Working Paper 92-23
May 1992

División de Economía
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (341) 624-9849

# THE BOOTSTRAP - A REVIEW

Wenceslao González Manteiga, José Manuel Prada Sánchez and Juan Romo*

Abstract

The bootstrap, extensively studied during the last decade, has become a powerful tool in different areas of Statistical Inference. In this work, we present the main ideas of bootstrap methodology in several contexts, citing the most relevant contributions and illustrating with examples and simulation studies some interesting aspects.

Key words:
Bootstrap, jackknife, cross-validation, regression, censored data, smoothing, symmetrization, bayesian methods and prediction error.

AMS Classification (1990): 62G09, 62G07, 62G05.

*González Manteiga, Universidad de Santiago de Compostela; Prada Sánchez, Universidad de Santiago de Compostela; Romo, Universidad Carlos III de Madrid, research partially supported by DGICYT PB90-0266.

# 1 Introduction

The bootstrap, a resampling method introduced by Efron (1979), aims to reproduce from the sample the mechanism generating the data and to use it in the statistic of interest, replacing *everywhere* the unknown populational model. Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a random sample from a variable $X$ with unknown distribution $P$ and consider the statistic $R = R(X_1, \ldots, X_n; P)$ . If $P_n$ is the empirical probability corresponding to $\mathbf{X}$ (giving mass $\frac{1}{n}$ to each observation), the bootstrap version of $R$ is $R^* = R(X_1^*, \ldots, X_n^*; P_n)$ , where $\mathbf{X}^* = (X_1^*, \ldots, X_n^*)$ is a random sample drawn from $P_n$; this is known as *standard*, *naïve* or *nonparametric* bootstrap. Other resampling approaches have also been considered: instead of $P_n$, we can use a smoothed version $\check{P}_n$ (*smoothed bootstrap*) or, if it is known that $P$ belongs to a parametric family $\{P_\theta : \theta \in \Theta\}$ and $\hat{\theta}_n = \hat{\theta}_n(\mathbf{X})$ is an estimator of $\theta$, $\mathbf{X}^*$ can be taken as a sample from $P_{\hat{\theta}_n}$ (*parametric bootstrap*); in general, any other estimator $\tilde{P}_n$ of $P$ could be used.

The bootstrap technique proceeds in several steps:

(i) simulate artificially a random sample $\mathbf{X}^* = (X_1^*, \ldots, X_m^*)$ —m not necessarily equal to n— from the empirical probability $P_n$.

(ii) evaluate R at the bootstrap sample to obtain the bootstrap version of the statistic $R^* = R(X_1^*, \ldots, X_n^*; P_n)$ .

(iii) replicate (i) and (ii) a large number $B$ of times, in order to get $B$ values of $R^*$, $R_i^* = R(\mathbf{X}^{*\mathbf{i}}; P_n), i = 1, \ldots, B$.

Finally, a histogram (or, in general, any other estimate of the distribution of $R^*$) is obtained from $R_i^*, i = 1, \ldots, B$. This is an approximation to the distribution of $R^*$ which in turn is the bootstrap estimation of the unknown distribution of $R$.

There exists an extensive literature on bootstrap. This paper is focused on the basic ideas in several statistical contexts; we quote the most relevant contributions and illustrate with examples and simulation studies some interesting topics. The contents are the following:

10. Other topics and applications.

Previous review or introductory papers on bootstrap or applications of the bootstrap in specific areas of statistical inference are Efron (1982), Efron and Tibshirani (1986), Hinkley (1988), DiCiccio and Romano (1988), Härdle and Mammen (1991) and Léger, Politis and Romano (1992).

Since it is extremely difficult to exhaustively review such a growing literature, the choice of topics inevitably reflects the authors interests; moreover, some very recent emerging applications in several fields (e.g., binomial models or model checking in time series (Tsay, 1992)) are not covered. We don't either consider questions related to the computational efficiency of the resampling procedure (see Efron (1990), Davison, Hinkley and Schechtman (1986), Graham, Hinkley, John and Shi (1990), Do and Hall (1991) and references therein); Sánchez (1991) is a survey on this subject.

## 2 The bootstrap as an alternative to the jackknife. Estimation of bias and variance

Initially, the bootstrap was introduced by Efron(1979) as an alternative technique to the classical Quenouille-Tuckey jackknife (see Miller, 1974) to estimate either the bias $E_F T(\mathbf{X}) - T(F)$ or the variance $Var_F T(\mathbf{X})$ of some estimator $T(\mathbf{X})$ of $T(F)$, where $\mathbf{X} = (X_1, \ldots, X_n)$ is a random sample from the distribution F. Taking $R(\mathbf{X}; F) = T(\mathbf{X}) - T(F)$, this entails to estimate $E_F R(\mathbf{X}; F)$ or $Var_F R(\mathbf{X}; F)$. If $F_n$ is the empirical distribution obtained from the sample $\mathbf{X}$, $E_F R(\mathbf{X}; F)$ can be estimated by $E_{F_n} R(\mathbf{X}^*; F_n) = E_* R(\mathbf{X}^*; F_n)$, which in turn can be approximated by

$$\bar{R}^* = \frac{1}{B} \sum_{i=1}^{B} R(\mathbf{X}^{*i}; F_n).$$

Analogously, a bootstrap estimator for $Var_F R(\mathbf{X}; F)$ is $Var_{F_n} R(\mathbf{X}^*; F_n) = Var_* R(\mathbf{X}^*; F_n)$ which can be also approximated by

$$\frac{1}{B-1} \sum_{i=1}^{B} (R_i^* - \bar{R}^*)^2.$$

When either it is difficult to obtain explicit estimations using $E_*$ or $V_*$ or when the approximatting computational mechanism is too complicated, a Taylor expansion can be used. Indeed, if $(x_1, \ldots, x_n)$ is a particular value of $\mathbf{X}$ and $N_i^* = \#\{X_i^* = x_i\}, i = 1, \ldots, n$ and $p_i^* = \frac{N_i^*}{n}$, then the bootstrap resampling vector $\mathbf{P}^* = (p_1^*, \ldots, p_n^*)$ is distributed as $\frac{1}{n}$ times a multinomial distribution with mean $E_* \mathbf{P}^* = \frac{e}{n} = (\frac{1}{n}, \ldots, \frac{1}{n})$ and covariance matrix given by

$$cov_* \mathbf{P}^* = \frac{I}{n^2} - \frac{e^t e}{n^3},$$

where I is the identity matrix and $e^t$ is the transpose of vector $e$. So, assuming that $R$ is invariant with respect to sample permutations, we can identify $R(\mathbf{X}^*, F_n)$ and $R(P^*)$ in the following second order Taylor expansion at $\frac{e}{n}$:

$$R(\mathbf{P}^*) \simeq R(\frac{e}{n}) + (\mathbf{P}^* - \frac{e}{n})U + \frac{1}{2}(\mathbf{P}^* - \frac{e}{n})V(\mathbf{P}^* - \frac{e}{n})^t$$

with $U^t = (U_1, \ldots, U_n)$ and $V = (V_{ij})$, where

$$U_i = \frac{\partial R(\mathbf{P}^*)}{\partial p_i^*} |_{\mathbf{P}^* = \frac{e}{n}}, i = 1, \ldots, n;$$

and

$$V_{ij} = \frac{\partial^2 R(\mathbf{P}^*)}{\partial p_i^* \partial p_j^*} |_{\mathbf{P}^* = \frac{e}{n}}, i = 1, \ldots, n.$$

Calculating mean and variance in the previous Taylor expansion, we get

$$E_* R^* \simeq E^* R(P^*) \simeq R(\frac{e}{n}) + \frac{1}{2} trace V(\frac{I}{n} - \frac{e^t e}{n^3}) = R(\frac{e}{n}) + \frac{1}{2n^2} \sum_{i=1}^{n} V_{ii} \tag{2.1}$$

$$Var_* R^* \simeq Var^* R(P^*) \simeq U^t(\frac{I}{n} - \frac{e^t e}{n^3})U = \frac{1}{n^2} \sum_{i=1}^{n} U_i^2. \tag{2.2}$$

Considering expresions (2.1) and (2.2), we can compare all known interesting estimates of bias and variance of a functional $T(F_n)$; taking $R(\mathbf{X}, F) = T(\mathbf{X}) - T(F) = T(F_n) - T(F)$, we get

a) *Bootstrap estimations* of bias and variance of $T(\mathbf{X})$:

$$Bias_{boot} T(\mathbf{X}) = E_* R^* = E_* T(\mathbf{X}^*) - T(F_n) \tag{2.3}$$

$$Var_{boot} T(\mathbf{X}) = Var_* R^* = Var_* T(\mathbf{X}^*) \tag{2.4}$$

b) *Approximations to bootstrap distributions* of bias and variance of $T(\mathbf{X})$:

$$\widehat{Bias}_{boot} T(\mathbf{X}) = \frac{1}{B} \sum_{i=1}^{B} T(\mathbf{X}^{*\mathbf{i}}) - T(F_n) = \overline{T^*} - T(F_n) \tag{2.5}$$

$$\widehat{Var}_{boot} T(\mathbf{X}) = \frac{1}{B-1} \sum_{i=1}^{B} (T(\mathbf{X}^{*\mathbf{i}}) - \overline{T^*})^2 \tag{2.6}$$

c) *Jaeckel's (1972) infinitesimal jackknife* estimations:

$$Bias_{jack-i} T(\mathbf{X}) = \frac{1}{2n^2} \sum_{i=1}^{n} V_{ii} \tag{2.7}$$

$$Var_{jack-i} T(\mathbf{X}) = \frac{1}{n^2} \sum_{i=1}^{n} U_i^2 \tag{2.8}$$

where

$$U_i = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon}(T(F_n + \varepsilon(\delta_{x_i} - F_n)) - T(F_n)) = \frac{d}{d\varepsilon} T(F_n + \varepsilon(\delta_{x_i} - F_n)) \mid_{\varepsilon=0},$$

$$V_{ii} = \frac{d^2}{d\varepsilon^2} T(F_n + \varepsilon(\delta_{x_i} - F_n)) \mid_{\varepsilon=0},$$

and $\delta_{x_i}$ is the corresponding Dirac's delta at $x_i, i = 1, \ldots, n$.

In several interesting cases, these estimations coincide with the ones given by the classical delta method and with the corresponding ones obtained through the empirical influence function (see Efron (1982) for further details).

d) *Ordinary jackknife estimations* (Quenouille, 1956, and Tukey, 1958):

$$Bias_{jack} T(\mathbf{X}) = \frac{1}{2n(n-1)} \sum_{i=1}^{n} \tilde{V}_{ii} \tag{2.9}$$

$$Var_{jack} T(\mathbf{X}) = \frac{1}{n^2} \sum_{i=1}^{n} \tilde{U}_i^2, \tag{2.10}$$

where $\tilde{V}_{ii}$ and $\tilde{U}_i$ are approximations for $V_{ii}$ and $U_i$ with $\epsilon = -\frac{1}{1-n}$.

Thus, the infinitesimal jackknife estimations of bias and variance coincide, respectively, with the linear and quadratic approximations of the bootstrap estimations, whereas the ordinary jackknife estimations approach the corresponding infinitesimal jackknife ones.

Expressions (2.9) and (2.10) should be compared with (2.7) and (2.8); the original ordinary jackknife expressions were

$$Bias_{jack} T(\mathbf{X}) = \frac{n-1}{n} \sum_{i=1}^{n} (T(\mathbf{X}_{(i)}) - T(\mathbf{X})) \tag{2.11}$$

$$Var_{jack} T(\mathbf{X}) = \frac{n-1}{n} \sum_{i=1}^{n} (T(\mathbf{X}_{(i)}) - \bar{T})^2, \tag{2.12}$$

where $\bar{T} = \frac{1}{n} \sum_{i=1}^{n} T(\mathbf{X}_{(i)})$ and $\mathbf{X}_{(i)}$ is the sample when the i-th observation is deleted. From $Bias_{jack}$, we can construct a new estimator of $T(F)$ (the so called corrected bias estimate)

$$T^{jack}(\mathbf{X}) = T(\mathbf{X}) - Bias_{jack} T(\mathbf{X}) = nT(\mathbf{X}) - \frac{n-1}{n} \sum_{i=1}^{n} T(\mathbf{X}_{(i)}) =$$

$$= \frac{1}{n} \sum_{i=1}^{n} (nT(\mathbf{X}) - (n-1)T(\mathbf{X}_{(i)})) = \frac{1}{n} \sum_{i=1}^{n} T^{(i)},$$

where $T^{(i)}$ is the *i-th jackknife pseudovalue* (which is also an estimate of $T(F)$). It can be shown that

$$Var_{jack}T(\mathbf{X}) = \frac{1}{n(n-1)}\sum_{i=1}^{n}(T^{(i)} - T^{jack}(\mathbf{X}))^2. \tag{2.13}$$

To end this section, we present a simulation study comparing jackknife and bootstrap as estimators of the variance of a given statistic. Let $F$ be a distribution which is symmetric about its mean (and median) $\theta = T(F) = \int x dF(x)$. To estimate $\theta$, we will consider the sample mean, $T(\mathbf{X}) = \bar{\mathbf{X}}$, the $\alpha$-trimmed mean

$$T_\alpha(\mathbf{X}) = \frac{1}{n-2[n\alpha]}\sum_{i=[n\alpha]+1}^{n-[n\alpha]} X_{(i)},$$

where $[t]$ denotes the greatest integer less than or equal to t and $(X_{(1)}, \ldots, X_{(n)})$ is the ordered sample, and the sample median $T_{1/2}(\mathbf{X})$. It is easy to check that the jackknife and bootstrap estimations of the variance of the sample mean are $S^2/n$ and $S_n^2/n$, where $S^2$ and $S_n^2$ are the sample quasi-variance and variance, respectively (so, the jackknife estimation is unbiased). The exact expression for the bootstrap variance of the sample median is

$$Var_{boot}T_{1/2}(\mathbf{X}) = \sum_{k=1}^{n}(X_{(k)} - E_* T_{1/2}(\mathbf{X}^*))^2 p_k,$$

$$p_k = P_*\{T_{1/2}(\mathbf{X}^*) = x_{(k)}\} =$$

$$= \sum_{j=0}^{m-1}\left(\binom{n}{j}\left(\frac{k-1}{n}\right)^j\left(1-\frac{k-1}{n}\right)^{n-j} - \binom{n}{j}\left(\frac{k}{n}\right)^j\left(1-\frac{k}{n}\right)^{n-j}\right)$$

if $n = 2m-1$ (resampling distribution of the bootstrap median (see Efron 1979, 1982)). Ghosh *et al* (1984) have shown the consistency, under certain conditions, of the bootstrap estimation of the variance of the sample median, in opposition to the asymptotic behavior of the jackknife (Miller, 1974):

$$nVar_{jack}T_{1/2}(\mathbf{X}) \longrightarrow_w \frac{1}{4f^2(\theta)}\left(\frac{\chi_2^2}{2}\right)^2$$

and (see, e.g., Kendall and Stuart, 1958)

$$nVarT_{1/2}(\mathbf{X}) \longrightarrow \frac{1}{4f^2(\theta)},$$

where $f$ is the density corresponding to $F$, $f(\theta) > 0$. Since the mean of the variable $(\frac{\chi_2^2}{2})^2$ is 2, the jackknife estimate is inconsistent.

The bootstrap estimator of the variance of the sample median was introduced by Maritz and Jarret (1978) and Efron (1979). Babu (1986) obtained the consistency of the bootstrap estimator of the variance of any sample quantile and Hall and Martin (1988) have proved that its rate of convergence is slow (of order $n^{-\frac{1}{4}}$).

Tables 1.1 and 1.2 present the results of a simulation study comparing the bootstrap and jackknife approximations to the variance of the sample mean, median and $\alpha$-trimmed means for a $N(0,1)$ population. We have used (2.6) and (2.12) and the fact that, in this case, the theoretical variance of the mean is $\frac{1}{n}$ and the asymptotic variances of the sample median and $\alpha$-trimmed means are $\frac{\pi}{2n}$ and

$$\frac{F_{\gamma(1,3/2)}\left(\frac{(\Phi^{-1}(1-\alpha))^2}{2}\right)+2\alpha(\Phi^{-1}(1-\alpha))^2}{n(1-2\alpha)^2},$$

respectively, where $F_{\gamma(1,3/2)}$ and $\Phi$ are the distribution functions for the gamma and normal distributions, respectively. Both tables give the exact variance of the estimates and the means and mean squared errors —between parentheses— of the jackknife and bootstrap approximations for 1000 simulated samples of size 11 (Table 1.1) and 21 (Table 1.2) with $B = 200$ bootstrap replications.

| | Exact | Jackknife | Bootstrap |
|---|---|---|---|
| Mean Var. | 0.0909 | 0.0904 (0.0016) | 0.0821 (0.0014) |
| Median Var. | 0.1427 | 0.1685 (0.0415) | 0.1738 (0.0148) |
| 0.1-mean Var. | 0.0963 | 0.0978 (0.0025) | 0.0989 (0.0021) |
| 0.2-mean Var. | 0.1040 | 0.1103 (0.0052) | 0.1099 (0.0031) |
| 0.3-mean Var. | 0.1138 | 0.1267 (0.0110) | 0.1222 (0.0050) |
| 0.4-mean Var. | 0.1263 | 0.1685 (0.0426) | 0.1400 (0.0084) |

Table 1.1

| | Exact | Jackknife | Bootstrap |
|---|---|---|---|
| Mean Var. | 0.0476 | 0.0475 (0.0002) | 0.0452 (0.0002) |
| Median Var. | 0.0747 | 0.1078 (0.0227) | 0.0895 (0.0031) |
| 0.1-mean Var. | 0.0504 | 0.0506 (0.0003) | 0.0520 (0.0003) |
| 0.2-mean Var. | 0.0545 | 0.0554 (0.0006) | 0.0563 (0.0005) |
| 0.3-mean Var. | 0.0596 | 0.0636 (0.0014) | 0.0625 (0.0008) |
| 0.4-mean Var. | 0.0661 | 0.0793 (0.0054) | 0.0718 (0.0016) |

Table 1.2

As expected, the results show a better behavior of the bootstrap approximation in mean squared error, improving with $\alpha$.

## 3 Bootstrap approximation to the distribution of a statistic

As we have seen, a general goal of bootstrap resampling is to approximate the distribution $P_F\{R(\mathbf{X},F)\leq x\}$ of the statistic $R(\mathbf{X},F)$ by using the distribution

$$P_{F_n}\{R(\mathbf{X}^*,F_n)\leq x\} = P^*\{R(\mathbf{X}^*,F_n)\leq x\}$$

of $R(\mathbf{X}^*,F_n)$. This can be expressed in several ways. If $R(\mathbf{X},F)$ converges weakly to a distribution $S(F)$, it suffices to show that $R(\mathbf{X}^*,F_n)$ converges weakly to $S(F)$ for almost all samples $X_1,\ldots X_n\ldots$ ($R(\mathbf{X}^*,F_n)\rightarrow_w S(F)$ *a.s.*) or to establish that the distance between the law of $R(\mathbf{X}^*, F_n)$ and the law of $S(F)$ tends to zero in probability for any distance metrizing weak convergence ($R(\mathbf{X}^*, F_n) \to_w S(F)$ *in probability*). Also, the discrepancy between the sampling and the bootstrap distributions can be measured by using different functional distances: the supremum distance

$$d_\infty(F, G) \equiv \sup_{x \in \mathcal{R}} |F(x) - G(x)|$$

or the Mallows metric, defined on the class of distribution functions with finite second moment, given by

$$d_2(F, G) \equiv \inf_{\{(X,Y):X \equiv_d F, Y \equiv_d G\}} (E|X - Y|^2)^{1/2} = \left( \int_0^1 (F^{-1}(t) - G^{-1}(t))^2 dt \right)^{1/2},$$

where $F^{-1}(t) \equiv \inf\{x : F(x) \geq t\}, t \in (0, 1)$. This metric characterizes weak and second order moments convergence. Discrepancies at a fixed point $x$ are usually treated by means of Edgeworth expansions.

Bickel and Freedman (1981) show that if $E_F X^2 < \infty$ then

$$d_2(P_{Fn}\{n^{1/2}(\bar{X}_n^* - \bar{X}_n) \leq \cdot\}, P_F\{n^{1/2}(\bar{X}_n - \mu) \leq \cdot\}) \to 0 \quad a.s.$$

Later on, Bickel and Freedman (1984) extend his result to stratified sampling. Under the same hypotheses, Singh (1981) proves the result for the distance $d_\infty$. Moreover, under different hypotheses, he gets the corresponding rates of convergence; in particular, if $E_F|X|^3 < \infty$, he obtains that

$$d_\infty(P_{Fn}\{n^{1/2}(\bar{X}_n^* - \bar{X}_n) \leq \cdot\}, P_F\{n^{1/2}(\bar{X}_n - \mu) \leq \cdot\}) = O\left(\frac{1}{n^{1/2}}\right) \quad a.s.$$

This rate of convergence is larger than the one obtained in the same paper for the sample quantiles, $R(\mathbf{X}, F) = n^{1/2}(F_n^{-1}(t) - F^{-1}(t))$ : assuming that the second order derivative of $F$ does exist, that the first order derivative is bounded in a neighborhood of $F^{-1}(t)$ and that $F'(F^{-1}(t)) > 0$, Singh shows that

$$d_\infty(P_{Fn}\{n^{1/2}(F_n^{-1*}(t) - F_n^{-1}(t)) \leq \cdot\}, P_F\{n^{1/2}(F_n^{-1}(t) - F^{-1}(t)) \leq \cdot\}) = O\left(\frac{\log\log n}{n^{1/4}}\right) \quad a.s.$$

One of the most studied cuestions over the eighties is comparing the bootstrap approximation rate with the approximation rate of previously existing methods, such as the normal approximation in the central limit theorem. Edgeworth expansions are the fundamental tool for this purpose. In this way, following, e. g., Hall (1988-a), if $F$ is not reticular and $E|X|^3 < \infty$, it is possible to compare rates of convergence for the bootstrap and the normal approximation using Edgeworth expansions for $R(\mathbf{X}, F) = n^{1/2}(\bar{X}_n - \mu)$:

$$P_F\{n^{1/2}(\bar{X}_n - \mu) \leq x\} = \Phi\left(\frac{x}{\sigma}\right) + n^{-1/2}\left(\frac{\mu_3}{6\sigma^3}\left(1 - \frac{x^2}{\sigma^2}\right)\right)\phi\left(\frac{x}{\sigma}\right) + O(n^{-1}) \qquad (3.14)$$

uniformly in $x$, and for $R(\mathbf{X}, F) = \frac{n^{1/2}(\bar{X}_n - \mu)}{S_n}$:

$$P_F \left\{ \frac{n^{1/2}(\bar{X}_n - \mu)}{S_n} \leq x \right\} = \Phi(x) + n^{-1/2} \left( \frac{\mu_3}{6\sigma^3}(1 + 2x^2) \right) \phi(x) + O(n^{-1}) \qquad (3.15)$$

uniformly in $x$, where $\sigma$ is the standard deviation of $X$, $\mu_3$ is the third order central moment and $\phi$ and $\Phi$ are, respectively, the density and distribution function of the normal distribution with zero mean and variance equals to one. Indeed, let us consider the Edgeworth expansions for the bootstrap distributions

$$P_{F_n} \left\{ n^{1/2}(\bar{X}_n^* - \bar{X}_n) \leq x \right\}$$

and

$$P_{F_n} \left\{ \frac{n^{1/2}(\bar{X}_n^* - \bar{X}_n)}{S_n} \leq x \right\},$$

obtained from (3.14) and (3.15) by replacing the theoretical parameters by the sampling ones. The bootstrap behaviour is better for the studentized statistic: the bootstrap and normal approximations are equivalent for $n^{1/2}(\bar{X}_n - \mu)$ since they are given by $\Phi\left(\frac{x}{S_n}\right)$ but the bootstrap approximation will be better for $\frac{n^{1/2}(\bar{X}_n - \mu)}{S_n}$ because we have

$$\Phi(x) + n^{-1/2} \left( \frac{\mu_{3n}}{6s_n^3}(1 + 2x^2) \right) \phi(x) + O(n^{-1}) = P_F \left\{ \frac{n^{1/2}(\bar{X}_n - \mu)}{S_n} \leq x \right\} + O(n^{-1}) \quad a.s.,$$

meanwhile the normal approximation error is $O(n^{-1/2})$. This is due to the equivalence between the rate of convergence of the bootstrap approximation and the first order correction of the Edgeworth expansion (see Abramovitch and Singh (1985) for further details). Also in this context, Bose and Babu (1991) have obtained probabilistic bounds for the deviation of the sampling distribution from the bootstrap distribution and they give the rate of convergence to one of the probability that the bootstrap approximation outperform the normal approximation.

The results by Bickel and Freedman (1981) and Singh (1981) have been extended to more general statistics like, for instance, M-statistics, statistics defined by Fréchet differentiable functionals, U-statistics or V-statistics (see section 7). Some asymptotic minimax properties of the bootstrap can be seen in Beran (1984).

There exist, however, some situations where this naïve bootstrap does not work. Babu (1984) showed that this happens for the appropriately normalized sample mean if the population $X$ is a symmetric stable random variable with index $p$, $1 < p < 2$; more specifically, Athreya (1987) proved that if $X$ belongs to the domain of attraction of a nonnormal stable law (and so $EX^2 = \infty$) then the bootstrap version of the correspondingly normalized mean has a random distribution and it does not converge to the stable law (see also Knight (1989) for a different proof of this result). Giné and Zinn (1989) have proved that $EX^2 < \infty$ is a necessary condition for the bootstrap of the sample mean to converge weakly almost surely; they have also established that if the bootstrap of the mean converges weakly in probability then the populational distribution has to be in the domain of attraction of a normal law. Hall (1989-a) characterizes weak convergence in probability for the bootstrap sample mean in terms of the tail behaviour of the populational distribution $F$. All this results give evidence of the naïve bootstrap failing for heavy-tailed distributions. A feasible modification of the bootstrap method is changing the resample size from $n$ to $m_n$ with $m_n = o(n)$; Arcones and Giné (1989) prove that if $X$ is in the domain of attraction of the normal law then the bootstrap of the sample mean converges weakly in probability to the Gaussian distribution for any resampling size $m_n \longrightarrow \infty$ and they also show that if $m_n = o(\frac{n}{\log\log n})$ and $X$ is in the domain of attraction of a stable law, then the bootstrap sample mean converges weakly almost surely but it does not happen if $EX^2 = \infty$ and $\inf_n \frac{m_n(\log\log n)}{n} > 0$. Wu, Carlstein and Cambanis (1989) apply blockwise bootstrap for the mean in the nonnormal stable case.

A different example where the naïve bootstrap is not asymptotically correct is presented by Beran and Srivastava (1985) for the distribution of the eigenvalues of the sample covariance matrix when the populational covariance matrix has multiple eigenvalues; another related situation where inconsistency of the bootstrap is related with ties between the parameters is considered by Hall, Härdle and Simar (1991). Bickel and Freedman (1981) prove that the bootstrap does not work when the statistic of interest is the largest or smallest value of the sample. Basawa *et al.* (1991) give also an example of inconsistent bootstrap in the context of AR(1) models with a unit root (see Section 5).

Efron (1992) uses jackknife-after-bootstrap techniques to study the accuracy and sensitivity of the bootstrap approximations, avoiding further bootstrap resampling. Cuevas and Romo (1991) consider the qualitative robustness for the bootstrap approximation to the distribution of a statistic.

## 4 Bootstrap confidence intervals

One of the main applications of bootstrap methodology is to calculate confidence intervals. Depending on the bootstrap mechanism being used, the differences between the resulting intervals can be important (see, e.g., Efron and Tibshirani (1986), DiCiccio and Romano (1988), Hall (1988-a) and Swanopoel (1990)).

In what follows, the distribution $\hat{F}_n$ producing the resampling data $\mathbf{X}^*$ is either the empirical distribution $F_n$ or (in the parametric bootstrap case) the corresponding parametrically estimated distribution $F_{\hat{\theta}_n}$.

1. *Standard method.* Assume that $\theta = T(F)$ is the parameter of interest. Let $\frac{\sigma^2}{n} = Var_F\hat{\theta} = Var_F T(F_n)$. The standard method is to approximate the distribution of

$$\frac{n^{1/2}(\hat{\theta}-\theta)}{\hat{\sigma}}$$

by a N(0,1) distribution and the interval is given by $\hat{\theta} \pm \frac{\hat{\sigma}}{n^{1/2}} z_{\alpha/2}$, where $\Phi(z_{\alpha/2}) = 1-\alpha/2$. The estimated parameter $\hat{\sigma}$ can be obtained using the delta method, the estimated Fisher information, the bootstrap, the jackknife, etc.

2. *Percentile method.* It is based on approximating $H(s) = P_F\{\hat{\theta} \leq s\}$ by

$$H^*(s) \equiv P_{F_n}\{\hat{\theta}^* \leq s\} = P^*\{\hat{\theta}^* \leq s\};$$

so, the interval is given by $(H^{*-1}(\alpha/2), H^{*-1}(1-\alpha/2))$. In practice, one simulates $B$ bootstrap samples of size $n$, $\mathbf{X}^{*i}, i = 1, \ldots, B$ and consider the ordered values $\hat{\theta}^*_{(1)} \leq, \ldots, \leq \hat{\theta}^*_{(B)}$. The interval $(\hat{\theta}^*_{(r)}, \hat{\theta}^*_{(s)})$ is an approximation to the previous interval, where $r = \left[\frac{B\alpha}{2}\right]$ and $s = \left[\frac{B(1-\alpha)}{2}\right]$.

3. *Bias-corrected percentile method.* It is assumed that there exists a monotonically increasing transformation $g$ such that $\tau(g(\hat{\theta}) - g(\theta)) + z_0$ has a known distribution $K$ symmetric about zero, for some unknown values $\tau$ and $z_0$. Then a theoretical interval will be

$$(H^{-1}(K(2z_0 - z_{\alpha/2})), H^{-1}(K(2z_0 + z_{\alpha/2})))$$

with $z_0 = K^{-1}(H(\theta))$. The value $z_o$ is called bias-correction parameter; there is no correction when $H(\theta) = 1/2$. The corresponding bootstrap interval is given by

$$(H^{*-1}(K(2\hat{z}_0 - z_{\alpha/2})), H^{*-1}(K(2\hat{z}_0 + z_{\alpha/2}))),$$

where $\hat{z}_0 = K^{-1}(H^*(\hat{\theta}))$ and it is known as bias-corrected percentile interval. This method was introduced by Efron (1982), who studied the case $K = \Phi$. The calculation in practice is analogous to the one for the percentil method $(\hat{\theta}^*_{(r)}, \hat{\theta}^*_{(s)})$ but with

$$r = \left[BK(2\hat{z}_0 - z_{\alpha/2})\right], \qquad s = \left[BK(2\hat{z}_0 + z_{\alpha/2})\right]$$

and

$$\hat{z}_0 \simeq K^{-1}\left(\frac{1}{B}\sum_{i=1}^{B} I_{\{\hat{\theta}^{*i} \leq \hat{\theta}\}}\right).$$

If $\hat{z}_0 = 0$ and $K = \Phi$, we get the percentile method.

4. *Accelerated bias-corrected percentile method.* Introduced by Efron (1987), it is assumed that there exists a monotonically increasing transformation $g$ and constants $\tau, z_0$ and $a$ such that

$$\tau\left(\frac{g(\hat{\theta}) - g(\theta)}{1 + a\tau g(\theta)}\right) + z_0$$

has a known distribution $K$ symmetric about zero. A theoretical interval is given by

$$(H^{-1}(K(2z_0 - b_{\alpha/2})), H^{-1}(K(2z_0 + c_{\alpha/2})))$$

with

$$b_{\alpha/2} = \frac{z_{\alpha/2} - z_0}{1 - a(z_0 - z_{\alpha/2})} + z_0, \qquad c_{\alpha/2} = \frac{z_{\alpha/2} + z_0}{1 - a(z_0 + z_{\alpha/2})} - z_0.$$

$z_0$ is the bias-correction parameter and $a$ is the acceleration constant; as in 3, $z_0 = K^{-1}(H(\theta))$. If $a = 0$, it becomes case 3. The corresponding bootstrap interval is given by

$$(H^{*-1}(K(2\hat{z}_0 - \hat{b}_{\alpha/2})), H^{*-1}(K(2\hat{z}_0 + \hat{c}_{\alpha/2}))),$$

where $\hat{z}_0 = K^{-1}(H^*(\hat{\theta}))$ and $\hat{b}_{\alpha/2}$ and $\hat{c}_{\alpha/2}$ are estimated from $\hat{z}_0$ and $\hat{a}$. Of course, the main difficulty with this method relies on the estimation of $a$. Efron (1987) gives the choices:

$$\hat{a} = \frac{1}{6}\gamma_3 \frac{\partial}{\partial\theta} \log f_\theta(X) \mid_{\theta=\hat{\theta}}$$

($\gamma_3$ is the skewness coefficient) for parametric bootstrap, and

$$\hat{a} \simeq \frac{1}{6} \frac{\sum_{i=1}^n V_i^3}{(\sum_{i=1}^n V_i^2)^{3/2}},$$

where $V_i = \lim_{\epsilon \to 0} \frac{1}{\epsilon}(T((1-\epsilon)F_n + \epsilon\delta x_i) - T(F_n)), i = 1, \ldots, n$, for nonparametric bootstrap.

As in 2 and 3, in practice, the bootstrap interval will be given by $(\hat{\theta}^*_{(r)}, \hat{\theta}^*_{(s)})$ with $r = \left[BK(2\hat{z}_0 - \hat{b}_{\alpha/2})\right]$ and $s = \left[BK(2\hat{z}_0 + \hat{c}_{\alpha/2})\right]$.

5. *Pivotal method.* Is was introduced by Beran (1987) and studied in Beran (1988, 1990). If $Q(\hat{\theta}, \theta)$ is a pivot for $\theta$ with distribution $H$ under the populational distribution $F$, it is known that $H(Q(\hat{\theta}^*, \hat{\theta}))$ is approximately uniform. If $Q(\hat{\theta}, \theta)$ is monotone in $\theta$, using the bootstrap quantiles of order $\alpha/2$ and $1-\alpha/2$, a confidence interval can be calculated. For instance, let $\frac{\hat{\sigma}^2}{n}$ be an estimation of $Var_F\hat{\theta}$. If $H$ is the distribution of

$$Q(\hat{\theta}, \theta) = \frac{n^{1/2}(\hat{\theta} - \theta)}{\hat{\sigma}},$$

then $H\left(\frac{n^{1/2}(\hat{\theta}-\theta)}{\hat{\sigma}}\right)$ follows a uniform distribution $H_1$ and

$$P_F\left\{\alpha/2 \le H\left(\frac{n^{1/2}(\hat{\theta} - \theta)}{\hat{\sigma}}\right) \le 1 - \alpha/2\right\} = 1 - \alpha$$

would give a pivotal confidence interval which cannot be calculated since $H$ is not known. In the bootstrap context,

$$\frac{n^{1/2}(\hat{\theta}^* - \hat{\theta})}{\hat{\sigma}^*}$$

and

$$H^*\left(\frac{n^{1/2}(\hat{\theta}^* - \hat{\theta})}{\hat{\sigma}^*}\right)$$

have distributions $H^*$ and $H_1^*$, respectively; so, $H^{-1}(\alpha/2)$ and $H^{-1}(1-\alpha/2)$ can be approximated by $H^{*-1}(H_1^{*-1}(\alpha/2))$ and $H^{*-1}(H_1^{*-1}(1-\alpha/2))$. The difficulty with this method is that the bootstrap mechanism has to be carried out twice. The required steps are the following:

(a) Obtain the bootstrap sample $\mathbf{X}^{*j}, j = 1, \ldots, B$ and consider the values

$$\frac{n^{1/2}(\hat{\theta}^{*j} - \hat{\theta})}{\hat{\sigma}^{*j}} \qquad , \qquad j = 1, \ldots, B.$$

Its empirical distribution tends to $H^*$ when $B \to \infty$.

(b) From each $\mathbf{X}^{*j}, j = 1, \ldots, B$ get $N$ bootstrap samples $\mathbf{X}^{*jk}, k = 1, \ldots, N$— wich will be conditionally independent given $\mathbf{X}$ and $\mathbf{X}^*, j = 1, \ldots, B$. If $Z_j, j = 1, \ldots, B$ is the fraction of values

$$\frac{n^{1/2}(\hat{\theta}^{*jk} - \hat{\theta}^{*j})}{\hat{\sigma}^{*jk}}, \qquad k = 1, \ldots, N$$

which are less than or equal to

$$\frac{n^{1/2}(\hat{\theta}^{*j} - \hat{\theta})}{\hat{\sigma}^{*j}},$$

the empirical distribution of $Z_j, j = 1, \ldots, B$ is used to approximate the distribution $H_1^*$.

6. *Percentile-t method.* Considered by Hall (1986, 1988-a), this technique applies the percentile method idea to the statistic

$$\frac{n^{1/2}(\hat{\theta} - \theta)}{\hat{\sigma}}.$$

Two versions can be considered. The *symmetrized percentile-t*, which tries to estimate the value $\theta(\alpha)$ satisfying

$$P_F\left\{n^{1/2}\left|\frac{\hat{\theta} - \theta}{\hat{\sigma}}\right| \leq \theta(\alpha)\right\} = 1 - \alpha,$$

using the quantity $\theta^*(\alpha)$ such that

$$P_F\left\{n^{1/2}\left|\frac{\hat{\theta}^*-\hat{\theta}}{\hat{\sigma}^*}\right|\leq\theta^*(\alpha)\right\}=1-\alpha,$$

and the *equal tails percentile-t* whose goal is to estimate $\theta_{1\alpha}$ and $\theta_{2\alpha}$ verifying

$$P_F\left\{\theta_{1\alpha}\leq\frac{n^{1/2}(\hat{\theta}-\theta)}{\hat{\sigma}}\leq\theta_{2\alpha}\right\}=1-\alpha$$

through $\theta^*_{1\alpha}$ and $\theta^*_{2\alpha}$. The values $\theta_{1\alpha}$ and $\theta_{2\alpha}$ are the order $\alpha/2$ and $1-\alpha/2$ quantiles of

$$H(s)=P_F\left\{\frac{n^{1/2}(\hat{\theta}-\theta)}{\hat{\sigma}}\leq s\right\}$$

and $\theta^*_{1\alpha}$ and $\theta^*_{2\alpha}$ are the corresponding quantiles for

$$H^*(s)=P_{F_n}\left\{\frac{n^{1/2}(\hat{\theta}^*-\hat{\theta})}{\hat{\sigma}^*}\leq s\right\}.$$

In practice, one has to proceed in the same way as for the percentile method but considering the values

$$\frac{n^{1/2}(\hat{\theta}^{*i}-\hat{\theta})}{\hat{\sigma}^{*i}},\qquad i=1,\ldots,B$$

or

$$\left|\frac{n^{1/2}(\hat{\theta}^{*i}-\hat{\theta})}{\hat{\sigma}^{*i}}\right|,\qquad i=1,\ldots,B,$$

depending on the considered percentile-t.

7. *Methods based on Edgeworth expansions.* Assuming that the statistic

$$\frac{n^{1/2}(\hat{\theta}-\theta)}{\hat{\sigma}}$$

admits an Edgeworth expansion of the form

$$P_F\left\{\frac{n^{1/2}(\hat{\theta}-\theta)}{\hat{\sigma}}\leq x\right\}=\Phi(x)+n^{-1/2}\phi(x)\pi_1(x)+n^{-1}\phi(x)\pi_2(x)+\ldots,$$

where $\pi_i(x), i=1,2,\ldots,$ are polynomials in $x$ with coefficients depending on the populational moments, several confidence intervals can be calculated, inverting this expansion (Hall, 1983) or using the bootstrap approximation to these inversions (Abramovitch and Singh (1985) and Rayner(1989)).

From the available results, the following conclusions can be drawn:

(a) Among all percentile methods (2, 3 and 4), the accelerated bias-corrected percentile is the one performing best.

(b) Methods 4 and 6 give intervals with theoretical covering error of order $O(n^{-1})$, and methods 1, 2 and 3 are of order $O(n^{-1/2})$.

(c) Methods 4 (Efron, 1987) and 6 (Hall, 1988-a) present different properties; the former is scale invariant and the later has good computational properties.

(d) Hall (1988-a) gives a covering error of order $O(n^{-2})$ for the symmetrized percentile-t and $O(n^{-1})$ for the equal tails percentile-t. The pivotal method 5 applied to the symmetrized percentile-t leads to an error $O(n^{-3})$ (Beran, 1987).

(e) Rayner (1987) presents simulation results for several methods of type 7.

| | $\rho = 0.1$ | $\rho = 0.3$ | $\rho = 0.5$ | $\rho = 0.7$ | $\rho = 0.9$ |
|---|---|---|---|---|---|
| TWO-SIDED | | | | | |
| $1+\rho$ | | | | | |
| Stand | 1.304(1.00) | 1.449(1.00) | 1.656(1.00) | 1.874(1.00) | 2.100(1.00) |
| Per | 0.514(0.96) | 0.554(0.96) | 0.547(0.94) | 0.410(0.92) | 0.167(0.93) |
| A-Per | 0.717(0.83) | 0.626(0.85) | 0.548(0.93) | 0.410(0.93) | 0.169(0.93) |
| Sim-t | 0.552(0.96) | 0.604(0.92) | 0.626(0.89) | 0.461(0.91) | 0.178(0.91) |
| Eq-t | 0.460(0.95) | 0.552(0.69) | 0.593(0.81) | 0.445(0.85) | 0.173(0.85) |
| $1-\rho$ | | | | | |
| Stand | 0.912(0.94) | 0.768(0.91) | 0.561(0.89) | 0.342(0.87) | 0.117(0.84) |
| Per | 0.516(0.96) | 0.555(0.96) | 0.546(0.94) | 0.408(0.92) | 0.166(0.93) |
| A-Per | 0.391(0.83) | 0.512(0.85) | 0.541(0.93) | 0.407(0.93) | 0.167(0.94) |
| Sim-t | 1.029(0.98) | 0.923(0.97) | 0.760(0.96) | 0.535(0.96) | 0.213(0.94) |
| Eq-t | 0.788(0.77) | 0.768(0.77) | 0.683(0.88) | 0.486(0.92) | 0.192(0.93) |
| ONE-SIDED | | | | | |
| $1+\rho$ | | | | | |
| Stand | 0.629(1.00) | 0.698(1.00) | 0.798(1.00) | 0.904(1.00) | 1.012(1.00) |
| Per | 1.037(0.91) | 1.095(0.93) | 1.241(0.92) | 1.491(0.92) | 1.810(0.91) |
| A-Per | 0.753(0.91) | 1.004(0.93) | 1.234(0.93) | 1.485(0.93) | 1.804(0.93) |
| t | 0.934(0.83) | 1.085(0.85) | 1.310(0.84) | 1.562(0.83) | 1.843(0.83) |
| $1-\rho$ | | | | | |
| Stand | 0.440(1.00) | 0.370(1.00) | 0.270(1.00) | 0.165(0.98) | 0.056(0.94) |
| Per | 0.510(1.00) | 0.420(1.00) | 0.292(0.96) | 0.167(0.96) | 0.053(0.96) |
| A-Per | 0.632(0.84) | 0.462(0.87) | 0.295(0.95) | 0.170(0.96) | 0.055(0.96) |
| t | 0.714(0.77) | 0.542(0.77) | 0.342(0.88) | 0.190(0.93) | 0.060(0.93) |

Table 4.1.

To compare the previous methods, we have carried out a simulation study divided into two parts. First, we have obtained $N = 500$ samples with size 25 from a bidimensional normal variable $X$ with distribution

$$N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right).$$

The parameters we consider are the eigenvalues of the covariance matrix $\lambda_1 = 1 + \rho$ and $\lambda_2 = 1 - \rho$. From the corresponding sample values, $\hat{\lambda}_1$ and $\hat{\lambda}_2$, using the statistics

$$\frac{n^{1/2}(\hat{\lambda}_i - \lambda_i)}{(2\hat{\lambda}_i)^{1/2}}, \qquad i = 1, 2,$$

we have calculated right one-sided and two-sided confidence intervals for methods 1 (Stand. in the table), 2 (Per in the tables), 4 (A-Per in the tables) and symmetrized and equal tails 6 (Sim-t and Eq-t in the "two-sided" tables and t in the "one-sided" ones), with $B = 1000$. Table 4.1 shows the mean lengths and mean coverings (between parentheses) for a theoretical covering value of 0.95 and different values of $\rho$; in the "one-sided" case, the mean lengths are replaced by the mean critical points.

Next, to estimate the mean $a$ of a variable X of the form

$$\frac{a\chi_k^2}{k}, \qquad a \in \mathcal{R},$$

we have carried out a simulation study with the same technical features. Table 4.2 presents the results for different values of $k$ and $a = 2$.

| | $k=1$ | $k=2$ | $k=3$ | $k=4$ | $k=5$ |
|---|---|---|---|---|---|
| TWO-SIDED | | | | | |
| ($a = 2$) | | | | | |
| Stand | 2.124(0.89) | 1.546(0.91) | 1.231(0.91) | 1.085(0.94) | 0.970(0.92) |
| Per | 2.061(0.89) | 1.504(0.90) | 1.199(0.92) | 1.056(0.94) | 0.945(0.92) |
| A-Per | 2.217(0.90) | 1.576(0.91) | 1.237(0.93) | 1.083(0.93) | 0.963(0.94) |
| Sim-t | 3.002(0.94) | 1.912(0.94) | 1.453(0.95) | 1.245(0.96) | 1.096(0.95) |
| Eq-t | 2.839(0.93) | 1.856(0.94) | 1.419(0.95) | 1.227(0.95) | 1.083(0.96) |
| ONE-SIDED | | | | | |
| ($a = 2$) | | | | | |
| Stand | 1.149(0.97) | 1.384(0.97) | 1.482(0.97) | 1.554(0.97) | 1.589(0.97) |
| Per | 1.230(0.95) | 1.432(0.96) | 1.516(0.96) | 1.582(0.96) | 1.611(0.97) |
| A-Per | 1.337(0.92) | 1.498(0.94) | 1.560(0.96) | 1.618(0.94) | 1.641(0.95) |
| t | 1.289(0.94) | 1.464(0.95) | 1.534(0.96) | 1.594(0.96) | 1.622(0.97) |

Table 4.2.

In the first problem, the *bootstrap-t* does not behave as well as expected; in fact, Efron has remarked that the *bootstrap-t* is troublesome in the correlation coefficent problem. However, the percentile method behaves very well, in opposition to the very conservative standard method. An asymptotic study for this example could be carried out using Edgeworth expansions (see Beran and Srivastava, 1985).

# 5 THE BOOTSTRAP IN LINEAR REGRESSION MODELS

We will consider three different models: **(a)** $Y_i = \mathbf{x}_i^t\beta + \varepsilon_i, i = 1,\dots,n$, where $\mathbf{x}_i \in \mathcal{R}^k$, $\beta$ is the unknown k-dimensional parameter and $\varepsilon_i, i = 1,\dots,n$ are independent and identically distributed errors with zero mean, **(b)** $(\mathbf{X}_i^t, Y_i), i = 1,\dots,n$ are independent and identically distributed random variables with distribution function $F$ and we look for the vector $\beta$ minimizing $E_F|Y - \mathbf{X}\beta|^2$, and **(c)** as in (a), but allowing for independent errors $\varepsilon_i, i = 1,\dots,n$, with distribution depending on $\mathbf{x}_i^t, i = 1,\dots,n$—e. g., through the variance (heteroskedastic models). For each of them, the bootstrap should be adapted to capture the specific features of the model.

For model **(a)**, we present the bootstrap as an alternative to the jackknife. Let $\hat{\beta} = (X^tX)^{-1}X^t\mathbf{Y}$ be the least squares estimate of $\beta$, where

$$X \equiv \begin{pmatrix} \mathbf{x}_1^t \\ \vdots \\ \mathbf{x}_n^t \end{pmatrix} \qquad \textit{and} \qquad \mathbf{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix},$$

and let us consider the estimation of $Cov_F\hat{\beta}$, where $F$ is the distribution of $\varepsilon_i, i = 1,\dots,n$. The following jackknife estimates have been proposed:

1. *Ordinary jackknife.* From (2.9) and (2.10), with $r_i = Y_i - \mathbf{x}_i^t\hat{\beta}$,

$$Cov_{jack}\hat{\beta} = \frac{n-1}{n}\sum_{i=1}^{n}(\hat{\beta}_{(i)} - \overline{\hat{\beta}})(\hat{\beta}_{(i)} - \overline{\hat{\beta}})^t \simeq \frac{n-1}{n}(X^tX)^{-1}(\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i^t r_i^2)(X^tX)^{-1} \tag{5.16}$$

   where $\hat{\beta}_{(i)}$ is the least squares estimate deleting the $i-th$ observation, $i = 1,\dots,n$ and

$$\overline{\hat{\beta}} = \frac{\sum_{i=1}^{n}\hat{\beta}_{(i)}}{n}.$$

2. *Hinkley's (1977) jackknife.* It is based on a modification of the previous jackknife. We have that

$$Cov_{jack}\hat{\beta} = \frac{1}{n(n-1)}\sum_{i=1}^{n}(\hat{\beta}^{(i)} - \overline{\hat{\beta}}')(\hat{\beta}^{(i)} - \overline{\hat{\beta}}')^t,$$

   where $\hat{\beta}^{(i)} = n\hat{\beta} - (n-1)\hat{\beta}_{(i)} = \hat{\beta} + (n-1)(\hat{\beta} - \hat{\beta}_{(i)})$ is the $i-th$ jackknife pseudovalue and

$$\overline{\hat{\beta}}' = \frac{\sum_{i=1}^{n}\hat{\beta}^{(i)}}{n}.$$

   So, taking $\hat{\beta}^{(i)H} = \hat{\beta} + n(1 - \omega_i)(\hat{\beta} - \hat{\beta}_{(i)})$, where

$$\omega_i = \mathbf{x}_i^t(X^tX)^{-1}\mathbf{x}_i, \qquad i = 1, \ldots, n,$$

we have

$$Cov_{jack-H}\hat{\beta} = \frac{1}{n(n-k)}\sum_{i=1}^{n}(\hat{\beta}^{(i)H} - \overline{\hat{\beta}}^{H})(\hat{\beta}^{(i)H} - \overline{\hat{\beta}}^{H})^t$$
$$\simeq (X^tX)^{-1}(\sum_{i=1}^{n}\frac{r_i^2}{1-\frac{k}{n}}\mathbf{x}_i\mathbf{x}_i^t)(X^tX)^{-1}$$

with $\overline{\hat{\beta}}^{H} = \frac{\sum_{i=1}^{n}(\hat{\beta}^{(i)H})}{n}$. This mechanism reflects the weighted character of regression by replacing $\frac{1}{n}$ in method (a) by $\omega_i$.

3. *Wu's (1986) jackknife.* It is a generalization of the ordinary jackknife, giving

$$Cov_{jack-W}\hat{\beta} = (X^tX)^{-1}(\sum_{i=1}^{n}\frac{r_i^2}{1-\omega_i}\mathbf{x}_i\mathbf{x}_i^t)(X^tX)^{-1}. \tag{5.17}$$

Since $Cov_F\hat{\beta} = \sigma^2(X^tX)^{-1}$, assuming that $E\varepsilon^2 = \sigma^2 < \infty$, it follows that methods (a) and (b) are biased and (c) is unbiased. Hinkley (1977) proves robustness properties of (b) in the heterokedastic case. Compared to these three jackknife procedures, the bootstrap proceeds in the following way:

(a) Get a bootstrap sample $\varepsilon_i^*, i = 1, \ldots, n$ from the empirical distribution of

$$\hat{\varepsilon}_i = r_i = y_i - \mathbf{x}_i^t\hat{\beta}, \quad i = 1, \ldots, n; \qquad \hat{\varepsilon}_i - \overline{\hat{\varepsilon}}, \quad i = 1, \ldots, n$$
$$or \quad \frac{\hat{\varepsilon}_i}{(\frac{n-k}{n})^{1/2}}, \quad i = 1, \ldots, n$$

ordinary, centered or standardized residuals, respectively.

(b) Obtain the bootstrap values $y_i^* = \mathbf{x}_i^t\hat{\beta} + \varepsilon_i^*, i = 1, \ldots, n$ and calculate the bootstrap least squares estimate $\hat{\beta}^*$.

(c) Replicate (a) and (b) a large number $B$ of times.

For each type of residuals obtained in (a), we get:

$$E^*\hat{\beta}^* \neq \hat{\beta}, \quad Cov^*\hat{\beta}^* = Cov_{boot}\hat{\beta} = \frac{1}{n}\sum_{i=1}^{n}r_i^2(X^tX)^{-1} \quad (ordinary \;\; residuals)$$

$$E^*\hat{\beta}^* = \hat{\beta}, \quad Cov^*\hat{\beta}^* = \frac{1}{n}\sum_{i=1}^{n}(r_i - \overline{r})^2(X^tX)^{-1} \qquad (centered \;\; residuals)$$

$$E^*\hat{\beta}^* \neq \hat{\beta}, \quad Cov^*\hat{\beta}^* = \frac{1}{n-k}\sum_{i=1}^{n}r_i^2(X^tX)^{-1}. \qquad (standardized \;\; residuals)$$

The bootstrap can be also used to approximate the distribution of $n^{1/2}(\hat{\beta}-\beta)$,

$$P_F\{n^{1/2}(\hat{\beta}-\beta)\leq x\}, \tag{5.18}$$

by

$$P_{F_n}\{n^{1/2}(\hat{\beta}^*-\hat{\beta})\leq x\}, \tag{5.19}$$

where $F_n$ is any of the three previous empirical distributions. Under regularity conditions, Freedman (1981) has established that (5.19) is a good approximation to (5.18) and Navidi (1989) has shown that it is better than the normal approximation. Hall (1989-b) has obtained confidence intervals with excellent approximations to the theoretical covering probability.

For model **(b)**, the resampling has to be carried out from the empirical distribution of $(\mathbf{X}_i^t, \mathbf{Y}_i), i=1,\ldots,n$ to get $(\mathbf{X}_i^{t*}, Y_i^*), i=1,\ldots,n$. Repeating $B$ times this process, we will have

$$\hat{\beta}^{*j}=(X^{t*j}X^{*j})^{-1}X^{t*j}\mathbf{Y}^{*j},$$

with

$$X^{*j}=\begin{pmatrix}\mathbf{X}_1^{t*j}\\ \vdots \\ \mathbf{X}_n^{t*j}\end{pmatrix} \qquad \textit{and} \qquad \mathbf{Y}^{*j}=\begin{pmatrix}\mathbf{Y}_1^{*j}\\ \vdots \\ \mathbf{Y}_n^{*j}\end{pmatrix}, \qquad j=1,\ldots,B,$$

to approximate $Cov_{boot}\hat{\beta}$. This resampling mechanism is well suited to model **(b)**; when applied to model **(a)**, leads to inconsistent procedures (Wu, 1986). Recently, Stute (1990) has shown, under very weak conditions, that

$$d_\infty(P_F\{n^{1/2}(\hat{\beta}-\beta)\leq\cdot\}, P_{F_n}\{n^{1/2}(\hat{\beta}^*-\hat{\beta})\leq\cdot\})\to 0 \qquad a.s.$$

Finally, if we consider heteroskedasticity in model **(c)**, $E\epsilon_i^2=\sigma_i^2$, it holds that

$$Cov_{F_1,\ldots,F_n}\hat{\beta}=(X^tX)^{-1}(\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i^t\sigma_i^2)(X^tX)^{-1},$$

where $F_1,\ldots,F_n$ are the distributions of the errors $\varepsilon_1,\ldots,\varepsilon_n$. The bootstrap estimators given for model **(a)** are not consistent in this set up; so, the bootstrap behaves worse than Wu's (1986) jackknife (5.17) ,which is unbiased under certain conditions.

Wu (1986) has proposed a bootstrap technique adapted to model **(c)**:

(a) For each $i$, sample a value $t_i^*$ from a distribution with zero mean and variance one.

(b) Get $Y_i^*=\mathbf{x}_i^t\hat{\beta}+\frac{r_i}{\sqrt{1-\omega_i}}t_i^*, i=1,\ldots,n$, and the corresponding estimate $\hat{\beta}^*$.

For this bootstrap mechanism, $E^*\hat{\beta}^*=\hat{\beta}$ and

$$Cov^*\hat{\beta}^*=(X^tX)^{-1}(\sum_{i=1}^{n}\frac{r_i^2}{1-w_i}\mathbf{x}_i\mathbf{x}_i^t)(X^tX)^{-1},$$

which coincides with Wu's estimation.

Next, we present an application of the bootstrap to linear regression from González Manteiga *et al.* (1990). The model expresses the dependence between fusion temperatures of certain ashes and their chemical composition. The goal of this study was to prevent the growing of incrustations on the chimney walls of the As Pontes thermic power station. From 144 observations with dimension 14 (4 corresponding to temperatures and 10 to chemical components), a new variable $C$ based on the principal component of the 4 temperatures was introduced and, through stepwise regression, the two most important chemical variables, $U = Al_2O_3(\%)$ and $V = Fe_2O_3(\%)$, were chosen. A correlation model $((U, V), C)$ was considered. The density of the residuals

$$\hat{\varepsilon}_i = C_i - \hat{\beta}_1 - \hat{\beta}_2 U_i - \hat{\beta}_3 V_i, \qquad i = 1, \dots, 144,$$

was estimated nonparametrically and the hypothesis of normality was clearly accepted (see Figure 5.1).

Figure 5.1

Then, classical (normal theory) and bootstrap confidence intervals were calculated for $\beta_i, i = 1, 2, 3$ and for the multiple correlation coefficient $\rho$, with $B = 1000$ and confidence level 0.05 (see Table 5.2). Although the most simple bootstrap version —ordinary percentile— was used, it was fairly competitive in a situation where the classical confidence intervals are optimal.

| | Classic | Bootstrap |
|---|---|---|
| $\beta_1$ | (-1.056, -0.156) | (-1.084, -0.211) |
| $\beta_2$ | (-0.132, -0.088) | (-0.134, -0.085) |
| $\beta_3$ | (0.305, 0.384) | (0.296, 0.405) |
| $\rho$ | (0.794, 0.890) | (0.810, 0.881) |

Table 5.2

Bootstrap for regression models with binary response has been considered by Sauermann (1989); Huet and Jolivet (1989) apply bootstrap and Edgeworth expansions to nonlinear regression models and Huet *et al.* (1990) present simulation results in this case.

The study of the bootstrap for time series and dynamic regression models was started by Freedman (1984). Bose (1988) has shown that, under some regularity conditions, the bootstrap approximation to the distribution of the least-squares estimator in stationary autoregressive models is of order $o(n^{-\frac{1}{2}})$ a.s., improving on the normal approximation (which is $O(n^{-\frac{1}{2}})$); Thombs and Schucany (1990) give bootstrap prediction intervals in this case. The validity of the bootstrap for the least squares estimator in explosive AR(1) models

$$X_t = \beta X_{t-1} + \varepsilon_t, \qquad |\beta| > 1,$$

has been established by Basawa *et al.* (1989) and Stute and Gründer (1990) have obtained bootstrap approximations to prediction intervals in this case. Basawa *et al.* (1991) prove that the bootstrapped least squares estimator has a random limit distribution for the unstable first-order autoregressive model ($\beta = 1$). Under no model assumptions, Künsch (1989) investigates blockwise bootstrap for stationary observations.

## 6 Bootstrap Prediction Error Estimation

Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a random sample, with $X_i = (T_i^t, Y_i)$, where $T_i$ is a p-dimensional predictors vector and $Y_i$ is a one-dimensional response variable having $(p+1)$-dimensional joint distribution $F$. The goal is to predict $Y_0$, given a future observation $T_0$; this means to give a value to the response variable corresponding to the predictors, by using a decision rule

$$\eta = \eta(T_0^t, \mathbf{X}) = \eta(T_0^t, F_n),$$

and minimizing

$$E_F Q(Y_0, \eta(T_0^t, \mathbf{X})).$$

The function $Q(y, \eta)$ is the incurrred loss when predicting $y$ with $\eta$; our goal is to estimate its expectation with respect to the distribution $F$ of $(T_0^t, Y_0)$ —the prediction error.

We will focus on the following situations:

(a) *Linear regression.* Consider, e.g., model **(b)** in Section 5 (analogously model **(a)**, replacing $T$ by a deterministic $t$). We have

$$\eta(T_0^t, \mathbf{X}) = T_0^t \hat{\beta},$$

where

$$\hat{\beta} = (\mathbf{T}^t\mathbf{T})^{-1}\mathbf{T}^t\mathbf{Y},$$

with

$$\mathbf{T} = \begin{pmatrix} T_1^t \\ \vdots \\ T_n^t \end{pmatrix} \qquad \textit{and} \qquad \mathbf{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix},$$

is the least squares estimator of $\beta$. If $Q(y,\eta) = (y-\eta)^2$ , we are estimating the least squares prediction error of the linear prediction rule.

(b) *Discriminant analysis.* In this set up, $Y_i$ is dichotomous, taking values 0 or 1 when the corresponding predictor variable $T_i^t$ belongs to each of the populations we are trying to discriminate. Let

$$\eta(T_0^t, \mathbf{X}) = \begin{cases} 0, & if \quad T_0^t\hat{\beta} < c \\ 1, & if \quad T_0^t\hat{\beta} \geq c, \end{cases}$$

where $c \in \mathcal{R}$ and $\hat{\beta} = (\overline{T}_1 - \overline{T}_2)S^{-1}$, with

$$S = \frac{1}{n}\left(\sum_{j=1}^{2} T_j T_j^t - n_1 \bar{T}_1 \bar{T}_1^t - n_2 \bar{T}_2 \bar{T}_2^t\right),$$

$$n_y = \#\{Y_j = y\}, \qquad \bar{T}_y = \frac{1}{n_y}\sum_{Y_i = y} T_i,$$

and $Q(y,\eta) = 1_{\{\eta \neq y\}}$. The goal is to estimate the probability of misclassification of the Fisher linear discriminant rule.

(c) *Logistic regression model.* If we put

$$\pi_i = P\{Y_i = 1|T_i\} = \frac{1}{1 + \exp\{-T_i^t\beta\}}$$

in model (b), we get a logistic regression model. In this case, usually is

$$\eta(T_0^t, \mathbf{X}) = \begin{cases} 0, & \text{if } \hat{\pi}_0 \geq c \\ 1, & \text{if } \hat{\pi}_0 < c \end{cases},$$

$c \in (0,1)$, where

$$\hat{\pi}_0 = \frac{1}{1 + \exp\{-T_0^t\hat{\beta}\}}$$

and $\hat{\beta}$ is the maximum likelihood estimator corresponding to the pseudolikelihood

$$\prod_{i=1}^{n} \pi_i^{Y_i}(1-\pi_i)^{1-Y_i}.$$

The objective will be to estimate the misclassification probability for the logistic discrimination rule.

The natural way to estimate the prediction error

$$E(\mathbf{X}, F) = E_F Q(Y_0, \eta(T_0^t, \mathbf{X}))$$

is the so called *apparent error*

$$E_{ap} = E_{ap}(\mathbf{X}) = E_{F_n} Q(Y_0, \eta(T_0^t, \mathbf{X}))$$

For (a) and (b), $E_{ap}$ is, respectively,

$$\frac{1}{n}\sum_{i=1}^{n}(Y_i - T_i^t\hat{\beta})^2 = \frac{1}{n}\sum_{i=1}^{n} r_i^2$$

and

$$\frac{\#\{\eta(T_0^t, \mathbf{X}) \neq Y_i\}}{n}.$$

Unfortunately, this procedure leads, in general, to underestimate the errors because the same sample is used for the estimation and for the prediction rule validation. Thus, it is important to estimate the *prediction error excess*,

$$R = R(\mathbf{X}, F) = E(\mathbf{X}, F) - E_{ap}(\mathbf{X}) \tag{6.20}$$

(see Efron (1982, 1983, 1986), Gong (1986), Stine (1985) and Bunke and Droge(1984)).

Let $w = w(F) = E_{\mathbf{X}} R(\mathbf{X}, F)$. It seems reasonable to correct the optimism of the apparent error by estimating $w$ and then defining a new estimation of the prediction error using $E_{ap} + \hat{w}$.

The jackknife estimation is

$$\begin{aligned} E_{ap} + \hat{w}^{jack} &= E_{ap} + \frac{1}{n}\sum_{i=1}^{n} Q(Y_i, \eta(T_i^t, \mathbf{X}_{(i)})) - \\ &- \frac{1}{n^2}\sum_{i,j=1}^{n} Q(Y_j, \eta(T_j^t, \mathbf{X}_{(i)})) \end{aligned} \tag{6.21}$$

and the cross-validation estimation is

$$E_{ap} + \hat{w}^{CV} = \frac{1}{n}\sum_{i=1}^{n} Q(Y_i, \eta(T_i^t, \mathbf{X}_{(i)})). \tag{6.22}$$

Efron (1982, 1983, 1986) introduced the bootstrap estimation

$$\begin{aligned} E_{ap} + \hat{w}^{boot} &= E_{ap} + E^* R(\mathbf{X}^*, F_n) = \\ &= E_{ap} + E^*\left\{\sum_{i=1}^{n}(\frac{1}{n} - p_i^*) Q(Y_i, \eta(T_i^t, \mathbf{X}^*))\right\}, \end{aligned} \tag{6.23}$$

where $\mathbf{X}^* = (X_1^*, \ldots, X_n^*) = ((T_1^{t*}, Y_1^*), \ldots, (T_n^{t*}, Y_n^*))$ is the bootstrap sample and

$$p_i^* = \frac{\#\{X_j^* = X_i\}}{n}, \qquad i = 1, \ldots, n.$$

The bootstrap estimation (6.23) is usually obtained by simulating $B$ samples $\mathbf{X}^{*b}, b = 1, \ldots, B$, for a large $B$, and calculating

$$E_{ap} + \frac{1}{B}\sum_{i=1}^{B}\left\{\sum_{i=1}^{n}(\frac{1}{n} - p_i^{*b})Q(Y_i, \eta(T_i^t, \mathbf{X}^{*b}))\right\}.$$

A comparative study for methods (6.21), (6.22) and (6.23) has been carried out, from both the theoretical and the applied points of view, in the linear regression set up (Efron, 1982, Bunke and Droge, 1984 and Stine, 1985), for logistic models (Gong, 1986 and Efron, 1986) and for discriminant analysis (Efron, 1983). In this last context, Prada Sánchez and Otero Cepeda (1989) present the following simulation results. Consider a population $X = (T^t, Y)$, where $T|Y = y$ has a distribution $N_2((y - \frac{1}{2}), I)$, with $P\{Y = 0\} = P\{Y = 1\} = \frac{1}{2}$. A sample with size 14 is simulated and the Fisher discriminant linear optimal rule is estimated. The true prediction error, $E$, the apparent one, $E_{ap}$, and the corresponding excess $R(\mathbf{X}, F)$ are calculated from it. This process is iterated 100 times, considering different estimations of the estimated expected excess in the process ($B = 200$). Table 6.1 gives the first ten estimations and the average final results.The last three columns in the table present alternative bootstrap techniques which will be introduced in Section 8. Also, the mean squared error is given for the estimations of $E$ through $E_{ap} + \hat{w}$, with five different values for $\hat{w}$. Since

$$\begin{aligned} M.S.E. &= E\left((E_{ap} + \hat{w}) - (E_{ap} + R)\right)^2 = \\ &= (E\hat{w} - w)^2 + Var\ \hat{w} - 2\ cov\ (\hat{w}, R) + Var\ R, \end{aligned}$$

the sign of $cov\ (\hat{w}, R)$ is very relevant. Comparing the methods in this case, the bootstrap behaves better than cross-validation: a larger bias, but smaller overall mean squared error.

| | $E$ | $E_{ap}$ | $R(X,F)$ | $\hat{w}^{CV}$ | $\hat{w}^{boot}$ | $\hat{w}^{s-bt}$ | $\hat{w}^{b-bt}$ | $\hat{w}^{bs-bt}$ |
|---|---|---|---|---|---|---|---|---|
| | 0.319 | 0.142 | 0.176 | 0.000 | 0.052 | 0.063 | -0.000 | 0.032 |
| | 0.316 | 0.142 | 0.173 | 0.000 | 0.033 | 0.051 | 0.001 | 0.095 |
| | 0.330 | 0.214 | 0.116 | 0.071 | 0.076 | 0.055 | 0.000 | 0.021 |
| | 0.337 | 0.357 | -0.019 | 0.071 | 0.093 | 0.048 | -0.001 | -0.060 |
| | 0.309 | 0.142 | 0.166 | 0.142 | 0.061 | 0.086 | -0.000 | 0.133 |
| | 0.338 | 0.214 | 0.124 | 0.142 | 0.081 | 0.115 | 0.001 | 0.095 |
| | 0.380 | 0.071 | 0.308 | 0.071 | 0.056 | 0.136 | -0.002 | 0.158 |
| | 0.324 | 0.357 | -0.033 | 0.000 | 0.081 | 0.078 | -0.020 | -0.046 |
| | 0.321 | 0.071 | 0.249 | 0.071 | 0.048 | 0.124 | -0.001 | 0.142 |
| | 0.360 | 0.422 | -0.067 | 0.071 | 0.104 | 0.097 | 0.006 | -0.086 |
| Mean | 0.349 | 0.252 | 0.097 | 0.100 | 0.077 | 0.091 | 0.000 | 0.050 |
| Dev. | 0.057 | 0.128 | 0.126 | 0.089 | 0.027 | 0.027 | 0.007 | 0.073 |
| Corr. | | | | 0.005 | -0.663 | -0.245 | 0.011 | 0.749 |
| M.S.E. | | | | 0.023 | 0.021 | 0.018 | 0.025 | 0.009 |

Table 6.1

## 7 The bootstrap for empirical processes. Applications

The validity of the bootstrap for empirical processes on the real line was first studied by Bickel and Freedman (1981). Gaenssler (1986) considered empirical processes indexed by Vapnik-Cervonenkis classes of sets . The definitive result on the bootstrap for general empirical processes has been obtained by Giné and Zinn (1990). Let $(S, \mathcal{S}, P)$ be a probability space and let $X_1, \ldots, X_n$ be a random sample of variables with distribution $P$ and let $P_n$ be the corresponding empirical probability. If $X_1^*, \ldots, X_n^*$ is the bootstrap sample, $P_n^*$ will be the associated empirical measure. Let $\nu_n = n^{\frac{1}{2}}(P_n - P)$ and let $\nu_n^* = n^{\frac{1}{2}}(P_n^* - P_n)$. Consider the empirical process $\{\nu_n(f) : f \in \mathcal{F}\}$ ( and its bootstrap version $\{\nu_n^*(f) : f \in \mathcal{F}\}$) indexed by a class $\mathcal{F}$ of measurable functions with envelope $F(s) = \sup_{f \in \mathcal{F}} \mid f(s) \mid$, finite for all $s \in S$. Giné and Zinn (1990) prove, under some technical measurability conditions on $\mathcal{F}$, that, in the space $l^\infty(\mathcal{F})$ of real bounded functions on $\mathcal{F}$,

$$n^{\frac{1}{2}}(P_n - P) \rightarrow_w G_P$$

if , and only if,

$$n^{\frac{1}{2}}(P_n^* - P_n) \rightarrow_w G_P, \qquad \textit{in probability},$$

where $G_P$ is a Gaussian centered process; they also show that the bootstrap central limit theorem holds almost surely if, and only if, the central limit theorem holds and, moreover, $\int F^2 dP < \infty$. This gives the validity of the bootstrap in many situations, without any local uniformity about P for the central limit theorem. The asymptotic correctness of the bootstrap for a large class of statistics follows from this theorem; e. g. , for continuous functions of the empirical measure viewed as an element of $l^\infty(\mathcal{F})$.

As a consequence of this result, Romo (1990) proves that the bootstrap works in probability for maximization estimators under nonstandard conditions (in the setup of Huber (1967)) assuming Pollard's (1985) "stochastic differentiability" hypothesis and also that the bootstrap approximation holds in probability for the k-means algorithm in clustering analysis. Under somewhat stronger conditions, Arcones and Giné (1990-a) established the almost sure validity of the bootstrap in these situations.

The asymptotic bootstrap approximation for statistical functionals which are differentiable in a (generalized) Fréchet sense has been obtained by Dudley (1990). Gill (1989) and Sheehy and Wellner (1988) get the correctness of the bootstrap under Hadamard (compact) differentiability of the statistical functional.

The bootstrap for $U$ and $V$ statistics has been considered by several authors. Let $h(x_1, \ldots, x_k)$ be a symmetric measurable function. The $U$ and $V$ statistics based on $h$ and $P$ are

$$U_k^n(h, P) = \binom{n}{k}^{-1} \sum_{1 \le i_1 < \ldots < i_k \le n} h(X_{i_1}, \ldots, X_{i_k})$$

and

$$V_k^n(h, P) = n^{-k} \sum_{i_1, \ldots, i_k = 1}^{n} h(X_{i_1}, \ldots, X_{i_k}).$$

These statistics satisfy a central limit theorem; Bickel and Freedman (1981) proved the bootstrapped central limit theorem for non-degenerate $U$ and $V$ statistics of order $k = 2$. Bretagnolle (1983) established the validity of the bootstrap in probability if the resampling sample size $m_n$ verifies $\frac{m_n}{n} \to 0$ and almost surely if $\frac{m_n(\log n)^b}{n} \to 0$ for some $b > 1$, in the general case with $k = 2$; he also notes that the naïve bootstrap does not work for $h(x,y) = xy$ if $EX_1 = 0$. Arcones and Giné (1990-b) propose a different resampling scheme which works almost surely for any $k$ and for any sample size $m_n \to \infty$. Klenk and Stute (1987) study the bootstrap for L-estimators.

Giné and Zinn (1991) obtain sufficient conditions for the asymptotic validity in probability of the bootstrap under parametric resampling. Arcones and Giné (1991) give tests of symmetry for continuous distributions based on the bootstrap version of the Kolmogorov distance between the empirical distribution and its symmetrization.

## 8 Alternative bootstrap resampling: smoothed, symmetrized and Bayesian bootstrap

In this section, we present several techniques that are alternative to the standard or naïve bootstrap.

(a) *Smoothed bootstrap.* In many statistical problems, we are interested in a populational local characteristic, for example the density function at a fixed point $x$, $T(F) = f(x)$. In this situation, it seems natural to resample from a distribution $\hat{F}_n$ having a density. A candidate is the smoothed distribution estimator,

$$\hat{F}_n(x) = \int_{-\infty}^{x} \hat{f}_n(t)dt$$

(see González Manteiga and Prada-Sánchez (1985)) with $\hat{f}_n$ a nonparametric normal density estimator,

$$\hat{f}_n(x) = \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right),$$

where $K$ is a density function, $h_n$ is the smoothing parameter and $\mathbf{X} = (X_1, \ldots, X_n)$ is a random sample from $F$. Since $\hat{F}_n$ is the convolution of the empirical distribution and the distribution corresponding to $K$, it is computationally easy to resample from it.

The smoothed bootstrap gives good results in practice (see the $s - bt$ column in Table 6.1); also, it improves on the standard bootstrap when we are interested on a local property. For example, if we want to estimate the variance of a sample quantile, $\sigma^2(F) = Var_F(X_{(r)})$, with $X_{(r)} = F_n^{-1}(p), r = [np], p \in (0,1)$, we have, under regularity conditions,

$$\sigma^2(F) = \frac{p(p-1)}{n(f(\xi_p))^2} + O\left(\frac{1}{n^2}\right),$$

where $\xi_p = F^{-1}(p)$ is the populational quantile. Hall and Martin (1988) have shown that, with standard bootstrap,

$$n(\hat{\sigma}^2_{boot} - \sigma^2) = O_P(n^{-1/4}),$$

and Hall, DiCiccio and Romano (1989) have proved that, for somoothed bootstrap,

$$n(\hat{\sigma}^2_{s-boot} - \sigma^2) = O_P(n^{-2/5}),$$

with $h_n \approx n^{-1/5}$. Similar differences in the rates of convergence can be observed when we try to approximate

$$P_F\left\{n^{1/2}(F_n^{-1}(p) - F^{-1}(p)) \leq x\right\}$$

by

$$P_{\tilde{F}_n}\left\{n^{1/2}(F_n^{-1*}(p) - F_n^{-1}(p)) \leq x\right\},$$

taking $\tilde{F}_n = F_n$ or $\hat{F}_n$ (Falk and Reiss, 1989-a). Also Falk and Reiss (1989-b) study the Kolmogorov and variational distances between the distribution of the sample p-quantile and the corresponding smooth bootstrap distribution and they show that this random distance can be again consistently estimated by using the bootstrap.

(b) *Symmetrized bootstrap.* If additional information on the populational distribution is avalaible, it can be used in the bootstrap mechanism. Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a sample from a population with distribution $F$, mean $\mu$ and variance $\sigma^2$. Assume that $F$ is symmetric with respect to $\mu$. If we try to approximate the distribution of

$$\frac{n^{1/2}(\bar{X}_n - \mu)}{S_n},$$

with an Edgeworth expansion given by

$$P_F\left\{\frac{n^{1/2}(\bar{X}_n - \mu)}{S_n} \leq x\right\} = \Phi(x) + n^{-1/2}\left(\frac{\mu_3}{6\sigma^3}(1 + 2x^2)\right)\phi(x) + O(n^{-1}),$$

we find that the error with respect to the normal distribution is $O(n^{-1})$ since $\mu_3 = 0$; this is the same rate that we get using standard bootstrap. This can be corrected by symmetrizing the bootstrap. Cao-Abad and Prada-Sánchez (1991) propose to symmetrize the sample to get

$$Y_i = \begin{cases} X_i, & i = 1, 2, \ldots, n \\ 2\bar{X}_n - X_{i-n}, & i = n+1, \ldots, 2n, \end{cases}$$

and then resampling from $Y_1, Y_2, \ldots, Y_{2n}$. Since the first order moments and the centered second order ones coincide for both samples and the centered moments of odd order for the new sample are zero, they get that the approximation rate is now $O(n^{-3/2})$.

(c) *Bayesian bootstrap.* Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a sample from a population $F$; the aim is to estimate the distribution of $\theta(F)|\mathbf{X} = x$. In bayesian bootstrap, we simulate $B$ random vectors $V_j$ considering a Dirichlet distribution with parameters $(n; 1, \ldots, 1)$ as a prior over all possible populational distributions and we weight $X_{(i)}$ by using the $i-th$ component of $V_j$; these weights add to one and determine a random distribution $F_n^{*j}$. The empirical distribution corresponding to the values $\theta(F_n^{*j})$, $j = 1, \ldots, B$ approximates the posterior distribution of $\theta(F)$. A smoothed bayesian bootstrap, can give excellent results (see last column of Table 6.1 for $\theta(F) = w(F)$). In practice, usually the $i-th$ component of $V_j$, $j = 1, \ldots, B$ is taken as $U_{(i)} - U_{(i-1)}$, $i = 1, \ldots, n(U_0 = 0, U_n = 1)$, where $U_i$ are independent and identically distributed uniform random variables on (0,1).

The bayesian bootstrap was introduced by Rubin (1981) and has been studied by Lo (1987, 1988) and Weng (1989). Boos and Monahan (1986) apply the bootstrap in a bayesian context, replacing the posterior distribution by an estimated posterior distribution using bootstrap.

## 9 The bootstrap in curve estimation

The bootstrap has been incorporated only very recently to the curve estimation literature. Next, we describe its applications to density estimation, regression estimation and smoothing parameter estimation.

(a) *Density estimation.* Let $F$ be a populational distribution with density $f$. If we consider a nonparametric kernel estimate,

$$\hat{f}_{h_n}(x) = \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right),$$

it is interesting to approximate the distribution of

$$R(\mathbf{X}, F) = (nh_n)^{1/2}(\hat{f}_{h_n}(x) - f(x)) =$$
$$(nh_n)^{1/2}\left(\int \frac{1}{h_n} K\left(\frac{x-u}{h_n}\right) dF_n(u) - F'(x)\right). \tag{9.24}$$

Since $F_n$ is not differentiable, $R(\mathbf{X}^*, F_n)$ is not well defined and the smoothed bootstrap seems to be a sensible option to approximate the distribution of (9.24) by using

$$\begin{aligned} R(\mathbf{X}^*, \hat{F}_{g_n}) &= (nh_n)^{(1/2)}\left(\hat{f}^*_{h_n}(x) - \hat{f}_{g_n}(x)\right) = \\ &= (nh_n)^{1/2}\left(\int \frac{1}{h_n} K(\frac{x-u}{h_n}) dF^*_n(u) - \hat{F}'_{g_n}(x)\right), \end{aligned}$$

where

$$\hat{F}_{g_n}(x) = \int_{-\infty}^{x} \hat{f}_{g_n}(t)dt$$

with

$$\hat{f}_{g_n}(t) = \frac{1}{ng_n} \sum_{i=1}^{n} K\left(\frac{t - X_i}{g_n}\right),$$

and $\mathbf{X}^*$ is a sample simulated from $\hat{F}_{g_n}$.

If we take $h_n$ minimising the estimated integrated squared error (ISE) or the estimated mean integrated squared error (MISE), verifying

$$n^{1/5}h_n \to_P constant,$$

it is possible to approximate the distribution function of (9.24) at a point $z$ by means of the normal approximation

$$\Phi\left(\frac{z-C}{V^{1/2}}\right),$$

where $C = \frac{1}{2}c_K^{5/2}d_K f''(x)$ and $V = c_K f(x)$, with

$$c_K = \int K^2(t)dt \ and \ d_K = \int t^2 K(t)dt$$

(see Parzen, 1962). Cao Abad (1990-a, 1990-b), using Berry-Esseen bounds, gets the following rates of convergence:

$$d_\infty\left(P\left\{(nh_n)^{1/2}(\hat{f}_{h_n}(x)-f(x)) \le \cdot\right\},\ \Phi\left(\frac{\cdot - C}{V^{1/2}}\right)\right) = O(n^{-1/5})$$

for the previous normal approximation, not attainable in practice,

$$d_\infty\left(P\left\{(nh_n)^{1/2}(\hat{f}_{h_n}(x)-f(x)) \le \cdot\right\},\ \Phi\left(\frac{\cdot - \hat{C}}{\hat{V}^{1/2}}\right)\right) = O(n^{-1/5}),$$

$\hat{C} = \frac{1}{2}c_K^{5/2}d_K f''_{g_n}(x)$ (with $g_n \approx n^{-1/9}$, optimal rate to estimate the second derivate $f''$) and $\hat{V} = c_K\hat{f}_{h_n}(x)$, for the so called "plug-in" normal approximation, and

$$d_\infty\left(P\left\{(nh_n)^{1/2}(\hat{f}_{h_n}(x)-f(x)) \le \cdot\right\},\ P^*\left\{(nh_n)^{1/2}(\hat{f}^*_{h_n}(x)-f_{g_n}(x)) \le \cdot\right\}\right) =$$
$$= O_P(n^{-2/9}),$$

($g_n \approx n^{-1/9}$), for the bootstrap approximation. This shows that the bootstrap is the one performing best.

Hall (1992-a, 1991), following Hall (1988-a) carries out a different approach. Taking the statistic $\frac{n^{1/2}(\hat{\theta}-\theta)}{\hat{\sigma}}$, he considers

$$\frac{\hat{f}_{h_n}(x) - E\hat{f}_{h_n}(x)}{\left(\hat{Var}\hat{f}_{h_n}(x)\right)^{1/2}},$$

where

$$E\hat{f}_{h_n}(x) = \frac{1}{h_n}\int K\left(\frac{x-u}{h_n}\right)dF(u)$$

and

$$\hat{Var}\hat{f}_{h_n}(x) = \frac{1}{nh_n}\left(\int K^2\left(\frac{x-u}{h_n}\right)\frac{1}{h_n}dF_n(u) - \frac{1}{nh_n}\left(\int K\left(\frac{x-u}{h_n}\right)dF_n(u)\right)^2\right).$$

Resampling from $F_n$, one obtains

$$d_\infty\left(P^*\left\{\frac{\hat{f}^*_{h_n}(x) - E^*\hat{f}_{h_n}(x)}{\left(\hat{Var}^*\hat{f}^*_{h_n}(x)\right)^{1/2}} \leq \cdot\right\},\ P\left\{\frac{\hat{f}_{h_n}(x) - E\hat{f}_{h_n}(x)}{\left(\hat{Var}\hat{f}_{h_n}(x)\right)^{1/2}} \leq \cdot\right\}\right) = O(n^{-4/5})$$

with probability one when $h_n \approx n^{-1/5}$. Even though Hall's rate of convergence is better than Cao Abad's, it does not consider the bias $E\hat{f}_n(x) - f(x)$, which has to be explicity estimated, with the corresponding loss in the speed of convergence (see Hall 1992-a, 1991)

(b) *Regression estimation.* Härdle and Mammen (1991) is an overview on this topic. Following the same pattern as in Section 5, the models we consider for $Y_i = m(\mathbf{X}_i) + \epsilon_i,\ i = 1,\ldots,n$ are:

(i) The errors $\epsilon_i,\ i = 1,\ldots,n$ are independent and identically distributed with zero mean and $\mathbf{X}_i = \mathbf{x}_i,\ i = 1,\ldots,n$ are deterministic values.

(ii) The variables $(\mathbf{X}_i, Y_i),\ i = 1,\ldots,n$ are independent and identically distributed such that $E(\epsilon_i|\mathbf{X}_i) = 0$, i.e., $m(\mathbf{x}) = E(Y_i|\mathbf{X}_i = \mathbf{x}),\ i = 1,\ldots,n$.

(iii) The errors $\epsilon_i, i = 1,\ldots,n$ are independent with zero mean and the distribution of $\epsilon_i$ depending on $\mathbf{x}_i,\ i = 1,\ldots,n$.

The function $m$ is usually estimated by the Nadaraya (1964) and Watson (1964) estimator

$$\hat{m}_{h_n}(\mathbf{x}) = \frac{\sum_{i=1}^n Y_i K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)}{\sum_{r=1}^n K\left(\frac{\mathbf{x}-\mathbf{X}_r}{h_n}\right)},$$

where $h_n$ is the smoothing parameter and $K$ is the kernel function. The aim is to approximate the distribution of

$$(nh_n)^{1/2}\left(\hat{m}_{h_n}(\mathbf{x}) - m(\mathbf{x})\right).$$

For model (i), the resampling procedure follows the next steps:

- Calculate the residuals $\hat{\epsilon}_i = Y_i - \hat{m}_{h_n}(\mathbf{x}_i),\ i = 1,\ldots,n$.
- Center the residuals $\tilde{\epsilon}_i = \hat{\epsilon}_i - \bar{\hat{\epsilon}},\ \bar{\hat{\epsilon}} = \frac{1}{n}\sum_{i=1}^n \hat{\epsilon}_i,\ i = 1,\ldots,n$.
- Resample from $\tilde{\epsilon}_i,\ i = 1,\ldots,n$ to get $\tilde{\epsilon}_i^*,\ i = 1,\ldots,n$.
- Obtain the new observations $Y_i^* = \hat{m}_{g_n}(\mathbf{x}_i) + \tilde{\epsilon}_i^*,\ i = 1,\ldots,n$.

- Calculate

$$\hat{m}^*_{h_n}(\mathbf{x}) = \frac{\sum_{i=1}^n Y_i^* K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)}{\sum_{r=1}^n K\left(\frac{\mathbf{x}-\mathbf{X}_r}{h_n}\right)},$$

and

- Get the bootstrap approximation by repeating $B$ times this process.

In this way, the distribution of

$$(nh_n)^{1/2}(\hat{m}_{h_n}(\mathbf{x}) - m(\mathbf{x}))$$

is approximated by the distribution of

$$(nh_n)^{1/2}(\hat{m}^*_{h_n}(\mathbf{x}) - \hat{m}_{g_n}(\mathbf{x})).$$

Härdle and Bowman (1988) proved that

$$d_\infty\left(P^*\left\{(nh_n)^{1/2}\left((\hat{m}^*_{h_n}(\mathbf{x}) - \hat{m}_{g_n}(\mathbf{x})\right) \leq \cdot\right\},\right.$$
$$\left.P\left\{(nh_n)^{1/2}(\hat{m}_{h_n}(\mathbf{x}) - m(\mathbf{x})) \leq \cdot\right\}\right)$$

tends to zero in probability, assuming $Var\ \epsilon = \sigma^2$ and some other regularity conditions. Later on, Hall (1992-b) got good rates of convergence regardless of any considerations on bias.

Two resampling procedures have been proposed for model (ii): the one by Dikta (1990), where $(\mathbf{X}_i^*, Y_i^*)$, $i = 1, \ldots, n$ is obtained from the original sample and then

$$\hat{m}^*_{h_n}(\mathbf{x}) = \frac{\sum_{i=1}^n Y_i^* K\left(\frac{\mathbf{x}-\mathbf{X}_i^*}{h_n}\right)}{\sum_{r=1}^n K\left(\frac{\mathbf{x}-\mathbf{X}_r^*}{h_n}\right)}$$

is calculated from it, and the smoothed bootstrap proposed by Cao Abad and González Manteiga (1990), where the sample is obtained from

$$\hat{F}_n(\mathbf{x}, y) = \frac{1}{n}\sum_{i=1}^n I_{\{Y_i \leq y\}} \int_{-\infty}^{\mathbf{x}} \frac{1}{g_n} K\left(\frac{t - \mathbf{X}_i}{g_n}\right) dt.$$

For optimal choices of $h_n$, the smoothed bootstrap leads to good results,

$$d_\infty\left(P^*\left\{(nh_n)^{1/2}\left(\hat{m}^*_{h_n}(\mathbf{x}) - \hat{m}_{g_n}(\mathbf{x})\right) \leq \cdot\right\},\right.$$
$$\left.P\left\{(nh_n)^{1/2}(\hat{m}_{h_n}(\mathbf{x}) - m(\mathbf{x})) \leq \cdot\right\}\right) =$$
$$= O_P(n^{-2/9})$$

when $g_n \approx n^{-1/9}$, opposite to Dikta's bootstrap, which is in trouble with bias,

$$E(\hat{m}_{h_n}(\mathbf{x}) - m(\mathbf{x})) = O(n^{-2/5}),$$

meanwhile

$$E^*(\hat{m}^*_{h_n}(\mathbf{x}) - m_{g_n}(\mathbf{x})) = O_P(n^{-4/5}),$$

as can be seen in Härdle and Mammen (1990-b).

The best suited bootstrap for model (iii) is the one known as "wild bootstrap", introduced by Härdle and Mammen (1990-a), based on ideas used by Wu (1986) for linear regression. The steps are the following:

- Obtain the residuals $\hat{\epsilon}_i = Y_i - \hat{m}_{h_n}(\mathbf{x}_i),\ i = 1, \ldots, n$.
- For each $i$, resample $\hat{\epsilon}^*_i$ from a distribution $\hat{F}_i$ verifying $E_{\hat{F}_i} Z = 0,\ E_{\hat{F}_i} Z^2 = \hat{\epsilon}^2_i$ and $E_{\hat{F}_i} Z^3 = \hat{\epsilon}^3_i$ ($Z$ has distribution $\hat{F}_i$).
- Calculate $Y^*_i = \hat{m}_{g_n}(\mathbf{x}_i) + \hat{\epsilon}^*_i,\ i = 1, \ldots, n$.
- Calculate $\hat{m}^*_{h_n}(\mathbf{x})$ and repeat this process $B$ times.

Cao Abad (1991) has shown that, under regularity conditions, the rate of convergence for the "wild bootstrap" is

$$\begin{aligned} & d_\infty\left(P^*\left\{(nh_n)^{1/2}\left(\hat{m}^*_{h_n}(\mathbf{x}) - \hat{m}_{g_n}(\mathbf{x})\right) \leq \cdot\right\},\right. \\ & \left. P\left\{(nh_n)^{1/2}\left(\hat{m}_{h_n}(\mathbf{x}) - m(\mathbf{x})\right) \leq \cdot\right\}\right) = \\ & = O_P(n^{-2/9}), \end{aligned}$$

meanwhile the rate of convergence for the normal approximation is $O_P(n^{-1/5})$.

The "wild bootstrap" idea is also used by Härdle and Marron (1991) to obtain bootstrap simultaneous error bars for model (ii). Härdle, Huet and Jolivet (1991) prove that "wild bootstrap" with explicit bias estimation gives improved coverage accuracy for confidence intervals.

The cuestion of detecting a difference between two mean functions in the setting of model (iii) by using bootstrap has been addressed by Hall and Hart (1990).

Franke and Wendel (1992) considers bootstrap for nonparametric autoregressive time series and Franke and Härdle (1990) show that the bootstrap works for kernel spectral density estimates.

(c)*Smoothing parameter estimation.* Another important application of bootstrap in curve estimation is the bandwidth choice. Thus, when we are using the kernel method for density estimation, an important choice is the $h_{MISE}$ which is the bandwidth minimising

$$\begin{aligned} MISE(h_n) &= E\left\{\int (\hat{f}_{h_n}(x) - f(x))^2 dx\right\} = V(h_n) + B^2(h_n) = \\ &= \frac{1}{nh_n}\int K^2(x)dx + h_n^4\left(\int f''^2(x)dx\right)\left(\frac{1}{2}\int x^2K(x)dx\right)^2 + \\ &+ O\left(\frac{1}{h_n}\right) + O(h_n^4), \end{aligned}$$

where $V(h_n)$ is the variance and $B^2(h_n)$ is the squared bias.

Among the several proposed methods to estimate $h_{MISE}$ (see Cao Abad, Cuevas and González Manteiga, 1991), it is important the bandwidth $h^*_{MISE}$, minimising

$$MISE^*(h_n) = E^*\left\{\int (\hat{f}^*_{h_n} - \hat{f}_{g_n}(x))^2 dx\right\}.$$

The case $g_n = h_n$ was considered by Taylor (1989), but, as pointed out by Cao Abad (1990-b), it can present some inconsistencies. Cao Abad (1990-b), Marron (1990), Falk (1990) and Jones, Marron and Park (1991) present exhaustive studies proving that, for some choices of $g_n$, it is possible to get

$$\frac{h^*_{MISE}}{h_{MISE}} - 1 = O_P(n^{-1/2}),$$

the usual "root-n" rate. Hall (1990) considers an alternative approach, resampling from subsamples of the original sample.

Other applications of smoothed bootstrap in curve estimation are the estimation of a density mode (Romano, 1988) and checking the number of modes (Mammen, Marron and Fisher, 1992).

## 10 Other topics and applications

In this section, we briefly sketch the application of bootstrap techniques to censored data and to the construction of new estimators and tests.

(a) *The bootstrap for censored data.* Let $T$ be the variable of interest (usually a lifetime: related to a patient, to the reliability of a system, etc.). Sometimes, $T$ is not observed because it is censored at a random time $C$. Thus the initial sample is given by $(X_i, \delta_i)$, $i = 1, \ldots, n$, where $X_i = \min\{T_i, C_i\}$ and $\delta_i = I_{\{T_i \le C_i\}}$ (random censorship on the right).

Two bootstrap resampling schemes have been initially proposed for this situation:

1. *Efron's (1981) resampling.* A sample $(X_i, \delta_i)^*, i = 1, \ldots, n$ (also denoted by $(X_i^*, \delta_i^*), i = 1, \ldots, n$) can be obtained simulating from the original sample. Alternatively, let $S^0(t) = P\{T > t\}$, $R(t) = P\{C > t\}$ and $S(t) = P\{X > t\}$ be the survival functions of $T, C$ and $X$, respectively. If

$$\hat{S}^0(t) = \prod_{\{X_{(j)} \le t\}} \left(\frac{n-j}{n-j+1}\right)^{\delta_{(j)}}$$

and

$$\hat{R}(t) = \prod_{\{X_{(j)} \le t\}} \left(\frac{n-j}{n-j+1}\right)^{(1-\delta_{(j)})}$$

are the Kaplan-Meier (1958) estimators of $S^0$ and $R$, where $(j)$ indicates order for the data $X_i$ and $\delta_{(j)}$ is the corresponding $\delta$, we can generate samples from $(1-\hat{S}^0)$, $(T_1^*,\ldots,T_n^*)$, and from $(1-\hat{R}),(C_1^*,\ldots,C_n^*)$, and then calculate $X_i^* = \min\{T_i^*, C_i^*\}$ and $\delta_i^* = I_{\{T_i^* \leq C_i^*\}}$, to get the bootstrap sample $(X_i^*, \delta_i^*), i = 1,\ldots,n$. Efron (1981) has shown, under regularity conditions on the distributions, that both resampling schemes are equivalent in estimating the variance of statistics of the form $\hat{\theta} = \theta(\hat{S}^0)$.

2. *Reid's (1981) resampling* . The resampling is done directly form $1-\hat{S}^0$ and then $\theta$ is evaluated at the empirical distribution corresponding to the sample $\{X_1^*,\ldots,X_n^*\}$ obtained from $1-\hat{S}^0$.

Akritas (1986) has carried out an interesting study to compare both resampling strategies. Since the empirical process $n^{1/2}(\hat{S}^0(\cdot) - S^0(\cdot))$ converges in law to $B^0(K(\cdot))\frac{S^0(\cdot)}{1-K(\cdot)}$ in $[0,\tau)$, with $\tau < \sup\{t : 1 - S(t) < 1\}$, where $T$ and $C$ are variables with positive support, $K(x) = \frac{C(x)}{1+C(x)}$ with

$$C(x) = \int_0^x \frac{1}{S^2(t)} d(1-\tilde{S})(t),$$

$1-\tilde{S}(x) = P\{X \leq x, \delta = 1\}$ and $B^0$ is the Brownian bridge (see Hall and Wellner (1980)); it follows from Akritas (1986) that the correct bootstrap to mimic the populational model is Efron's (1981). Indeed, for Efron's (1981) resampling, we have that

$$n^{1/2}(\hat{S}^{0*}(\cdot) - \hat{S}^0(\cdot)) \longrightarrow_w B^0(K(\cdot))\frac{S^0(\cdot)}{1-K(\cdot)} \qquad a.s.,$$

and, for Reid's (1981), it holds that

$$n^{1/2}(\hat{S}^{0*}(\cdot) - \hat{S}^0(\cdot)) \longrightarrow_w B^0(K(\cdot)) \qquad a.s.$$

The reason for this different behavior is that Reid's resampling is done from non-censored data. Asymptotic results for this resampling can be also found in Lo and Singh (1985), Horvath and Yandell (1987) and Chung (1989). The bidimensional censoring case is developped in Dabrowske (1989).

Dikta and Ghorai (1990) study the model with proportional hazard rate censoring. In this context,

$$R(t) = S^0(t)^\beta \qquad \beta \in \mathcal{R}_+,$$

with $P\{\delta = 1\} = \frac{1}{1+\beta} = E(\delta)$, and so, a natural estimate for $S^0$ is

$$\hat{S}^0(t) = S_n^{\gamma_n}(t) = \left(\frac{1}{n}\sum_{i=1}^n I_{\{X_i > t\}}\right)^{\frac{1}{n}\sum_{i=1}^n \delta_i},$$

since $S(t) = R(t)S^0(t) = S^0(t)^{1+\beta}$. The resampling procedure in this case is the following:

(a) Draw a random sample $X_1^*, \ldots, X_n^*$ from $1 - S_n$.

(b) Draw a random sample $\delta_1^*, \ldots, \delta_n^*$ from a Bernoulli distribution with parameter $\gamma_n$, independent from $X_1^*, \ldots, X_n^*$ (in this model, $\gamma$ and $X$ are independent).

(c) Calculate the bootstrap estimate $(S_n^*)^{\gamma_n^*} = \left(\frac{1}{n}\sum_{i=1}^n I_{\{X_i^* > t\}}\right)^{\frac{1}{n}\sum_{i=1}^n \gamma_i^*}$.

(d) Repeat this process $B$ times.

Under regularity conditions, Dikta and Gorai (1990) have established that

$$d_\infty \Bigg( P^*\{\sup_{0 \le t \le \tau} |n^{1/2}(S_n^{*\gamma_n^*}(t) - S_n^{\gamma_n}(t))| \le \cdot\},$$
$$P\{\sup_{0 \le t \le \tau} |n^{1/2}(S_n^{\gamma_n}(t) - S^0(t))| \le \cdot\} \to 0 \qquad \textit{almost surely.}$$

(b)*Some applications of the bootstrap to the construction of new estimators and tests.*

An area very recently explored is the construction of estimators using bootstrap techniques. Suppose we want to obtain an estimate of the parameter $\theta = \theta(F)$. The natural nonparametric maximum likelihood estimater would be obtained through $\theta(F_n)$; in some situations, this estimation may be meaningless and a previous approximation,

$$\theta_m(F) \approx \theta(F), \qquad m = m_n \to \infty,$$

is needed. Finally, the bootstrap approximation $\theta_m(F_n)$ is used.

Swanopoel (1986) presents this idea for density estimation. Let $X_1, \ldots, X_n$ be a sample from the distribution $F$ with density $f$ and let $X_{(1)} \le, \ldots, \le X_{(n)}$ be the corresponding ordered sample. If $a = [nF(x)] + 1$, we have that

$$n^{1/2}(X_{(a)} - x) \longrightarrow_d N\left(0, \frac{F(x)(1 - F(x))}{f^2(x)}\right),$$

and, so,

$$\theta_m(F) = \left(\frac{F(x)(1 - F(x))}{mVar_F(X_{(a)})}\right)^{1/2} \approx f(x) = \theta(F),\ m \to \infty,$$

can be used for a bootstrap estimation of the density function:

$$\theta_m(F_n) = \hat{f}(x) = \left(\frac{F_n(x)(1 - F_n(x))}{mVar_{F_n}(X_{(a)})}\right)^{1/2}.$$

Another application of this technique is given by Swanopoel (1990): if we consider $T_m(X_1, X_2, \ldots, X_m)$ as a previous estimate of $\theta = \theta(F)$, taking

$$\theta_m(F) = median_F(T_m(X_1, \ldots, X_m)),$$

we get a general family of bootstrap estimators

$$\theta_m(F_n) = median_{F_n}(T_m(X_1, \ldots, X_m)).$$

For instance, if $T_m(X_1, \ldots, X_m) = \bar{X}_m$ and $m = 1$,

$$\theta_m(F_n) = median(X_1, \ldots, X_n);$$

if $m = 2, \theta_m(F_n) = median\left(\frac{X_i+X_j}{2}\right)$ is the Hodges-Lehmann estimate, etc.

Léger and Romano (1989, 1990) show very interesting applications of this principle. For example, let the $\beta - trimmed$ functional

$$\theta_\beta(F) = \frac{1}{1-2\beta}\int_\beta^{1-\beta} F^{-1}(t)dt, \quad 0 \leq \beta \leq 1.$$

A natural bootstrap estimation is given by

$$\theta_\beta(F_n) = \frac{1}{1-2\beta}\int_\beta^{1-\beta} F_n^{-1}(t)dt$$

and the adaptive bootstrap estimator is $\hat{\theta}_{\hat{\beta}}$ minimising the risk bootstrap estimator $E_{F_n}(\theta_\beta(F_n^*) - \theta_\beta(F_n))^2$. Other extensions of this methodology to more general loss functions can be found in the above mentioned papers.

Romano (1988) propose bootstrap nonparametric tests (independence of variables, goodness of fit, equality of distributions and rotational invariance).

Recently, Boos and Brownie (1989) and Boss, Jensen and Veraberbeke (1989) have used bootstrap resampling for variance homogeneity tests obtaining encouraging results compared to the classical Bartlett test.

# REFERENCES


ABRAMOVITCH, L. and SINGH, K. (1985). "Edgeworth corrected pivotal statistics and the bootstrap". *Ann. Statist.* 13, 1, 116-132.

AKRITAS, M.G. (1986). "Bootstrapping the Kaplan-Meier Estimator". *J.A.S.A.* 81, 396, 1032-1038.

ATHREYA, K.B. (1987). "Bootstrap of the mean in the infinite variance case". *Ann. Statist.* 15, 724-731.

ARCONES, M. A. and GINÉ, E. (1989). "The bootstrap of the mean with arbitrary sample size". *Ann. Ins. Henry Poincaré* 25, 457-481.

ARCONES, M. A. and GINÉ, E. (1990-a). "On the bootstrap of M-estimators and other statistical functionals". To appear in *Proc. IMS Bootstrap Conference, East Lansing 1990.* Wiley, New York.

ARCONES, M. A. and GINÉ, E. (1990-b). "The bootstrap central limit theorem for U and V statistics". To appear in *Ann. Statist.*

ARCONES, M. A. and GINÉ, E. (1991). "Some bootstrap tests of symmetry for univariate continuous distributions". *Ann. Statist.* 19, 3, 1496-1511.

BABU, G.J. (1984). "Bootstrapping statistics with linear combinations of chi-squares as weak limit". *Sankhya* 46, 86-93.

BABU, G.J. (1986). "A note on bootstrapping the variance of sample quantile". *Ann. Inst. Stat. Math.* 38, 439-443.

BASAWA, I.V., MALLIK, A.K. McCORMICK, W.P. and TAYLOR, R.L. (1989). "Bootstrap explosive autoregressive processes". *Ann. Statist.* 17, 1479-1486.

BASAWA, I.V., MALLIK, A.K., McCORMICK, W.P. , REEVES, J.H. and TAYLOR, R.L. (1991). "Bootstrapping unstable first-order autorregressive processes". *Ann. Stat.* 19, 1098-1101.

BERAN, R. (1982). "Estimated sampling distributions. The bootstrap and competitors". *Ann. Statist.* 10, 1, 212-225.

BERAN, R. (1984). "Jackknife approximations to bootstrap estimates". *Ann. Statist.* 12, 1, 101-118.

BERAN, R. (1984). "Bootstrap methods in statistics". *Jber. d. Dt. Math.-Verein* 86, 14-30.

BERAN, R. (1987). "Prepivoting to reduce level error of confidence sets" *Biometrika.* 74, 457-468.

BERAN, R. (1988). "Balanced simultaneous confidence sets". *J.A.S.A.* 83, 679-697.

BERAN, R. (1990). "Refining bootstrap simultaneous confidence sets". *J.A.S.A.* 85, 410, 417-426.

BERAN, R. and SRIVASTAVA, M.S. (1985). "Bootstrap tests and confidence regions for functions of a covariance matrix". *Ann. Statist.* 13, 1, 95-115.

BICKEL, P.J. and FREEDMAN, D.A. (1981). "Some asymptotic theory for the bootstrap". *Ann. Statist.* 9, 6, 1196-1217.

BICKEL, P.J. and FREEDMAN, D.A. (1984). "Asymptotic normality and the bootstrap in stratified sampling". *Ann. Statist.* 12, 2, 470-482.

BOOS, D.D. and BROWNIE, C. (1989). "Bootstrap methods for testing homogeneity of variances". *Technometrics* 31, 1, 69-82.

BOOS, D.D., JANSSEN, P. and VERAVERBEKE, N. (1989). "Resampling from centered data in the two-sample problem". *Journal of Stat. Planning and Inference* 21, 327-345.

BOOS, D.D. and MONAHAN, J.F. (1986). "Bootstrap methods using prior information". *Biometrika* 73, 77-83.

BOSE, A. (1988). "Edgeworth correction by bootstrap in autoregressions". *Ann. Statist.* 16, 4, 1709-1722.

BOSE,A. and BABU,G.J.(1991). "Accuracy of the bootstrap approximation".*Probab. Theory Relat. Fields* 90, 301-316.

BRETAGNOLLE, J. (1983). "Lois limites du bootstrap de certaines fonctionnelles". *Ann. Inst. H. Poincaré* 3, 281-296.

BUNKE, O. and DROGE, B. (1984). "Bootstrap and cross-validation estimates of the prediction error for linear models". *Ann. Statist.* 12, 4, 1400-1424.

CAO ABAD, R. (1990-a). "Ordenes de convergencia para las aproximaciones normal y bootstrap en la estimación no paramétrica de la función de densidad". *Trabajos de Estadística* 5, 2, 23-32.

CAO ABAD, R. (1990-b). "Aplicaciones y nuevos resultados del método bootstrap en la estimación no paramétrica de curvas". *Ph.D. dissertation.* Universidad de Santiago de Compostela.

CAO ABAD, R. (1991). "Rates of convergence for the wild bootstrap in nonparametric regression". *Ann. Statist.* 19, 4, 2226-2231.

CAO ABAD, R. and GONZALEZ MANTEIGA, W. (1990). "Bootstrap methods in regression smoothing: An alternative procedure to the wild bootstrap". Preprint.

CAO ABAD, R. and PRADA SANCHEZ, J. M. (1991). "Bootstrapping the mean of a symmetric population ". Preprint.

CAO ABAD, R., CUEVAS, A. and GONZALEZ MANTEIGA, W. (1991). "A comparative study of several smoothing methods in density estimation".

CUEVAS, A. and ROMO, J. (1991). "On robustness properties of bootstrap approximations". Preprint.

CUEVAS, A. and ROMO, J. (1992). "On the estimation of the influence curve". Preprint.

CHUNG, C. (1989). "Confidence bands for percentile residual lifetime under random censorship model" *J. Multiv. Analysis.* 29, 94-126.

DAVISON, A. C., HINKLEY, D. V. and SCHECHTMAN, E. (1986). "Efficient bootstrap simulation". *Biometrika*, 73, 555-561.

DABROWSKA, D.H. (1989). "Kaplan-Meier estimate on the plane: weak convergence, LIL and the bootstrap". *J. Multiv. Analysis*, 29, 308-325.

DICICCIO, T.J. and ROMANO, J.P. (1988). "A review of bootstrap confidence intervals". *J. R. Statist. Soc. B* 50, 3, 338-354.

DIKTA, G. (1990). "Bootstrap approximation of nearest neighborhood regression function estimates". *J. Multiv. Analysis.* 32, 213-229.

DIKTA, G. and GHORAI, J.K. (1990). "Bootstrap approximation with censored data under the proportional hazard model". *Commun. Stat.Theory Meth.* 19, 573-581.

DO, K.-A. and HALL, P. (1991). "On importance resampling for the bootstrap". *Biometrika.* 78, 1, 161-167.

DUDLEY, R.M. (1990). "Nonlinear functionals of empirical measures and the bootstrap".*Probability in Banach spaces VII, 63-82.* Progress in Probability Series. Birkhauser, Boston.

EFRON, B. (1979). "Bootstrap methods: another look at the jackknife". *Ann. Statist.* 7, 1-26.

EFRON, B. (1981). "Censored data and the bootstrap". *J.A.S.A.* 76, 374, 312-319.

EFRON, B. (1982). "The jackknife, the bootstrap and other resampling plans". *Regional conference series in applied mathematics.* CBMS-NSF.

EFRON, B. (1983). "Estimating the error rate of a prediction rule: improvements on cross-validation". *J. A. S. A.* 78, 382, 316-331.

EFRON, B. (1986). "How biased is the apparent error rate of a prediccion rule?". *J.A.S.A.* 81, 394, 461-470.

EFRON, B. (1987). "Better bootstrap confidence intervals". *J.A.S.A.* 82, 397, 171-200.

EFRON, B. (1990). "More efficient bootstrap computations". *J. A. S. A.* 85, 409, 79-89.

EFRON, B. (1992). "Jackknife-after-bootstrap standard errors and influence functions". *J. R. Statist. Soc.* B, 54, 1, 83-127.

EFRON, B. and TIBSHIRANI, R. (1986). "Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy". *Stat. Science.* 1, 1, 54-77.

FALK, M. (1990). "Bootstrap optimal bandwidth selection for kernel density estimates". *Journal of Stat. Planning and Inference.* To appear.

FALK, M. and REISS, R.D. (1989-a). "Weak convergence of smoothed and nonsmoothed bootstrap quantile estimates". *Ann. Probab.* 17, 362-371.

FALK, M. and REISS, R.D. (1989-b). " Bootstrapping the distance between smooth bootstrap and sample quantile distribution". *Probab. Th. Rel. Fields.* 82, 177-186.

FRANKE, J. and WENDEL, H. (1992). "A bootstrap approach for nonlinear autoregressions. Some preliminary results". *Proceedings of the Bootstrap Conference in Trier, Germany* 101-106. Springer Verlag.

FRANKE, J. and HÄRDLE, W.(1990). "On bootstrapping kernel spectral estimates". *Ann. Statist.*. To appear.

FREEDMAN, D.A. (1981). "Bootstrapping regression models". *Ann. Statis.* 9, 6, 1218-1228.

FREEDMAN, D.A. (1984). "On bootstrapping two-stage least squares estimates in stationary linear models". *Ann. Statist.* 12, 3, 827-842.

GAENSSLER, P. (1986). "Bootstrapping empirical processes indexed by Vapnik-Cervo -nenkis classes of sets ". *Probability Theory and Mathematical Statistics.* Prohorov et al. (ed.), VNU Press, The Netherlands, pp. 467-481.

GHOSH, M., PARR, W., SINGH, K. and BABU, J. (1984). "A note on bootstrapping the sample median". *Ann. Statist.* 12, 1130- 1135.

GILL, R. D. (1989). "Non- and semi-parametric maximum likelihood estimators and the von Mises method (Part I)". *Scand. J. Statist.* 16, 97-128.

GINE, E. and ZINN, J. (1989). "Necessary conditions for the bootstrap of the mean". *Ann. Statist.* 17, 684-691.

GINE, E. and ZINN, J. (1990). "Bootstrapping general empirical measures".*Ann. Probab.* 18, 851-869.

GINE, E. and ZINN, J. (1991). "Gaussian characterization of uniform Donsker classes of functions".*Ann Probab.* 19, 3.

GONG, G. (1986). "Cross-validation, the jackknife, and the bootstrap: excess error estimation in forward logistic regression". *J.A.S.A.* 81, 393, 108-113.

GONZALEZ MANTEIGA, W. and PRADA SANCHEZ, J.M. (1985). "Una aplicación de los métodos de suavización no paramétricos en la técnica bootstrap". *Proceedings Jornadas Hispano-Lusas de Matemáticas*, 337-346. Murcia.

GONZALEZ MANTEIGA, W., PRADA SANCHEZ, J.M., FIESTRAS JANEIRO, M.G. and GARCIA JURADO, I. (1990). "Dependence between fusion temperatures and chemical components of a certain type of coal using classical, non parametric and bootstrap techniques". *Journal of Chemometrics.* 4, 6, 429-440.

GRAHAM, R.L., HINKLEY, D.V., JOHN, P.W.M. and SHI, S. (1990). "Balanced design of bootstrap simulations". *J.R. Statist. Soc.* 52, 1, 185-202.

HALL, P. (1983). "Inverting an Edgeworth expansion". *Ann. Statist.* 11, 2, 569-576.

HALL, P. (1986). "On the bootstrap and confidence intervals". *Ann. Statist.* 14, 4, 1431-1452.

HALL, P. (1988-a). "Theoretical comparison of bootstrap confidence intervals". *Ann. Statist.* 16, 927-953.

HALL, P. (1988-b). "Rate of convergence in bootstrap approximations". *Ann. Probab.* 16, 4, 1665-1684.

HALL, P. (1989-a). "Asymptotic properties of the bootstrap for heavy-tailed distributions". *Ann. Probab.* 18, 3, 1342-1360.

HALL, P. (1989-b). "Unusual properties of bootstrap confidence intervals in regression problems". *Prob. Th. Rel. Fields.* 81, 247-273.

HALL, P. (1990). "Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems". *Journal of Multiv. Analysis.* 32, 177-203.

HALL, P. (1991). "Edgeworth expansions for nonparametric density-estimators, with applications". *Statistics* 22, 2, 215-232.

HALL, P. (1992-a). "Effect of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability density". *Ann. Statist.* 20, 2.

HALL, P. (1992-b). "On bootstrap confidence intervals in nonparametric regression". *Ann. Statist.* 20, 2.

HALL, P., DICICCIO, T.J. and ROMANO, J. (1988). "On smoothing and the bootstrap". *Ann. Statist.* 17, 2, 692-704.

HALL, P., HÄRDLE, W. and SIMAR, L. (1991). "On the inconsistency of bootstrap distribution estimators". Preprint.

HALL, P. and HART, J. D. (1990). "Bootstrap test for difference bettween means in parametric regression". *Journ. Amer. Stat. Assoc.*, 85, 412, 1039-1049.

HALL, P. and MARTIN, M.A. (1988). "Exact convergence rate of bootstrap quantile variance estimator". *Prob. Th. Rel. Fields.* 80, 261-268.

HALL, W.L. and WELLNER, J.A. (1980). "Confidence bands for a survival curve from, censored data". *Biometrika* 67, 133- 143.

HÄRDLE, W. and BOWMAN, A. (1988). "Bootstrapping in nonparametric regression: local adaptive smoothing and confidence bands". *J.A.S.A.* 83, 102-110.

HÄRDLE, W., HUET, S. and JOLIVET, E. "Better confidence intervals for regression curve estimation". Preprint.

HÄRDLE, W. and MAMMEN, E. (1991). "Bootstrap methods in nonparametric regression". *Nonparametric Functional Estimation and Related Topics.* Kluwer Ac. Pub., 111-123.

HÄRDLE, W. and MAMMEN, E. (1990-b). "Comparing nonparametric versus parametric regression fits". Preprint.

HÄRDLE W. and MARRON, J.S. (1991). "Bootstrap simultaneous error bars for nonparametric regression". *Ann. Statist.* 19, 2, 778-796.

HINKLEY, D.V. (1977). "Jackknifing in unbalanced situations". *Technometrics* 19, 285-292.

HINKLEY, D.V. (1988)."Bootstrap methods". *J.R. Statist. Soc. B* 50, 3, 321-337.

HORVATH, L. and YANRELL, B.S. (1987). "Convergence rates for the bootstrapped product-limit process". *Ann. Statist.* 15, 3, 1155-1173.

HUBER, P. J. (1964)."The behavior of maximum likelihood estimates under non-standard conditions".*Proceedings Fifth Berkeley Symposium on Mathematical Statistics and Probability* Vol 1, 221-233. University of California Press, Berkeley.

HUET, S. and JOLIVET, E. (1989). "Bootstrap and Edgeworth expansion: the nonlinear regression as an example". Preprint.

HUET, S., JOLIVET, E. and MESSEAN, A. (1990). "Some simulation results about confidence intervals and bootstrap methods in nonlinear regression". *Statistics* 21, 3, 369-432.

JAECKEL, L. (1972). "The infinitesimal jackknife". Memorandum MM72-1215-11. Bell Laboratories, Murray Hill, N.J.

JONES, M.C., MARRON, J.S. and PARK, B.U. (1991). "A simple root- n bandwidth selector". *Ann. Statist.* 19, 4, 1919-1932.

KENDALL, M. and STUART, A. (1985). "The Advanced Theory of Statistics". Griffin, London.

KLENK, A. and STUTE, W. (1987). "Bootstrapping of L-estimates". *Stat. and Decisions*

KNIGHT, K. (1989). "On the bootstrap of the sample mean in the infinite variance case". *Ann. Statist.* 17, 3, 1168-1175.

KÜNSCH, H.R. (1989). "The jackknife and the bootstrap for general stationary observations". *Ann. Stat.* 17, 3, 1217-1241.

LEGER, C., POLITIS, D.N. and ROMAÑO, J.P. (1992). "Bootstrap technology and applications". *Technometrics.* To appear.

LEGER, C. and ROMANO, J. (1989). "Bootstrap adaptive trimmed means". Technical Report 314. Stanford.

LEGER, C. and ROMANO, J. (1990). "Bootstrap choice of tunning parameters". *Ann. Inst. Stat. Math.* 42, 4, 709-735.

LO, A.Y. (1987). "A large sample study of the Bayesian bootstrap". *Ann. Statist.* 15, 1, 360-375.

LO, A.Y. (1988). "A Bayesian bootstrap for finite populations". *Ann. Statist.* 16, 4, 1684-1695.

LO, S. and SINGH, K. (1985). "The product-limit estimator and the bootstrap: some asymptotic representations". *Prob. Th. Rel. Fields*, 71, 455-465.

LOH, W. (1984). "Estimating an endpoint of a distribution with resampling methods". *Ann. Statist.* 12, 4, 1543-1550.

MAMMEN, E., MARRON, J.S. and FISHER, N.I. (1992). "Some asymptotics for multimodality tests based on kernel density estimates". *Prob. Th. Rel. Fields* 91, 1, 115-132.

MARITZ, J.S. and JARRET, R.G. (1978). "A note on estimating the variance of the sample median". *J. Am. Stat. Assoc.* 73, 194-196.

MARRON, J.S. (1990). "Bootstrap bandwidth selection". (Mimeo Series #2027) Dept. of Statistics. Chapel-Hill Univ.

MILLER, R.G. (1974). "The jackknife- a review". *Biometrika* 61, 1-17.

NADARAYA, E.A. (1964). "On estimating regression". *Theory Prob. Appl.* 10, 186-190.

NAVIDI, W. (1989). "Edgeworth expansions for bootstrapping regression models". *Ann. Statist.* 17, 4, 1472-1478.

PARR, W.C. (1985). "The bootstrap: some large sample theory and connections with robustness". *Stat. Prob. letters.* 5, 3, 97-100.

PARZEN, E. (1962). "On estimating probability density and mode". *Ann. Math. Statist.* 35, 2065-2076.

POLLARD, D. (1985). "New ways to prove central limit theorems". *Econometric Theory* 1, 259-314.

PRADA SANCHEZ, J.M. and OTERO CEPEDA, X.L. (1989). "The use of smooth bootstrap techniques for estimating the error rate of a prediction rule". *Commun. Stat. Simulation.* 18, 3, 1169- 1186.

QUENOUILLE, M. (1956). "Notes on bias in estimation". *Biometrika* 43, 353-360.

RAYNER, K. (1989). "Bootstrap inversion of Edgeworth expansions for nonparametric confidence intervals". *Stat. Prob. Letters* 8, 201-206.

REID, N. (1981). "Estimating the median survival time". *Biometrika* 68, 3, 601-8.

ROMANO, J. (1988). "On weak convergence and optimality of kernel density estimates of the mode". *Ann. Statist.* 16, 2, 629- 647.

ROMO, J. (1990). "Bootstrapping maximization estimates under non-standard conditions". Preprint.

RUBIN, D.B. (1981). "The Bayesian bootstrap". *Ann. Statist.* 9, 1, 130-134.

SANCHEZ, A. (1991). "Efficient bootstrap simulation: an overview". *Qüestiió* 14, 43-88.

SAUERMANN, W. (1989). "Bootstrapping the maximun likelihood estimator in high-dimensional log-linear models". *Ann. Statist.* 17, 3, 1198-1216.

SHEEHY, A. and WELLNER, J. A. (1988). "Uniformity in $P$ of some limit theorems for empirical measures and processes". Technical Report #134, Department of Statistics, University of Washington, Seattle.

SINGH, K. (1981). "On the asymptotic accuracy of Efron's bootstrap". *Ann. Statist.* 9, 6, 1187-1195.

STINE, R.A. (1985). "Bootstrap prediction intervals for regression". *J.A.S.A.* 80, 392, 1026-1031.

STUTE, W. (1990). "Bootstrap of the linear correlation model". *Statistics* 21, 3, 433-436.

STUTE, W. and GRÜNDER (1990). "Prediction intervals for explosive AR(1) processes". *Bootstrapping and Related Techniques.* (Univ. of Trier), p. 121-130. Springer Verlag.

SWANOPOEL, J. (1986). "On the construction of nonparametric density function estimators using the bootstrap". *Commum. Stat. Theor. Meth.* 15, 5, 1399-1415.

SWANOPOEL, J. (1990). "A review of bootstrap methods". *South African Stat. J.* 24, 1-34.

TAYLOR, C.C. (1989). "Bootstrap choice of the smoothing parameter in kernel density estimation". *Biometrika* 76, 705-712.

THOMBS, L.A. and SCHUCANY, W.R. (1990). "Bootstrap prediction intervals for autoregression". *J.A.S.A.* 85, 410, 486-492.

TSAY, R. S. (1992). "Model checking via parametric bootstrap in time series analysis'. *Applied Statistics* 41, 1, 1-15.

TUKEY, J. (1958). "Bias and confidence in not quite large samples, abstract". *Ann. Math. Statist.* 29, 614.

WATSON, G.S. (1964). "Smooth regression analysis". *Sankhya*, Series A. 359-372.

WENG, C.S. (1989). "On a second-order asymptotic property of the Bayesian bootstrap mean". *Ann. Statist.* 17, 2, 705-710.

WU, C.F.J. (1986). "Jackknife, bootstrap and other resampling methods in regression analysis". *Ann. Statist.* 14, 4, 1261- 1295.

WU, W., CARLSTEIN, E. and CAMBANIS, S. (1990). "Bootstrapping the sample mean for data with infinite variance". Technical Report 296. Chapel-Hill.