Centrum voor Wiskunde en Informatica

REPORT*RAPPORT*

Bootstrap resampling: a survey of recent research in the Netherlands

R. Helmers and H. Putter

Department of Operations Reasearch, Statistics, and System Theory

BS-R9517 1995

Report BS-R9517
ISSN 0924-0659

CWI
P.O. Box 94079
1090 GB Amsterdam
The Netherlands

Bootstrap Resampling: a Survey of Recent Research in the Netherlands

Roelof Helmers
CWI
P.O.Box 94079, 1090 GB Amsterdam, The Netherlands

Hein Putter
Department of Mathematics, University of Amsterdam
Plantage Muidergracht 24
1018 TV Amsterdam, The Netherlands

Abstract

A survey of recent research in the Netherlands in the general area of bootstrap resampling methods is presented.

AMS Subject Classification (1991): 62G09, 62E20, 62E25.
Keywords and Phrases: bootstrap resampling, Edgeworth expansions, Monte Carlo.
Note: This paper will be submitted for publication in the Proceedings of the Regional Conference on Mathematical Analysis and Statistics, Gadjah Mada University, Yogyakarta, Indonesia, July 10-13, 1995.

1. Introduction

The bootstrap is a computer-intensive method for estimating the variability of statistical quantities and for setting confidence regions. Efron's bootstrap is to resample the data. Given observations $X_1, \dots, X_n$ artificial samples are drawn with replacement from $X_1, \dots, X_n$, putting equal probability mass $\frac{1}{n}$ at each X_i. Bootstrap resampling often gives much better estimates than traditional statistics usually provide us with. The bootstrap can also be an effective tool in many problems of statistical inference, which otherwise would have been too complicated to handle; e.g., the construction of a confidence band in nonparametric regression, testing for the number of modes of a density, or the calibration of confidence bounds. For an excellent introduction to the bootstrap we refer the interested reader to Efron and Tibshirani (1993).

In this paper we survey recent research in the Netherlands in the general area of bootstrap resampling methods. In section 2 we describe Efron's nonparametric bootstrap in a simple setting and address briefly the important question: when does Efron's bootstrap work and when does it fail? The benificial effect of 'Studentization' before bootstrapping is discussed in section 3. Some Monte Carlo results, which support this claim in finite samples, are presented in section 4. Finally, in section 5, a short overview of miscellaneous topics of current interest in bootstrap theory is given.

2. Efron's nonparametric bootstrap

We consider a classical problem in statistics: Suppose $X_1, \dots, X_n$ is a random sample of size n from a population with unknown distribution (df)F on the real line. Let, in addition,

$$\theta = \theta(F) \tag{2.1}$$

denote a real-valued parameter which we want to estimate.

Let $T_n = T_n(X_1, \dots, X_n)$ denote an estimator of θ, based on the data $X_1, \dots, X_n$. Our object of interest in this section is the distribution of $n^{\frac{1}{2}}(T_n - \theta)$, i.e. we define

$$G_n(x) = P(n^{\frac{1}{2}}(T_n - \theta) \leq x), \ -\infty < x < \infty, \tag{2.2}$$

where P denotes 'probability' corresponding to F. Clearly G_n, the exact df of $n^{\frac{1}{2}}(T_n - \theta)$, is unknown, because F is not known to us, but we can try to estimate it. The bootstrap estimator (approximation) of G_n is given by

$$G_n^*(x) = P_n^*(n^{\frac{1}{2}}(T_n^* - \theta_n) \leq x), \ -\infty < x < \infty. \tag{2.3}$$

Here $T_n^* = T_n(X_1^*, \ldots, X_n^*)$, where $X_1^*, \ldots, X_n^*$ denotes an artifical random sample - the bootstrap sample - from $\hat{F}_n$, the empirical df of the original observations $X_1, \ldots, X_n$, and $\theta_n = \theta(\hat{F}_n)$. Note that $\hat{F}_n$ is the random df - in fact a step function - which puts probability mass $\frac{1}{n}$ at each of the X_i's $(1 \leq i \leq n)$, sometimes referred to as the resampling distribution.

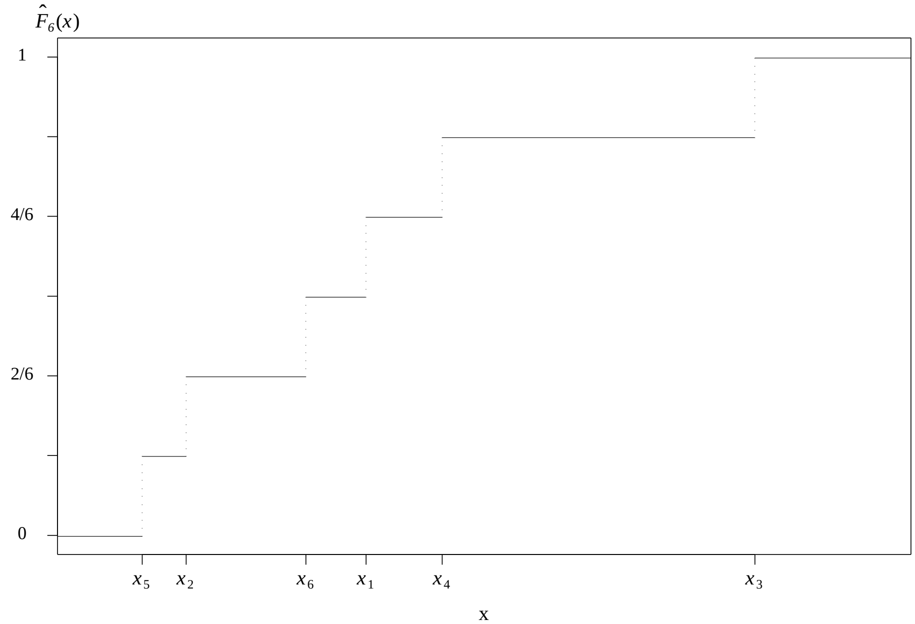

Figure 1: Empirical distribution function based on observations $x_1, \ldots, x_6$.

Finally, P_n^* denotes 'probability' corresponding to $\hat{F}_n$, conditionally given $\hat{F}_n$, i.e. given the observations $X_1, \ldots, X_n$. To emphasize the fact that G_n^* is a conditional df, one may as well write

$$G_n^*(x) = P_n^*(n^{\frac{1}{2}}(T_n^* - \theta_n) \leq x | X_1, \ldots, X_n), \ -\infty < x < \infty, \tag{2.4}$$

instead of (2.3). Obviously, given the observed values $X_1, \ldots, X_n$ in our sample, $\hat{F}_n$ is completely known, and - at least in principle - G_n^* is also completely known. We may view G_n^* as the empirical counterpart (in the 'bootstrap world') to G_n in the 'real world'. In practice, exact computation of G_n^* by complete enumeration is usually impossible (even in our sophisticated computer age), but one may rely on Monte Carlo simulations to obtain accurate numerical estimates of G_n^*. This topic will be the subject of section 4.

When does Efron's bootstrap work? Consistency of the bootstrap approximation G_n^*, viewed as an estimate of G_n is generally viewed as an absolute prerequisite for Efron's bootstrap to work. In other words we require

$$\sup_x |G_n(x) - G_n^*(x)| \to 0, \ \text{as } n \to \infty \tag{2.5}$$

to hold, with P-probability one (i.e., for almost all sequences $X_1, X_2, \ldots$), or a slightly weaker version of it, namely that (2.5) holds only in P-probability, rather than P-almost surely. Of course, the assertion (2.5) is only a first order asymptotic result, and the error committed, when the bootstrap is applied in finite samples - say, with sample size $n = 20$ - may still be quite large.
In the important special case where $\theta(F) = \int x dF(x)$, the population mean, and $T_n = \bar{X}_n = n^{-1}\sum_{i=1}^n X_i$, the sample mean, the by now classical result of Bickel and Freedman (1981) asserts that (2.5) holds true, i.e. Efron's bootstrap works, provided the variance σ^2 of the underlying df F is finite. If σ^2 is infinite the situation becomes more complex: Athreya (1987) has proved that Efron's bootstrap still works, provided F is in the domain of attraction of the normal law; otherwise Efron's bootstrap fails. We also refer the interested reader to chapter 1 of Mammen (1992).

Van Zwet (1994) has recently studied the performance of Efron's bootstrap estimate of variance for arbitrary symmetric statistics $T_n = T_n(X_1, \ldots, X_n)$ with finite second moment using the Hoeffding decomposition. He showed that Efron's bootstrap will typically work, provided $\sum_{i=1}^n E(T_n|X_i)$, the linear term in the Hoeffding decomposition of T_n, is the dominant one and the higher order terms in the Hoeffding decomposition tend to zero rather fast. The requirement concerning the linear term is also shown to be a necessary condition for the consistency of Efron's bootstrap; otherwise (2.5) generally fails to hold. A specific example for which Efron's bootstrap fails is a 'degenerate U-statistic', and a (slightly) different bootstrap resampling scheme is in fact needed here, to ensure consistency (cf. Dehling & Mikosch (1994)). We also refer to Bickel, Götze, van Zwet (1994) and Mammen (1992) for a discussion of this and other examples of bootstrap failure. A specific case for which Efron's bootstrap works - namely Serfling's class of generalized L-statistics - is considered in Helmers, Janssen, Serfling (1990).

Gill (1989) has proved (in a fairly abstract framework involving Banach spaces and Hadamard differentiability) that "the bootstrap works if the δ-method works". In our simple setting this means that if (2.5) holds in P-probability (i.e. $n^{\frac{1}{2}}(T_n - \theta)$ can be bootstrapped), then $n^{\frac{1}{2}}(f(T_n) - f(\theta))$ can also be bootstrapped, provided f is differentiable at θ. This result is a very useful one, as it can be easily applied in many different situations.

In Putter and van Zwet (1993) (cf. also chapter 2 of Putter (1994)) the importance of a proper choice of the resampling distribution (not necessarily the empirical distribution $\hat{F}_n$, as in Efron's nonparametric bootstrap) is emphasized. Let $\tau_n(F)$ denote the distribution of a statistical quantity $R_n = R_n(X_1, \ldots, X_n; F)$. If $\tilde{F}_n$ denotes the resampling distribution, $\tilde{F}_n = \tilde{F}_n(X_1, \ldots, X_n)$ being an estimate of F, then the bootstrap estimate of $\tau_n(F)$ becomes $\tau_n(\tilde{F}_n)$. Note that $\tilde{F}_n$ may be very different from the empirical distribution $\hat{F}_n$ of $X_1, \ldots, X_n$, e.g. one may consider $\tilde{F}_n = F_{\hat{\theta}_n}$, when it is a priori known that F belongs to a parametric model $\{F_\theta, \theta \in \Theta\}$, the finite dimensional parameter θ is estimated by $\hat{\theta}_n = \hat{\theta}_n(X_1, \ldots, X_n)$, a consistent estimator of θ (parametric bootstrap). Putter and van Zwet (1993) have proved the following general result: if (i) for every n the map $F \to \tau_n(F)$ is continuous, (ii) for every $F \in \mathcal{F}$ ($\mathcal{F}$ is the collection of distributions, in which F takes its values) $\tau_n(F)$ converges to a limit $\tau(F)$ (in an appropriate metric) and (iii) there exists an estimator $\tilde{F}_n$ of F with values in $\mathcal{F}$, which is consistent for $F \in \mathcal{F}$ (with respect to a suitable topology Π), then there exists a "topologically small" set D (a set of the first category) in the topological space $(\mathcal{F},\Pi)$, such that the sequence $\{\tau_n\}$ is equicontinuous at every point $F \in \mathcal{F}\backslash D$ and hence the bootstrap estimator $\tau_n(\tilde{F}_n)$ with resampling distribution $\tilde{F}_n$ is (weakly) consistent for every $F \in \mathcal{F}\backslash D$.

3. The accuracy of bootstrap estimates

In section 2 we have seen that Efron's bootstrap is consistent for the case of the sample mean $\bar{X}_n = n^{-1}\sum_{i=1}^n X_i$, if the underlying distribution F of the observations has a finite second moment. With

$$G_n(x) = P(n^{\frac{1}{2}}(\bar{X}_n - \mu) \leq x) \tag{3.1}$$

and

$$G_n^*(x) = P_n^*(n^{\frac{1}{2}}(\bar{X}_n^* - \bar{X}_n) \leq x) \tag{3.2}$$

we have, with P-probability 1,

$$\sup_x |G_n(x) - G_n^*(x)| \to 0, \text{ as } n \to \infty \tag{3.3}$$

whenever $0 < \int x^2 dF(x) < \infty$. However, the question remains: how well does Efron's bootstrap estimate G_n^* approximates G_n? To answer this question we argue as follows:

Because of the central limit theorem we know that

$$\sup_x |G_n(x) - \Phi(\frac{x}{\sigma})| \to 0, \text{ as } n \to \infty \tag{3.4}$$

where Φ is the standard normal df, and σ^2 denotes the variance of F. Similarly, we also know that, with P-probability 1,

$$\sup_x |G_n^*(x) - \Phi(\frac{x}{s_n})| \to 0, \text{ as } n \to \infty \tag{3.5}$$

where $S_n^2 = n^{-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \int (x - \bar{X}_n)^2 d\hat{F}_n$; the variance σ^2 of F is replaced by its bootstrap counterpart S_n^2.

The famous Berry-Esseen theorem now asserts that the speed of the convergence towards normality in (3.4) and (3.5) is of the order $n^{-\frac{1}{2}}$, provided $\int |x|^3 dF(x) < \infty$. Also $\Phi(\frac{x}{\sigma})$ and $\Phi(\frac{x}{s_n})$ typically differ by an amount of order $n^{-\frac{1}{2}}$ in P-probability. So, we can conclude from this that the rate of convergence in (3.3) is also of the classical order $n^{-\frac{1}{2}}$. The next step is: what can we do to improve the accuracy of the bootstrap estimator G_n^*? One possibility is to employ 'Studentization'. That is, instead of the statistical quantity $n^{\frac{1}{2}}(\bar{X}_n - \mu)$ and its bootstrap version $n^{\frac{1}{2}}(\bar{X}_n^* - \bar{X}_n)$, we consider the *Student t-statistic* $n^{\frac{1}{2}}(\bar{X}_n - \mu)/S_n$ and its bootstrap counterpart $n^{\frac{1}{2}}(\bar{X}_n^* - \bar{X}_n)/S_n^*$, with respective distribution functions

$$G_{ns}(x) = P(n^{\frac{1}{2}}(\bar{X}_n - \mu)/S_n \leq x), \ -\infty < x < \infty, \tag{3.6}$$

and

$$G_{ns}^*(x) = P_n^*(n^{\frac{1}{2}}(\bar{X}_n^* - \bar{X}_n)/S_n^* \leq x), \ -\infty < x < \infty. \tag{3.7}$$

Note that S_n^{*2} is nothing but S_n^2, with the X_i's replaced by the X_i^*'s. Similarly, as in (3.4) and (3.5), one can show that

$$\sup_x |G_{ns}(x) - \Phi(x)| \to 0, \text{ as } n \to \infty, \tag{3.8}$$

and, with P-probability 1,

$$\sup_x |G_{ns}^*(x) - \Phi(x)| \to 0, \text{ as } n \to \infty, \tag{3.9}$$

provided $0 < \int x^2 dF(x) < \infty$. Clearly, while G_n and G_n^* were approximated by different normal df's, G_{ns} and G_{ns}^* are both approximated by the standard normal df. Thus we have already eliminated one source of error leading to the $n^{-\frac{1}{2}}$ rate. The second contribution to the $n^{-\frac{1}{2}}$ rate of $G_n^* - G_n$ to zero, is a consequence of the Berry-Esseen theorem. However, it is possible to refine this theorem, and replace (3.8) and (3.9) by one-term Edgeworth expansions:

$$G_{ns}(x) = \Phi(x) + 6^{-1} n^{-\frac{1}{2}} \kappa_3 (2x^2 + 1)\phi(x) + o(n^{-\frac{1}{2}}) \tag{3.10}$$

and, similarly, in P-probability,

$$G^*_{ns}(x) = \Phi(x) + 6^{-1} n^{-\frac{1}{2}} \kappa_{3n} (2x^2 + 1)\phi(x) + o(n^{-\frac{1}{2}}) \tag{3.11}$$

where ϕ denotes the standard normal density. (To validate these asymptotic expansions one needs the additional requirement that F is a non-lattice distribution, as well as a finite absolute moment of F of order $p > 3$; A proof of these assertions under the present set of conditions can be found in Putter (1994); see also Helmers (1991) who requires $p > 4$ and Hall (1992), where the even stronger assumption $p = 6$ is made. The question whether $p = 3$ would already be sufficient is still open). The quantities κ_3 and κ_{3n} are defined by

$$\kappa_3 = \frac{E(X_1 - \mu)^3}{\sigma^3} \tag{3.12}$$

and

$$\kappa_{3n} = \frac{n^{-1}\sum_{i=1}^n (X_i - \bar{X}_n)^3}{(n^{-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2)^{\frac{3}{2}}} \tag{3.13}$$

i.e. κ_3 is the third cumulant of F (measuring skewness) and κ_{3n} is the empirical (bootstrap) counterpart of κ_3. Now one simply compares (3.10) and (3.11), and finds that

$$\sup_x |G_n(x) - G^*_n(x)| \sim \frac{|\kappa_3 - \kappa_{3n}|}{6n^{\frac{1}{2}}} \sup_x (2x^2 + 1)\phi(x) \tag{3.14}$$

where $a_n \sim b_n$ means $a_n/b_n \to 1$, as $n \to \infty$. Note that $\sup_x (2x^2 + 1)\phi(x)$ is finite. Because, by the weak law of large numbers, $\kappa_{3n} \to \kappa_3$, as $n \to \infty$, we can now easily conclude that in P-probability

$$n^{\frac{1}{2}} \sup_x |G_{ns}(x) - G^*_{ns}(x)| \to 0, \text{ as } n \to \infty \tag{3.15}$$

provided $0 < \int |x|^p dF(x) < \infty$, for some $p > 3$, and F is non-lattice. In fact, under somewhat more stringent conditions, one can show that $\sup_x |G_{ns}(x) - G^*_{ns}(x)|$, the accuracy of the bootstrap approximation, is of the exact order n^{-1} in P-probability. In other words: the bootstrap estimate G^*_{ns} is asymptotically closer to G_{ns} than the standard normal df. This 'bootstrap is better than normal' property of Efron's bootstrap for Student's t statistic clearly suggests the benificial effect of 'Studentization' before bootstrapping for this important special case.

The above result for Student's t is in fact already known for about 10 years (see e.g. the references in Hall (1992)) and it is generally viewed as an important argument in favor of Efron's bootstrap. In Helmers (1991) it was proved that the 'bootstrap is better than normal' property also holds true for more complicated non linear statistics like Hoeffding's famous class of U-statistics. Here we note that the moment conditions needed in Helmers (1991) can be somewhat relaxed. This is an easy consequence of a recent result of Putter (1994). A specific U-statistic of interest is the statistic $\binom{n}{2}^{-1} \sum\sum_{1 \le i < j \le n} |X_i - X_j|$, which is Gini's mean difference, an estimator of scale.

We conclude this section with the remark that the extension of the 'bootstrap is better than normal' property to arbitrary Studentized symmetric statistics is still an interesting open problem at present. In any case, however, the quadratic and higher order terms in the Hoeffding decomposition for a symmetric statistic $T_n = T_n(X_1, \ldots, X_n)$ should be of a required order of magnitude, otherwise the speed of bootstrap convergence asserted in (3.15) fails to hold. An important specific example of the latter is the case of the median, and more generally, quantiles. In such 'non-smooth' cases (the parameter of interest, e.g., $\theta = \theta(F) = F^{-1}(\frac{1}{2})$ is a much less smooth functional of F, then the parameter $\theta = \theta(F) = \int x dF(x)$) we have - instead of (3.15) - a slower rate (roughly of order $n^{-\frac{1}{4}}$) of convergence of Efron's bootstrap approximation. We refer to section 4 for a small sample computer simulation which clearly supports this claim. The same phenomenon also shows up in the case of U-quantiles considered in Helmers, Janssen, Veraverbeke (1992) (cf. Helmers, Hušková (1994) for an extension to multivariate U-quantiles). A specific U-quantile of interest is the well known Hodges-Lehmann estimator of location , which is given by the median of all pairwise averages. Young (1994), p392, in a prominent recent review paper remarks that establishing the consistency and the accuracy of the bootstrap approximation for statistics like the Hodges-Lehmann estimator is important, because 'the contexts to which the results apply are highly relevant to precisely the sort of circumstances - when there is limited knowledge about the underlying distribution - for which bootstrap was designed'.

4. Monte Carlo

Given a random sample $X_1, \dots, X_n$ from a distribution F, bootstrap estimates require bootstrap samples $X_1^*, \dots, X_n^*$ from the empirical distribution $\hat{F}_n$. For the distribution of a statistical quantity $R_n = R_n(X_1, \dots, X_n; F)$ (we considered the important special case $R_n = n^{\frac{1}{2}}(T_n - \theta)$ in section 2), its bootstrap estimate is the conditional distribution of $R_n^* = R_n(X_1^*, \dots, X_n^*; \hat{F}_n)$, given $X_1, \dots, X_n$. In principle, this distribution is known. For a sample $X_1, \dots, X_n$ of n distinct numbers there are $\binom{2n-1}{n}$ distinct bootstrap samples, so the distribution of R_n^* can be retrieved by complete enumeration. For $n = 10$ already however, near to 100,000 bootstrap samples have to be enumerated, so very soon this method becomes unfeasible and we have to turn to another solution: Monte Carlo simulation. In a sense, this boils down to repeatedly drawing a random bootstrap sample from all possible bootstrap samples. We fix a large number B. With the use of the computer, we generate a bootstrap sample and calculate the resulting value of R_n^*. By repeating this procedure B times, we obtain B values, say $R_{n,1}^*, \dots, R_{n,B}^*$, which give an approximation to the distribution of R_n^*. Monte Carlo simulation was already well established before the invention of the bootstrap in 1979, but it finds a very natural place here. Generating a bootstrap sample amounts to randomly drawing a sample of size n with replacement from $X_1, \dots, X_n$.

The Monte Carlo procedure introduces a second source of randomness. The values $R_{n,1}^* \dots, R_{n,B}^*$ can be seen as independent realizations from the distribution of R_n^*. Thus, if G_n^* is the distribution function of R_n^* and $G_{n,B}^*$ is the empirical distribution function of $R_{n,1}^*, \dots, R_{n,B}^*$, then for fixed $X_1, \dots, X_n$ and fixed n,

$$\sqrt{B} \sup_x \left| G_{n,B}^*(x) - G_n^*(x) \right| \to \sup_x \mathcal{B}(G_n^*(x)) \ , \quad \text{as } B \to \infty \ , \tag{4.1}$$

where $\mathcal{B}(\cdot)$ is a Brownian bridge (cf. section 5). This limiting distribution has been investigated extensively in the context of goodness of fit tests (Kolmogorov-Smirnov test). Roughly speaking, this implies that the Monte Carlo error decreases at the rate of $\frac{1}{\sqrt{B}}$ as B tends to infinity.

The above analysis indicates that by choosing B suitably large, we can control the second source of randomness, and make sure that the Monte Carlo error is small in comparison with the bootstrap approximation error. In the bootstrap literature, a remarkably low number of bootstrap samples is advised, usually around 100–1000. Of course, this number depends first of all on the kind of inference we want to draw. For estimation of standard errors a lower number suffices (B=200). For construction of confidence intervals B should be taken considerably larger. If we are aiming for second order accuracy as in section 3 (cf. (3.14) and (3.15)), we should take care that too low a choice for B doesn't ruin the second order correctness. This implies that the Monte Carlo error, $\frac{1}{\sqrt{B}}$, should be of a smaller order than $\frac{1}{n}$, the accuracy of the bootstrap approximation, i.e. B should be of a larger order than n^2. So with larger sample sizes we should not decrease B (as is common practice to save computing time) but increase B to keep the accuracy comparably small.

We conclude this section with two simulations. In Figure 2 (borrowed from Putter (1994)), we consider bootstrapping Student's t, the subject of section 3. We generated a sample $X_1, \dots, X_{20}$

of size $n = 20$ from a standard exponential distribution. The distribution G_{ns} was approximated by Monte Carlo using 10^7 samples. Next, on the basis of the sample $X_1, \ldots, X_{20}$, the distribution G_{ns} was estimated in three ways, first using the classical normal approximation, secondly using the bootstrap G^*_{ns} (as in (3.7), using Monte Carlo simulation with $B = 10^6$). With this choice of B, we are pretty certain that the Monte Carlo error is negligible. The third way uses empirical Edgeworth expansion (EEE). Here the expansion (3.10) is taken and the constant κ_3 is estimated directly by its empirical counterpart κ_{3n} in (3.13). To make the differences between these three methods discernible we have plotted for each of the three methods the resulting estimate minus the target distribution G_{ns}. The graph that lies closest to zero corresponds therefore to the best approximation. It is clearly seen that both bootstrap and Edgeworth expansion outperform the normal approximation. The bootstrap performs slightly better than EEE, due to the fact that the bootstrap also implicitly estimates higher order terms in the Edgeworth expansion consistently.

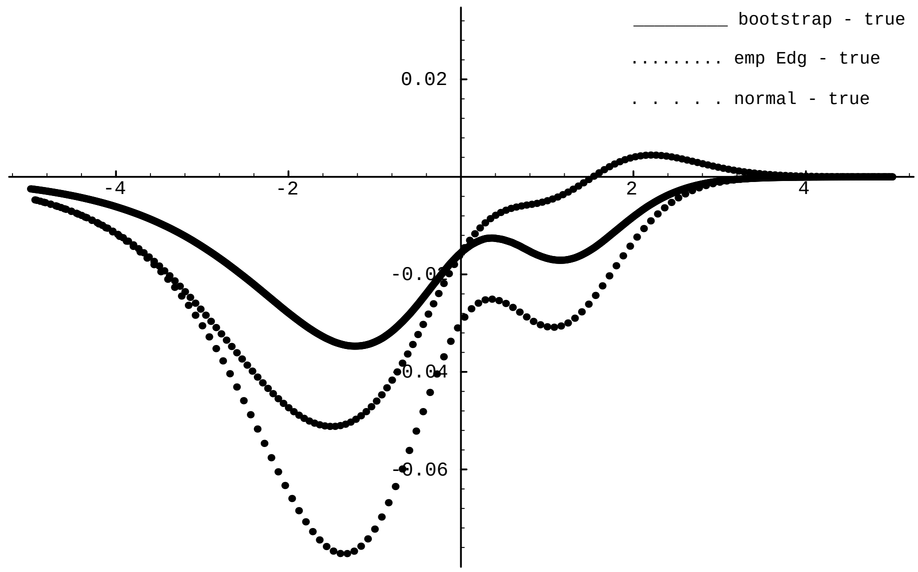

Figure 2: n=20; exponential

The second example concerns the median. Let $X_1, \ldots, X_n$ be an i.i.d. sample from a distribution F, let $\theta = F^{-1}(\frac{1}{2})$ be the median of F, and let us suppose that F is differentiable at θ, with derivative $f(\theta) > 0$. Here F^{-1} is the left continuous version of the inverse of F, i.e. $F^{-1}(t) = \inf\{x : F(x) \geq t\}$. Let $T_n = \theta(\hat{F}_n) = \hat{F}_n^{-1}(\frac{1}{2})$ denote the sample median and let $R_n = n^{\frac{1}{2}}(T_n - \theta)$. It is well known that the exact distribution of R_n converges to a normal distribution with variance $(4(f(\theta))^2)^{-1}$. Given $X_1, \ldots, X_n$, let $X_1^*, \ldots, X_n^*$ be an i.i.d. sample from $\hat{F}_n$, the empirical distribution of $X_1, \ldots, X_n$, and define $T_n^* = \hat{F}_n^{*-1}(\frac{1}{2})$, the median of the bootstrap sample ($\hat{F}_n^*$ is the empirical distribution function of $X_1^*, \ldots, X_n^*$). The bootstrap estimator of the exact df of R_n is the conditional df of $R_n^* = n^{\frac{1}{2}}(T_n^* - T_n)$, conditionally given $\hat{F}_n$. Asymptotic theory (Singh (1981), Hall and Martin(1988); see also Reiss (1989)) tells us that the accuracy of the bootstrap approximation is of the order $n^{-\frac{1}{4}}$ in probability. It is also well known that the smoothed bootstrap, where resampling is done from a smoothed version of the empirical distribution, results in a better approximation; depending on the kernel and the bandwidth chosen in the smoothing procedure one can obtain a rate of the order $\mathcal{O}(n^{-\frac{1}{2}+\varepsilon})$, for any $\varepsilon > 0$. For details we refer to Hall, DiCiccio and Romano (1989). In the computer simulations that led to Figure 3 we have generated a sample of size $n = 100$ from a standard normal distribution. For the smoothed bootstrap we used a normal kernel and a bandwidth $h = 0.1$. Both Efron's bootstrap and the smoothed bootstrap could be evaluated exactly, i.e. without Monte Carlo simulation, by using formula (6.4.2), page 222 of Reiss(1989). Note that, although theoretically Efron's bootstrap is consistent, it is worthless in practice, even for a sample size n as large as 100. The difference with the true distribution function is maximized at $x = 0.39$; at that point the true df equals 0.657 while the bootstrap approximation yields 0.935, which means a relative error of more than 40%. A maximum difference of 0.278 between the true distribution and Efron's bootstrap approximation is in line with the results of a computer simulation in Reiss(1989), see page 225. The smoothed bootstrap seems to perform much better. A more sophisticated choice of kernel and bandwidth will presumably lead to a further reduction of the error of the smooth bootstrap approximation. (cf. also D.Hinkley's comment (page 401) on Young's 1994 review paper).

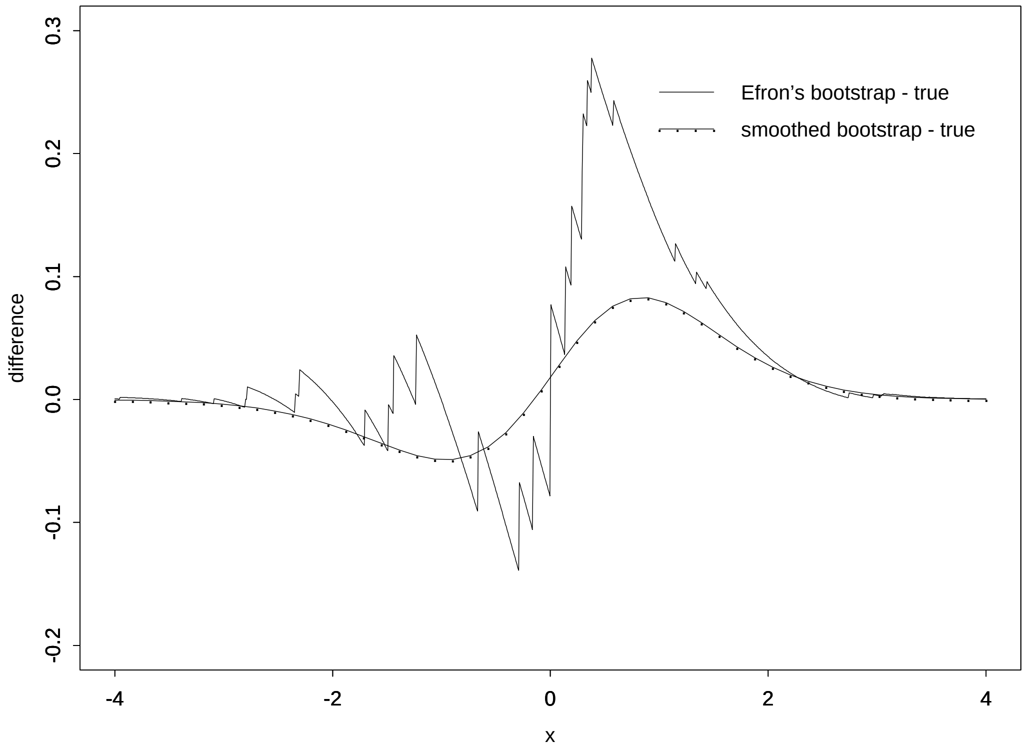

Figure 3: median; n=100; normal

5. MISCELLANEOUS

In this section we briefly discuss a few selected topics of current interest in bootstrap theory and its applications: bootstrapping (functionals of) empirical processes, resampling methods for finite populations, and (semi) parametric & smooth bootstrapping.

The general area of empirical processes and bootstrap resampling has received a lot of attention in the past decade (cf. Wellner (1992) for an excellent survey of recent developments). The starting point of all this was the influential paper by Bickel and Freedman (1981). These authors proved the following by now classical result: the empirical process U_n, defined by

$$U_n(x) = n^{\frac{1}{2}}(\hat{F}_n(x) - F(x)),$$

for all real x, and the bootstrapped empirical process U_n^*, defined by

$$U_n^*(x) = n^{\frac{1}{2}}(\hat{F}_n^*(x) - \hat{F}_n(x)),$$

for all real x, converge both weakly to the same Gaussian process $\mathcal{B}(F)$ (with probability 1, i.e. for almost all sequences $X_1, X_2, \ldots$). Here F denotes the common *df* of the independent random variables $X_1, X_2, \ldots$, $\hat{F}_n$ is the empirical *df* based on $X_1, \ldots, X_n$, and $\hat{F}_n^*$ denotes the empirical *df* based on $X_1^*, \ldots, X_n^*$, a bootstrap (re)sample of size n, drawn from $\hat{F}_n$; $\mathcal{B}$ is the well-known Brownian bridge process. A beautiful and farreaching extension of the Bickel & Freedman result was obtained by Giné & Zinn (1990). Helmers (1994) shows (by means of the classical method of moments) that the local time of the empirical process (which is nothing but $n^{-\frac{1}{2}}$ times the number of zero-crossings of the U_n-process) can also be bootstrapped. However, this functional is not at all continuous and so it appears that our problem can not be settled easily by an application of an 'extended continuous mapping' theorem.

Resampling methods for finite populations is another topic of current interest. We refer to Booth, Butler and Hall (1994) for a survey. In Helmers & Wegkamp (1995) the situation is considered where the finite population is viewed as a realization of a certain superpopulation model (heteroscedastic linear regression, without intercept). This enables us to incorporate auxiliary information (past experience) in the statistical analysis. The authors first came across this problem in a 1994 statistical consultation project at CWI with the Netherlands postal services PTT Post. In this setup a new resampling scheme called 'two-stage wild bootstrapping' is proposed and studied.

In Venetiaan (1994) smooth bootstrapping is employed to estimate the performance of translation-invariant estimators in the nonparametric location problem. This is achieved by obtaining bootstrap analogues for the spread of the quantiles of the distributions of these estimators and for theoretical lower bounds - Klaassen's spread inequality and a 'confidence interval inequality' - for this spread. To bootstrap these lower bounds one has to resample from a smoothed empirical distribution, since e.g. the spread inequality requires the existence of a density. Efron's naive bootstrap would cause serious problems here.

Resampling with spatial data is very clearly an important area for future research. We briefly describe here a practical application in which spatial bootstrapping is used. In a (still ongoing) statistical consultation project at CWI with the North Sea Directorate, Ministry of Public Works the problem is to estimate the intensity of oil-pollution in the North Sea. A planar inhomogeneous Poisson process with intensity function $\lambda(., \theta)$ - parametrized by a finite dimensional parameter θ - was used as a spatial (parametric) model for the locations of (the centres of) oil spots. The parametrization enables one to incorporate the available a priori knowledge about oil pollution, such as the location of sources of oil-pollution (i.e. shipping areas or off-shore locations) and the intensity of shipping in various regions. However, nothing seems to be known about the distribution of the sizes (volumes) of oil spots, but we can of course use the sizes of the observed oil spots to estimate it (nonparametric approach). In this setup a relatively simple form of semiparametric bootstrapping was developed in order to estimate the accuracy of the estimated total amount of oil-pollution in the North Sea. We refer to Helmers (1995) for the details of all this.

References

[1] K.B. Athreya (1987), Bootstrap of the mean in the infinite variance case. *Ann. Statist.* 15, 724–731.

[2] P.J. Bickel and D.A. Freedman (1981), Some asymptotic theory for the bootstrap. *Ann.Statist.* 9, 1196–1217.

[3] P.J. Bickel, F. Götze and W.R. van Zwet (1994), Resampling fewer than n observations: gains, losses, and remedies for losses. Preprint 94-084, SFB 343 Diskrete Strukturen in der Mathematik, University of Bielefeld.

[4] J.G. Booth, R.W. Butler and P. Hall (1994), Bootstrap methods for finite populations. *Jour. Amer. Stat. Ass.*, vol 89, 428, 1282–1289.

[5] H. Dehling and T. Mikosch (1994), Random quadratic forms and the bootstrap for U-statistics. *J. Mult. Anal.* 51, 392–413.

[6] B. Efron and R.J. Tibshirani (1993), *An Introduction to the Bootstrap.* Chapman and Hall, New York.

[7] R.D. Gill (1989), Non- and Semi-parametric maximum likelihood estimators and the von Mises-Method (Part I). *Scand. J.Statist.* 16, 97–128.

[8] E. Giné and J. Zinn (1990), Bootstrapping general empirical measures. *Ann. Probab.* 18, 851–869.

[9] P. Hall (1992), *The Bootstrap and Edgeworth expansions.* Springer, New York.

[10] P. Hall and M.A. Martin (1988), Exact convergence rate of bootstrap quantile variance estimator. *Prob. Theory. Rel. Fields* 80, 261–268.

[11] P. Hall, T.J. DiCiccio and J.P. Romano (1989), On smoothing and the bootstrap. *Ann.Statist.* 17, 692–702.

[12] R. Helmers, P. Janssen, R. Serfling (1990), Berry-Esseen and bootstrap results for generalized L-statistics. *Scand. J. Statist.* 17, 65–77.

[13] R. Helmers (1991), On the Edgeworth expansion and the bootstrap approximation for a Studentized U-statistic. *Ann. Statist.* 19, 470–484.

[14] R. Helmers, P. Janssen, N. Veraverbeke (1992), Bootstrapping U-quantiles. In *Exploring the Limits of Bootstrap* (R.LePage and L.Billard, eds.), pp 145–155. Wiley, New York.

[15] R. Helmers, M. Hušková (1994), Bootstrapping multivariate U-quantiles and related statistics. *J.Mult. Anal.* 49, 97–109.

[16] R. Helmers (1994), A note on bootstrapping the local time of the empirical process. CWI report BS-R9422. Submitted for publication.

[17] R. Helmers and M.H. Wegkamp (1995), Wild bootstrapping in finite populations with auxiliary information. CWI report BS-R95xx.

[18] R. Helmers (1995), On estimating the intensity of oil pollution in the North Sea. CWI report BS-R95xx.

[19] E. Mammen (1992), *When Does Bootstrap Work? Asymptotic Results and Simulations.* Lecture Notes in Statistics 77, Springer, New York.

[20] H. Putter (1994), *Consistency of Resampling Methods.* PH.D. Thesis, University of Leiden.

[21] H. Putter and W.R. van Zwet (1993), Resampling, II: Consistency of substitution estimators with applications to the bootstrap. Submitted to *Ann. Statist.*

[22] R.D. Reiss (1989), *Approximate Distributions of Order Statistics, with Applications to Nonparametric Statistics.* Springer, New York.

[23] K. Singh (1981), On the asymptotic accuracy of Efron's bootstrap. *Ann. Statist.* 9, 1187–1195.

[24] S.A. Venetiaan (1994), *Bootstrap bounds.* PH.D. Thesis, University of Amsterdam.

[25] J.A. Wellner (1992), Bootstrap Limit Theorems: a partial survey. In *Nonparametric Statistics and Related Topics*, 313–329 (A.K.Md.E.Saleh, ed.). Elsevier Science Publishers, Amsterdam.

[26] A. Young (1994), Bootstrap: more than a stab in the dark? *Statistical Science* 9, no 3, 382–415.

[27] W.R. van Zwet (1994), Resampling, I: The jackknife and the naive bootstrap. To be submitted to *Ann. Statist.*