1

INTRODUCTION

COPYRIGHTED MATERIAL

1.1 HISTORICAL BACKGROUND

The "bootstrap" is one of a number of techniques that is now part of the broad umbrella of nonparametric statistics that are commonly called resampling methods. Some of the techniques are far older than the bootstrap. Permutation methods go back to Fisher (1935) and Pitman (1937, 1938), and the jackknife started with Quenouille (1949). Bootstrapping was made practical through the use of the Monte Carlo approximation, but it too goes back to the beginning of computers in the early 1940s.

However, 1979 is a critical year for the bootstrap because that is when Brad Efron's paper in the *Annals of Statistics* was published (Efron, 1979). Efron had defined a resampling procedure that he coined as bootstrap. He constructed it as a simple approximation to the jackknife (an earlier resampling method that was developed by John Tukey), and his original motivation was to derive properties of the bootstrap to better understand the jackknife. However, in many situations, the bootstrap is as good as or better than the jackknife as a resampling procedure. The jackknife is primarily useful for small samples, becoming computationally inefficient for larger samples but has become more feasible as computer speed increases. A clear description of the jackknife

---

An Introduction to Bootstrap Methods with Applications to R, First Edition. Michael R. Chernick, Robert A. LaBudde.
© 2011 John Wiley & Sons, Inc. Published 2011 by John Wiley & Sons, Inc.

and its connecton to the bootstrap can be found in the SIAM monograph Efron (1982). A description of the jackknife is also given in Section 1.2.1.

Although permutation tests were known in the 1930s, an impediment to their use was the large number (i.e., $n!$) of distinct permutations available for samples of size n. Since ordinary bootstrapping involves sampling with replacement n times for a sample of size n, there are n^n possible distinct ordered bootstrap samples (though some are equivalent under the exchangeability assumption because they are permutations of each other). So, complete enumeration of all the bootstrap samples becomes infeasible except in very small sample sizes. Random sampling from the set of possible bootstrap samples becomes a viable way to approximate the distribution of bootstrap samples. The same problem exists for permutations and the same remedy is possible. The only difference is that $n!$ does not grow as fast as n^n, and complete enumeration of permutations is possible for larger n than for the bootstrap.

The idea of taking several Monte Carlo samples of size n with replacement from the original observations was certainly an important idea expressed by Efron but was clearly known and practiced prior to Efron (1979). Although it may not be the first time it was used, Julian Simon laid claim to priority for the bootstrap based on his use of the Monte Carlo approximation in Simon (1969). But Simon was only recommending the Monte Carlo approach as a way to teach probability and statistics in a more intuitive way that does not require the abstraction of a parametric probability model for the generation of the original sample. After Efron made the bootstrap popular, Simon and Bruce joined the campaign (see Simon and Bruce, 1991, 1995).

Efron, however, starting with Efron (1979), first connected bootstrapping to the jackknife, delta method, cross-validation, and permutation tests. He was the first to show it to be a real competitor to the jackknife and delta method for estimating the standard error of an estimator. Also, quite early on, Efron recognized the broad applicability of bootstrapping for confidence intervals, hypothesis testing, and more complex problems. These ideas were emphasized in Efron and Gong (1983), Diaconis and Efron (1983), Efron and Tibshirani (1986), and the SIAM monograph (Efron 1982). These influential articles along with the SIAM monograph led to a great deal of research during the 1980s and 1990s. The explosion of bootstrap papers grew at an exponential rate. Key probabilistic results appeared in Singh (1981), Bickel and Freedman (1981, 1984), Beran (1982), Martin (1990), Hall (1986, 1988), Hall and Martin (1988), and Navidi (1989).

In a very remarkable paper, Efron (1983) used simulation comparisons to show that the use of bootstrap bias correction could provide better estimates of classification error rate than the very popular cross-validation approach (often called leave-one-out and originally proposed by Lachenbruch and Mickey, 1968). These results applied when the sample size was small, and classification was restricted to two or three classes only, and the predicting features had multivariate Gaussian distributions. Efron compared several variants of the bootstrap with cross-validation and the resubstitution methods. This led to several follow-up articles that widened the applicability and superiority of a version of the bootstrap called 632. See Chatterjee and Chatterjee (1983), Chernick et al. (1985, 1986, 1988a, b), Jain et al. (1987), and Efron and Tibshirani (1997).

Chernick was a graduate student at Stanford in the late 1970s when the bootstrap activity began on the Stanford and Berkeley campuses. However, oddly the bootstrap did not catch on with many graduate students. Even Brad Efron's graduate students chose other topics for their dissertation. Gail Gong was the first student of Efron to do a dissertation on the bootstrap. She did very useful applied work on using the bootstrap in model building (particularly for logistic regression subset selection). See Gong (1986). After Gail Gong, a number of graduate students wrote dissertations on the bootstrap under Efron, including Terry Therneau, Rob Tibshirani, and Tim Hesterberg. Michael Martin visited Stanford while working on his dissertation on bootstrap confidence intervals under Peter Hall. At Berkeley, William Navidi did his thesis on bootstrapping in regression and econometric models under David Freedman.

While exciting theoretical results developed for the bootstrap in the 1980s and 1990s, there were also negative results where it was shown that the bootstrap estimate is not "consistent" in the probabilistic sense (i.e., approaches the true parameter value as the sample size becomes infinite). Examples included the mean when the population distribution does not have a finite variance and when the maximum or minimum is taken from a sample. This is illustrated in Athreya (1987a, b), Knight (1989), Angus (1993), and Hall et al. (1993). The first published example of an inconsistent bootstrap estimate appeared in Bickel and Freedman (1981). Shao et al. (2000) showed that a particular approach to bootstrap estimation of individual bioequivalence is also inconsistent. They also provide a modification that is consistent. Generally, the bootstrap is consistent when the central limit theorem applies (a sufficient condition is Lyapanov's condition that requires existence of the $2 + \delta$ moment of the population distribution). Consistency results in the literature are based on the existence of Edgeworth expansions; so, additional smoothness conditions for the expansion to exist have also been assumed (but it is not known whether or not they are necessary).

One extension of the bootstrap called m-out-of-n was suggested by Bickel and Ren (1996) in light of previous research on it, and it has been shown to be a method to overcome inconsistency of the bootstrap in several instances. In the m-out-of-n bootstrap, sampling is with replacement from the original sample but with a value of m that is smaller than n. See Bickel et al. (1997), Gine and Zinn (1989), Arcones and Gine (1989), Fukuchi (1994), and Politis et al. (1999).

Some bootstrap approaches in time series have been shown to be inconsistent. Lahiri (2003) covered the use of bootstrap in time series and other dependent cases. He showed that there are remedies for the m-dependent and moving block bootstrap cases (see Section 5.5 for some coverage of moving block bootstrap) that are consistent.

1.2 DEFINITION AND RELATIONSHIP TO THE DELTA METHOD AND OTHER RESAMPLING METHODS

We will first provide an informal definition of bootstrap to provide intuition and understanding before a more formal mathematical definition. The objective of bootstrapping is to estimate a parameter based on the data, such as a mean, median, or standard deviation. We are also interested in the properties of the distribution for the parameter's estimate and may want to construct confidence intervals. But we do not want to make overly restrictive assumptions about the form of the distribution that the observed data came from.

For the simple case of independent observations coming from the same population distribution, the basic element for bootstrapping is the empirical distribution. The empirical distribution is just the discrete distribution that gives equal weight to each data point (i.e., it assigns probability $1/n$ to each of the original n observations and shall be denoted F_n).

Most of the common parameters that we consider are functionals of the unknown population distribution. A functional is simply a mapping that takes a function F into a real number. In our case, we are only interested in the functionals of cumulative probability distribution functions. So, for example, the mean and variance of a distribution can be represented as functionals in the following way. Let μ be the mean for a distribution function F, then $\mu = \int x dF(x)$ Let σ^2 be the variance then $\sigma^2 = \int (x-\mu)^2 dF(x)$. These integrals over the entire possible set of x values in the domain of F are particular examples of functionals. It is interesting that the sample estimates most commonly used for these parameters are the same functionals applied to the F_n.

Now the idea of bootstrap is to use only what you know from the data and not introduce extraneous assumptions about the population distribution. The "bootstrap principle" says that when F is the population distribution and $T(F)$ is the functional that defines the parameter, we wish to estimate based on a sample of size n, let F_n play the role of F and F_n^*, the bootstrap distribution (soon to be defined), play the role of F_n in the resampling process. Note that the original sample is a sample of n independent identically distributed observations from the distribution F and the sample estimate of the parameter is $T(F_n)$. So, in bootstrapping, we let F_n play the role of F and take n independent and identically distributed observations from F_n. Since F_n is the empirical distribution, this is just sampling randomly with replacement from the original data.

Suppose we have $n = 5$ and the observations are $X_1 = 7$, $X_2 = 5$, $X_3 = 3$, $X_4 = 9$, and $X_5 = 6$ and that we are estimating the mean. Then, the sample estimate of the population parameter is the sample mean, $(7 + 5 + 3 + 9 + 6)/5 = 6.0$. Then sampling from the data with replacement generates what we call a bootstrap sample.

The bootstrap sample is denoted X_1^*, X_2^*, X_3^*, X_4^*, and X_5^*. The distribution for sampling with replacement from F_n is called the bootstrap distribution, which we previously denoted by F_n^*. The bootstrap estimate is then $T\left(F_n^*\right)$. So a bootstrap sample might be $X_1^* = 5$, $X_2^* = 9$, $X_3^* = 7$, $X_4^* = 7$, and $X_5^* = 5$, with estimate $(5 + 9 + 7 + 7 + 5)/5 = 6.6$.

Note that, although it is possible to get the original sample back typically some values get repeated one or more times and consequently others get omitted. For this bootstrap sample, the bootstrap estimate of the mean is $(5 + 9 + 7 + 7 + 5)/5 = 6.6$. Note that the bootstrap estimate differs from the original sample estimate, 6.0. If we take another bootstrap sample, we may get yet another estimate that may be different from the previous one and the original sample. Assume for the second bootstrap sample we get in this case the observation equal to 9 repeated once. Then, for this bootstrap sample, $X_1^* = 9$, $X_2^* = 9$, $X_3^* = 6$, $X_4^* = 7$, and $X_5^* = 5$, and the bootstrap estimate for the mean is 7.2.

If we repeat this many times, we get a histogram of values for the mean, which we will call the Monte Carlo approximation to the bootstrap distribution. The average of all these values will be very close to 6.0 since the theoretical mean of the bootstrap distribution is the sample mean. But from the histogram (i.e., resampling distribution), we can also see the variability of these estimates and can use the histogram to estimate skewness, kurtosis, standard deviation, and confidence intervals.

In theory, the exact bootstrap estimate of the parameter could be calculated by averaging appropriately over all possible bootstrap samples, and in this example for the mean, that value would be 6.0. As noted before, there can be n^n distinct bootstrap samples (taking account of the ordering of the observations), and so even for $n = 10$, this becomes very large (i.e., 10 billion). So, in practice, a Monte Carlo approximation is used.

If you randomly generate $M = 10{,}000$ or 100,000 bootstrap samples, the distribution of bootstrap estimates will approximate the bootstrap distribution for the estimate. The larger M is the closer the histogram approaches the true bootstrap distribution. Here is how the Monte Carlo approximation works:

1. Generate a sample with replacement from the empirical distribution for the data (this is a bootstrap sample).
2. Compute $T\left(F_n^*\right)$ the bootstrap estimate of $T(F)$. This is a replacement of the original sample with a bootstrap sample and the bootstrap estimate of $T(F)$ in place of the sample estimate of $T(F)$.
3. Repeat steps 1 and 2 M times where M is large, say 100,000.

Now a very important thing to remember is that with the Monte Carlo approximation to the bootstrap, there are two sources of error:

1. the Monte Carlo approximation to the bootstrap distribution, which can be made as small as you like by making M large;
2. the approximation of the bootstrap distribution F_n^* to the population distribution F.

If $T\left(F_n^*\right)$ converges to $T(F)$ as $n \to \infty$, then bootstrapping works. It is nice that this works out often, but it is not guaranteed. We know by a theorem called the Glivenko–Cantelli theorem that F_n converges to F uniformly. Often, we know that the sample estimate is consistent (as is the case for the sample mean). So, (1) $T(F_n)$ converges to $T(F)$ as $n \to \infty$. But this is dependent on smoothness conditions on the functional T. So we also need (2) $T\left(F_n^*\right) - T\left(F_n\right)$ to tend to 0 as $n \to \infty$. In proving that bootstrapping works (i.e., the bootstrap estimate is consistent for the population parameter), probability theorists needed to verify (1) and (2). One approach that is commonly used is by verifying that smoothness conditions are satisfied for expansions like the Edgeworth and Cornish–Fisher expansions. Then, these expansions are used to prove the limit theorems.

The probability theory associated with the bootstrap is beyond the scope of this text and can be found in books such as Hall (1992). What is important is that we know that consistency of bootstrap estimates has been demonstrated in many cases and examples where certain bootstrap estimates fail to be consistent are also known. There is a middle ground, which are cases where consistency has been neither proved nor disproved. In those cases, simulation studies can be used to confirm or deny the usefulness of the bootstrap estimate. Also, simulation studies can be used when the sample size is too small to count on asymptotic theory, and its use in small to moderate sample sizes needs to be evaluated.

1.2.1 Jackknife

The jackknife was introduced by Quenouille (1949). Quenouille's aim was to improve an estimate by correcting for its bias. Later on, Tukey (1958) popularized the method and found that a more important use of the jackknife was to estimate standard errors of an estimate. It was Tukey who coined the name jackknife because it was a statistical tool with many purposes. While bootstrapping uses the bootstrap samples to estimate variability, the jackknife uses what are called pseudovalues.

First, consider an estimate $\tilde{u}$ based on a sample of size n of observations independently drawn from a common distribution F. Here, just as with the bootstrap, we again let F_n be the empirical distribution for this data set and assume that the parameter $u = T(F)$, a functional; $\tilde{u} = T(F_n)$, and $\tilde{u}_{(i)} = T(F_{n(i)})$, where $F_{n(i)}$ is the empirical distribution function for the $n - 1$ observations obtained by leaving the ith observation out. If $\tilde{u}$ is the population variance, the jackknife estimate of variance of σ^2 is obtained as follows:

$$\sigma^2_{\text{JACK}} = n\sum_{i=1}^{n}\left(\tilde{u}_{(i)} - u^*\right)^2 \big/ (n-1),$$

where $u^* = \sum_{i=1}^{n} \tilde{u}_{(i)} \big/ n$. The jackknife estimate of standard error for $\tilde{u}$ is just the square root of σ^2_{JACK}. Tukey defined the pseudovalue as $\tilde{u}_i = \tilde{u} + (n-1)\left(\tilde{u} - \tilde{u}_{(i)}\right)$. Then the jackknife estimate of the parameter u is $u_{\text{JACK}} = \sum_{i=1}^{n} \tilde{u}_i \big/ n$. So the name pseudovalue comes about because the estimate is the average of the pseudovalues. Expressing the estimate of the variance of the estimate $\tilde{u}$ in terms of the pseudovalues we get

$$\sigma^2_{\text{JACK}} = \sum_{i=1}^{n}\left(\tilde{u}_i - u_{\text{JACK}}\right)^2 \big/ [n(n-1)].$$

In this form, we see that the variance is the usual estimate for variance of a sample mean. In this case, it is the sample mean of the pseudovalues. Like the bootstrap, the jackknife has been a very useful tool in estimating variances for more complicated estimators such as trimmed or Winsorized means.

One of the great surprises about the bootstrap is that in cases like the trimmed mean, the bootstrap does better than the jackknife (Efron, 1982, pp. 28–29). For the sample median, the bootstrap provides a consistent estimate of the variance but the jackknife does not! See Efron (1982, p. 16 and chapter 6). In that monograph, Efron also showed, using theorem 6.1, that the jackknife estimate of standard error is essentially the bootstrap estimate with the parameter estimate replaced by a linear approximation of it. In this way, there is a close similarity between the two methods, and if the linear approximation is a good approximation, the jackknife and the bootstrap will both be consistent. However, there are complex estimators where this is not the case.

1.2.2 Delta Method

It is often the case that we are interested in the moments of an estimator. In particular, for these various methods, the variance is the moment we are most interested in. To illustrate the delta method, let us define $\varphi = f(\alpha)$ where the parameters φ and α are both one-dimensional variables and f is a function differentiable with respect to α. So there exists a Taylor series expansion for f at a point say α_0. Carrying it out only to first order, we get $\varphi = f(\alpha) = f(\alpha_0) + (\alpha - \alpha_0)f'(\alpha_0)$ + remainder terms and dropping the remainder terms leaves

$$\varphi = f(\alpha) = f(\alpha_0) + (\alpha - \alpha_0) f'(\alpha_0)$$

or

$$f(\alpha) - f(\alpha_0) = (\alpha - \alpha_0) f'(\alpha_0).$$

Squaring both sides of the last equation gives us $[f(\alpha) - f(\alpha_0)]^2 = (\alpha - \alpha_0)^2[f'(\alpha_0)]^2$. Now we want to think of $\varphi = f(\alpha)$ as a random variable, and upon taking expectations of the random variables on each side of the equation, we get

$$E[f(\alpha) - f(\alpha_0)]^2 = E(\alpha - \alpha_0)^2 [f'(\alpha_0)]^2. \tag{1.1}$$

Here, α and $f(\alpha)$ are random variables, and α_0, $f(\alpha_0)$, and $f'(\alpha_0)$ are all constants. Equation 1.1 provides the delta method approximation to the variance of $\varphi = f(\alpha)$ since the left-hand side is approximately the variance of φ and the right-hand side is the variance of α multiplied by the constant $[f'(\alpha_0)]^2$ if we choose α_0 to be the mean of α.

1.2.3 Cross-Validation

Cross-validation is a general procedure used in statistical modeling. It can be used to determine the best model out of alternative choices such as order of an autoregressive time series model, which variables to include in a logistic regression or a multiple linear regression, number of distributions in a mixture model, and the choice of a parametric classification model or for pruning classification trees.

The basic idea of cross-validation is to randomly split the data into two subsets. One is used to fit the model, and the other is used to test the model. The extreme case would be to fit all the data except for a single observation and see how well that model predicts the value of the observation left out. But a sample of size 1 is not very good for assessment. So, in the case of classification error rate estimation, Lachenbruch and Mickey (1968) proposed the leave-one-out method of assessment. In this case, a model is fit to the $n - 1$ observations that are included and is tested on the one left out. But the model fitting and prediction is then done separately for all n observations by testing the model fit without observation i for predicting the class for the case i. Results are obtained from each i and then averaged. Efron (1983) included a simulation study that showed for bivariate normal distributions the "632" variant of the bootstrap does better than leave-one-out. For pruning classification trees, see Brieman et al. (1984).

1.2.4 Subsampling

The idea of subsampling goes back to Hartigan (1969), who developed a theory of confidence intervals for random subsampling. He proved a theorem called the typical value theorem when M-estimators are used to estimate parameters. We shall see in the chapter on confidence intervals that Hartigan's results were motivating factors for Efron to introduce the percentile method bootstrap confidence intervals.

More recently the theory of subsampling has been further developed and related to the bootstrap. It has been applied when the data are independent observations and also when there are dependencies among the data. A good summary of the current literature along with connections to the bootstrap can be found in Politis et al. (1999), and consistency under very minimal assumptions can be found in Politis and Romano (1994). Politis, Romano, and Wolf included applications when the observations are independent and also for dependent situations such as stationary and nonstationary time series, random fields, and marked point processes. The dependent situations are also well covered in section 2.8 of Lahiri (2003).

We shall now define random subsampling. Let $S_1, S_2, \ldots, S_{B-1}$ be $B - 1$ of the $2^n - 1$ nonempty subsets of the integers $1, 2, \ldots, n$. These $B - 1$ subsets are selected at random without replacement. So a subset of size 3 might be drawn, and it would contain $\{1, 3, 5\}$. Another subset of size 3 that could be drawn could be $\{2, 4, n\}$. Subsets of other sizes could also be drawn. For example, a subset of size 5 is $\{1, 7, 9, 12, 13\}$. There are many subsets to select from. There is only 1 subset of size n, and it contains all the integers from 1 to n. There are n subsets of size $n - 1$. Each distinct subset excludes one and only one of the integers from 1 to n. For more details on this and M-estimators and the typical value theorem see sections 3.1.1 and 3.1.2 of Chernick (2007).

1.3 WIDE RANGE OF APPLICATIONS

There is a great deal of temptation to apply the bootstrap in a wide variety of settings. But as we have seen, the bootstrap does not always work. So how do we know when it will work? We either have to prove a consistency theorem under a set of assumptions or we have to verify that it is well behaved through simulations.

In regression problems, there are at least two approaches to bootstrapping. One is called "bootstrapping residuals," and the other is called "bootstrapping vectors or cases." In the first approach, we fit a model to the data and compute the residuals from the model. Then we generate a bootstrap sample by resampling with replacement from the model residuals. In the second approach, we resample with replacement from the n, $k + 1$ dimensional vectors:

$$(y_i, X_{1i}, X_{2i}, \dots, X_{ki}) \quad \text{for} \quad I = 1, 2, \dots, n.$$

In the first approach, the model is fixed. In the second, it is redetermined each time. Both methods can be applied when a parametric regression model is assumed. But in practice, we might not be sure that the parametric form is correct. In such cases, it is better to use the bootstrapping vectors approach.

The bootstrap has also been successfully applied to the estimation of error rates for discriminant functions using bias adjustment as we will see in Chapter 2. The bootstrap and another resampling procedure called "permutation tests," as described in Good (1994), are attractive because they free the scientists from restrictive parametric assumptions that may not apply in their particular situation.

Sometimes the data can have highly skewed or heavy-tailed distributions or multiple modes. There is no need to simplify the model by, say, a linear approximation when the appropriate model is nonlinear. The estimator can be defined through an algorithm and there does not need to be an analytic expression for the parameters to be estimated.

Another feature of the bootstrap is its simplicity. For almost any problem you can think of, there is a way to construct bootstrap samples. Using the Monte Carlo approximation to the bootstrap estimate, all the work can be done by the computer. Even though it is a computer-intensive method, with the speed of the modern computer, most problems are feasible, and in many cases, up to 100,000 bootstrap samples can be generated without consuming hours of CPU time. But care must be taken. It is not always apparent when the bootstrap will fail, and failure may not be easy to diagnose.

In recent years, we are finding that there are ways to modify the bootstrap so that it will work for problems where the simple (or naïve) bootstrap is known to fail. The "m-out-n" bootstrap is one such example.

In many situations, the bootstrap can alert the practitioner to variability in his procedures that he otherwise would not be aware of. One example in spatial statistics is the development of pollution level contours based on a smoothing method called "kriging." By generating bootstrap samples, multiple kriging contour maps can be generated, and the differences in the contours can be determined visually.

Also, the stepwise logistic regression problem that is described in Gong (1986) shows that variable selection can be somewhat of a chance outcome when there are many competing variables. She showed this by bootstrapping the entire stepwise selection procedure and seeing that the number of variables and the choice of variables selected can vary from one bootstrap sample to the next.

Babu and Feigelson (1996) applied the bootstrap to astronomy problems. In clinical trials, the bootstrap is used to estimate individual bioequivalence, for P-value adjustment with multiple end points, and even to estimate mean differences when the sample size is not large enough for asymptotic theory to take hold or the data are very non-normal and statistics other that the mean are important.

1.4 THE BOOTSTRAP AND THE R LANGUAGE SYSTEM

In subsequent chapters of this text, we will illustrate examples with calculations and short programs using the R language system and its associated packages.

R is an integrated suite of an object-oriented programming language and software facilities for data manipulation, calculation, and graphical display. Over the last decade, R has become the statistical environment of choice for academics, and probably is now the most used such software system in the world. The number of specialized packages available in R has increased exponentially, and continues to do so. Perhaps the best thing about R (besides its power and breadth) is this: It is completely free to use. You can obtain your own copy of the R system at http://www.cran.r-project.org/.

From this website, you can get not only the executable version of R for Linux, Macs, or Windows, but also even the source programs and free books containing documentation. We have found *The R Book* by Michael J. Crawley a good way to learn how to use R, and have found it to be an invaluable reference afterword.

There are so many good books and courses from which you can learn R, including courses that are Internet based, such as at http://statistics.com. We will not attempt to teach even the basics of R here. What we will do is show those features of direct applicability, and give program snippets to illustrate examples and the use of currently available R packages for bootstrapping. These snippets will be presented in the Courier typeface to distinguish them from regular text and to maintain spacing in output generated.

At the current time, using R version 2.10.1, the R query (“>” denotes the R command line prompt)

```
    > ?? bootstrap
    or
    > help.search('bootstrap')
    results in
    agce::resamp.std Compute the standard
deviation by bootstrap.
    alr3::boot.case Case bootstrap for
regression models
    analogue::RMSEP Root mean square error of
prediction
    analogue::bootstrap Bootstrap estimation and
errors
    analogue::bootstrap.waBootstrap estimation and
errors for WA models
    analogue::bootstrapObject Bootstrap object
description
```

analogue::getK Extract and set the number of analogues

analogue::performance Transfer function model performance statistics

analogue::screeplot.mat Screeplots of model results

analogue::summary.bootstrap.mat Summarise bootstrap resampling for MAT models

animation::boot.iid Bootstrapping the i.i.d data

ape::boot.phylo Tree Bipartition and Bootstrapping Phylogenies

aplpack::slider.bootstrap.lm.plot interactive bootstapping for lm

bnlearn::bn.boot Parametric and nonparametric bootstrap of Bayesian networks

bnlearn::boot.strength Bootstrap arc strength and direction

boot::nested.corr Functions for Bootstrap Practicals

boot::boot Bootstrap Resampling

boot::boot.array Bootstrap Resampling Arrays

boot::boot.ci Nonparametric Bootstrap Confidence Intervals

boot::cd4.nested Nested Bootstrap of cd4 data

boot::censboot Bootstrap for Censored Data

boot::freq.array Bootstrap Frequency Arrays

boot::jack.after.boot Jackknife-after-Bootstrap Plots

boot::linear.approx Linear Approximation of Bootstrap Replicates

boot::plot.boot Plots of the Output of a Bootstrap Simulation

boot::print.boot Print a Summary of a Bootstrap Object

boot::print.bootci Print Bootstrap Confidence Intervals

boot::saddle Saddlepoint Approximations for Bootstrap Statistics

boot::saddle.distn Saddlepoint Distribution Approximations for Bootstrap Statistics boot::tilt.boot Non-parametric Tilted Bootstrap

boot::tsboot Bootstrapping of Time Series

BootCL::BootCL.distribution Find the bootstrap distribution

BootCL::BootCL.plot Display the bootstrap distribution and p-value

BootPR::BootAfterBootPI Bootstrap-after-Bootstrap Prediction

BootPR::BootBC Bootstrap bias-corrected estimation and forecasting for AR models

BootPR::BootPI Bootstrap prediction intevals and point forecasts with no bias-correction

BootPR::BootPR-package Bootstrap Prediction Intervals and Bias-Corrected Forecasting

BootPR::ShamanStine.PI Bootstrap prediction interval using Shaman and Stine bias formula

bootRes::bootRes-package The bootRes Package for Bootstrapped Response and Correlation Functions

bootRes::dendroclim Calculation of bootstrapped response and correlation functions.

bootspecdens::specdens Bootstrap for testing equality of spectral densities

bootStepAIC::boot.stepAIC Bootstraps the Stepwise Algorithm of stepAIC() for Choosing a Model by AIC

bootstrap::bootpred Bootstrap Estimates of Prediction Error

bootstrap::bootstrap Non-Parametric Bootstrapping

bootstrap::boott Bootstrap-t Confidence Limits

bootstrap::ctsub Internal functions of package bootstrap

bootstrap::lutenhorm Luteinizing Hormone

bootstrap::scor Open/Closed Book Examination Data

bootstrap::spatial Spatial Test Data

BSagri::BOOTSimpsonD Simultaneous confidence intervals for Simpson indices

cfa::bcfa Bootstrap-CFA

ChainLadder::BootChainLadder Bootstrap-Chain-Ladder Model

CircStats::vm.bootstrap.ci Bootstrap Confidence Intervals

circular::mle.vonmises.bootstrap.ci Bootstrap Confidence Intervals

clue::cl_boot Bootstrap Resampling of Clustering Algorithms

CORREP::cor.bootci Bootstrap Confidence Interval for Multivariate Correlation

Daim::Daim.data1 Data set: Artificial bootstrap data for use with Daim

DCluster::achisq.boot Bootstrap replicates of Pearson's Chi-square statistic

DCluster::besagnewell.boot Generate boostrap replicates of Besag and Newell's statistic

DCluster::gearyc.boot Generate bootstrap replicates of Moran's I autocorrelation statistic

DCluster::kullnagar.boot Generate bootstrap replicates of Kulldorff and Nagarwalla's statistic

DCluster::moranI.boot Generate bootstrap replicates of Moran's I autocorrelation statistic

DCluster::pottwhitt.boot Bootstrap replicates of Potthoff-Whittinghill's statistic

DCluster::stone.boot Generate boostrap replicates of Stone's statistic

DCluster::tango.boot Generate bootstrap replicated of Tango's statistic

DCluster::whittermore.boot Generate bootstrap replicates of Whittermore's statistic

degreenet::rplnmle Rounded Poisson Lognormal Modeling of Discrete Data

degreenet::bsdp Calculate Bootstrap Estimates and Confidence Intervals for the Discrete Pareto Distribution

degreenet::bsnb Calculate Bootstrap Estimates and Confidence Intervals for the Negative Binomial Distribution

degreenet::bspln Calculate Bootstrap Estimates and Confidence Intervals for the Poisson Lognormal Distribution

degreenet::bswar Calculate Bootstrap Estimates and Confidence Intervals for the Waring Distribution degreenet::bsyule Calculate Bootstrap Estimates and Confidence Intervals for the Yule Distribution

degreenet::degreenet-internal Internal degreenet Objects

delt::eval.bagg Returns a bootstrap aggregation of adaptive histograms

delt::lstseq.bagg Calculates a scale of bootstrap aggregated histograms

depmix::depmix Fitting Dependent Mixture Models

Design::anova.Design Analysis of Variance (Wald and F Statistics)

Design::bootcov Bootstrap Covariance and Distribution for Regression Coefficients

Design::calibrate Resampling Model Calibration

Design::predab.resample Predictive Ability using Resampling

Design::rm.impute Imputation of Repeated Measures

Design::validate Resampling Validation of a Fitted Model's Indexes of Fit

Design::validate.cph Validation of a Fitted Cox or Parametric Survival Model's Indexes of Fit

Design::validate.lrm Resampling Validation of a Logistic Model

Design::validate.ols Validation of an Ordinary Linear Model

dynCorr::bootstrapCI Bootstrap Confidence Interval

dynCorr::dynCorrData An example dataset for use in the example calls in the help files for the dynamicCorrelation and bootstrapCI functions

e1071::bootstrap.lca Bootstrap Samples of LCA Results

eba::boot Bootstrap for Elimination-By-Aspects (EBA) Models

EffectiveDose::Boot.CI Bootstrap confidence intervals for ED levels

EffectiveDose::EffectiveDose-package Estimation of the Effective Dose including Bootstrap confindence intervals

el.convex::samp sample from bootstrap

equate::se.boot Bootstrap Standard Errors of Equating

equivalence::equiv.boot Regression-based TOST using bootstrap

extRemes::boot.sequence Bootstrap a sequence.

FactoMineR::simule Simulate by bootstrap

FGN::Boot Generic Bootstrap Function

FitAR::Boot Generic Bootstrap Function

FitAR::Boot.ts Parametric Time Series Bootstrap

fitdistrplus::bootdist Bootstrap simulation of uncertainty for non-censored data

fitdistrplus::bootdistcens Bootstrap simulation of uncertainty for censored data

flexclust::bootFlexclust Bootstrap Flexclust Algorithms

fossil::bootstrap Bootstrap Species Richness Estimator

fractal::surrogate Surrogate data generation

FRB::FRBmultiregGS GS-Estimates for multivariate regression with bootstrap confidence intervals

FRB::FRBmultiregMM MM-Estimates for Multivariate Regression with Bootstrap Inference

FRB::FRBmultiregS S-Estimates for Multivariate Regression with Bootstrap Inference

FRB::FRBpcaMM PCA based on Multivariate MM-estimators with Fast and Robust Bootstrap

FRB::FRBpcaS PCA based on Multivariate S-estimators with Fast and Robust Bootstrap

FRB::GSboot_multireg Fast and Robust Bootstrap for GS-Estimates

FRB::MMboot_loccov Fast and Robust Bootstrap for MM-estimates of Location and Covariance

FRB::MMboot_multireg Fast and Robust Bootstrap for MM-Estimates of Multivariate Regression

FRB::MMboot_twosample Fast and Robust Bootstrap for Two-Sample MM-estimates of Location and Covariance

FRB::Sboot_loccov Fast and Robust Bootstrap for S-estimates of location/covariance

FRB::Sboot_multireg Fast and Robust Bootstrap for S-Estimates of Multivariate Regression

FRB::Sboot_twosample Fast and Robust Bootstrap for Two-Sample S-estimates of Location and Covariance

ftsa::fbootstrap Bootstrap independent and identically distributed functional data

gmvalid::gm.boot.coco Graphical model validation using the bootstrap (CoCo).

gmvalid::gm.boot.mim Graphical model validation using the bootstrap (MIM)

gPdtest::gPd.test Bootstrap goodness-of-fit test for the generalized Pareto distribution

hierfstat::boot.vc Bootstrap confidence intervals for variance components

Hmisc::areg Additive Regression with Optimal Transformations on Both Sides using Canonical Variates

Hmisc::aregImpute Multiple Imputation using Additive Regression, Bootstrapping, and Predictive Mean Matching

Hmisc::bootkm Bootstrap Kaplan-Meier Estimates

Hmisc::find.matches Find Close Matches

Hmisc::rm.boot Bootstrap Repeated Measurements Model

Hmisc::smean.cl.normal Compute Summary Statistics on a Vector

Hmisc::transace Additive Regression and Transformations using ace or avas

Hmisc::transcan Transformations/Imputations using Canonical Variates

homtest::HOMTESTS Homogeneity tests

hopach::boot2fuzzy function to write MapleTree files for viewing bootstrap estimated cluster membership probabilities based on hopach clustering results

hopach::bootplot function to make a barplot of bootstrap estimated cluster membership probabilities

hopach::boothopach functions to perform non-parametric bootstrap resampling of hopach clustering results

ICEinfer::ICEcolor Compute Preference Colors for Outcomes in a Bootstrap ICE Scatter within a Confidence Wedge

ICEinfer::ICEuncrt Compute Bootstrap Distribution of ICE Uncertainty for given Shadow Price of Health, lambda

ICEinfer::plot.ICEcolor Add Economic Preference Colors to Bootstrap Uncertainty Scatters within a Confidence Wedge

ICEinfer::plot.ICEuncrt Display Scatter for a possibly Transformed Bootstrap Distribution of ICE Uncertainty

ICEinfer::print.ICEuncrt Summary Statistics for a possibly Transformed Bootstrap Distribution of ICE Uncertainty

ipred::bootest Bootstrap Error Rate Estimators

maanova::consensus Build consensus tree out of bootstrap cluster result

Matching::ks.boot Bootstrap Kolmogorov-Smirnov

MBESS::ci.reliability.bs Bootstrap the confidence interval for reliability coefficient

MCE::RProj The bootstrap-then-group implementation of the Bootstrap Grouping Prediction Plot for estimating R.

MCE::groupbootMCE The group-then-bootstrap implementation of the Bootstrap Grouping Prediction Plot for estimating MCE

MCE::groupbootR The group-then-bootstrap implementation of the Bootstrap Grouping Prediction Plot for estimating R

MCE::jackafterboot Jackknife-After-Bootstrap Method of MCE estimation

MCE::mceBoot Bootstrap-After-Bootstrap estimate of MCE

MCE::mceProj The bootstrap-then-group implementation of the Bootstrap Grouping Prediction Plot for estimating MCE.

meboot::meboot Generate Maximum Entropy Bootstrapped Time Series Ensemble

meboot::meboot.default Generate Maximum Entropy Bootstrapped Time Series Ensemble

meboot::meboot.pdata.frame Maximum Entropy Bootstrap for Panel Time Series Data meifly::lmboot Bootstrap linear models

mixreg::bootcomp Perform a bootstrap test for the number of components in a mixture of regressions.

mixstock::genboot Generate bootstrap estimates of mixed stock analyses

mixstock::mixstock.boot Bootstrap samples of mixed stock analysis data

mixtools::boot.comp Performs Parametric Bootstrap for Sequentially Testing the Number of Components in Various Mixture Models

mixtools::boot.se Performs Parametric Bootstrap for Standard Error Approximation

MLDS::simu.6pt Perform Bootstrap Test on 6-point Likelihood for MLDS FIT

MLDS::summary.mlds.bt Method to Extract Bootstrap Values for MLDS Scale Values

msm::boot.msm Bootstrap resampling for multi-state models

mstate::msboot Bootstrap function in multi-state models

multtest::boot.null Non-parametric bootstrap resampling function in package ′multtest′

ncf::mSynch the mean (cross-)correlation (with bootstrap CI) for a panel of spatiotemporal data

nFactors::eigenBootParallel Bootstrapping of the Eigenvalues From a Data Frame

nlstools::nlsBoot Bootstrap resampling

np::b.star Compute Optimal Block Length for Stationary and Circular Bootstrap

nsRFA::HOMTESTS Homogeneity tests

Oncotree::bootstrap.oncotree Bootstrap an oncogenetic tree to assess stability

ouch::browntree Fitted phylogenetic Brownian motion model

ouch::hansentree-methods Methods of the ″hansentree″ class

pARccs::Boot_CI Bootstrap confidence intervals for (partial) attributable risks (AR and PAR) from case-control data

PCS::PdCSGt.bootstrap.NP2 Non-parametric bootstrap for computing G-best and d-best PCS

PCS::PdofCSGt.bootstrap5 Parametric bootstrap for computing G-best and d-best PCS

PCS::PofCSLt.bootstrap5 Parametric bootstrap for computing L-best PCS

peperr::complexity.ipec.CoxBoost Interface function for complexity selection for CoxBoost via integrated prediction error curve and the bootstrap

peperr::complexity.ipec.rsf_mtry Interface function for complexity selection for random survival forest via integrated prediction error curve and the bootstrap

pgirmess::difshannonbio Empirical confidence interval of the bootstrap of the difference between two Shannon indices

pgirmess::piankabioboot Bootstrap Pianka’s index

pgirmess::shannonbioboot Boostrap Shannon’s and equitability indices

phangorn::bootstrap.pml Bootstrap

phybase::bootstrap Bootstrap sequences

phybase::bootstrap.mulgene Bootstrap sequences from multiple loci

popbio::boot.transitions Bootstrap observed census transitions

popbio::countCDFxt Count-based extinction probabilities and bootstrap confidence intervals

prabclus::abundtest Parametric bootstrap test for clustering in abundance matrices

prabclus::prabtest Parametric bootstrap test for clustering in presence-absence matrices

pvclust::msfit Curve Fitting for Multiscale Bootstrap Resampling

qgen::dis Bootstrap confidence intervals

qpcR::calib2 Calculation of qPCR efficiency by dilution curve analysis and bootstrapping of dilution curve replicates

qpcR::pcrboot Bootstrapping and jackknifing qPCR data

qtl::plot.scanoneboot Plot results of bootstrap for QTL position

qtl::scanoneboot Bootstrap to get interval estimate of QTL location qtl::summary.scanoneboot Bootstrap confidence interval for QTL location

QuantPsyc::distInd.ef Complex Mediation for use in Bootstrapping

QuantPsyc::proxInd.ef Simple Mediation for use in Bootstrapping

quantreg::boot.crq Bootstrapping Censored Quantile Regression

quantreg::boot.rq Bootstrapping Quantile Regression

r4ss::SS_splitdat Split apart bootstrap data to make input file.

relaimpo::boot.relimp Functions to Bootstrap Relative Importance Metrics

ResearchMethods::bootSequence A demonstration of how bootstrapping works, taking multiple bootstrap samples and watching how the means of those samples begin to normalize.

ResearchMethods::bootSingle A demonstration of how bootstrapping works step by step for one function.

rms::anova.rms Analysis of Variance (Wald and F Statistics)

rms::bootcov Bootstrap Covariance and Distribution for Regression Coefficients

rms::calibrate Resampling Model Calibration

rms::predab.resample Predictive Ability using Resampling

rms::validate Resampling Validation of a Fitted Model’s Indexes of Fit

rms::validate.cph Validation of a Fitted Cox or Parametric Survival Model’s Indexes of Fit

rms::validate.lrm Resampling Validation of a Logistic Model

rms::validate.ols Validation of an Ordinary Linear Model

robust::rb Robust Bootstrap Standard Errors

rqmcmb2::rqmcmb Markov Chain Marginal Bootstrap for Quantile Regression

sac::BootsChapt Bootstrap (Permutation) Test of Change-Point(s) with One-Change or Epidemic Alternative

sac::BootsModelTest Bootstrap Test of the Validity of the Semiparametric Change-Point Model

SAFD::btest.mean One-sample bootstrap test for the mean of a FRV

SAFD::btest2.mean Two-sample bootstrap test on the equality of mean of two FRVs

SAFD::btestk.mean Multi-sample bootstrap test for the equality of the mean of FRVs

scaleboot::sboptions Options for Multiscale Bootstrap

scaleboot::plot.scaleboot Plot Diagnostics for Multiscale Bootstrap

scaleboot::sbconf Bootstrap Confidence Intervals

scaleboot::sbfit Fitting Models to Bootstrap Probabilities

scaleboot::scaleboot-package Approximately Unbiased P-values via Multiscale Bootstrap

scaleboot::scaleboot Multiscale Bootstrap Resampling

scaleboot::summary.scaleboot P-value Calculation for Multiscale Bootstrap

sem::boot.sem Bootstrap a Structural Equation Model

shapes::resampletest Tests for mean shape difference using complex arithmetic, including bootstrap and permutation tests.

shapes::iglogl Internal function(s)

shapes::testmeanshapes Tests for mean shape difference, including permutation and bootstrap tests

simpleboot::hist.simpleboot Histograms for bootstrap sampling distributions.

simpleboot::lm.boot Linear model bootstrap.

simpleboot::summary.lm.simpleboot Methods for linear model bootstrap.

simpleboot::loess.boot 2-D Loess bootstrap.

simpleboot::fitted.loess.simpleboot Methods for loess bootstrap.

simpleboot::one.boot One sample bootstrap of a univariate statistic.

simpleboot::pairs.boot Two sample bootstrap.

simpleboot::perc Extract percentiles from a bootstrap sampling distribution.

simpleboot::plot.lm.simpleboot Plot method for linear model bootstraps.

simpleboot::plot.loess.simpleboot Plot method for loess bootstraps.

simpleboot::samples Extract sampling distributions from bootstrapped linear/loess models.

simpleboot::two.boot Two sample bootstrap of differences between univariate statistics.

sm::sm.binomial.bootstrap Bootstrap goodness-of-fit test for a logistic regression model.

sm::sm.poisson.bootstrap Bootstrap goodness-of-fit test for a Poisson regression model

spls::ci.spls Calculate bootstrapped confidence intervals of SPLS coefficients

spls::correct.spls Correct the initial SPLS coefficient estimates based on bootstrapped confidence intervals

Stem::covariates2 Stem internal objects

Stem::Stem.Bootstrap Parametric bootstrap

survey::bootweights Compute survey bootstrap weights

tractor.base::angleBetweenVectors Undocumented functions

TSA::arima.boot Compute the Bootstrap Estimates of an ARIMA Model

tsDyn::TVAR.sim Simulation and bootstrap of multivariate Threshold Autoregressive model

tsDyn::TVECM.sim Simulation and bootstrap of bivariate VECM/TVECM

tsDyn::extendBoot extension of the bootstrap replications

tsDyn::setar.sim Simulation and bootstrap of Threshold Autoregressive model

tseries::tsbootstrap Bootstrap for General Stationary Data

ttrTests::bootstrap Generates a Bootstrap Sample from Raw Data

ttrTests::generateSample Generates a Bootstrap Sample from Price Data

TWIX::bootTWIX Bootstrap of the TWIX trees

UsingR::cfb Bootstrap sample from the Survey of Consumer Finances

varSelRF::varSelRFBoot Bootstrap the variable selection procedure in varSelRF

```
   vegetarian::bootstrap Estimates
Uncertainties with Bootstrapping
   verification::table.stats.boot Percentile
bootstrap for 2 by 2 table
   vrtest::AutoBoot.test Wild Bootstrapping of
Automatic Variance Ratio Test
   vrtest::Boot.test Bootstrap Variance Ratio
Tests
   waveslim::dwpt.boot Bootstrap Time Series
Using the DWPT
   wmtsa::wavBootstrap Adaptive wavelet-based
bootstrapping
   wmtsa::wavDWPTWhitest Seeks the whitest
transform of a discrete wavelet packet transform
(DWPT)
```

The part of the name before "`::`" is the "package" name (class), which installs a library that has the "`function`" object whose name follows the "`::`." The above list should indicate both the breadth of applications of the bootstrap and the breadth of its implementation in the R system.

R comes with some basic packages preinstalled. Most special application packages have to be downloaded by the user via the menu line command Packages | Install Packages. This makes the chosen packages(s) part of the R software on your computer. To actually bring the package into use in your environment, you will also need the require() or library() functions. Two packages of note related to bootstrapping are the "`bootstrap`" package, which is documented by the book *An Introduction to the Bootstrap* by B. Efron and R. J. Tibshirani, and the "`boot`" package, which is documented by *Bootstrap Methods and Their Application* by A. C. Davison and D. V. Hinkley. For example, you can require the "`boot`" library by

```
> require('boot')
Loading required package: boot
```

R is a vectorized and object-oriented language. Most operations are most efficient when done as vector operations instead of on individual elements. For example,

```
> x<- 1:10
> y<- 21:30
> x
[1] 1 2 3 4 5 6 7 8 9 10
> y
[1] 21 22 23 24 25 26 27 28 29 30
> x+y
[1] 22 24 26 28 30 32 34 36 38 40
```

```
    > x/y
    [1] 0.04761905 0.09090909 0.13043478
0.16666667 0.20000000 0.23076923 0.25925926
0.28571429 0.31034483 0.33333333
    > x*y
    [1] 21 44 69 96 125 156 189 224 261 300
    > sqrt(x)
    [1] 1.000000 1.414214 1.732051 2.000000
2.236068 2.449490 2.645751 2.828427 3.000000
3.162278
    > exp(x)
    [1] 2.718282 7.389056 20.085537 54.598150
148.413159 403.428793 1096.633158 2980.957987
8103.083928
    [10] 22026.465795
    > x[2]
    [1] 2
    > x[2]+y[3]
    [1] 25
```

Note that individual elements are indicated by subscripts within brackets "`[]`," and "`n:m`" is shorthand for the vector whose elements are the sequence of integers from *n* to *m*.

One function in the basic R packages that lies at the heart of resampling is the sample() function, whose syntax is

```
sample(x, size, replace = FALSE, prob = NULL)
```

The first argument "`x`" is the vector of data, that is, the original sample. "`size`" is the size of the resample desired. "`replace`" is "`TRUE`" if resampling is with replacement, and "`FALSE`" if not (the default). "`prob`" is a vector of probability weights if the equal-weight default is not used. Any arguments omitted will assume the default. If "`size`" is omitted, it will default to the length of "`x`."

For our purposes, it will usually be easiest to resample the indices of the data from a sample of size *n*, rather than the data itself. For example, if we have five data in our set, say

```
> x<- c(-0.3, 0.5, 2.6, 1.0, -0.9)
> x
[1] -0.3 0.5 2.6 1.0 -0.9
then
> i<- sample(1:5, 5, replace=TRUE)
> i
[1] 3 2 3 2 2
```

```
> x[i]
[1] 2.6 0.5 2.6 0.5 0.5
```

is the resample of the original data.

As we move through the text, more features of R related to the bootstrap will be illustrated in context, as we need them.

1.5 HISTORICAL NOTES

Bootstrap research began in earnest in the late 1970s, although some key developments can be traced back to earlier times. The theory took off in the early 1980s after Efron (1979). The first proofs of the consistency of the bootstrap estimate of the sample mean came in 1981 with the papers of Singh (1981) and Bickel and Freedman (1981).

The significance of Efron (1979) is best expressed in Davison and Hinkley (1997) who wrote " The publication in 1979 of Bradley Efron's first article on bootstrap methods was a major event in Statistics, at once synthesizing some of the earlier resampling ideas and establishing a new framework for simulation-based statistical analysis. The idea of replacing complicated and often inaccurate approximations to biases, variances and other measures of uncertainty by computer simulation caught the imagination of both theoretical researchers and users of statistical methods."

Regarding the precursors of the bootstrap, Efron pointed out some of the early work of R. A. Fisher (in the 1920s on maximum likelihood estimation) as the inspiration for many of the basic ideas. The jackknife was introduced by Quenouille (1949) and popularized by Tukey (1958). Miller (1974) provided an excellent review of jackknife methods. Extensive coverage of the jackknife as developed up to 1972 can be found in Gray and Schucany (1972).

As noted earlier, Bickel and Freedman (1981) and Singh (1981) were the first to show consistency of the bootstrap estimate of the sample mean under certain regularity conditions. In their paper, Bickel and Freedman (1981) also provided an example where the bootstrap estimate is not consistent. Gine and Zinn (1989) provided necessary but not sufficient conditions for the consistency of the bootstrap mean.

Athreya (1987a, b), Knight (1989), and Angus (1993) all provided examples where the bootstrap fails due to the fact that the necessary conditions were not satisfied. In some of these cases the inconsistency is shown by deriving the actual limiting distribution for the bootstrap estimator normalized and by showing that it is not degenerate but differs from the limiting distribution for the original parameter estimate.

Subsampling methods began with Hartigan (1969, 1971, 1975) and McCarthy (1969). Diaconis and Holmes (1994) showed how the Monte Carlo approximation can sometimes be avoided through the use of Gray codes.

Efron (1983) compared several variations with the bootstrap estimate when estimating classification error rates for linear discriminant functions. Other papers that showed through simulation the advantage of the bootstrap 632 estimate include Chernick et al. (1985, 1986, 1988a, b). For the Pearson VII family, the 632 is not always the best bootstrap estimator when the parameter controlling the tail behavior increases and the first moment no longer exists (Chernick et al., 1988b). Other related papers include Chatterjee and Chatterjee (1983), McLachlan (1980), Snapinn and Knoke (1984, 1985a, b, 1988), Jain et al. (1987), and Efron and Tibshirani (1997). Many other references for the historical development of the bootstrap can be found in Chernick (2007) and Chernick and LaBudde (2010).

1.6 EXERCISES

1. Suppose three mice who are littermates have weights 82, 107, and 93 g.
   (a) What is the mean weight of the mice?
   (b) How many possible bootstrap samples of this sample are there?
   (c) List all of the possible bootstrap samples as triples.
   (d) Compute the mean of each bootstrap sample.
   (e) Compute the mean of the resample means. How does this compare with the original sample mean?
   (f) What are the high and low values of the resample means?

2. Suppose three mice in Exercise 1.6.1 have maze transit times of 27, 36, and 22 s.
   (a) What is the mean transit time of the mice?
   (b) How many possible bootstrap samples of this sample are there?
   (c) List all of the possible bootstrap samples as triples.
   (d) Compute the mean of each bootstrap resample.
   (e) Compute the mean of the resample means. How does this compare with the original sample mean?
   (f) What are the high and low values of the resample means?

3. Install the R system on your computer, and install the package "`bootstrap`." Then enter the following commands at the prompt:

   ```
   >require('bootstrap')
   >help('bootstrap')
   ```

   Note the webpage that appears in your browser.

4. Install the R system on your computer, and install the package "`boot`." Then enter the following commands at the prompt:

   ```
   >require('boot')
   >help('boot')
   ```

   Note the webpage that appears in your browser.

5. Aflatoxin residues in peanut butter: In actual testing, 12 lots of peanut butter had aflatoxin residues in parts per billion of 4.94, 5.06, 4.53, 5.07, 4.99, 5.16, 4.38, 4.43, 4.93, 4.72, 4.92, and 4.96.
   (a) How many possible bootstrap resamples of these data are there?
   (b) Using R and the sample() function, or a random number table or generator, generate five resamples of the integers from 1 to 12.
   (c) For each of the resamples in Exercise 1.6.5b, find the mean of the corresponding elements of the aflatoxin data set.

(d) Find the mean of the resample means. Compare this with the mean of the original data set.

(e) Find the minimum and the maximum of the five resample means. This a crude bootstrap confidence interval on the mean. (If you had used 1000 resamples, and used the 25th and 975th largest means, this would have given a reasonable 95% confidence interval.)

6. Sharpness of a razor blade: In a filament cut test, a razor blade was tested six different times with ultimate forces corresponding to 8.5, 13.9, 7.4, 10.3, 15.7, and 4.0 g.

(a) How many possible bootstrap resamples of these data are there?

(b) Using R and the sample() function, or a random number table or generator, generate 10 resamples of the integers from 1 to 6.

(c) For each of the resamples in Exercise 1.6.6b, find the mean of the corresponding elements of the sharpness data set.

(d) Find the mean of the resample means. Compare this with the mean of the original data set.

(e) Find the minimum and maximum of the 10 resample means. This is a crude bootstrap confidence interval on the mean. (If you had used 1000 resamples, and used the 25th and 975th largest means, this would have given a reasonable 95% confidence interval.)

REFERENCES

Angus, J. E. (1993). Asymptotic theory for bootstrapping extremes. Commun. Stat. Theory Methods 22, 15–30.

Arcones, M. A., and Gine, E. (1989). The bootstrap of the mean with arbitrary bootstrap sample size. Ann. Inst. Henri Poincare 25, 457–481.

Athreya, K. B. (1987a). Bootstrap of the mean in the infinite variance case. In Proceedings of the First World Congress of the Bernoulli Society (Y. Prohorov, and V. Sazonov, eds.), Vol. 2, pp. 95–98. VNU Science Press, The Netherlands.

Athreya, K. B. (1987b). Bootstrap estimate of the mean in the infinite variance case. Ann. Stat. 15, 724–731.

Babu, G. J., and Feigelson, E. (1996). Astrostatistics. Chapman & Hall, New York.

Beran, R. J. (1982). Estimated sampling distributions: The bootstrap and competitors. Ann. Stat. 10, 212–225.

Bickel, P. J., and Freedman, D. A. (1981). Some asymptotic theory for the bootstrap. Ann. Stat. 9, 1196–1217.

Bickel, P. J., and Freedman, D. A. (1984). Asymptotic normality and the bootstrap in stratified sampling. Ann. Stat. 12, 470–482.

Bickel, P. J., and Ren, J.-J. (1996). The *m* out of *n* bootstrap and goodness fit of tests with populations. Comput. Math. Appl. 15, 29–37.

Bickel, P. J., Goetze, F., and van Zwet, W. R. (1997). Resampling fewer than *n* observations, gains, losses, and remedies for losses. Stat. Sin. 7, 1–32.

Brieman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. (1984). Classification and Regression Trees. Wadsworth, Belmont.

Chatterjee, S., and Chatterjee S. (1983). Estimation of misclassification probabilities by bootstrap methods. Commun. Statist. Simul. Comput. 11, 645–656.

Chernick, M. R. (2007). Bootstrap Methods: A Guide for Practitioners and Researchers, Second Edition, Wiley, Hoboken.

Chernick, M. R., Murthy V. K., and Nealy, C. D. (1985). Applications of bootstrap and other resampling techniques: Evaluation of classifier performance. Pattern Recog. Lett. 3, 167–178.

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

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.

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. Appl. 15, 897–902.

Chernick, M. R. and LaBudde, R. A. (2010). Revisiting qualms about bootstrap confidence intervals. Am. J. Manage. Sci. 29, 437–456.

Davison, A. C., and Hinkley, D. V. (1997). Bootstrap Methods and Their Applications. Cambridge University Press, Cambridge.

Diaconis, P., and Efron, B. (1983). Computer-intensive methods in statistics. Sci. Am. 248, 116–130.

Diaconis, P., and Holmes, S. (1994). Gray codes for randomization procedures. Stat. Comput. 4, 287–302.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. Ann. Stat. 7, 1–26.

Efron, B. (1982). The Jackknife, the Bootstrap and Other Resampling Plans. SIAM, Philadelphia.

Efron, B. (1983). Estimating the error rate of a prediction rule: Improvements on cross-validation. J. Am. Stat. Assoc. 78, 316–331.

Efron, B., and Gong, G. (1983). A leisurely look at the bootstrap, the jackknife and cross-validation. Am. Stat. 37, 36–48.

Efron, B., and Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals and other measures of statistical accuracy. Stat. Sci. 1, 54–77.

Efron, B., and Tibshirani, R. (1997). Improvements on cross-validation: The 632+ rule. J. Am. Stat. Assoc. 92, 548–560.

Fisher, R. A. (1935). The Design of Experiments. Hafner, New York.

Fukuchi, J. I. (1994). Bootstrapping extremes of random variables. PhD dissertation, Iowa State University, Ames.

Gine, E., and Zinn, J. (1989). Necessary conditions for bootstrap of the mean. Ann. Stat. 17, 684–691.

Gong, G. (1986). Cross-validation, the jackknife, and the bootstrap: Excess error in forward logistic regression. J. Am. Statist. Assoc. 81, 108–113.

Good, P. I. (1994). Permutation Tests. Springer-Verlag, New York.

Gray, H. L., and Schucany, W. R. (1972). The Generalized Jackknife Statistic. Marcel Dekker, New York.

Hall, P. (1986). On the bootstrap and confidence intervals. Ann. Stat. 14, 1431–1452.

Hall, P. (1988). Theoretical comparison of bootstrap confidence intervals (with discussion). Ann. Stat. 16, 927–985.

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

Hall, P., Hardle, W., and Simar, L. (1993). On the consistency of bootstrap distribution estimators. Comput. Stat. Data Anal. 16, 11–18.

Hall, P., and Martin, M. A. (1988). On bootstrap resampling and iteration. Biometrika 75, 661–671.

Hartigan, J. A. (1969). Using subsample values as typical values. J. Am. Stat. Assoc. 64, 1303–1317.

Hartigan, J. A. (1971). Error analysis by replaced samples. J. R. Stat. Soc. B 33, 98–113.

Hartigan, J. A. (1975). Necessary and sufficient conditions for the asymptotic joint normality of a statistic and its subsample values. Ann. Stat. 3, 573–580.

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

Knight, K. (1989). On the bootstrap of the sample mean in the infinite variance case. Ann. Stat. 17, 1168–1175.

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

Lahiri, S. N. (2003). Resampling Methods for Dependent Data. Springer-Verlag, New York.

Martin, M. A. (1990). On the bootstrap iteration for coverage correction in confidence intervals. J. Am. Stat. Assoc. 85, 1105–1118.

McCarthy, P. J. (1969). Pseudo-replication: Half-samples. Int. Stat. Rev. 37, 239–263.

McLachlan, G. J. (1980). The efficiency of Efron's bootstrap approach applied to error rate estimation in discriminant analysis. J. Stat. Comput. Simul. 11, 273–279.

Miller, R. G., Jr. (1974). The jackknife – A review. Biometrika, 61, 1–17.

Navidi, W. (1989). Edgeworth expansions for bootstrapping in regression models. Ann. Stat. 17, 1472–1478.

Pitman, E. J. G. (1937). Significance tests which may be applied to samples from any population. J. R. Stat. Soc. Suppl. 4, 119–130 and 225–232 (parts I and II).

Pitman, E. J. G. (1938). Significance tests which may be applied to samples from any population. Part III. The analysis of variance test. Biometrika 29, 322–335.

Politis, D. N., and Romano, J. P. (1994). Large sample confidence intervals based on subsamples under minimal assumptions. Ann. Stat. 22, 2031–2050.

Politis, D. N., Romano, J. P., and Wolf, M. (1999). Subsampling. Springer-Verlag, New York.

Quenouille, M. H. (1949). Approximate tests of correlation in time series. J. R. Stat. Soc. B 11, 18–84.

Shao, J., Kubler, J., and Pigeot, I. (2000). Consistency of the bootstrap procedure in individual bioequivalence. Biometrika 87, 573–585.

Simon, J. L. (1969). Basic Research Methods in Social Science. Random House, New York.

Simon, J. L., and Bruce, P. (1991). Resampling: A tool for Everyday Statistical Work. Chance 4, 22–32.

Simon, J. L., and Bruce, P. (1995). The new biostatistics of resampling. M. D. Comput. 12, 115–121.

Singh, K. (1981). On the asymptotic accuracy of Efron's bootstrap. Ann. Stat. 9, 1187–1195.

Snapinn, S. M., and Knoke, J. D. (1984). Classification error rate estimators evaluated by unconditional mean square error. Technometrics 26, 371–378.

Snapinn, S. M., and Knoke, J. D. (1985a). An evaluation of smoothed classification error rate estimators. Technometrics 27, 199–206.

Snapinn, S. M., and Knoke, J. D. (1985b). Improved classification error rate estimation. Bootstrap or smooth? Unpublished report.

Snapinn, S. M., and Knoke, J. D. (1988). Bootstrap and smoothed classification error rate estimates. Commun. Stat. Simul. Comput. 17, 1135–1153.

Tukey, J. W. (1958). Bias and confidence in not quite large samples (abstract). Ann. Math. Stat. 29, 614.