymptotic theory (e.g., an applicable central limit theorem) to help estimate the distribution of the statistics of interest. This is because bootstrap methods can apply to most random quantities, e.g., the ratio of variance and mean. There are at least two ways of performing case resampling.

1. The Monte Carlo algorithm for case resampling is quite simple. First, we resample the data with replacement, and the size of the resample must be equal to the size of the original data set. Then the statistic of interest is computed from the resample from the first step. We repeat this routine many times to get a more precise estimate of the Bootstrap distribution of the statistic.^[1]

2. The 'exact' version for case resampling is similar, but we exhaustively enumerate every possible resample of the data set. This can be computationally expensive as there are a total of

$$\binom{2n-1}{n} = \frac{(2n-1)!}{n!(n-1)!} \text{ different resamples, where } n \text{ is the size of the data set. Thus for}$$

$n = 5, 10, 20, 30$ there are 126, 92378, 6.89×10^{10} and 5.91×10^{16} different resamples respectively.^[23]

Estimating the distribution of sample mean

Consider a coin-flipping experiment. We flip the coin and record whether it lands heads or tails. Let $X = x_1, x_2, \dots, x_{10}$ be 10 observations from the experiment. $x_i = 1$ if the i th flip lands heads, and 0 otherwise. By invoking the assumption that the average of the coin flips is normally distributed, we can use the t-statistic to estimate the distribution of the sample mean,

$$\bar{x} = \frac{1}{10}(x_1 + x_2 + \cdots + x_{10}).$$

Such a normality assumption can be justified either as an approximation of the distribution of each *individual* coin flip or as an approximation of the distribution of the *average* of a large number of coin flips. The former is a poor approximation because the true distribution of the coin flips is Bernoulli instead of normal. The latter is a valid approximation in *infinitely large* samples due to the central limit theorem.

However, if we are not ready to make such a justification, then we can use the bootstrap instead. Using case resampling, we can derive the distribution of \bar{x} . We first resample the data to obtain a *bootstrap resample*. An example of the first resample might look like this $X_1^* = x_2, x_1, x_{10}, x_{10}, x_3, x_4, x_6, x_7, x_1, x_9$. There are some duplicates since a bootstrap resample comes from sampling with replacement from the data. Also the number of data points in a bootstrap resample is equal to the number of data points in our original observations. Then we compute the mean of this resample and obtain the first *bootstrap mean*: μ_1^* . We repeat this process to obtain the second resample X_2^* and compute the second bootstrap mean μ_2^* . If we repeat this 100 times, then we have $\mu_1^*, \mu_2^*, \dots, \mu_{100}^*$. This represents an *empirical bootstrap distribution* of sample mean. From this empirical distribution, one can derive a *bootstrap confidence interval* for the purpose of hypothesis testing.

Regression

In regression problems, *case resampling* refers to the simple scheme of resampling individual cases – often rows of a data set. For regression problems, as long as the data set is fairly large, this simple scheme is often acceptable.^{[24][25][26]} However, the method is open to criticism.^[16]

In regression problems, the explanatory variables are often fixed, or at least observed with more control than the response variable. Also, the range of the explanatory variables defines the information available from them. Therefore, to resample cases means that each bootstrap sample will lose some information. As such, alternative bootstrap procedures should be considered.

Bayesian bootstrap

Bootstrapping can be interpreted in a Bayesian framework using a scheme that creates new data sets through reweighting the initial data. Given a set of N data points, the weighting assigned to data point i in a new data set \mathcal{D}^J is $w_i^J = x_i^J - x_{i-1}^J$, where \mathbf{x}^J is a low-to-high ordered list of $N - 1$ uniformly distributed random numbers on $[0, 1]$, preceded by 0 and succeeded by 1. The distributions of a parameter inferred from considering many such data sets \mathcal{D}^J are then interpretable as posterior distributions on that parameter.^[27]

Smooth bootstrap

Under this scheme, a small amount of (usually normally distributed) zero-centered random noise is added onto each resampled observation. This is equivalent to sampling from a kernel density estimate of the data. Assume K to be a symmetric kernel density function with unit variance. The standard kernel estimator $\hat{f}_h(\mathbf{x})$ of $f(\mathbf{x})$ is

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right), \quad [28]$$

where h is the smoothing parameter. And the corresponding distribution function estimator $\hat{F}_h(x)$ is

$$\hat{F}_h(x) = \int_{-\infty}^x \hat{f}_h(t) dt. \quad [28]$$

Parametric bootstrap

Based on the assumption that the original data set is a realization of a random sample from a distribution of a specific parametric type, in this case a parametric model is fitted by parameter θ , often by maximum likelihood, and samples of random numbers are drawn from this fitted model. Usually the sample drawn has the same sample size as the original data. Then the estimate of original function F can be written as $\hat{F} = F_{\hat{\theta}}$. This sampling process is repeated many times as for other bootstrap methods. Considering the centered sample mean in this case, the random sample original distribution function F_{θ} is replaced by a bootstrap random sample with function $F_{\hat{\theta}}$, and the probability distribution of $\bar{X}_n - \mu_{\theta}$ is approximated by that of $\bar{X}_n^* - \mu^*$, where $\mu^* = \mu_{\hat{\theta}}$, which is the expectation corresponding to $F_{\hat{\theta}}$.^[29] The use of a parametric model at the sampling stage of the bootstrap methodology leads to procedures which are different from those obtained by applying basic statistical theory to inference for the same model.

Resampling residuals

Another approach to bootstrapping in regression problems is to resample residuals. The method proceeds as follows.

1. Fit the model and retain the fitted values \hat{y}_i and the residuals $\hat{\varepsilon}_i = y_i - \hat{y}_i, (i = 1, \dots, n)$.
2. For each pair, (x_i, y_i) , in which x_i is the (possibly multivariate) explanatory variable, add a randomly resampled residual, $\hat{\varepsilon}_j$, to the fitted value \hat{y}_i . In other words, create synthetic response variables $y_i^* = \hat{y}_i + \hat{\varepsilon}_j$ where j is selected randomly from the list $(1, \dots, n)$ for every i .
3. Refit the model using the fictitious response variables y_i^* , and retain the quantities of interest (often the parameters, $\hat{\mu}_i^*$, estimated from the synthetic y_i^*).
4. Repeat steps 2 and 3 a large number of times.

This scheme has the advantage that it retains the information in the explanatory variables. However, a question arises as to which residuals to resample. Raw residuals are one option; another is studentized residuals (in linear regression). Although there are arguments in favor of using studentized residuals; in practice, it often makes little difference, and it is easy to compare the results of both schemes.

Gaussian process regression bootstrap

When data are temporally correlated, straightforward bootstrapping destroys the inherent correlations. This method uses Gaussian process regression (GPR) to fit a probabilistic model from which replicates may then be drawn. GPR is a Bayesian non-linear regression method. A Gaussian process (GP) is a collection of random variables, any finite number of which have a joint Gaussian (normal) distribution. A GP is defined by a mean function and a covariance function, which specify the mean vectors and covariance matrices for each finite collection of the random variables.^[30]

Regression model:

$$y(x) = f(x) + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2), \quad \varepsilon \text{ is a noise term.}$$

Gaussian process prior:

For any finite collection of variables, x_1, \dots, x_n , the function outputs $f(x_1), \dots, f(x_n)$ are jointly distributed according to a multivariate Gaussian with mean $\mathbf{m} = [\mathbf{m}(x_1), \dots, \mathbf{m}(x_n)]^\top$ and covariance matrix $(K)_{ij} = k(x_i, x_j)$.

Assume $f(x) \sim \mathcal{GP}(\mathbf{m}, k)$. Then $y(x) \sim \mathcal{GP}(\mathbf{m}, l)$,

where $l(x_i, x_j) = k(x_i, x_j) + \sigma^2 \delta(x_i, x_j)$, and $\delta(x_i, x_j)$ is the standard Kronecker delta function.^[30]

Gaussian process posterior:

According to GP prior, we can get

$$[y(x_1), \dots, y(x_r)] \sim \mathcal{N}(\mathbf{m}_0, K_0),$$

where $\mathbf{m}_0 = [\mathbf{m}(x_1), \dots, \mathbf{m}(x_r)]^\top$ and $(K_0)_{ij} = k(x_i, x_j) + \sigma^2 \delta(x_i, x_j)$.

Let x_1^*, \dots, x_s^* be another finite collection of variables, it's obvious that

$$[y(x_1), \dots, y(x_r), f(x_1^*), \dots, f(x_s^*)]^\top \sim \mathcal{N}\left(\begin{pmatrix} \mathbf{m}_0 \\ \mathbf{m}_* \end{pmatrix}, \begin{pmatrix} K_0 & K_* \\ K_*^\top & K_{**} \end{pmatrix}\right),$$

where $\mathbf{m}_* = [\mathbf{m}(x_1^*), \dots, \mathbf{m}(x_s^*)]^\top$, $(K_{**})_{ij} = k(x_i^*, x_j^*)$, $(K_*)_{ij} = k(x_i, x_j^*)$.

According to the equations above, the outputs y are also jointly distributed according to a multivariate Gaussian. Thus,

$$[f(x_1^*), \dots, f(x_s^*)]^\top \mid ([y(x)]^\top = y) \sim \mathcal{N}(\mathbf{m}_{\text{post}}, K_{\text{post}}),$$

where $y = [y_1, \dots, y_r]^\top$, $\mathbf{m}_{\text{post}} = \mathbf{m}_* + K_*^\top (K_0 + \sigma^2 I_r)^{-1} (y - \mathbf{m}_0)$, $K_{\text{post}} = K_{**} - K_*^\top (K_0 + \sigma^2 I_r)^{-1} K_*$, and I_r is $r \times r$ identity matrix.^[30]

Wild bootstrap

The wild bootstrap, proposed originally by Wu (1986),^[31] is suited when the model exhibits heteroskedasticity. The idea is, as the residual bootstrap, to leave the regressors at their sample value, but to resample the response variable based on the residuals values. That is, for each replicate, one

computes a new \mathbf{y} based on

$$\mathbf{y}_i^* = \hat{\mathbf{y}}_i + \hat{\varepsilon}_i \mathbf{v}_i$$

so the residuals are randomly multiplied by a random variable \mathbf{v}_i with mean 0 and variance 1. For most distributions of \mathbf{v}_i (but not Mammen's), this method assumes that the 'true' residual distribution is symmetric and can offer advantages over simple residual sampling for smaller sample sizes. Different forms are used for the random variable \mathbf{v}_i , such as

- The standard normal distribution
- A distribution suggested by Mammen (1993).^[32]

$$\mathbf{v}_i = \begin{cases} -(\sqrt{5} - 1)/2 & \text{with probability } (\sqrt{5} + 1)/(2\sqrt{5}), \\ (\sqrt{5} + 1)/2 & \text{with probability } (\sqrt{5} - 1)/(2\sqrt{5}) \end{cases}$$

Approximately, Mammen's distribution is:

$$\mathbf{v}_i = \begin{cases} -0.6180 & \text{(with a 0 in the units' place) with probability 0.7236,} \\ +1.6180 & \text{(with a 1 in the units' place) with probability 0.2764.} \end{cases}$$

- Or the simpler distribution, linked to the Rademacher distribution:

$$\mathbf{v}_i = \begin{cases} -1 & \text{with probability } 1/2, \\ +1 & \text{with probability } 1/2. \end{cases}$$

Block bootstrap

The block bootstrap is used when the data, or the errors in a model, are correlated. In this case, a simple case or residual resampling will fail, as it is not able to replicate the correlation in the data. The block bootstrap tries to replicate the correlation by resampling inside blocks of data (see Blocking (statistics)). The block bootstrap has been used mainly with data correlated in time (i.e. time series) but can also be used with data correlated in space, or among groups (so-called cluster data).

Time series: Simple block bootstrap

In the (simple) block bootstrap, the variable of interest is split into non-overlapping blocks.

Time series: Moving block bootstrap

In the moving block bootstrap, introduced by Künsch (1989),^[33] data is split into $n - b + 1$ overlapping blocks of length b : Observation 1 to b will be block 1, observation 2 to $b + 1$ will be block 2, etc. Then from these $n - b + 1$ blocks, n/b blocks will be drawn at random with replacement. Then aligning these n/b blocks in the order they were picked, will give the bootstrap observations.

This bootstrap works with dependent data, however, the bootstrapped observations will not be stationary anymore by construction. But, it was shown that varying randomly the block length can avoid this problem.^[34] This method is known as the *stationary bootstrap*. Other related modifications

of the moving block bootstrap are the *Markovian bootstrap* and a stationary bootstrap method that matches subsequent blocks based on standard deviation matching.

Time series: Maximum entropy bootstrap

Vinod (2006),^[35] presents a method that bootstraps time series data using maximum entropy principles satisfying the Ergodic theorem with mean-preserving and mass-preserving constraints. There is an R package, **meboot**,^[36] that utilizes the method, which has applications in econometrics and computer science.

Cluster data: block bootstrap

Cluster data describes data where many observations per unit are observed. This could be observing many firms in many states or observing students in many classes. In such cases, the correlation structure is simplified, and one does usually make the assumption that data is correlated within a group/cluster, but independent between groups/clusters. The structure of the block bootstrap is easily obtained (where the block just corresponds to the group), and usually only the groups are resampled, while the observations within the groups are left unchanged. Cameron et al. (2008) discusses this for clustered errors in linear regression.^[37]

Methods for improving computational efficiency

The bootstrap is a powerful technique although may require substantial computing resources in both time and memory. Some techniques have been developed to reduce this burden. They can generally be combined with many of the different types of Bootstrap schemes and various choices of statistics.

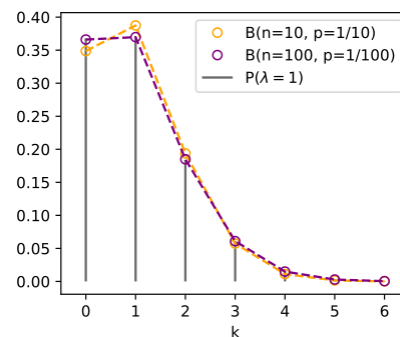
Poisson bootstrap

The ordinary bootstrap requires the random selection of n elements from a list, which is equivalent to drawing from a multinomial distribution. This may require a large number of passes over the data and is challenging to run these computations in parallel. For large values of n , the Poisson bootstrap is an efficient method of generating bootstrapped data sets.^[38] When generating a single bootstrap sample, instead of randomly drawing from the sample data with replacement, each data point is assigned a random weight distributed according to the Poisson distribution with $\lambda = 1$. For large sample data, this will approximate random sampling with replacement. This is due to the following approximation:

$$\lim_{n \rightarrow \infty} \text{Binomial}(n, 1/n) = \text{Poisson}(1)$$

This method also lends itself well to streaming data and growing data sets, since the total number of samples does not need to be known in advance of beginning to take bootstrap samples.

For large enough n , the results are relatively similar to the original bootstrap estimations.^[39]



Graph showing the convergence of the Binomial Distribution to the Poisson as the Binomial parameters are $n \cdot p = 1$ and n grows

A way to improve on the poisson bootstrap, termed "sequential bootstrap", is by taking the first samples so that the proportion of unique values is ≈ 0.632 of the original sample size n . This provides a distribution with main empirical characteristics being within a distance of $O(n^{3/4})$.^[40] Empirical investigation has shown this method can yield good results.^[41] This is related to the reduced bootstrap method.^[42]

Bag of Little Bootstraps

For massive data sets, it is often computationally prohibitive to hold all the sample data in memory and resample from the sample data. The Bag of Little Bootstraps (BLB)^[43] provides a method of pre-aggregating data before bootstrapping to reduce computational constraints. This works by partitioning the data set into b equal-sized buckets and aggregating the data within each bucket. This pre-aggregated data set becomes the new sample data over which to draw samples with replacement. This method is similar to the Block Bootstrap, but the motivations and definitions of the blocks are very different. Under certain assumptions, the sample distribution should approximate the full bootstrapped scenario. One constraint is the number of buckets $b = n^\gamma$ where $\gamma \in [0.5, 1]$ and the authors recommend usage of $b = n^{0.7}$ as a general solution.

Choice of statistic

The bootstrap distribution of a point estimator of a population parameter has been used to produce a bootstrapped confidence interval for the parameter's true value if the parameter can be written as a function of the population's distribution.

Population parameters are estimated with many point estimators. Popular families of point-estimators include mean-unbiased minimum-variance estimators, median-unbiased estimators, Bayesian estimators (for example, the posterior distribution's mode, median, mean), and maximum-likelihood estimators.

A Bayesian point estimator and a maximum-likelihood estimator have good performance when the sample size is infinite, according to asymptotic theory. For practical problems with finite samples, other estimators may be preferable. Asymptotic theory suggests techniques that often improve the performance of bootstrapped estimators; the bootstrapping of a maximum-likelihood estimator may often be improved using transformations related to pivotal quantities.^[44]

Deriving confidence intervals from the bootstrap distribution

The bootstrap distribution of a parameter-estimator has been used to calculate confidence intervals for its population-parameter.^[1]

Bias, asymmetry, and confidence intervals

- **Bias:** The bootstrap distribution and the sample may disagree systematically, in which case bias may occur.

If the bootstrap distribution of an estimator is symmetric, then percentile confidence-interval are often used; such intervals are appropriate especially for median-unbiased estimators of minimum risk (with respect to an absolute loss function). Bias in the bootstrap distribution will lead to bias in the confidence interval.

Otherwise, if the bootstrap distribution is non-symmetric, then percentile confidence intervals are often inappropriate.

Methods for bootstrap confidence intervals

There are several methods for constructing confidence intervals from the bootstrap distribution of a real parameter:

- **Basic bootstrap**,^[44] also known as the **Reverse Percentile Interval**.^[45] The basic bootstrap is a simple scheme to construct the confidence interval: one simply takes the empirical quantiles from the bootstrap distribution of the parameter (see Davison and Hinkley 1997, equ. 5.6 p. 194):

$(2\hat{\theta} - \theta^*_{(1-\alpha/2)}, 2\hat{\theta} - \theta^*_{(\alpha/2)})$ where $\theta^*_{(1-\alpha/2)}$ denotes the $1 - \alpha/2$ percentile of the bootstrapped coefficients θ^* .

- **Percentile bootstrap**. The percentile bootstrap proceeds in a similar way to the basic bootstrap, using percentiles of the bootstrap distribution, but with a different formula (note the inversion of the left and right quantiles):

$(\theta^*_{(\alpha/2)}, \theta^*_{(1-\alpha/2)})$ where $\theta^*_{(1-\alpha/2)}$ denotes the $1 - \alpha/2$ percentile of the bootstrapped coefficients θ^* .

See Davison and Hinkley (1997, equ. 5.18 p. 203) and Efron and Tibshirani (1993, equ 13.5 p. 171).

This method can be applied to any statistic. It will work well in cases where the bootstrap distribution is symmetrical and centered on the observed statistic^[46] and where the sample statistic is median-unbiased and has maximum concentration (or minimum risk with respect to an absolute value loss function). When working with small sample sizes (i.e., less than 50), the basic / reversed percentile and percentile confidence intervals for (for example) the variance statistic will be too narrow. So that with a sample of 20 points, 90% confidence interval will include the true variance only 78% of the time.^[47] The basic / reverse percentile confidence intervals are easier to justify mathematically^{[48][45]} but they are less accurate in general than percentile confidence intervals, and some authors discourage their use.^[45]

- **Studentized bootstrap**. The studentized bootstrap, also called *bootstrap-t*, is computed analogously to the standard confidence interval, but replaces the quantiles from the normal or student approximation by the quantiles from the bootstrap distribution of the Student's t-test (see Davison and Hinkley 1997, equ. 5.7 p. 194 and Efron and Tibshirani 1993 equ 12.22, p. 160):

$(\hat{\theta} - t^*_{(1-\alpha/2)} \cdot \widehat{se}_{\theta}, \hat{\theta} - t^*_{(\alpha/2)} \cdot \widehat{se}_{\theta})$ where $t^*_{(1-\alpha/2)}$ denotes the $1 - \alpha/2$ percentile of the bootstrapped Student's t-test $t^* = (\hat{\theta}^* - \hat{\theta}) / \widehat{se}_{\hat{\theta}^*}$, and \widehat{se}_{θ} is the estimated standard error of the coefficient in the original model.

The studentized test enjoys optimal properties as the statistic that is bootstrapped is pivotal (i.e. it does not depend on nuisance parameters as the t-test follows asymptotically a $N(0,1)$

distribution), unlike the percentile bootstrap.

- **Bias-corrected bootstrap** – adjusts for bias in the bootstrap distribution.
- **Accelerated bootstrap** – The bias-corrected and accelerated (BCa) bootstrap, by Efron (1987),^[14] adjusts for both bias and skewness in the bootstrap distribution. This approach is accurate in a wide variety of settings, has reasonable computation requirements, and produces reasonably narrow intervals.^[14]

Bootstrap hypothesis testing

Efron and Tibshirani^[1] suggest the following algorithm for comparing the means of two independent samples: Let $\mathbf{x}_1, \dots, \mathbf{x}_n$ be a random sample from distribution F with sample mean $\bar{\mathbf{x}}$ and sample variance $\sigma_{\mathbf{x}}^2$. Let $\mathbf{y}_1, \dots, \mathbf{y}_m$ be another, independent random sample from distribution G with mean $\bar{\mathbf{y}}$ and variance $\sigma_{\mathbf{y}}^2$

1. Calculate the test statistic $t = \frac{\bar{\mathbf{x}} - \bar{\mathbf{y}}}{\sqrt{\sigma_{\mathbf{x}}^2/n + \sigma_{\mathbf{y}}^2/m}}$
2. Create two new data sets whose values are $\mathbf{x}'_i = \mathbf{x}_i - \bar{\mathbf{x}} + \bar{\mathbf{z}}$ and $\mathbf{y}'_i = \mathbf{y}_i - \bar{\mathbf{y}} + \bar{\mathbf{z}}$, where $\bar{\mathbf{z}}$ is the mean of the combined sample.
3. Draw a random sample (\mathbf{x}^*_i) of size n with replacement from \mathbf{x}'_i and another random sample (\mathbf{y}^*_i) of size m with replacement from \mathbf{y}'_i .
4. Calculate the test statistic $t^* = \frac{\bar{\mathbf{x}}^* - \bar{\mathbf{y}}^*}{\sqrt{\sigma_{\mathbf{x}}^{*2}/n + \sigma_{\mathbf{y}}^{*2}/m}}$
5. Repeat 3 and 4 B times (e.g. $B = 1000$) to collect B values of the test statistic.
6. Estimate the p-value as $p = \frac{\sum_{i=1}^B I\{t^*_i \geq t\}}{B}$ where $I(\text{condition}) = 1$ when *condition* is true and 0 otherwise.

Example applications

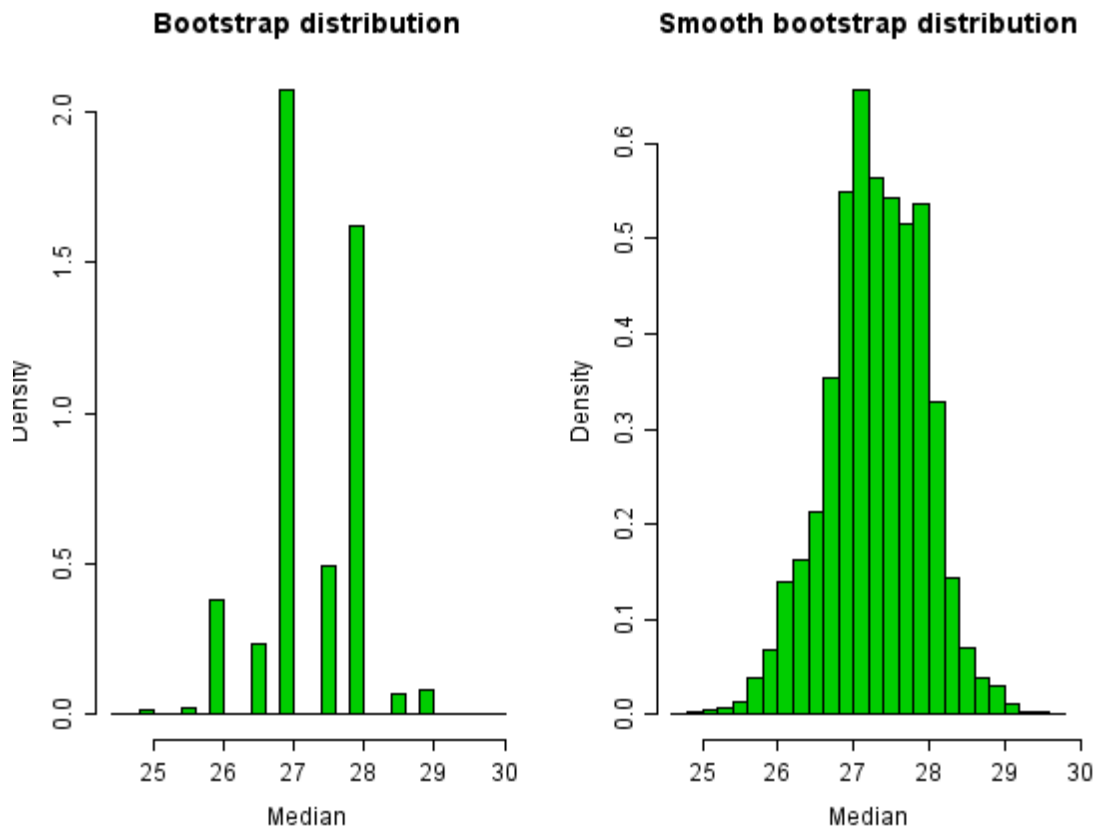
Smoothed bootstrap

In 1878, Simon Newcomb took observations on the speed of light.^[49] The data set contains two outliers, which greatly influence the sample mean. (The sample mean need not be a consistent estimator for any population mean, because no mean needs to exist for a heavy-tailed distribution.) A well-defined and robust statistic for the central tendency is the sample median, which is consistent and median-unbiased for the population median.

The bootstrap distribution for Newcomb's data appears below. We can reduce the discreteness of the bootstrap distribution by adding a small amount of random noise to each bootstrap sample. A conventional choice is to add noise with a standard deviation of σ/\sqrt{n} for a sample size n ; this noise is often drawn from a Student-t distribution with $n-1$ degrees of freedom.^[50] This results in an

approximately-unbiased estimator for the variance of the sample mean. This means that samples taken from the bootstrap distribution will have a variance which is, on average, equal to the variance of the total population.

Histograms of the bootstrap distribution and the smooth bootstrap distribution appear below. The bootstrap distribution of the sample-median has only a small number of values. The smoothed bootstrap distribution has a richer support. However, note that whether the smoothed or standard bootstrap procedure is favorable is case-by-case and is shown to depend on both the underlying distribution function and on the quantity being estimated.^[51]



In this example, the bootstrapped 95% (percentile) confidence-interval for the population median is (26, 28.5), which is close to the interval for (25.98, 28.46) for the smoothed bootstrap.

Relation to other approaches to inference

Relationship to other resampling methods

The bootstrap is distinguished from:

- the jackknife procedure, used to estimate biases of sample statistics and to estimate variances, and
- cross-validation, in which the parameters (e.g., regression weights, factor loadings) that are estimated in one subsample are applied to another subsample.

For more details see resampling.

Bootstrap aggregating (bagging) is a meta-algorithm based on averaging model predictions obtained from models trained on multiple bootstrap samples.

U-statistics

In situations where an obvious statistic can be devised to measure a required characteristic using only a small number, r , of data items, a corresponding statistic based on the entire sample can be formulated. Given an r -sample statistic, one can create an n -sample statistic by something similar to bootstrapping (taking the average of the statistic over all subsamples of size r). This procedure is known to have certain good properties and the result is a U-statistic. The sample mean and sample variance are of this form, for $r = 1$ and $r = 2$.

See also

- Accuracy and precision
- Bootstrap aggregating
- Bootstrapping
- Empirical likelihood
- Imputation (statistics)
- Reliability (statistics)
- Reproducibility
- Resampling

References

1. Efron, B.; Tibshirani, R. (1993). *An Introduction to the Bootstrap*. Boca Raton, FL: Chapman & Hall/CRC. ISBN 0-412-04231-2. software (<http://lib.stat.cmu.edu/S/bootstrap.funs>) Archived (<https://archive.is/20120712124533/http://lib.stat.cmu.edu/S/bootstrap.funs>) 2012-07-12 at archive.today
2. Second Thoughts on the Bootstrap – Bradley Efron, 2003
3. Varian, H.(2005). "Bootstrap Tutorial". *Mathematica Journal*, 9, 768–775.
4. Weisstein, Eric W. "Bootstrap Methods." From MathWorld--A Wolfram Web Resource. <http://mathworld.wolfram.com/BootstrapMethods.html>
5. Notes for Earliest Known Uses of Some of the Words of Mathematics: Bootstrap (<http://www.economics.soton.ac.uk/staff/aldrich/Mathematical%20Words.htm#boots>) (John Aldrich)
6. Earliest Known Uses of Some of the Words of Mathematics (B) (<http://jeff560.tripod.com/b.html>) (Jeff Miller)
7. Efron, B. (1979). "Bootstrap methods: Another look at the jackknife" (<https://doi.org/10.1214/aos/1176344552>). *The Annals of Statistics*. 7 (1): 1–26. doi:10.1214/aos/1176344552 (<https://doi.org/10.1214/aos/1176344552>).
8. Quenouille M (1949) Approximate tests of correlation in time-series. J Roy Statist Soc Ser B 11 68–84
9. Tukey J (1958) Bias and confidence in not-quite large samples (abstract). Ann Math Statist 29 614
10. Jaeckel L (1972) The infinitesimal jackknife. Memorandum MM72-1215-11, Bell Lab
11. Bickel P, Freeman D (1981) Some asymptotic theory for the bootstrap. Ann Statist 9 1196–1217

12. Singh K (1981) On the asymptotic accuracy of Efron's bootstrap (https://projecteuclid.org/download/pdf_1/euclid.aos/1176345636). *Ann Statist* 9 1187–1195
13. Rubin D (1981). The Bayesian bootstrap. *Ann Statist* 9 130–134
14. Efron, B. (1987). "Better Bootstrap Confidence Intervals". *Journal of the American Statistical Association*. *Journal of the American Statistical Association*, Vol. 82, No. 397. **82** (397): 171–185. doi:10.2307/2289144 (<https://doi.org/10.2307%2F2289144>). JSTOR 2289144 (<https://www.jstor.org/stable/2289144>).
15. Diccio T, Efron B (1992) More accurate confidence intervals in exponential families (<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.998.7855&rep=rep1&type=pdf>). *Biometrika* 79 231–245
16. "21 Bootstrapping Regression Models" (https://www.sagepub.com/sites/default/files/upm-binaries/21122_Chapter_21.pdf) (PDF). Archived (https://web.archive.org/web/20150724033201/http://www.sagepub.com:80/sites/default/files/upm-binaries/21122_Chapter_21.pdf) (PDF) from the original on 2015-07-24.
17. DiCiccio TJ, Efron B (1996) Bootstrap confidence intervals (with Discussion). *Statistical Science* 11: 189–228
18. Hinkley, David (1994-08-01). "[Bootstrap: More than a Stab in the Dark?]: Comment" (<https://dx.doi.org/10.1214/ss/1177010387>). *Statistical Science*. **9** (3). doi:10.1214/ss/1177010387 (<https://doi.org/10.1214%2Fss%2F1177010387>). ISSN 0883-4237 (<https://www.worldcat.org/issn/0883-4237>).
19. Goodhue, D.L., Lewis, W., & Thompson, R. (2012). Does PLS have advantages for small sample size or non-normal data? (https://www.misq.org/skin/frontend/default/misq/pdf/appendices/2012/V36I3_Appendices/GoodhueResearchNoteAppendix.pdf) *MIS Quarterly*, 36(3), 981–1001.
20. Efron, B., Rogosa, D., & Tibshirani, R. (2004). Resampling methods of estimation. In N.J. Smelser, & P.B. Baltes (Eds.). *International Encyclopedia of the Social & Behavioral Sciences* (pp. 13216–13220). New York, NY: Elsevier.
21. Adèr, H. J., Mellenbergh G. J., & Hand, D. J. (2008). *Advising on research methods: A consultant's companion*. Huizen, The Netherlands: Johannes van Kessel Publishing. ISBN 978-90-79418-01-5.
22. Bootstrap of the mean in the infinite variance case Athreya, K.B. *Ann Stats* vol 15 (2) 1987 724–731 (https://projecteuclid.org/download/pdf_1/euclid.aos/1176350371)
23. "How many different bootstrap samples are there? Statweb.stanford.edu" (<https://web.archive.org/web/20190914102512/http://statweb.stanford.edu/~susan/courses/s208/node37.html>). Archived from the original (<http://statweb.stanford.edu/~susan/courses/s208/node37.html>) on 2019-09-14. Retrieved 2019-12-09.
24. Jenkins, David G.; Quintana-Ascencio, Pedro F. (2020-02-21). "A solution to minimum sample size for regressions" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7034864>). *PLOS ONE*. **15** (2): e0229345. Bibcode:2020PLoSO..1529345J (<https://ui.adsabs.harvard.edu/abs/2020PLoSO..1529345J>). doi:10.1371/journal.pone.0229345 (<https://doi.org/10.1371%2Fjournal.pone.0229345>). ISSN 1932-6203 (<https://www.worldcat.org/issn/1932-6203>). PMC 7034864 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7034864>). PMID 32084211 (<https://pubmed.ncbi.nlm.nih.gov/32084211>).
25. Lumley, Thomas (2002). "The Importance of the Normality Assumption in Large Public Health Data Sets". *Annual Review of Public Health*. **23**: 151–169. doi:10.1146/annurev.publhealth.23.100901.140546 (<https://doi.org/10.1146%2Fannurev.publhealth.23.100901.140546>). PMID 11910059 (<https://pubmed.ncbi.nlm.nih.gov/11910059>).
26. Li, Xiang; Wong, Wanling; Lamoureux, Ecosse L.; Wong, Tien Y. (2012-05-01). "Are Linear Regression Techniques Appropriate for Analysis When the Dependent (Outcome) Variable Is Not Normally Distributed?" (<https://iovs.arvojournals.org/article.aspx?articleid=2128171>). *Investigative Ophthalmology & Visual Science*. **53** (6): 3082–3083. doi:10.1167/iovs.12-9967 (<https://doi.org/10.1167%2Fiovs.12-9967>). ISSN 1552-5783 (<https://www.worldcat.org/issn/1552-5783>). PMID 22618757 (<https://pubmed.ncbi.nlm.nih.gov/22618757>).

27. Rubin, D. B. (1981). "The Bayesian bootstrap". *Annals of Statistics*, 9, 130.
28. WANG, SUOJIN (1995). "Optimizing the smoothed bootstrap". *Ann. Inst. Statist. Math.* **47**: 65–80. doi:10.1007/BF00773412 (<https://doi.org/10.1007%2FBF00773412>). S2CID 122041565 (<https://api.semanticscholar.org/CorpusID:122041565>).
29. *A modern introduction to probability and statistics : understanding why and how*. Dekking, Michel, 1946-. London: Springer. 2005. ISBN 978-1-85233-896-1. OCLC 262680588 (<https://www.worldcat.org/oclc/262680588>).
30. Kirk, Paul (2009). "Gaussian process regression bootstrapping: exploring the effects of uncertainty in time course data" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2677737>). *Bioinformatics*. **25** (10): 1300–1306. doi:10.1093/bioinformatics/btp139 (<https://doi.org/10.1093%2Fbioinformatics%2Fbtp139>). PMC 2677737 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2677737>). PMID 19289448 (<https://pubmed.ncbi.nlm.nih.gov/19289448>).
31. Wu, C.F.J. (1986). "Jackknife, bootstrap and other resampling methods in regression analysis (with discussions)" (http://dml.cz/bitstream/handle/10338.dmlcz/135473/Kybernetika_38-2002-4_1.pdf) (PDF). *Annals of Statistics*. **14**: 1261–1350. doi:10.1214/aos/1176350142 (<https://doi.org/10.1214%2Faos%2F1176350142>).
32. Mammen, E. (Mar 1993). "Bootstrap and wild bootstrap for high dimensional linear models" (<https://doi.org/10.1214%2Faos%2F1176349025>). *Annals of Statistics*. **21** (1): 255–285. doi:10.1214/aos/1176349025 (<https://doi.org/10.1214%2Faos%2F1176349025>).
33. Künsch, H. R. (1989). "The Jackknife and the Bootstrap for General Stationary Observations" (<https://doi.org/10.1214%2Faos%2F1176347265>). *Annals of Statistics*. **17** (3): 1217–1241. doi:10.1214/aos/1176347265 (<https://doi.org/10.1214%2Faos%2F1176347265>).
34. Politis, D. N.; Romano, J. P. (1994). "The Stationary Bootstrap". *Journal of the American Statistical Association*. **89** (428): 1303–1313. doi:10.1080/01621459.1994.10476870 (<https://doi.org/10.1080%2F01621459.1994.10476870>). hdl:10983/25607 (<https://hdl.handle.net/10983%2F25607>).
35. Vinod, HD (2006). "Maximum entropy ensembles for time series inference in economics". *Journal of Asian Economics*. **17** (6): 955–978. doi:10.1016/j.asieco.2006.09.001 (<https://doi.org/10.1016%2Fj.asieco.2006.09.001>).
36. Vinod, Hrishikesh; López-de-Lacalle, Javier (2009). "Maximum entropy bootstrap for time series: The meboot R package" (<https://doi.org/10.18637%2Fjss.v029.i05>). *Journal of Statistical Software*. **29** (5): 1–19. doi:10.18637/jss.v029.i05 (<https://doi.org/10.18637%2Fjss.v029.i05>).
37. Cameron, A. C.; Gelbach, J. B.; Miller, D. L. (2008). "Bootstrap-based improvements for inference with clustered errors" (<http://www.nber.org/papers/t0344.pdf>) (PDF). *Review of Economics and Statistics*. **90** (3): 414–427. doi:10.1162/rest.90.3.414 (<https://doi.org/10.1162%2Frest.90.3.414>).
38. Chamandy, N; Muralidharan, O; Najmi, A; Naidu, S (2012). "Estimating Uncertainty for Massive Data Streams" (<http://www.unofficialgoogledatascience.com/2015/08/an-introduction-to-poisson-bootstrap26.html>).
39. Hanley, James A., and Brenda MacGibbon. "Creating non-parametric bootstrap samples using Poisson frequencies." *computer methods and programs in biomedicine* 83.1 (2006): 57-62. pdf (<http://www.med.mcgill.ca/epidemiology/Hanley/Reprints/bootstrap-hanley-macgibbon2006.pdf>)
40. Babu, G. Jogesh, P. K. Pathak, and C. R. Rao. "Second-order correctness of the Poisson bootstrap." *The Annals of Statistics* 27.5 (1999): 1666-1683. link (<https://www.jstor.org/stable/2674086>)
41. Shoemaker, Owen J., and P. K. Pathak. "The sequential bootstrap: a comparison with regular bootstrap." *Communications in Statistics-Theory and Methods* 30.8-9 (2001): 1661-1674. link (<https://www.tandfonline.com/doi/abs/10.1081/STA-100105691>)
42. Jiménez-Gamero, María Dolores, Joaquín Muñoz-García, and Rafael Pino-Mejías. "Reduced bootstrap for the median." *Statistica Sinica* (2004): 1179-1198. link (<https://www.jstor.org/stable/24307226>)

43. Kleiner, A; Talwalkar, A; Sarkar, P; Jordan, M. I. (2014). "A scalable bootstrap for massive data". *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*. **76** (4): 795–816. arXiv:1112.5016 (<https://arxiv.org/abs/1112.5016>). doi:10.1111/rssb.12050 (<https://doi.org/10.1111%2Frssb.12050>). ISSN 1369-7412 (<https://www.worldcat.org/issn/1369-7412>). S2CID 3064206 (<http://api.semanticscholar.org/CorpusID:3064206>).
44. Davison, A. C.; Hinkley, D. V. (1997). *Bootstrap methods and their application*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press. ISBN 0-521-57391-2. software (<http://statwww.epfl.ch/davison/BMA/library.html>).
45. Hesterberg, Tim C (2014). "What Teachers Should Know about the Bootstrap: Resampling in the Undergraduate Statistics Curriculum". arXiv:1411.5279 (<https://arxiv.org/abs/1411.5279>) [stat.OT (<https://arxiv.org/archive/stat.OT>)].
46. Efron, B. (1982). *The jackknife, the bootstrap, and other resampling plans*. Vol. 38. Society of Industrial and Applied Mathematics CBMS-NSF Monographs. ISBN 0-89871-179-7.
47. Scheiner, S. (1998). *Design and Analysis of Ecological Experiments* (<https://archive.org/details/designanalysisof0000unse>). CRC Press. ISBN 0412035618.
48. Rice, John. *Mathematical Statistics and Data Analysis* (2 ed.). p. 272. "Although this direct equation of quantiles of the bootstrap sampling distribution with confidence limits may seem initially appealing, it's rationale is somewhat obscure."
49. Data from examples in *Bayesian Data Analysis* (<http://www.stat.columbia.edu/~gelman/book/data/>)
50. Chihara, Laura; Hesterberg, Tim (3 August 2018). *Mathematical Statistics with Resampling and R* (<https://onlinelibrary.wiley.com/doi/book/10.1002/9781119505969>) (2nd ed.). John Wiley & Sons, Inc. doi:10.1002/9781119505969 (<https://doi.org/10.1002%2F9781119505969>). ISBN 9781119416548. S2CID 60138121 (<https://api.semanticscholar.org/CorpusID:60138121>).
51. Young, G. A. (July 1990). "Alternative Smoothed Bootstraps" (<https://dx.doi.org/10.1111/j.2517-6161.1990.tb01801.x>). *Journal of the Royal Statistical Society, Series B (Methodological)*. **52** (3): 477–484. doi:10.1111/j.2517-6161.1990.tb01801.x (<https://doi.org/10.1111%2Fj.2517-6161.1990.tb01801.x>). ISSN 0035-9246 (<https://www.worldcat.org/issn/0035-9246>).

Further reading

- Diaconis, P.; Efron, B. (May 1983). "Computer-intensive methods in statistics" (<https://web.archive.org/web/20160313214811/https://statistics.stanford.edu/sites/default/files/BIO%2083.pdf>) (PDF). *Scientific American*. **248** (5): 116–130. Bibcode:1983SciAm.248e.116D (<https://ui.adsabs.harvard.edu/abs/1983SciAm.248e.116D>). doi:10.1038/scientificamerican0583-116 (<https://doi.org/10.1038%2Fscientificamerican0583-116>). Archived from the original (<https://statistics.stanford.edu/sites/default/files/BIO%2083.pdf>) (PDF) on 2016-03-13. Retrieved 2016-01-19. popular-science
- Efron, B. (1981). "Nonparametric estimates of standard error: The jackknife, the bootstrap and other methods". *Biometrika*. **68** (3): 589–599. doi:10.1093/biomet/68.3.589 (<https://doi.org/10.1093%2Fbiomet%2F68.3.589>).
- Hesterberg, T. C.; D. S. Moore; S. Monaghan; A. Clipson & R. Epstein (2005). "Bootstrap methods and permutation tests" (https://web.archive.org/web/20060215221403/http://bcs.whfreeman.com/ips5e/content/cat_080/pdf/moore14.pdf) (PDF). In David S. Moore & George McCabe (eds.). *Introduction to the Practice of Statistics*. software (<http://www.insightful.com/Hesterberg/bootstrap>). Archived from the original (http://bcs.whfreeman.com/ips5e/content/cat_080/pdf/moore14.pdf) (PDF) on 2006-02-15. Retrieved 2007-03-23.
- Efron, Bradley (1979). "Bootstrap methods: Another look at the jackknife" (http://projecteuclid.org/DPubS/Repository/1.0/Disseminate?view=body&id=pdf_1&handle=euclid.aos/1176344552). *The Annals of Statistics*. **7**: 1–26. doi:10.1214/aos/1176344552 (<https://doi.org/10.1214%2Faos%2F1176344552>).