

Resampling Method

Related terms:

Credible Interval, Jackknife, Bootstrap, Bootstrap Method, Maximum Likelihood Estimator, Multivariate, Parametric

View all Topics >

 Download as PDF

 Set alert

 About this page

An Overview of Statistics in Education

S. Sinharay, in [International Encyclopedia of Education \(Third Edition\)](#), 2010

Resampling Methods

Resampling methods (see, e.g., Efron, 1982) draw samples from the observed data to draw certain conclusions about the population of interest. Two of the most popular resampling methods are the jackknife and bootstrap. Both of these are examples of nonparametric statistical methods.

Jackknife is used in statistical inference to estimate the bias and standard error of a test statistic. The basic idea behind jackknife lies in systematically recomputing the statistic a large number of times, leaving out one observation or a group of observations at a time from the sample. Estimates of the bias and variance of the statistic can be calculated from this set of jackknife replications of the statistic. The jackknife finds several applications in complex sampling schemes, such as multistage sampling with varying sampling weights – an example of such application is NAEP, where the **jackknife method** is employed to compute standard errors of estimates.

Bootstrap is a statistical method for estimating the sampling distribution of an estimator by sampling with replacement from the original sample, most often with the purpose of deriving robust estimates of standard errors and confidence intervals of a population parameter like a mean, median, and **correlation coefficient**. It is often used as a robust alternative to procedures based on **parametric** assumptions, especially when those assumptions are in doubt, or where parametric inference is impossible or requires very complicated formulas for the calculation of standard errors. See, for example, Hanson *et al.* (1993), who applied the **bootstrap method** to compute the standard error of an equating method.

[View chapter](#)

[Purchase book](#)

Volume 2

F. Westad, M. Kermit, in [Comprehensive Chemometrics](#), 2009

2.14.5.1 Uncertainty Estimates

Uncertainties may be estimated from **resampling methods** such as jackknifing (JK) and bootstrapping.³⁵ Westad and Kermit³⁶ adapted uncertainty estimation by JK in ICA similar to **PCA**. The uncertainty of the ICA loadings (estimated sources), u_m , can be estimated from³⁶

$$\hat{\sigma}^2(\mathbf{u}_m) = \left(\sum_{l=1}^L (\mathbf{u}_m - \mathbf{u}_{m(-l)})^2 \right) \left(\frac{L-1}{L} \right) \quad (31)$$

where L is the number of segments in the cross-validation, $\hat{\sigma}^2(u_m)$ is the estimated uncertainty variance of each variable in the ICA loading for component m and \mathbf{u}_m the ICA loading for component a using all objects. The term $\mathbf{u}_{m(-l)}$ indicates the ICA loading for component m using all objects except the object(s) left out in cross-validation segment l . The quantities \mathbf{u}_m and $\hat{\sigma}^2(\mathbf{u}_m)$ may be subject to a t -test to give the significance values for individual variables for each component.

[View chapter](#)

[Purchase book](#)

Jackknife Methods

S. Sinharay, in [International Encyclopedia of Education \(Third Edition\)](#), 2010

Concluding Remarks

The **jackknife method** is a popular **resampling method** that provides estimates of the bias and a **standard error** of an estimate by recomputing the estimate from **subsamples** of the available sample. The method has some similarities to the **bootstrap method**, but may provide different results from bootstrap in real applications. The method is relatively straightforward and hence can be useful in applications where theoretical statistics cannot readily provide estimates of the bias or standard error of an estimate. Such an example, of an application of the jackknife method to NAEP, is briefly described in this article.

[View chapter](#)

[Purchase book](#)

Large-sample Statistical Methods

J.M. Singer, D.F. Andrade, in [International Encyclopedia of Education \(Third Edition\)](#), 2010

Extensions to Empirical Distribution Functions and Order Statistics

Empirical distribution functions and order statistics have important applications in nonparametric regression models, **resampling methods** such as the jackknife and bootstrap, sequential testing, as well as in survival and reliability analyses. In particular, they serve as the basis for the well-known **goodness-of-fit** Kolmogorov–Smirnov and Cramér–von Mises statistics and for L - and R -estimators such as trimmed or Winsorized means. Given the sample observations Y_1, \dots, Y_n assumed to follow some distribution function F , the empirical distribution function computed at a given real number y is

$$F_n(y) = n^{-1} \sum_{i=1}^n I(Y_i \leq y)$$

where $I(Y_i \leq y)$ is an indicator function assuming the value 1 if $Y_i \leq y$ and 0 otherwise. It is intimately related to the order statistics, $Y_{n:1} \leq Y_{n:2} \leq \dots \leq Y_{n:n}$ where $Y_{n:1}$ is the smallest among Y_1, \dots, Y_n , $Y_{n:2}$ is the second smallest, and so on. For each fixed sample, F_n is a distribution function when considered as a function of y . For every fixed y , when considered as a function of Y_1, \dots, Y_n , $F_n(y)$ is a random variable; in this context, since the $I(Y_i \leq y)$, $i = 1, \dots, n$, are independent and identically distributed zero-one valued random variables, we may apply the CLT to conclude that for each fixed y , the distribution of $F_n(y)$ may be approximated by an $N[F(y), F(y)[1 - F(y)]/n]$ distribution, provided that n is sufficiently large. To extend these results to the function F_n computed at all real values y , more sophisticated methods are needed as suggested in Jurečková and Sen (1996) among others.

[View chapter](#)

[Purchase book](#)

Model Evaluation and Enhancement

Robert Nisbet Ph.D., ... Ken Yale D.D.S., J.D., in [Handbook of Statistical Analysis and Data Mining Applications \(Second Edition\)](#), 2018

Jackknife

The jackknife is a method used to estimate the variance and bias of a large population. This was the earliest **resampling method**, introduced by Quenouille (1949) and named by Tukey (1958). It involves a leave-one-out strategy of the estimation of a parameter (e.g., the mean) in a data set of N observations (or records). Ideally, $N - 1$ models are built on the data set with different factors left out of each model. The estimates of all models are then aggregated into a single estimate of the parameter. The jackknife gets computationally intractable as $N \rightarrow \infty$. The success of the jackknife in academics and research led to the development of the **bootstrap method**.

[View chapter](#)

[Purchase book](#)

Time Series Analysis: Methods and Applications

Arthur Berg, ... Dimitris N. Politis, in [Handbook of Statistics](#), 2012

1 Introduction

Ever since the fundamental recognition of the potential role of the computer in modern statistics (Efron, 1979a,b), the bootstrap and other **resampling methods** have been extensively developed for inference in independent data settings; see, e.g., the works done by Davison and Hinkley (1997), Efron and Tibshirani (1993), Hall (1997), Shao and Tu (1995). Such methods are even more important in the context of dependent data where the distribution theory for estimators and test statistics may be difficult to obtain even asymptotically.

In the time series context, different resampling and subsampling methods have been proposed, and are currently receiving the attention of the statistical community. Reviews of the impact of **bootstrap methods** on time series analysis may be found in books (Lahiri, 2003; Politis *et al.*, 1999), papers (Bühlmann, 2002; Politis, 2003), and the review by J.-P. Kreiss and S. N. Lahiri in this volume of the Handbook.

In the paper at hand, we revisit the problem of assessing whether a given time series is linear versus nonlinear, or Gaussian versus non-Gaussian. In practice, a Gaussian classification would indicate an **Autoregressive Moving Average** (ARMA) model with Gaussian innovations is appropriate, whereas a linear classification would indicate that an ARMA model with independent but possibly non-Gaussian innovations can still be considered. However, the rejection of linearity typically requires the practitioner to carefully select an appropriate **nonlinear model** for the underlying time series, or even to proceed in a model-free, nonparametric manner.

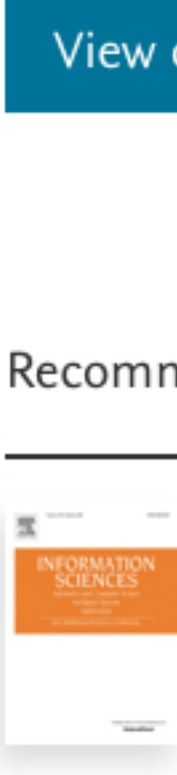
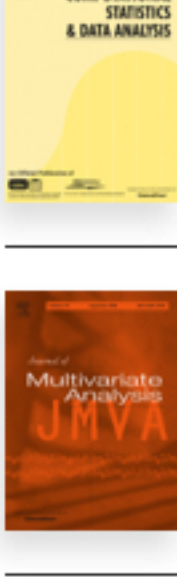


We review the traditional linearity and Gaussianity tests that are based on the normalized bispectrum. The critical regions of these tests have been traditionally determined via **asymptotic methods**. As an alternative, we describe how these critical regions can be determined via resampling (e.g., the AR-sieve bootstrap) and/or subsampling. One of the advantages of subsampling methodology is the generality under which it is valid. There are a number of examples where subsampling yields consistent estimation but the bootstrap fails (Politis *et al.*, 1999). Although subsampling is more widely applicable, it is noted that when the bootstrap is indeed valid it may possess second-order **asymptotic properties** (Hall, 1997) giving the bootstrap an advantage.

The literature on linearity and Gaussianity tests is reviewed in the next section. The concept of time series linearity is thoroughly described in Section 3. Sections 4 and 5 focus on the AR-sieve bootstrap and subsampling tests, respectively.

[View chapter](#)

[Purchase book](#)

Recommended publications

	Information Sciences Journal
	Computational Statistics & Data Analysis Journal
	Journal of Multivariate Analysis Journal
	Statistical Methodology Journal

[Browse books and journals](#)

Computational Statistics with R

Chaitra H. Nagaraja, in [Handbook of Statistics](#), 2014

4.3 For Loops

Especially useful for **resampling methods** and simulation studies, a for loop repeats a process for a preset (fixed) number of iterations. The basic format is for(*INDEX* in *VECTOR*){*PROCESS*}. *INDEX* is the variable name that is the counter for the loops, *VECTOR* is a vector object or a vector itself which gives the counter values to be looped through, and *PROCESS* is the set of operations to be repeated. For example, to compute the sample mean, we must start by adding the elements of a vector, best done by a loop. In the following code, we use the data faithful that contains observations on Old Faithful geyser activity at Yellowstone National Park in the United States. The first column lists the length of each eruption, and the second, the time until the next eruption, both in minutes. The code below utilizes a for loop to compute and print the sample mean of each column; observe that we again use indentation for clarity:

```
> for(i in names(faithful)){ # loop through each
  column of faithful
  >
  > sum.x <- 0
  > for(j in 1:nrow(faithful)){ # loop through each row
    of column i
    > # compute cumulative sum
    > sum.x <- sum.x + faithful[j,i]
  > } # end j for loop (through rows)
  >
  > # compute and print column average
  > print(paste("sample mean of column", i, "is:",
    + round(sum.x/nrow(faithful), digits=3), sep=" "))
  > } # end i for loop (through columns)
  >
  > # checking our code:
  > apply(faithful, 2, mean)
  >
  > # eruptions mean: 3.488
  > # waiting mean: 70.897
```

We have two loops that are nested in the example above. For loop i , *VECTOR* is composed of character strings, the column names, whereas loop j is indexed by the row numbers.

The function paste() was used above to format the output generated by print() (or cat()). A second use is to print how many iterations have been completed. For instance, say the loop is indexed by i through a numerical vector:

```
> i <- 100
> if(i%100==0) print(paste("iter",i,"is complete",
  sep=" "))
> # sep=" " indicates a space should separate each
  component in paste()
```

This prints i after every 100 iterations using the **modulo** operator $\%$. For loops with time intensive processes, generating such a statement is convenient to monitor your progress.

[View chapter](#)

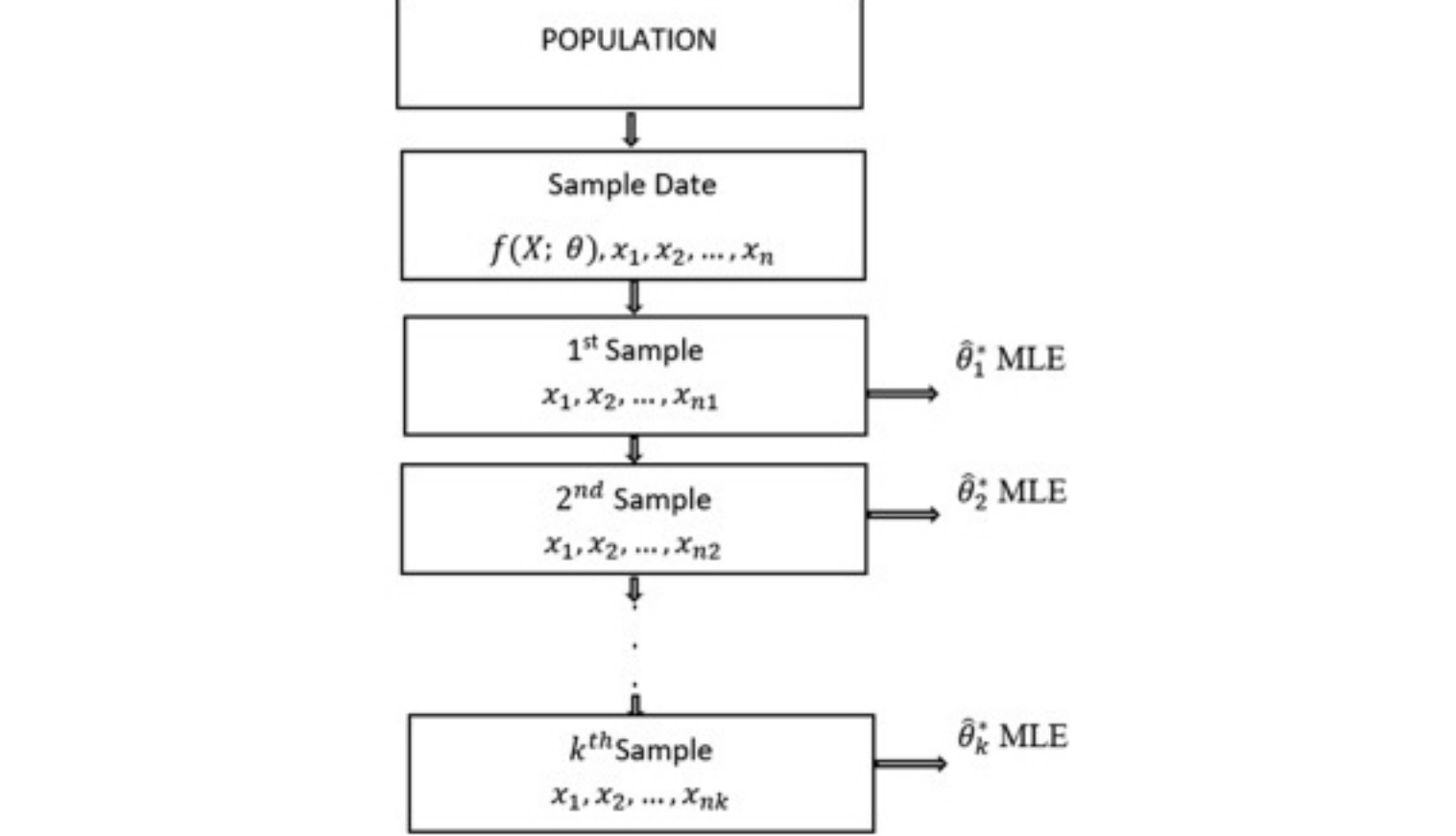
[Purchase book](#)

Bayesian estimation and inference

Kandethody M. Ramachandran, Chris P. Tsokos, in [Mathematical Statistics with Applications in R \(Third Edition\)](#), 2021

10.6.2 Bootstrap resampling

Bradley Efron in 1979 introduced the bootstrap **resampling method** for estimating the sampling distribution of an estimator. Given a set of data n , using the subject method, we generate k samples with replacement from the given data with $k < n$. The pdf of the k samples will follow the original pdf of the n independent and identically distributed observations. Consider the observation x_1, x_2, \dots, x_n ; by bootstrapping we obtain different subsets of our original sample, that is, a subsample of size k . There are several uses of this method, but in our present study of empirical Bayes, we shall use bootstrapping resampling to obtain an estimate of the prior pdf. For a given set of data x_1, x_2, \dots, x_n , we will proceed if possible to identify the pdf, $f(x|\theta)$, that follows the observations or the population that it is drawn from. Through bootstrap resampling we will obtain a sequence of estimates, $\hat{\theta}_1^*$, $\hat{\theta}_2^*$, \dots , $\hat{\theta}_k^*$, and through goodness-of-fit methods, we proceed to obtain an estimate of the prior pdf, $\pi(\theta)$, if possible. The following diagram illustrates the process we follow to resample using the bootstrap method:



[Sign in to download full-size image](#)

Thus, our objective is to use this sequence of estimates to obtain, through the goodness-of-fit method, if possible, the pdf that drives these estimated $\hat{\theta}_1^*$, $\hat{\theta}_2^*$, \dots , $\hat{\theta}_k^*$ and consider it as the prior pdf, $\pi(\theta)$, of the parameter θ , that is, $\pi(\hat{\theta}^*)$. We then proceed to obtain the Bayesian estimate of θ without having to guess it.

[View chapter](#)

[Purchase book](#)

Resampling methods combined with Rao-Blackwellized Monte Carlo Data Association algorithm

Soheil Sadat Hosseini, Mohsin M. Jamali, in [Handbook of Probabilistic Models](#), 2020

3.3 Residual resampling (remainder resampling)

The residual resampling algorithm developed by Beadle and Djuric (1997) is an alternative technique to systematic **resampling method**. It includes two steps of resampling. The first step is to locate which particle's weight is above $\frac{1}{B}$, and the second step is to draw randomly utilizing the remaining particles whose weights are below $\frac{1}{B}$ (referred to as residuals). For the first step, set $B_k^{(s)} = B\omega_k^{(s)}$, and this step is named the deterministic replication section. The second stage is residual resampling method, assigning the residual of the weight as follows:

$$\hat{\omega}_k^{(m)} = \omega_k^{(m)} - \frac{B_k^{(m)}}{B} \quad (18.16)$$

Then, the new particles are sampled by calling the CSW technique with the parameters. The total number of replicated particles in the first step is $B_k = \sum_{s=1}^M B_k^{(s)}$ and in the second step $R_k = B - B_k$.

In the second stage, CSW resampling is applied to choose the particles based on the residual weights. The probability for choosing $x_k^{(s)}$ is proportional to the residual weight of that particle (Probability $(x_k^n = x_k^s) = \hat{\omega}_k^{(s)}$). The first stage is a deterministic replication, and thus, the variation of the number of times $x_k^{(s)}$ resampled only happens on the second step. Thus, if the second stage is employed utilizing CSW resampling, the lower and upper limits of the number of times that the s th particle is resampled are $B\omega_k^{(s)}$ and $B\omega_k^{(s)} + R_k$, respectively (Beadle and Djuric, 1997) (Fig. 18.2).



[Sign in to download full-size image](#)

Figure 18.2. Pseudocode for residual resampling method.

[View chapter](#)

[Purchase book](#)

Volume 2

G.J. McLachlan, in [Comprehensive Chemometrics](#), 2009

2.30.10.2 Resampling Approach

A formal test of the null hypothesis $H_0 : g = g_0$ versus the alternative $H_1 : g = g_1 (g_1 > g_0)$ can be undertaken using a **resampling method**, as described in McLachlan.²⁶ With this approach, **bootstrap samples** are generated from the mixture model fitted under the null hypothesis of g_0 components. That is, the bootstrap samples are generated from the g_0 -component mixture model with the vector Ψ of unknown parameters replaced by its ML estimate $\hat{\Psi}_{g_0}$ computed by consideration of the log likelihood formed from the original data under H_0 . The value of $-2 \log \lambda$, where λ is the **likelihood ratio statistic**, is computed for each bootstrap sample after fitting mixture models for $g = g_0$ and g_1 to it in turn. The process is repeated independently B times, and the replicated values of $-2 \log \lambda$ formed from the successive bootstrap samples provide an assessment of the bootstrap, and hence of the true, null distribution of $-2 \log \lambda$. Other resampling approaches include that based on the Gap statistic of Tibshirani *et al.*²⁷ and the Clest method of Dudoit and Fridlyand.²⁸

[View chapter](#)

[Purchase book](#)

[FEEDBACK](#) 