Bootstrap, Wild Bootstrap and Generalized Bootstrap.

Enno Mammen

Abstract. Some modifications and generalizations of the bootstrap procedure have been proposed. In this note we will consider the wild bootstrap and the generalized bootstrap and we will give two arguments why it makes sense to use these modifications instead of the original bootstrap. The first argument is that there exist examples where generalized and wild bootstrap work, but where the original bootstrap fails and breaks down. The second argument will be based on higher order considerations. We will show that the class of generalized and wild bootstrap procedures offers a broad spectrum of possibilities for adjusting higher order properties of the bootstrap.

1. Introduction. Bootstrap procedures are a powerful tool in modern statistics. In this note we discuss two new proposals how to bootstrap: the wild bootstrap procedure and generalized bootstrap procedures. Let us first recall the definition of generalized bootstrap. We will do this for the context of observing an i.i.d. sample $(X_1, \dots, X_n)$ with distribution P. The statistical problem is the estimation (or approximation) of the distribution of $R(\hat{P}_n,P)$. Here R is a real valued function defined on a set of pairs of distributions including P and empirical measures. $\hat{P}_n$ is the empirical distribution, i.e. $\hat{P}_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}$, where δ_x is the point mass concentrated at point x. Important examples of $R(\hat{P}_n,P)$ are $T(\hat{P}_n) - T(P)$ or $[T(\hat{P}_n) - T(P)] / S(\hat{P}_n)$, where T is a real valued statistical functional and S is a scale functional. In these examples an approximation of the distribution $\mathcal{L}(R(P_n,P))$ can be used for constructing confidence intervals of T(P). The bootstrap idea is to approximate $\mathcal{L}(R(\hat{P}_n,P))$ by $\mathcal{L}^*(R(\hat{P}_n^*,\hat{P}_n))$ where $\mathcal{L}^*(\dots)$ is the conditional distribution $\mathcal{L}^*(\dots|X_1, \dots, X_n)$ given the sample and where $\hat{P}_n^*$ is a random measure.

In the bootstrap procedure, as it has been originally proposed in Efron (1979), the random measure $\hat{P}_n^*$ is chosen as empirical measure $\hat{P}_n^B$ of a sample with distribution $\hat{P}_n$, i.e. $\hat{P}_n^B = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i^*}$ where $(X_1^*, \dots, X_n^*)$ is a sample with distribution $\hat{P}_n$. Usually, here the calculation of $\mathcal{L}^*(R(\hat{P}_n^B,\hat{P}_n)$ is done by Monte Carlo. For this purpose, M i.i.d. samples $(X_1^*, \dots, X_n^*)$, with distribution $\hat{P}_n$ are generated. This gives M values $R_1, \dots, R_M$ of $R(\hat{P}_n^B,\hat{P}_n)$. The distribution $\mathcal{L}^*(R(\hat{P}_n^B,\hat{P}_n))$ can now be approximated by the empirical distribution of $R_1, \dots, R_M$. How large M has to be chosen depends on the application. Popular choices are M = 1000 or larger. The generation of the samples $(X_1^*, \dots X_n^*)$ is nothing else than n times drawing with replacement out of the set $\{X_1, \dots, X_n\}$. This interpretation was also the motivation for the nice name of bootstrap. The bootstrap procedure was been introduced in Efron (1979). Introductions to bootstrap are also given in the articles Efron and Gong (1983) and Efron and Tibshirani (1986) and in the book Helmers (1991). Many applications of bootstrap are discussed in Efron (1982). Survey articles are Beran (1984), Hinkley (1988) and Diciccio and Romano (1988). For an asymptotic treatment of bootstrap see also the books Beran and Ducharme (1991), Mammen (1992a) and the references cited therein. A detailed analysis of bootstrap based on higher-order Edgeworth expansions can be found in the book Hall (1992).

Let us now come to generalized bootstrap. In this approach the random measure $\widehat{P}_n^*$ used in the resampling is of the form

$$\widehat{P}_n^{GB} = n^{-1} \sum_{i=1}^{n} N_i \, \delta_{X_i} ,$$

where $(N_1, \dots, N_n)$ are random weights which are conditionally (given $X_1, \dots, X_n$)) exchangeable (i.e. $\mathcal{L}^*\big(N_1, \dots, N_n \big) = \mathcal{L}^*\big(N_{\pi(1)}, \dots, N_{\pi(n)} \big)$ for all permutations π of $\{1, \dots, n\}$). Conditions on $(N_1, \dots, N_n)$ under which generalized bootstrap works (i.e. is consistent) are given in Mason and Newton (1992), see also section 3 and Häusler, Mason and Newton (1992), Husková and Janssen (1992a,b). Resampling of empirical processes using generalized bootstrap have been studied in Praestgaard and Wellner (1992). We get bootstrap as a special case of generalized bootstrap if we use multinomialy distributed weights, i.e. $(N_1, \dots, N_n) \sim \mathrm{MULT}\,(1/n, \dots, 1/n; n)$. Other exchangeable weights have been also proposed in a Bayesian context (see Rubin (1981), Lo (1987), Weng (1989)). Let us mention here also the following examples of generalized bootstrap.

EXAMPLE 1. (BOOTSTRAP SUBSAMPLING).

$$\frac{n}{m} \, (N_1, \dots, N_n) \sim \mathrm{MULT}\,(1/n, \dots, 1/n; m) \quad \text{for an } m < n.$$

EXAMPLE 2. (I.I.D. RESAMPLING).

$$N_i = M_i \quad \text{or} \quad N_i = M_i \,/\, [n^{-1} \sum_{i=1}^{n} M_i] \qquad (i = 1, \dots, n),$$

where $M_1, \dots, M_n$ are i.i.d..

EXAMPLE 3. (GENERALIZED JACKKNIFE). For a deterministic vector $\rho = (\rho(1), \dots, \rho(n))$ in $\mathbf{R}^n$ put $N_i = \rho(\Pi(i))$, where Π is a random permutation of $(1, \dots, n)$.

Another resampling scheme is wild bootstrap. The idea of wild bootstrap can be better explained in the set up of a regression model

$$Y_i = m(x_i) + \varepsilon_i \qquad (i = 1, \dots, n).$$

Here $x_1, \dots, x_n$ are deterministic design vectors. $\varepsilon_1, \dots, \varepsilon_n$ are independent with mean zero. However, it is not assumed that they must have the same distribution. $Y_1, \dots, Y_n$ are the observations and m is an unknown regression function. Suppose a (parametric or nonparametric) estimate $\hat{m}$ of m is given and we are interested in the distribution $\mathcal{L}(R(\hat{m},m))$. The wild bootstrap estimate is $\mathcal{L}^*(R(\hat{m}^{WB};\hat{m}))$, where again $\mathcal{L}^*(\quad)$ is the conditional distribution $\mathcal{L}(\quad | Y_1, \dots, Y_n)$ given the observations. The estimate $\hat{m}^*$ is based on the sample (x_i, Y_i^*), where

$$Y_i^* = \hat{m}(x_i) + \varepsilon_i^*$$

and $\varepsilon_1^*, \dots, \varepsilon_n^*$ are conditionally independent and are constructed such that

$$E^* \varepsilon_i^* = 0 , \qquad E^*(\varepsilon_i^*)^2 = \hat{\varepsilon}_i^2 .$$

Here $\hat{\varepsilon}_i$ is the i-th residual $Y_i - \hat{m}(x_i)$ and $E^*(\quad)$ is the conditional expectation. Sometimes (in particular, if one expects that in the considered example higher order Edgeworth expansions are accurate) one requires additionaly that $E^*(\varepsilon_i^*)^3 = \hat{\varepsilon}_i^3$. The random distribution $\mathcal{L}^*(\varepsilon_i^*)$ could be interpreted as an estimate of $\mathcal{L}(\varepsilon_i)$. This is an estimate of a distribution which uses only *one* single real valued residual. Because this sounds a little bit silly wild bootstrap got its name. For most applications no consistent estimation of every single $\mathcal{L}(\varepsilon_i)$ is necessary. What is needed is the consistent estimation of distributions of certain averages of the ε_i. This makes the wild bootstrap working. Wild bootstrap variance estimates have been proposed for linear models in Wu (1986). That this resampling scheme could also be used for the estimation of distribution has been remarked by Beran (1986). An example where bootstrap breaks down and where wild bootstrap works is given in Härdle and Mammen (1993). Further discussions of wild bootstrap can be found in Liu (1988), Liu and Singh (1991, 1992), Zheng and Tu (1988) and Mammen (1992b, 1993), see also Mammen (1992a). In this note we consider only the following constructions of $\varepsilon_1^*, \dots, \varepsilon_n^*$. First generate $Z_1, \dots, Z_n$ i.i.d. ~ Q with $E(Z_i) = 0$, $EZ_i^2 = 1$ (and, occasionally, $EZ_i^3 = 1$). Then put $\varepsilon_i^* = \hat{\varepsilon}_i Z_i$. For choices of Q see Liu (1988) and Mammen (1992a).

2. Asymptotic equivalence of bootstrap, wild bootstrap and normal approximations. For the comparison of bootstrap and wild bootstrap let us consider a model which lies somehow between an i.i.d. model and a regression model. For the model of a triangular area of independent observations $X_{n,1}, \dots, X_{n,n}$ we are interested in estimating the distribution of $R = R(\hat{P}_n, P_{n,i}\ (1 \le i \le n)) = T(\hat{P}_n) - \frac{1}{n}\sum_{i=1}^{n} T(P_{n,i})$. Here $P_{n,i}$ is the distribution of $X_{n,i}$. To simplify matters we restrict ourselves here to linear statistical functionals $T(P) = \int x\, dP$. As bootstrap estimate of $\mathcal{L}(R)$ we get then $\mathcal{L}^*_{n,B} = \mathcal{L}^*(T(\hat{P}_n^B) - T(\hat{P}_n))$.

The wild bootstrap estimate $\mathcal{L}^*_{WB}$ can be written as

$$\mathcal{L}^*_{n,WB} = \mathcal{L}^*\Big(\frac{1}{n}\sum_{i=1}^{n} Z_i\, [\, X_{n,i} - T(\hat{P}_n)] \Big),$$

where Z_i has a fixed conditional distribution: $\mathcal{L}^*(Z_i) = Q$.

Bootstrap, wild bootstrap and normal approximations are closely related in this model. For instance, here wild bootstrap turns out to be a special case of generalized bootstrap. Indeed, the random factors Z_i can also be interpreted as random weights. Therefore, $\mathcal{L}^*_{n,WB}$ is nothing else than i.i.d. resampling which is an example of generalized bootstrap, see example 2 in the introduction. On the other hand we will see in the next section that consistency of generalized bootstrap can be proved by constructing an accompaning asymptotically equivalent sequence of wild bootstrap schemes. Furthermore, the normal approximation is a special case of wild bootstrap. To see this one has to put $Q = N(0,1)$. Then we get $\mathcal{L}^*_{n,WB} = N(0,S_n^2)$, where S_n^2 is the usual variance estimate $S_n^2 = \frac{1}{n^2}\sum_{i=1}^{n} (X_{n,i} - T(\hat{P}_n))^2$.

The following theorem shows that bootstrap, wild bootstrap and normal approximation work under the same conditions. This may look a little bit disappointing because in the introduction we have promised to show that wild bootstrap works under weaker conditions than bootstrap. The essential point is here that in the following theorem the distribution Q of Z_i is fixed. Things become quite different if Q is allowed to depend on n (see section 4).

To avoid any scaling we measure the accuracy of $\mathcal{L}^*_{n,B}$, $\mathcal{L}^*_{n,WB}$ and $N(0,S_n^2)$ by the Kolmogorov-distance d_∞:

$$d_\infty(\mu,\nu) = \sup_{-\infty < t < +\infty} |\, \mu(X \le t) - \nu(X \le t)\, | .$$

<u>THEOREM 1</u> (Mammen, 1992b). *For a sequence* $t(n)$ *consider* $\mathcal{L}_{t(n),n} = \mathcal{L}(T(\hat{P}_n) - t(n))$.

Then the following statements are equivalent.

(i) $d_\infty(\mathcal{L}^*_B, \mathcal{L}_{t(n),n}) \to 0$ *(in probability),*

(ii) $d_\infty(\mathcal{L}^*_{n,WB}, \mathcal{L}_{t(n),n}) \to 0$ *(in probability),*

(iii) $d_\infty(N(0,S_n^2), \mathcal{L}_{t(n),n}) \to 0$ *(in probability).*

The essential point behind these equivalences is that for asymptotic normality of a sum of independent variables it is needed that the absolutely maximal variable is of smaller order than the sum. Exactly the same is needed for bootstrap and wild bootstrap. In section 4 we look more closely at the case where the absolutely maximal summand is of the same order as the sum.

3. Consistency of generalized bootstrap. As in the last section we consider again a triangular area of independent observations $X_{n,1}, \dots, X_{n,n}$ and the linear functional $T(P) = \int x\, dP$. Then the generalized bootstrap estimate of $\mathcal{L}(R)$ is

$$\mathcal{L}^*(n^{-1} \sum_{i=1}^{n} N_{n,i} X_{n,i} - T(\hat{P}_n)),$$

where for every n the weights $N_{n,1}, \dots, N_{n,n}$ are (conditionally) exchangeable. We do not want to assume that $\sum_{i=1}^{n} N_{n,i}$ is equal to n. Therefore it makes sense to consider the following modification of the generalized bootstrap estimate

$$\mathcal{L}^*_{n,GB} = \mathcal{L}^*(n^{-1} \sum_{i=1}^{n} N_{n,i} [X_{n,i} - T(\hat{P}_n)]).$$

The following theorem can be seen by application of the results in Mason and Newton (1992).

__**THEOREM 2.**__ *Put* $\overline{N}_n = \frac{1}{n} \sum_{i=1}^{n} N_{n,i}$ *and assume* $\frac{1}{n} \sum_{i=1}^{n} (N_{n,i} - \overline{N}_n)^2 \to 1$ *(in probability) and* $E^*([N_{n,i} - \overline{N}_n]^2\, \mathbf{1}(|N_{n,i} - \overline{N}_n| > \tau_n)) \to 0$ *(in probability) for every* $\tau_n \to \infty$.

Furthermore suppose that for a sequence t(n)

$$d_\infty(\mathcal{L}_{t(n),n}, N(0,S_n^2)) \to 0 \qquad \textit{(in probability).}$$

Then it holds that

$$d_\infty(\mathcal{L}_{t(n),n}, \mathcal{L}^*_{n,GB}) \to 0 \qquad \textit{(in probability).}$$

Let us shortly sketch the basic idea of the proof (see also Mason and Newton, 1992). The theorem follows from $d_\infty(\mathcal{L}_{t(n),n}, \mathcal{L}^{**}_{n,GB}) \to 0$ (in probability), where $\mathcal{L}^{**}_{n,GB}$ is the conditional distribution of $S = n^{-1} \sum_{i=1}^{n} N_{n,i} [X_{n,i} - T(\widehat{P}_n)]$, given the sample $(X_{n,1}, \dots, X_{n,n})$ ***and*** the order statistic of $N_{n,1}, \dots, N_{n,n}$. Given this order statistic and the sample, the statistic S is a rank statistic and classical techniques of the theory of rank statistics can be applied. The main tool is the following lemma of Hájek (1961). For real numbers $a_1, \dots, a_n, b_1, \dots, b_n$ with $b_1 + \dots + b_n = 0$ consider $S = \sum_{i=1}^{n} b_i a_{\Pi(i)}$ and $T = \sum_{i=1}^{n} b_i a_{U(i)}$, where Π is a random permutation of $(1, \dots, n)$ and $U(1), \dots, U(n)$ are i.i.d. and uniformly distributed on $\{1, \dots, n\}$.

<u>LEMMA</u> (Hájek, 1961). *With* $\bar{a} = \sum_{i=1}^{n} a_i$ *it holds that*

$$\frac{E(S - T)^2}{\operatorname{Var}(T)} \le \frac{2\sqrt{2}\, n}{n-1} \; \frac{\max_{1 \le i \le n} | a_i - \bar{a} |}{\sqrt{\sum_{i=1}^{n} (a_i - \bar{a})^2}} .$$

Applied to our situation (i.e. putting $b_i = X_{n,i} - T(\widehat{P}_n)$ and $a_i = N_{n,i}$) the right hand side of the inequality converges to zero (in probability) under our conditions. Therefore the lemma shows that the study of generalized bootstrap can be reduced to the study of wild bootstrap (or equivalently: of i.i.d. resampling). Consistency of wild bootstrap follows from the result of the last section and the theorem is proved. Again, this shows the strong relation between generalized and wild bootstrap.

4. Heavy tailed distributions. Generalized bootstrap works under weaker conditions than bootstrap. This follows from the following result of Arcones and Giné (1992). Remember that subsampling is an example of generalized bootstrap (see example 1).

<u>THEOREM</u> (Arcones and Giné, 1992). *For an i.i.d. sample* $(X_1, \dots, X_n)$ *with (now fixed) law in the domain of attraction of a* p*-stable law* $(1 < p \le 2)$ *the subsampling estimate* $\mathcal{L}^*\left(\left[\sum_{i=1}^{m(n)} (X_i^* - \overline{X}_n^*)^2 \right]^{-1/2} \sum_{i=1}^{m(n)} (X_i^* - \overline{X}_n)^2 \right)$ *has in probability the same weak limit as*

$\left[\sum_{i=1}^{n} (X_i - \overline{X}_n)^2\right]^{-1/2} \sum_{i=1}^{n} X_i - EX_i$ *if* $m(n) \to \infty$ *and* $m(n)/n \to 0$. *Here* $\overline{X}_n =$ $n^{-1} \sum_{i=1}^{n} X_i$ *and* $\overline{X}_n^* = m(n)^{-1} \sum_{i=1}^{m(n)} X_i^*$.

Let us shortly give some heuristics to indicate why this result holds and what is needed here for the accuracy of bootstrap for finite n. Suppose $1 < p < 2$ and without loss of generality $EX_i = 0$. We use the following two facts:

(1) $P(X_i > x) = c_1 x^{-p} + o(x^{-p})$,

$P(X_i < -x) = c_2 x^{-p} + o(x^{-p})$ for $x \to \infty$ with $c_1 + c_2 > 0$.

(2) Order $X_1, \dots, X_n$ by absolute value, i.e. $|X_{(1)}| \geq \dots \geq |X_{(n)}|$. From the results in Hall (1978) it follows

$$\lim_{r\to\infty} \overline{\lim_{n\to\infty}} P\left(\left| \frac{\sum_{i=1}^{r} X_{(i)} - d_r}{\left[\sum_{i=1}^{r} X_{(i)}^2\right]^{1/2}} - \frac{\sum_{i=1}^{n} X_i}{\left[\sum_{i=1}^{n} (X_i - \overline{X}_n)^2\right]^{1/2}} \right| > \varepsilon \right) = 0,$$

where $d_r = \frac{c_1 - c_2}{c_1 + c_2} \frac{p}{1-p} r^{1-1/p}$.

(1) shows that $|X_{(i)}|$ is of order $n^{1/p}$ for every fixed i. Because of (2), therefore the distribution of $\sum_{i=1}^{n} X_i \cdot \left[\sum_{i=1}^{n} (X_i - \overline{X}_n)^2\right]^{-1/2}$ is approximately determined by the behavior of the distribution function of X_i in the intervals $[-\beta n^{1/p}, -\alpha n^{1/p}]$ and $[\alpha n^{1/p}, \beta n^{1/p}]$ with α small enough and β large enough. A similar discussion applies in the bootstrap world. However now the behavior of the distribution function of X_i^* in the intervals $[-\beta m(n)^{1/p}, -\alpha m(n)^{1/p}]$ and $[\alpha m(n)^{1/p}, \beta m(n)^{1/p}]$ is determing. That implies: what is implicitely done by the subsampling here is an extrapolation. First the distribution function of X_i is estimated in the intervals $\pm [\alpha m(n)^{1/p}, \beta m(n)^{1/p}]$. Then it is assumed that (1) holds. This implies $P(X_i > x) / P(X_i > y) = (x/y)^{-p} (1 + o(1))$ and an analogue expansion for the lower tails. These formulas are now implicitely used for extrapolation from $m(n)^{1/p}$ to $n^{1/p}$.

The surprising fact here is that knowledge of p is not necessary for subsampling. It is implicitely estimated by the resampling. Besides the choice of m(n) everything is automatically done by the bootstrap. The optimal choice of m(n) depends on the accuracy of the expansion in (1). Preliminary calculations suggest that subsampling (even with optimal choice of m(n)) cannot achieve the rate of accuracy of the limiting stable law. This is the price for not knowing c_1,c_2 and p. At the end of this section let us mention that besides subsampling there exist also other versions of generalized and wild bootstrap which will work here. Other examples where wild bootstrap works and where classical bootstrap breaks down or is inaccurate are given in Härdle and Mammen (1993) and Mammen (1992a).

5. Higher order performance of generalized bootstrap. Generalized bootstrap offers a variety of possibilities for adjusting the higher order performance of the usual bootstrap. For the root $R = R(\hat{P}_n,P) = \int x\, d(\hat{P}_n - P)$ and for its studentized version $R^{STUD} = R^{STUD}(\hat{P}_n,P) = R(\hat{P}_n,P)\ [\int (x - \int x d\hat{P}_n)^2\, d\hat{P}_n\,]^{-1/2}$ we consider the generalized bootstrap estimates $F^*_{GB}(x)$ and $F^{STUD,*}_{GB}(x)$ of $P(R \le x)$ or $P(R^{STUD} \le x)$, resp.. Let $F^*_B(x)$ and $F^{STUD,*}_B(x)$ denote the usual bootstrap estimates. Then under conditions on the moments and mixed moments of the weights $N_1, \dots, N_n$ Hall and Mammen (1992) give asymptotic formulas for the first four cumulants of the distribution functions F^*_{GB} and $F^{STUD,*}_{GB}$. These formulas suggest (under additional conditions on higher order cumulants) that

$$F^*_{GB}(x) - F^*_B(x) = F^{STUD,*}_{GB}(x) - F^{STUD,*}_B(x) + o_p(1/n)$$

$$= n^{-1}\, \psi(x)\, \varphi(x) + o_p(1/n).$$

Here φ is the standard normal density and ψ is the function

$$\psi(x) = -\frac{1}{2}\, x\, (\beta_1 - \beta_3) - \frac{1}{24}\, (x^3 - 3x)\, [3(2\beta_3 + \beta_4 - 2\beta_1 - 2\beta_5) + \hat{\sigma}_n^{-4}\, \hat{\mu}_{4,n}\, \beta_2],$$

$\hat{\sigma}_n^2 = \int (x - \int x d\hat{P}_n)^2\, d\hat{P}_n$ and $\hat{\mu}_{4,n} = \int (x - \int x d\hat{P}_n)^4\, d\hat{P}_n$ are the empirical variance and the empirical centered fourth moment. The β_i's are defined as

$$\beta_1 = n[E^*(N_1 - 1)^2 - 1] + 1,$$

$$\beta_2 = E^*(N_1 - 1)^4 - 4,$$

$$\beta_3 = n\, E^*(N_1 - 1)(N_2 - 1) + 1,$$

$$\beta_4 = n[E^*(N_1 - 1)^{\,2}\ (N_2 - 1)^2 - [\ E^*(N_1 - 1)^2\ ]^2\ ] + 1,$$

$$\beta_5 = n\ E^*(N_1 - 1)^2\,(N_2 - 1)\,(N_3 - 1) + 1.$$

It is assumed that $E^*N_1 = 1$ and that the β_i's are bounded. This means that the mixed moments appearing in the definition of the β_i's are not too different from the i.i.d. resampling (from the wild bootstrap). The β_i's are chosen such that $\beta_1 = \ldots = \beta_5 = 0$ for the usual bootstrap. We allow that the conditional distribution of the weights may be adaptively chosen depending on the sample. Therefore also the β_i's may depend on the sample. For instance, for the generalized jackknife with $\sum_{i=1}^{n} \rho(i) = n$ we get $\beta_3 = \beta_5 = 0$ and $\beta_4 = -(\beta_2 + 2)$ with β_2 depending on ρ. For a detailed discussion of this expansion we refer to Hall and Mammen (1992). As a possible application let us mention here only that generalized bootstrap may be used to adjust coverage probabilities of bootstrap confidence intervalls.

References.

Arcones, M. A. and Giné, E. (1991). Additions and Correction to "the bootstrap of the mean with arbitrary sample size". *Ann. Inst. Henri Poincaré* **27** 583 - 595.

Beran, R. (1984). Bootstrap methods in statistics. *Jahresber. Math. Ver.* **86** 14 - 30.

Beran, R. (1986). Discussion to Wu, C.F.J.: Jackknife, bootstrap and other resampling methods in regression analysis. *Ann. Statist.* **14** 1295-1298.

Beran, R. and Ducharme, G. R. (1991). Asymptotic Theory for Bootstrap Methods in Statistics. Publication CRM, Universite de Montreal.

Diciccio, T. J. and Romano, J. P. (1988). A review of bootstrap confidence intervals (with discussion). *J. R. Statist. B* **50** 338-354.

Efron, B. (1979). Bootstrap methods: Another look at the Jackknife. *Ann. of Statist.* **7** 1-26.

Efron, B. (1982). The jackknife, the bootstrap and other resampling plans. *CBMS-NSF Regional Conference Series in Applied Mathem. Monogr.* **38**. *SIAM*, Philadelphia.

Efron, B. and Gong, G. (1983). A leisurly look at the bootstrap, the jackknife and crossvalidation. *Amer. Statistician* **37** 36 - 48.

Efron, B. and Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statist. Science* **1** 54 - 77.

Hájek, J. (1961). Some extensions of the Wold-Wolfowitz-Noether theorem. *Ann. Math. Statist.* **32** 506 - 523.

Hall, P. (1978). On the extreme terms of a sample from the domain of attraction of a stable law. *J. London Math. Soc.* **(2) 18** 181 - 191.

Hall, P. (1992). The bootstrap and Edgeworth expansions. Springer, New York.

Hall, P. and Mammen, E. (1992). On general resampling algorithms and their performance in distribution estimation. Preprint.

Härdle, W. and Mammen, E. (1993). Comparing non parametric versus parametric regression fits. *Ann. Statist., to appear*.

Häusler, E., Mason, D. M. and Newton, M. A. (1991). Weighted bootstrapping of means. *CWI Quarterly* **4**, no. 3 213 - 228.

Helmers, R. (1991). Bootstrap methods. *Unpublished book manuscript*.

Husková, M. and Janssen, P. (1992a). Consistency of the generalized bootstrap for degenerate U'-statistics. Preprint.

Husková, M. and Janssen, P. (1992b). Generalized bootstrap for studentized U-statistics. A rank statistic approach. Preprint.

Liu, R. (1988). Bootstrap procedure under some non i.i.d. models. *Ann. Stat.* **16** 1696 - 1708.

Liu, R. and Singh, K. (1992). Efficiency and robustness in resampling. *Ann. Statist.* **20** 370 - 384.

Liu, R. and Singh, K. (1991). Using i.i.d. bootstrap for general non i.i.d. models. *Unpublished manuscript*.

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

Mammen, E. (1992a). When does bootstrap work: asymptotic results and simulations. *Lecture Notes in Statistics* **77**. Springer New York Heidelberg Berlin.

Mammen, E. (1992b). Bootstrap, wild bootstrap, and asymptotic normality. *Prob. Theory and Rel. Fields* **93** 439 - 455.

Mason, D. M. and Newton, M. A. (1992). A rank statistic approach to the consistency of a general bootstrap. *Ann. Statist.* **20** 1611 - 1624.

Praestgaard, J. and Wellner, J. A. (1992). Exchangeably weighted bootstrap of the general empirical process. *Technical report* **230**, Dept. of Statistics, University of Washington.

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

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

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

Zheng, Z. and Tu, D. (1988). Random weighting method in regression models. *Scientia Sinica Series A* **XXXI** 1442 – 1459.