Jackknife and Bootstrap

G. Jogesh Babu

1 Introduction

The classical statistical methods of earlier sections concentrated mainly on the statistical properties of the estimators that have a simple closed form and which can be analyzed mathematically. Except for a few important but simple statistics, these methods involve often unrealistic model assumptions. It is often relatively simple to devise a statistic that measures the property of interest, but is almost always difficult or impossible to determine the distribution of that statistic. These limitations have been overcome in the last two decades of the 20th Century with advances in electronic computers. A class of computationally intensive procedures known as *resampling methods* provide inference on a wide range of statistics under very general conditions. Resampling methods involve constructing hypothetical 'populations' derived from the observations, each of which can be analyzed in the same way to see how the statistics depend on plausible random variations in the observations. Resampling the original data preserves whatever distributions are truly present, including selection effects such as truncation and censoring.

Perhaps the *half-sample method* is the oldest resampling method, where one repeatedly chooses at random half of the data points, and estimates the statistic for each resample. The inference on the parameter can be based on the histogram of the resampled statistics. It was used by Mahalanobis in 1946 under the name *interpenetrating samples*. An important variant is the Quenouille–Tukey *jackknife method*. For a dataset with n data points, one constructs exactly n hypothetical datasets each with $n-1$ points, each one omitting a different point. The most important of resampling methods is called the *bootstrap*. Bradley Efron introduced the bootstrap method, also known as resampling with replacement, in 1979. Here one generates a large number of datasets, each with n data points randomly drawn from the original data. The constraint is that each drawing is made from the entire dataset, so a simulated dataset is likely to miss some points and have duplicates or triplicates of others. Thus, bootstrap can be viewed as a *Monte Carlo method* to simulate from an existing data, without any assumption on the underlying population.

2 Jackknife

Jackknife method was introduced by Quenouille (1949) to estimate the bias of an estimator. The method is later shown to be useful in reducing the bias as well as in estimating the variance of an estimator. Let $\hat{\theta}_n$ be an estimator of θ based on n i.i.d. random vectors $X_1, \ldots, X_n$, i.e., $\hat{\theta}_n = f_n(X_1, \ldots, X_n)$, for some function f_n. Let

$$\hat{\theta}_{n,-i} = f_{n-1}(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$$

be the corresponding recomputed statistic based on all but the i-th observation. The jackknife estimator of bias $E(\hat{\theta}_n) - \theta$ is given by

$$bias_J = \frac{(n-1)}{n} \sum_{i=1}^{n} \big(\hat{\theta}_{n,-i} - \hat{\theta}_n\big). \tag{1}$$

Jackknife estimator θ_J of θ is given by

$$\theta_J = \hat{\theta}_n - bias_J = \frac{1}{n} \sum_{i=1}^{n} \big(n\hat{\theta}_n - (n-1)\hat{\theta}_{n,-i}\big). \tag{2}$$

Such a bias corrected estimator hopefully reduces the over all bias. The summands above

$$\theta_{n,i} = n\hat{\theta}_n - (n-1)\hat{\theta}_{n,-i}, \ i = 1, \dots, n$$

are called *pseudo-values.*

2.1 Bias Reduction

Jackknifing, indeed, helps in reducing bias of an estimator in many cases. Suppose the expected value of the estimator $\hat{\theta}_n$ is of the form

$$E(\hat{\theta}_n) = \theta + \frac{a}{n} + \frac{b}{n^2},$$

then clearly,

$$E(\hat{\theta}_{n,i}) = \theta - \frac{b}{n(n-1)}.$$

Consequently, the bias of the jackknife estimator is $E(\theta_J) - \theta = O(n^{-2})$, which is of lower order than the bias of $\hat{\theta}_n$.

2.2 Estimation of variance

In the case of the sample mean $\hat{\theta}_n = \bar{X}_n$, it is easy to check that the *pseudo-values* are simply,

$$\theta_{n,i} = n\hat{\theta}_n - (n-1)\hat{\theta}_{n,-i} = X_i, \ i = 1, \dots, n.$$

This provides motivation for the jackknife estimator of variance of $\hat{\theta}_n$,

$$\begin{aligned} var_J(\hat{\theta}_n) &= \frac{1}{n(n-1)} \sum_{i=1}^{n} (\theta_{n,i} - \theta_J)(\theta_{n,i} - \theta_J)' \\ &= \frac{n-1}{n} \sum_{i=1}^{n} (\hat{\theta}_{n,-i} - \bar{\theta}_n)(\hat{\theta}_{n,-i} - \bar{\theta}_n)', \end{aligned} \tag{3}$$

where $\bar{\theta}_n = \frac{1}{n} \sum_{i=1}^{n} \hat{\theta}_{n,-i}$. For most statistics, jackknife estimator of variance is consistent, i.e.,

$$Var_J(\hat{\theta}_n)/Var(\hat{\theta}_n) \to 1,$$

as $n \to \infty$ almost surely. In particular, this holds for a **smooth functional model**. To describe this, let the statistic of interest $\hat{\theta}_n$ based on n data points be defined by $H(\bar{Z}_n)$, where $\bar{Z}_n$ is the sample mean of random vectors $Z_1, \ldots, Z_n$ and H is continuously differentiable in a neighborhood of $E(\bar{Z}_n)$. Many commonly occurring statistics, including: Sample Means, Sample Variances, Central and Non-central t-statistics (with possibly non-normal populations), Sample Coefficient of Variation, Maximum Likelihood Estimators, Least Squares Estimators, Correlation Coefficients, Regression Coefficients, Smooth transforms of these statistics, fall under this model.

However, consistency does not always hold; for example the jackknife method fails for non-smooth statistics, such as the sample median. If $\hat{\theta}_n$ denotes the sample median in the univariate case, then in general,

$$Var_J(\hat{\theta}_n)/Var(\hat{\theta}_n) \to \left(\frac{1}{2}\chi_2^2\right)^2$$

in distribution, where χ_2^2 denotes a *chi-square* random variable with 2 degrees of freedom (see Efron 1982, §3.4). So in this case, the jackknife method does not lead to a consistent estimator of the variance. However, a resampling method called *bootstrap* discussed in the next section, would lead to a consistent estimator.

3 Bootstrap

Bootstrap resampling constructs datasets with n points (rather than $n - 1$for the jackknife) where each point was selected from the full dataset; that is, resampling with replacement. The importance of the bootstrap emerged during the 1980s when mathematical study demonstrated that it gives nearly optimal estimate of the distribution of many statistics under a wide range of circumstances. In several cases, the method yields better results than those obtained by the classical normal approximation theory. However, one should caution that bootstrap is not the solution for all problems. The theory developed in 1980s and 1990s, show that bootstrap fails in some 'non-smooth' situations. Hence, caution should be used and should resist the temptation to use the method inappropriately. Many of these methods work well in the case of **smooth functional model**. As described earlier, these estimators are smooth functions of sample mean of a random vectors. In view of this, the bootstrap method is first described here for special case of the sample mean.

3.1 Description of the bootstrap method

Let $\mathbf{X} = (X_1, \ldots, X_n)$ be data drawn from an unknown population distribution F. Suppose $\hat{\theta}_n$, based on data $\mathbf{X}$, is a good estimator of θ, a parameter of interest. The interest lies in assessing its accuracy in estimation. Determining the confidence intervals for θ requires knowledge of the sampling distribution G_n of $\hat{\theta}_n - \theta$, *i.e.* $G_n(x) = \mathrm{P}(\hat{\theta}_n - \theta \leq x)$, for all x.

For example, the sample mean $\bar{X}_n = n^{-1} \sum_{i=1}^{n} X_i$ is a good estimator of the population mean μ. To get the confidence interval for μ, we must find the sampling distribution of $\bar{X}_n - \mu$, which depends on the shape and other characteristics of the unknown distribution F.

Classical statistical theory uses the Central Limit Theorem (normal approximation) to the sampling distribution. Even if the sampling distribution is not symmetric, the central limit theorem gives an approximation by a symmetric normal distribution. This can be seen from the following example. If $(X_1, Y_1), \ldots, (X_n, Y_n)$ denote observations from a bivariate normal population, then the maximum likelihood estimator of the correlation coefficient ρ is given by Pearson's correlation coefficient,

$$\hat{\rho}_n = \frac{\sum_{i=1}^{n}(X_i Y_i - \bar{X}_n \bar{Y}_n)}{\sqrt{\left(\sum_{i=1}^{n}(X_i - \bar{X}_n)^2\right)\left(\sum_{i=1}^{n}(Y_i - \bar{Y}_n)^2\right)}}.$$

For statistics with asymmetrical distributions, such as that of $\hat{\rho}_n$, the classical theory suggests variable transformations. In this case, *Fisher's Z-transformation* Z given by

$$Z = \frac{\sqrt{(n-3)}}{2}\left(\ln\left(\frac{1+\hat{\rho}_n}{1-\hat{\rho}_n}\right) - \ln\left(\frac{1+\rho}{1-\rho}\right)\right)$$

gives a better normal approximation. This approximation corrects skewness and is better than the normal approximation of $\sqrt{n}(\hat{\rho}_n - \rho)$. The bootstrap method, when properly used, avoids such individual transformations by taking into account the skewness of the sampling distribution. It automatically corrects for skewness.

The bootstrap method presumes that if $\hat{F}_n$ is a good approximation to the unknown population distribution F, then the behavior of the samples from $\hat{F}_n$ closely resemble that of the original data. Here $\hat{F}_n$ can be the *empirical distribution function* (EDF, or a smoothed EDF) of the data $X_1, \ldots, X_n$, or a parametric estimator of the function F. Once $\hat{F}_n$ is provided, datasets $\mathbf{X}^* = (X_1^*, \ldots, X_n^*)$ are resampled from $\hat{F}_n$ and the statistic θ^* based on $\mathbf{X}^*$ is computed for each resample. Under very general conditions Babu & Singh (1984) have shown that the difference between the sampling distribution G_n of $\hat{\theta}_n - \theta$ and the 'bootstrap distribution' G_b [*i.e.* the distribution of $\theta^* - \hat{\theta}_n$] is negligible. G_b can be used to draw inferences about the parameter θ in place of the unknown G_n. In principle, Bootstrap distribution (Histogram) G_b is completely known, as it is constructed entirely from the original data. However, to get the complete bootstrap distribution, one needs to compute the statistics for nearly all of $M = n^n$ possible bootstrap samples. For the simple example of sample mean, presumably one needs to compute,

$$\begin{array}{ll} X_1^{*(1)}, \ldots, X_n^{*(1)}, & r_1 = \bar{X}^{*(1)} - \bar{X} \\ X_1^{*(2)}, \ldots, X_n^{*(2)}, & r_2 = \bar{X}^{*(2)} - \bar{X} \\ \ddots \quad \ddots\ddots & \ddots \\ X_1^{*(M)}, \ldots, X_n^{*(M)}, & r_M = \bar{X}^{*(M)} - \bar{X}. \end{array}$$

The bootstrap distribution is given by the histogram of $r_1, \ldots, r_M$. Even for $n = 10$ data points, M turns out to be ten billion. In practice, the statistic of interest, $\theta^* - \hat{\theta}_n$, is computed for a number N (say $N = n(\log n)^2$) of resamples, and its histogram is constructed. Asymptotic theory shows that the sampling distribution of $\theta^* - \hat{\theta}_n$, can be well-approximated by genrating $N \simeq n(\ln n)^2$ bootstrap resamples (Babu & Singh 1983). Thus, only $N \sim 50$ simulations are needed for $n = 10$ and $N \sim 50,000$ for $n = 1000$. The

Table 1: Statistics and their bootstrap versions

Statistic	Bootstrap Version
Mean, $\bar{X}_n$	$\bar{X}_n^*$
Variance, $\frac{1}{n}\sum_{i=1}^n (X_i - \bar{X}_n)^2$	$\frac{1}{n}\sum_{i=1}^n (X_i^* - \bar{X}_n^*)^2$
Ratio estimator, $\bar{X}_n/\bar{Y}_n$	$\bar{X}_n^*/\bar{Y}_n^*$
Correlation coefficient, $\frac{\sum_{i=1}^n (X_i Y_i - \bar{X}_n \bar{Y}_n)}{\sqrt{(\sum_{i=1}^n (X_i - \bar{X}_n)^2)(\sum_{i=1}^n (Y_i - \bar{Y}_n)^2)}}$	$\frac{\sum_{i=1}^n (X_i^* Y_i^* - \bar{X}_n^* \bar{Y}_n^*)}{\sqrt{(\sum_{i=1}^n (X_i^* - \bar{X}_n^*)^2)(\sum_{i=1}^n (Y_i^* - \bar{Y}_n^*)^2)}}$

distribution of the estimator for the original dataset is obtained from the histogram of the estimators obtained from the bootstrapped samples.

The most popular and simple bootstrap is the *nonparametric bootstrap*, where the resampling with replacement is based on the EDF of the original data. This gives equal weights to each of the original data points. Table 1 gives bootstrap versions of some commonly used statistics. In the case of ratio estimator and the correlation coefficient, the data pairs are resampled from the original data pairs (X_i, Y_i).

3.2 Confidence intervals

Bootstrap resampling is also widely used deriving confidence intervals for parameters. However, unless the limiting distribution of the point estimator is free from the unknown parameters, one can not invert it to get confidence intervals. Such quantities, with distributions that are free from unknown parameters, are called 'pivotal' statistics. It is thus important to focus on pivotal or approximately pivotal quantities in order to get reliable confidence intervals for the parameter of interest. For example, if $X_i \sim N(\mu, \sigma^2)$, then $\sqrt{n}(\bar{X} - \mu)/s_n$ has t distribution with $n-1$ degrees of freedom, and hence it is pivotal, where $s_n^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2$. In non-normal case, it is approximately pivotal. To obtain bootstrap confidence interval for μ, we compute $\sqrt{n}(\bar{X}^{*(j)} - \bar{X})/s_n$ for N bootstrap samples, arrange the values in increasing order

$$h_1 < h_2 < \cdots < h_N.$$

One can then read off from the histogram (say) the 90% confidence interval of the parameter. That is, the 90% confidence interval for μ is given by

$$\bar{X} - h_m \frac{s_n}{\sqrt{n}} \leq \mu < \bar{X} - h_k \frac{s_n}{\sqrt{n}},$$

where $k = [0.05N]$ and $m = [0.95N]$. Babu & Singh (1983) have shown that $N \sim n(\log n)^2$ bootstrap iterations would be sufficient.

It is important to note that even when σ is known the bootstrap version of $\sqrt{n}(\bar{X} - \mu)/\sigma$ is $\sqrt{n}(\bar{X}^* - \bar{X})/s_n$. One should not replace $\sqrt{n}(\bar{X}^* - \bar{X})/s_n$ by $\sqrt{n}(\bar{X}^* - \bar{X})/\sigma$.

3.3 Bootstrap at its best: *Smooth function model*

It is well established using Edgeworth expansions that the bootstrap provides a good approximation for a 'Studentized smooth functional model'. A broad class of commonly used statistics, including least squares estimators and some maximum likelihood estimators, can be expressed as smooth function of multivariate means. The model is illustrated using Pearson's well known estimator $\hat{\rho}_n$ of correlation coefficient ρ. The sample correlation coefficient $\hat{\rho}_n$ based on the data $(X_i, Y_i), i = 1, \ldots, n$, can be expressed as $\hat{\rho}_n = H(\bar{\mathbf{Z}}_n)$, and $\rho^* = H(\bar{\mathbf{Z}}_n^*)$, where

$$\mathbf{Z}_i = (X_iY_i, X_i^2, Y_i^2, X_i, Y_i), \quad \mathbf{Z}_i^* = (X_i^*Y_i^*, X_i^{*2}, Y_i^{*2}, X_i^*, Y_i^*)$$

and

$$H(a_1, a_2, a_3, a_4, a_5) = \frac{(a_1 - a_4a_5)}{\sqrt{((a_2 - a_4^2)(a_3 - a_5^2))}}.$$

Note that H is a differentiable function.

In general, if the standard deviation of $T_n(\mathbf{X}; F)$ is not known (which is often the case), the function may be divided by a good estimator of the standard deviation of the statistic. This makes it an 'approximate pivotal' quantity. Such a correction by a special type of estimator of standard deviation for the smooth function model refers to Studentization, as it is similar to the Student's t-statistic. The empirical distribution of the data is used to estimate the standard deviation of the statistic in a special way, making it an 'approximate pivotal' quantity. For the smooth function model, a good estimator of the variance of $\sqrt{n}H(\bar{\mathbf{Z}}_n)$ is given by $\hat{\sigma}^2 = \ell^T(\bar{\mathbf{Z}}_n)\Sigma_n\ell(\bar{\mathbf{Z}}_n)$, where $\ell(\mathbf{x})$ denotes the vector of first order partial derivatives of H at $\mathbf{x}$, ^{T} denotes transpose, and Σ_n denotes the variance-covariance matrix computed from the $\{\mathbf{Z}_i\}$. That is,

$$\Sigma_n = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{Z}_i - \bar{\mathbf{Z}}_n)(\mathbf{Z}_i - \bar{\mathbf{Z}}_n)^T. \tag{4}$$

This leads to Studentization or approximate pivotal function

$$t_n = \sqrt{n}(H(\bar{\mathbf{Z}}_n) - H(\mathrm{E}(\mathbf{Z}_1))/\hat{\sigma} \tag{5}$$

Its bootstrap version is

$$t_n^* = \sqrt{n}(H(\bar{\mathbf{Z}}_n^*) - H(\bar{\mathbf{Z}}_n))/\sqrt{\ell^T(\bar{\mathbf{Z}}_n^*)\Sigma_n^*\ell(\bar{\mathbf{Z}}_n^*)}, \tag{6}$$

where Σ_n^* denotes the variance-covariance matrix computed from the bootstrap sample $\{\mathbf{Z}_i^*\}$, *i.e.*

$$\Sigma_n^* = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{Z}_i^* - \bar{\mathbf{Z}}_n^*)(\mathbf{Z}_i^* - \bar{\mathbf{Z}}_n^*)^T. \tag{7}$$

If $H(\bar{\mathbf{Z}}_n)$ represents the sample mean $\bar{X}_n$, then $\hat{\sigma}^2 = s_n^2$, and if $H(\bar{\mathbf{Z}}_n)$ represents the ratio statistic $\hat{\theta} = \bar{X}_n/\bar{Y}_n$, then $\hat{\sigma}^2 = \bar{Y}^{-2}n^{-1}\sum_{i=1}^{n}(X_i - \hat{\theta}Y_i)^2$.

Under very general conditions, if $\ell(\mathrm{E}(\mathbf{Z}_1)) \neq 0$, then the approximation of the sampling distribution of t_n by the bootstrap distribution (the distribution of t_n^*) is better than the classical normal approximation. This is mainly because the bootstrap automatically corrects for the skewness factor. This is established using Edgeworth expansion (see Babu & Singh (1983), and Babu & Singh (1984)):

$$P(t_n \leq x) = \Phi(x) + \frac{1}{\sqrt{n}} p(x)\phi(x) + \text{ error}$$
$$P^*(t_n^* \leq x) = \Phi(x) + \frac{1}{\sqrt{n}} p_n(x)\phi(x) + \text{ error.}$$

The 'error' terms are so small that

$$\sqrt{n}|\mathrm{P}(t_n \leq x) - \mathrm{P}^*(t_n^* \leq x)| \to 0.$$

The theory above is applicable in very general set up that includes the statistics: *Sample Means, Sample Variances, Central and Non-central t-statistics (with possibly non-normal populations), Sample Coefficient of Variation, Maximum Likelihood Estimators, Least Squares Estimators, Correlation Coefficients, Regression Coefficients*, and Smooth transforms of these statistics.

Thus the sampling distribution of several commonly occurring statistics are closer to the corresponding bootstrap distribution than the normal distribution. These conditional approximations are suggestive of the unconditional ones, though one cannot be derived from the other by elementary methods. Babu & Bose (1988) provide theoretical justification for the accuracy of the bootstrap confidence intervals both in terms of the actual coverage probability achieved and also the limits of the confidence interval.

In spite of these positive results, one should use caution in using bootstrap methods. It is not a 'cure all' solution. There are cases where bootstrap method fails. These include, non-smooth statistics such as $\hat{\theta} = \max_{1 \leq i \leq n} X_i$ (see Bickel & Freedman (1981)), heavy tailed distributions, $\hat{\theta} = \bar{X}$ and $\mathrm{E}X_1^2 = \infty$ (see Babu (1984) and Athreya (1987)), and asymptotically non-linear statistics such as, $\hat{\theta} - \theta = H(\bar{\mathbf{Z}}_n) - H(\mathrm{E}(\mathbf{Z}_1)$ when $\partial H(\mathrm{E}(\mathbf{Z}_1)) = 0$. In the last case the limiting distribution is like that of linear combinations of Chi-squares, but here a modified version works (Babu (1984)).

3.4 Linear regression

Consider the simple linear regression model, where the data $(X_1, Y_1), \ldots, (X_n, Y_n)$ satisfies

$$Y_i = \alpha + \beta X_i + e_i, \tag{8}$$

where α and β are unknown parameters, $X_1, \ldots, X_n$ are often called the design points. The error variables e_i need not be Gaussian, but are assumed to be independent with zero mean and standard deviation σ_i. This model is called homoscedastic if all the σ_i are identical. Otherwise, the model is known as heteroscedastic. In what follows, for any sequence of pairs $\{(U_1, V_1), \ldots, (U_n, V_n)\}$ of numbers, we use the notation

$$S_{UV} = \sum_{i=1}^{n} (U_i - \bar{U}_n)(V_i - \bar{V}_n) \quad \text{and} \quad \bar{U}_n = \frac{1}{n} \sum_{i=1}^{n} U_i. \tag{9}$$

There are two conceptually separate models to consider, random and fixed design models. In the first case, the pairs $\{(X_1, Y_1), \ldots, (X_n, Y_n)\}$ are assumed to be random data points and the conditional mean and variance of e_i given X_i are assumed to be zero and σ_i^2. In the latter case, $X_1, \ldots, X_n$ are assumed to be fixed numbers (fixed design). In both the cases, the least squares estimators $\hat{\alpha}$ and $\hat{\beta}$ of α and β are given by

$$\hat{\beta} = S_{XY}/S_{XX} \quad \text{and} \quad \hat{\alpha} = \bar{Y}_n - \hat{\beta}\bar{X}_n. \tag{10}$$

However the variances of these estimators are different for a random and fixed designs, though the difference is very small for large n. We shall concentrate on the fixed design case here.

The variance of the slope $\hat{\beta}$ is given by

$$\mathrm{var}(\hat{\beta}) = \sum_{i=1}^{n} (X_i - \bar{X}_n)^2 \sigma_i^2 / S_{XX}^2, \tag{11}$$

and depends on the individual error deviations σ_i, which may or may not be known. Knowledge of $\mathrm{var}(\hat{\beta})$ provides the confidence intervals for β. Several resampling methods are available in the literature to estimate the sampling distribution and $\mathrm{var}(\hat{\beta})$. We consider three bootstrap procedures: a) the classical bootstrap, b) the paired bootstrap.

The classical bootstrap

Let $\hat{e}_i$ denote the residual of the i-th element of $\hat{e}_i = Y_i - \hat{\alpha} - \hat{\beta}X_i$ and define $\tilde{e}_i$ to be

$$\tilde{e}_i = \hat{e}_i - \frac{1}{n}\sum_{j=1}^{n} \hat{e}_j. \tag{12}$$

A bootstrap sample is obtained by randomly drawing $e_1^*, \ldots, e_n^*$ with replacement from $\tilde{e}_1, \ldots, \tilde{e}_n$. The bootstrap estimators β^* and α^* of the slope and the intercept are given by

$$\beta^* - \hat{\beta} = S_{Xe^*}/S_{XX} \quad \text{and} \quad \alpha^* - \hat{\alpha} = (\hat{\beta} - \beta^*)\bar{X}_n + \bar{e}_n^*. \tag{13}$$

To estimate the sampling distribution and variance, the procedure is repeated N times to obtain

$$\beta_1^*, \ldots, \beta_N^* \quad \text{where} \quad N \sim n(\log n)^2. \tag{14}$$

The histogram of these β^*s give a good approximation to the sampling distribution of $\hat{\beta}$ and the estimate of the variance $\hat{\beta}$ is given by

$$\mathrm{var}_{\mathrm{Boot}} = \frac{1}{N}\sum_{j=1}^{N} (\beta_j^* - \hat{\beta})^2. \tag{15}$$

This variance estimator is the best among the two methods proposed here, if the residuals are homoscedastic; *i.e.* if the variances of the residuals $\mathrm{E}(\epsilon_i^2) = \sigma_i^2 = \sigma^2$ are all the same. However if they are not, then the bootstrap estimator of the variance is an inconsistent estimator, and does not approach the actual variance. The *paired bootstrap* is robust against heteroscedasticity, giving consistent estimator of variance when the residuals have different standard deviations.

The paired bootstrap

The paired bootstrap are useful to handle heteroscedastic data. The paired bootstrap method treats the design points as random quantities. A simple random sample $(\tilde{X}_1, \tilde{Y}_1), \ldots,$ $(\tilde{X}_n, \tilde{Y}_n)$ is drawn from (X_1, Y_1), …, (X_n, Y_n) and the paired bootstrap estimators of slope and intercept are constructed as

$$\tilde{\beta} = \frac{\sum_{i=1}^{n}(\tilde{X}_i - \bar{\tilde{X}})(\tilde{Y}_i - \bar{\tilde{Y}})}{\sum_{i=1}^{n}(\tilde{X}_i - \bar{\tilde{X}})^2}, \quad \text{and} \quad \tilde{\alpha} = \bar{\tilde{Y}} - \tilde{\beta}\bar{\tilde{X}}$$

The variance is obtained by repeating the resampling scheme N times and applying equation (15).

Figure 1 provides a simple FORTRAN code for jackknife and paired bootstrap resampling.

```
C       PAIRED BOOTSTRAP RESAMPLING
        NSIM = INT(N * ALOG(FLOAT(N))**2)
        DO 20 ISIM = 1,NSIM
        DO 10 I = 1,N
           J = INT(RANDOM * N + 1.0)
           XBOOT(I) = X(J)
  10       YBOOT(I) = Y(J)
  20    CONTINUE

C       JACKKNIFE RESAMPLING
        DO 40 NSIM = 1,N
        DO 30 I = 1,N-1
           IF(I.LT.NSIM)
             XJACK(I) = X(I)
             YJACK(I) = Y(I)
           ELSE
             XJACK(I) = X(I+1)
             YJACK(I) = Y(I+1)
           ENDELSE
  30   CONTINUE
  40   CONTINUE
```

Figure 1: FORTRAN code illustrating the paired bootstrap and jackknife resampling for a two dimensional dataset $(x_i, y_i), i = 1, \ldots, N$.

The bootstrap methodology, mathematics and second order properties are reviewed in Babu & Rao (1993). A detailed account of second order asymptotics can be found in Hall

(1992). A less mathematical overview of the bootstrap is presented in Efron and Tibshirani (1993). The book by Zoubir & Iskander (2004) serves as a handbook on 'bootstrap' for engineers, to analyze complicated data with little or no model assumptions. Bootstrap has found many applications in engineering field including, artificial neural networks, biomedical engineering, environmental engineering, image processing, and Radar and sonar signal processing. Majority of the applications in the book are taken from signal processing literature.

References

[1] Athreya, K. B. (1987). Bootstrap of the mean in the infinite variance case. *Ann. Statist.,* **15**, no. 2, 724-731.

[2] Babu, G. J. (1984). Bootstrapping statistics with linear combinations of chi-squares as weak limit. *Sankhya*, Series A, **46**, 85-93.

[3] Babu, G. J. (1986). A note on bootstrapping the variance of the sample quantile. *Ann. Inst. Statist. Math.*, **38**, Part A, 439-443.

[4] Babu, G. J., and Bose, A. (1988). Bootstrap confidence intervals. *Statistics and Probability Letters*, **7**, 151-160.

[5] Babu, G. J., and Rao, C. R. (1993). Bootstrap methodology. In *Handbook of Statistics,* **Vol. 9** "Computational Statistics." C. R. Rao (Ed.), Elsevier Science Publishers B. V., Amsterdam, 627-659.

[6] Babu, G. J., and Singh, K. (1983). Inference on means using the bootstrap. *Annals of Statistics*, **11**, 999-1003.

[7] Babu, G. J., and Singh, K. (1984a). On one term Edgeworth correction by Efron's Bootstrap. *Sankhya*, Series A, **46**, 219-232.

[8] Babu, G. J., and Singh, K. (1984b). Asymptotic representations related to jackknifing and bootstrapping L-statistics. *Sankhyā*, Series A, **46**, 195-206.

[9] Bickel, P. J., and Freedman, D. A. Some asymptotic theory for the bootstrap. *Ann. Statist.*, **9**, no. 6, 1196–1217.

[10] Chernick, M. R. (2007). *Bootstrap Methods - A guide for Practitioners and Researchers*, (2nd Ed.) Wiley Inter-Science.

[11] Chernick, M. R., and LaBudde, R. A. (2011). *An Introduction to Bootstrap Methods with Applications to R*, Wiley.

[12] Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Ann. Statist.,* **7**, no. 1, 1-26.

[13] Efron, B. (1982). *The jackknife, the bootstrap and other resampling plans.* CBMS-NSF Regional Conference Series in Applied Mathematics, 38. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, Pa.

[14] Efron, B., and Tibshirani, R. J. (1993). *An introduction to the bootstrap.* Monographs on Statistics and Applied Probability, 57. Chapman and Hall, New York.

[15] Hall, P. (1992). *The bootstrap and Edgeworth expansion.* Springer Series in Statistics. Springer-Verlag, New York.

[16] Quenouille, M. (1949). Approximate tests of correlation in time series. *J. Roy. Statist. Soc.*, Ser. B, **11**, pp. 1884.

[17] Zoubir, A. M., and Iskander, D. R. (2004). *Bootstrap Techniques for Signal Processing.* Cambridge University Press, Cambridge, U.K.