Probab. Theory Relat. Fields 93, 439–455 (1992)

Probability Theory and Related Fields

Bootstrap, wild bootstrap, and asymptotic normality *

Enno Mammen

Institut für Angewandte Mathematik, Im Neuenheimer Feld 294, W-6900 Heidelberg 1, Federal Republic of Germany

Received February 26, 1990; in revised form February 5, 1992

Summary. We show for an i.i.d. sample that bootstrap estimates consistently the distribution of a linear statistic if and only if the normal approximation with estimated variance works. An asymptotic approach is used where everything may depend on n. The result is extended to the case of independent, but not necessarily identically distributed random variables. Furthermore it is shown that wild bootstrap works under the same conditions as bootstrap.

Mathematics Subject Classification (1991): 62 G 09, 62 G 05, 62 E 20

1 Introduction

In this paper consistency of bootstrap is compared with asymptotic normality. This is done for linear statistics of n i.i.d. observations. It is shown that bootstrap works asymptotically under the same assumptions as a normal approximation with estimated variance (Theorem 1). This result is extended to the case of independent but not necessarily identically distributed observations (Theorem 2). Moreover, bootstrap with a Poisson random sample size is considered. This bootstrap procedure is a special case of a class of resampling plans called wild bootstrap which have been proposed for the non i.i.d. case. We show that bootstrap works as long as the same holds for wild bootstrap (Theorems 3 and 4).

2 Results

Let us first consider samples $\mathbf{X}_n=(X_{n,1},\dots,X_{n,n})$ of n i.i.d. variables with unknown distribution P_n. We study the following bootstrap procedure for estimating $P(T_n(\hat{P}_n)-T_n(P_n)\leq t)$, where $\hat{P}_n$ is the empirical distribution based on the sample. The bootstrap estimate is $P^*(T_n(\hat{P}_n^*)-T_n(\hat{P}_n)\leq t)$. Here P^* denotes the condi-

* This work has been supported by the Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 123 "Stochastische Mathematische Modelle"

tional law $P(\dots|\mathbf{X_n})$. $\hat{P}_n^*$ is the empirical distribution of a bootstrap sample $\mathbf{X_n^*}=(X_{n,1}^*, \dots, X_{n,n}^*)$; that is, given X_n, the bootstrap sample $\mathbf{X_n^*}$ is an i.i.d. sample with distribution $\hat{P}_n$. For fixed $P_n=P$ and $T_n=T$ an elegant proof of the consistency of bootstrap can be found in Gill (1989) for asymptotically linear functionals. In Mammen (1992 a, b) a simple treatment of higher-order properties is given for smooth functionals.

In this paper we consider the case of a linear functional $T_n(Q)=\int g_n(x)\, Q(dx)$. Put $\hat{T}_n=T_n(\hat{P}_n)$. For fixed $T_n=T$ and $P_n=P$ it is known that bootstrap is (weakly) consistent if and only if $g(X_{n,1})$ belongs to the domain of attraction of the normal law (see Hall (1990); Giné and Zinn (1989); Csörgö and Mason (1989); Athreya (1987); and Knight (1989)). In our approach we allow everything to depend on n. There are two reasons for doing this: first, as a first step into the direction of studying an arbitrary sequence of statistics $\hat{T}_n$; second, because it makes it easier to judge if the asymptotic results apply for a finite sample size n. We will show in the following theorem that asymptotic normality is necessary and sufficient for consistency of bootstrap. The proof will be based on the following simple argument. If bootstrap works, the bootstrap estimate of the distribution of the standardized functional must be asymptotically equivalent to a sequence of *non-random* distributions. We will show that for this asymptotic equivalence it is necessary and sufficient that the absolute maximal summand $\max_{1\leq i\leq n} |g_n(X_{n,i})|$ is of smaller order than the sum $\hat{T}_n$. This Lindeberg-type condition implies the asymptotic normality of $\hat{T}_n$. Note that we do not have to treat the two cases $E g_n^2(X_{n,i})<\infty$ and $E g_n^2(X_{n,i})=\infty$ separately.

Theorem 1 *Consider a sequence $X_{n,1}, \dots, X_{n,n}$ of i.i.d. variables with distribution P_n. For a function g_n consider $\hat{T}_n=T_n(\hat{P}_n)=\frac{1}{n}\sum_{i=1}^n g_n(X_{n,i})$. Consider a bootstrap sample $X_{n,1}^*, \dots, X_{n,n}^*$ with empirical distribution $\hat{P}_n^*$. Denote $\hat{T}_n^*=T_n(\hat{P}_n^*)$. Then for every sequence t_n the following assertions are equivalent:*

(i) *There exist σ_n with*

$$d_\infty(\mathscr{L}(\hat{T}_n-t_n), N(0,\sigma_n^2))\to 0.$$

(ii) $d_\infty(\mathscr{L}(\hat{T}_n-t_n), N(0,S_n^2))\to 0$ *(in probability), where* $S_n^2=\frac{1}{n^2}\sum_{i=1}^n (g_n(X_{n,i})-\hat{T}_n)^2$.

(iii) $d_\infty(\mathscr{L}(\hat{T}_n-t_n), \mathscr{L}^*(\hat{T}_n^*-\hat{T}_n))\to 0$ *(in probability).*

Here d_∞ denotes the Kolmogorov distance and $\mathscr{L}^*(\dots)$ is the conditional law $\mathscr{L}(\dots|X_{n,1}, \dots, X_{n,n})$. If (i), (ii) *or* (iii) *hold then t_n can be chosen as $\mu+E(g_n(X_{n,i})-\mu)\, 1(|g_n(X_{n,i})-\mu|\leq n\sigma_n)$ where μ is a median of the distribution of $g_n(X_{n,1})$.*

Remark 1 Under the additonal assumption of

$$\sigma_n^{-1}\, E|g_n(X_{n,i})-t_n|\, 1(|g_n(X_{n,i})-t_n|>n\sigma_n)\to 0,$$

the sequence t_n can be replaced by $E(\hat{T}_n)=T_n(P_n)$. Then according to the theorem bootstrap works as an estimate of $P(T_n(\hat{P}_n)-T_n(P_n)\leq t)$ if and only if the same holds for the classical approximation $N(0,S_n^2)$. It can easily be shown that, under the conditons of the theorem, bootstrap of the studentized functional also works.

Remark 2 For $g_n(x)=x$ the sequence t_n is equal to $T(P_n')$ where P_n' is the distribution of $X'_{n,i}=\mu+(X_{n,i}-\mu)\ 1(|X_{n,i}-\mu|\leqq n\sigma_n)$ where now μ is a median of the distribution of $X_{n,1}$. If (i), (ii) or (iii) hold then one can show that $P(X'_{n,i}=X_{n,i}$ for $1\leqq i\leqq n)\to 1$, i.e. asymptotically one cannot distinguish if a sample comes from P_n or P_n'.

Remark 3 The equivalence of (i) and (ii) is well known and due to Raikov (1938) (see also below). We have included condition (ii) here because it is more natural to compare the bootstrap procedure and the normal approximation with estimated variance $N(0, S_n^2)$.

Remark 4 A simple example where the conditions of the theorem are not fulfilled is given by the following triangular array:

$$g_n(X_{n,i})=\begin{cases}0 & \text{with probability } 1-\dfrac{1}{n}\\ 1 & \text{with probability } \dfrac{1}{n}.\end{cases}$$

Then $\hat{T}_n$ converges weakly to a Poisson distribution with mean 1. This does not hold in the bootstrap world, because for instance

$$P(\mathscr{L}^*(\hat{T}_n^*-\hat{T}_n)=\delta_0)\geqq P(g_n(X_{n,1})=\ldots=g_n(X_{n,n})=0)=(1-n^{-1})^n\to e^{-1}>0$$

for $n\to\infty$.

We get another example where the conditions of the theorem are not fulfilled by taking fixed $g_n=g$ and $P_n=$the distribution of $g(X_{n,1})=P$, where P belongs to the domain of attraction of a stable law (see Athreya (1987); Knight (1989)).

Remark 5 In case of nonlinear functionals T, asymptotic normality does not imply consistency of bootstrap. A simple example is the Hodges estimate (Beran (1982)). Other examples are U-statistics $\hat{T}_n=\sum_{i\neq j} W_n(X_i, X_j)$ which are pure (i.e. $E(W_n(X_i, X_j)\,|X_i)=E(W_n(X_i, X_j)|\,X_j)=0$ for $i\neq j$) and where the kernel W_n may depend on n. For simple conditions under which U-statistics are asymptotically normal see de Jong (1987). An example is $\hat{T}_n=n\sqrt{h_n}\int(\hat{f}_{h_n}-K_{h_n}*f)^2\,dx$ where $f=F'$ is the underlying density and where $\hat{f}_h$ is a kernel estimate with kernel $K_h(u)=1/h\ K(u/h)$. Note that $\hat{T}_n$ is asymptotically a pure U-statistic:

$$\hat{T}_n=h_n^{-1/2}(K_{h_n}*K_{h_n})(0)+\sum_{i\neq j} U_n(X_i, X_j)-E(U_n(X_i, X_j)\,|\,X_i)$$
$$-E(U_n(X_i, X_j)\,|\,X_j)+EU_n(X_i, X_j)+o_P(1),$$

where

$$U_n(X_i, X_j)=h_n^{1/2}\,n^{-1}(K_{h_n}*K_{h_n})\,(X_i-X_j).$$

The asymptotic normality of $\hat{T}_n$ for $h_n\to 0$ has been proved by Hall (1984). Bootstrap does not work here. Indeed, one can easily show that: $\mathrm{var}^*(\hat{T}_n^*)-3\ \mathrm{var}(\hat{T}_n)\to 0$ (in probability), i.e. the bootstrap variance estimate is not consistent. This example for inconsistency of bootstrap may seem artificial, because the knowledge of the underlying density is used in the construction of the statistic, but not in the resampling step of the bootstrap procedure. A mathematically related example which makes more sense statistically can be found in Härdle and Mammen (1990). There the L^2-distance between a nonparametric kernel regression estimator and a parametric regression estimator is proposed as a goodness-of-fit test statistic of a parametric regression model. It is shown that bootstrap does not estimate the distribution of the test statistic consistently on the hypothesis. The test statistic is asymptotically equivalent to a pure U-statistic and asymptotically normal.

Although in the bootstrap procedure an i.i.d. model is mimicked, bootstrap works also for models with independent but not necessarily identically distributed observations. This has been observed first in Liu (1988) and Liu and Singh (1991). In the next theorem we show that here again bootstrap works as long as the normal approximation with estimated variance is asymptotically consistent.

Theorem 2 *Consider a sequence* $X_{n,1}, \ldots, X_{n,n}$ *of independent random variables with distribution* $P_{n,i}$. *For a function* g_n *we define* $\hat{T}_n$ *and* $\hat{T}_n^*$ *as in Theorem* 1. *Then for every sequence* t_n *the following assertions are equivalent*:

(i) *There exists a sequence* σ_n *such that for every* $\varepsilon > 0$

$$\sup_{1 \le i \le n} P\left(\left|\frac{g_n(X_{n,i}) - t_n}{n\sigma_n}\right| \ge \varepsilon\right) \to 0 \quad \textit{(asymptotic negligibility)}, \tag{2.1}$$

$$\sum_{i=1}^{n} \left(E\left[\frac{g_n(X_{n,i}) - t_n}{n\sigma_n} 1\left(\left|\frac{g_n(X_{n,i}) - t_n}{n\sigma_n}\right| \le \varepsilon\right)\right]\right)^2 \to 0, \tag{2.2}$$

and

$$d_\infty(\mathscr{L}(\hat{T}_n - t_n), N(0, \sigma_n^2)) \to 0. \tag{2.3}$$

(ii) *The normal approximation with estimated variance works*:

$$d_\infty(\mathscr{L}(\hat{T}_n - t_n), N(0, S_n^2)) \to 0 \quad \textit{(in probability)}, \tag{2.4}$$

where again $S_n^2 = \frac{1}{n^2} \sum_{i=1}^{n} (g_n(X_{n,i}) - \hat{T}_n)^2$.

(iii) *Bootstrap works:*

$$d_\infty(\mathscr{L}(\hat{T}_n - t_n), \mathscr{L}^*(\hat{T}_n^* - \hat{T}_n)) \to 0 \quad \textit{(in probability)}. \tag{2.5}$$

For understanding why bootstrap works also for non i.i.d. observations let us introduce another resampling plan. Consider a Poisson variable N with $EN = n$ which is independent of $(X_{n,1}, \ldots, X_{n,n})$. Given $X_{n,1}, \ldots, X_{n,n}$ and $N = k$ we generate (conditionally) independent variables $X_{n,1}^*, \ldots, X_{n,k}^*$ with conditional distribution $\hat{P}_n$. We write $Y_{n,i} = n^{-1} g_n(X_{n,i})$ and $Y_{n,i}^* = n^{-1} g_n(X_{n,i}^*)$. We consider

$$\hat{T}_n^{*,P} = Y_{n,1}^* + \ldots + Y_{n,k}^*.$$

Then $\mathscr{L}^*\left(\hat{T}_n^{*,P}-\frac{N}{n}\hat{T}_n\right)$ can be used as an estimate of $\mathscr{L}(\hat{T}_n-t_n)$. We show that this estimate works under the same assumptions as bootstrap.

Theorem 3 *For a sample $X_{n,1}, \ldots, X_{n,n}$ of independent observations and a functional T_n as in Theorems* 1 *and* 2 *the following assertions are equivalent for every sequence t_n:*

(i) $d_\infty(\mathscr{L}(\hat{T}_n-t_n), \mathscr{L}^*(\hat{T}_n^*-\hat{T}_n))\to 0$ *(in probability),*

(ii) $d_\infty\left(\mathscr{L}(\hat{T}_n-t_n), \mathscr{L}^*\left(\hat{T}_n^{*,P}-\frac{N}{n}\hat{T}_n\right)\right)\to 0$ *(in probability).*

We have introduced bootstrap resampling with Poisson random sample size, because this resampling plan has the following nice interpretation. Denote

$$N_j = \#\{1\leqq i\leqq N: X^*_{n,i}=X_{n,j}\}.$$

Suppose that the $X_{n,1}, \ldots, X_{n,n}$ are pairwise different (a.s.). Then the N_j's are independent Poisson variables with $EN_j=1$ (a.s.). Furthermore

$$\hat{T}_n^{*,P}-\frac{N}{n}\hat{T}_n=\sum_{j=1}^n N_j(Y_{n,j}-\hat{T}_n/n)$$
$$=\sum_{j=1}^n (N_j-1)(Y_{n,j}-\hat{T}_n/n) \qquad \text{(a.s.)}.$$

We write this as

$$\hat{T}_n^W=\sum_{j=1}^n Y_{n,j}^W, \tag{2.6}$$

where $Y_{n,j}^W=(N_j-1)(Y_{n,j}-\hat{T}_n/n)$.

Because of $E(N_j-1)=0$ and $E(N_j-1)^2=E(N_j-1)^3=1$ one gets for the conditional expectations of $Y_{n,j}^W$

$$E^*\,Y_{n,j}^W=0, \tag{2.7}$$

$$E^*(Y_{n,j}^W)^2=(Y_{n,j}-\hat{T}_n/n)^2, \tag{2.8}$$

and

$$E^*(Y_{n,j}^W)^3=(Y_{n,j}-\hat{T}_n/n)^3. \tag{2.9}$$

Resampling plans with (2.6), …, (2.8) have been introduced in Beran (1986) and Wu (1986). They have been called in Härdle and Mammen (1990) wild bootstrap for the following reason. $\mathscr{L}^*(Y_{n,j}^W)$ could be interpreted as an estimate of $\mathscr{L}(Y_{n,j}-E\,Y_{n,j})$. This estimate is based on only *one* residual $Y_{n,j}-\hat{T}_n/\mathrm{n}$. Condition (2.9) has been introduced by Liu (1988) and Härdle and Mammen (1990) to improve the rate of convergence of the wild bootstrap estimate. We consider the following class of wild bootstrap procedures:

(2.10) *Choose a distribution Q with $E(Z|Q)=0$ and $E(Z^2|Q)=1$.*

(2.11) *Generate i.i.d. variables* $Z_1^W, \ldots, Z_n^W$ *with distribution* Q *and put*

$$\hat{T}_n^W = \sum_{j=1}^n Y_{n,j}^W,$$

where $Y_{n,j}^W = Z_j^W (Y_{n,j} - \hat{T}_n/n)$.

(2.12) *Estimate* $\mathscr{L}(\hat{T}_n - t_n)$ *by* $\mathscr{L}^*(\hat{T}_n^W)$.

The next theorem shows that this class of wild bootstrap estimates works under the same conditions as bootstrap.

Theorem 4 *Consider a wild bootstrap procedure of the form* (2.10), ..., (2.12) *with fixed* Q. *Then for a sample* $X_{n,1}, \ldots, X_{n,n}$ *of independent observations and a functional* T_n *as in Theorem* 2 *the following assertions are equivalent for every sequence* t_n:

(i) $d_\infty(\mathscr{L}(\hat{T}_n - t_n), \mathscr{L}^*(\hat{T}_n^* - \hat{T}_n)) \to 0$ *(in probability),*
(ii) $d_\infty(\mathscr{L}(\hat{T}_n - t_n), \mathscr{L}^*(\hat{T}_n^W)) \to 0$ *(in probability).*

Note that this class of resampling procedures contains as special cases bootstrap with Poisson sample size $(Q = \mathscr{L}(N_j - 1))$ and the normal approximation $N(0, S_n^2)\,(Q = N(0, 1))$.

3 Proofs

We give only the proofs of Theorems 2 and 4. Theorems 1 and 3 follow immediately from Theorems 2 or 4, respectively. A direct and simpler proof of Theorem 1 can be found in Mammen (1992 c). In the proof we make repeated use of the following version of the central limit theorem (see Gnedenko and Kolmogorov (1954) and Araujo and Giné (1980)).

Theorem (Central limit theorem) *For samples* $(Y_{n,1}, \ldots, Y_{n,n})$ *of independent observations the following statements are equivalent:*

(i) *It holds that*

$$\sum_{i=1}^n Y_{n,i} \to N(0, 1) \quad \text{(in distribution)} \tag{3.1}$$

and

$$\sup_{1 \leq i \leq n} P(|Y_{n,i}| > \varepsilon) \to 0 \quad \forall \varepsilon > 0. \tag{3.2}$$

(ii) *There exists* $\tau > 0$ *with:*

$$\sum_{i=1}^n E\, Y_{n,i,\tau} \to 0, \tag{3.3}$$

$$\sum_{i=1}^n E(Y_{n,i,\tau} - E Y_{n,i,\tau})^2 \to 1, \tag{3.4}$$

$$\sum_{i=1}^n P(|Y_{n,i}| > \varepsilon) \to 0 \quad \forall \varepsilon > 0, \tag{3.5}$$

where $Y_{n,i,\tau} = Y_{n,i}\,1(|Y_{n,i}| \leq \tau)$.

Proof of Theorem 2

Proof of "(i)⇒(ii)". This implication follows from the following classical result due to Raikov (1938) (see also Theorem 5 on p. 143 in Gnedenko and Kolmogorov (1954)):

Theorem (Raikov): *Suppose* $Y_{n,1}, \ldots, Y_{n,n}$ *are independent random variables with distribution* $P_{n,i}$. *If the* $Y_{n,i}$ *are infinitesimal, i.e.*

$$\sup_{1 \leq i \leq n} P(|Y_{n,i}| \geq \varepsilon) \to 0 \qquad \forall \varepsilon > 0,$$

then the following assertions are equivalent:

(i) *There exists a sequence* u_n *with*

$$\sum_{i=1}^{n} Y_{n,i} - u_n \to N(0,1) \qquad (\textit{in distribution}), \tag{3.6}$$

(ii) *It holds that*

$$\sum_{i=1}^{n} (Y_{n,i} - E\,Y_{n,i}\,1(|Y_{n,i}| < 1))^2 \to 1 \qquad (\textit{in probability}). \tag{3.7}$$

We use Raikov's theorem with $u_n = 0$ and $Y_{n,i} = (g_n(X_{n,i}) - t_n)/(n\sigma_n)$. Then (2.2) can be written as

$$\sum_{i=1}^{n} (E\,Y_{n,i}\,1(|Y_{n,i}| \leq \varepsilon))^2 \to 0.$$

Therefore we get from Raikov's theorem

$$\sum_{i=1}^{n} Y_{n,i}^2 \to 1 \qquad \text{(in probability).}$$

This implies

$$S_n^2/\sigma_n^2 = \sum_{i=1}^{n} \left(Y_{n,i} - n^{-1} \sum_{j=1}^{n} Y_{n,j}\right)^2$$
$$= \sum_{i=1}^{n} Y_{n,i}^2 - n^{-1} \left(\sum_{i=1}^{n} Y_{n,i}\right)^2 = 1 + o_P(1).$$

Proof of "(ii)⇒(i)". (2.3) follows trivially: Note that according to (ii) there exist (nonrandom!) measures Q_n with

$$d_\infty(Q_n, N(0, S_n^2)) \to 0 \qquad \text{(in probability).}$$

This implies that for a sequence of positive numbers σ_n one gets $(S_n - \sigma_n)/\sigma_n \to 0$ (in probability). Therefore (2.3) follows. It remains to show (2.1) and (2.2). Put again $Y_{n,i} = (g_n(X_{n,i}) - t_n)/(n\sigma_n)$.

Proof of (2.1). We have to show for every $\varepsilon>0$

(3.8) $$\sup_{1\leqq i\leqq n} P(|Y_{n,i}|\geqq\varepsilon)\to 0.$$

Suppose that (3.8) does not hold and that without loss of generality for an $\varepsilon>0$

(3.9) $$P(|Y_{n,1}|\geqq\varepsilon)\to\delta>0.$$

We know $\sum_{i=1}^{n} Y_{n,i}^2\to 1$ (in probability) because of $S_n^2/\sigma_n^2\to 1$ (in probability). This implies $P(|Y_{n,1}|\leqq 1)\to 1$. Therefore, by going to subsequences of (n), we can assume that there exist measures Q_1 and Q_2 with

(3.10) $$Y_{n,1}\to Q_1 \quad \text{(in distribution)},$$

(3.11) $$\sum_{i=2}^{n} Y_{n,i}\to Q_2 \quad \text{(in distribution)}.$$

Because of $\sum_{i=1}^{n} Y_{n,i}\to N(0,1)$ (in distribution) this implies $Q_1 * Q_2=N(0,1)$. According to a result of Levy (see, e.g., Feller 1966), p. 498) therefore Q_1 and Q_2 are Gaussian:

(3.12) $$Q_1=N(m,s^2)$$

and

(3.13) $$Q_2=N(-m,1-s^2) \quad \text{for some } m \text{ and } s.$$

We apply now again $\sum_{i=1}^{n} Y_{n,i}^2\to 1$ (in probability). An application of the necessary conditions of the law of large numbers (see Sect. 27 in Gnedenko and Kolmogorov (1954)) shows that this implies:

(3.14) $$\sum_{i=1}^{n} P(|Y_{n,i}^2-\tau_{n,i}|\geqq 1)\to 0,$$

(3.15) $$\sum_{i=1}^{n} E|Y_{n,i}^2-\tau_{n,i}|^2\,\mathbf{1}(|Y_{n,i}^2-\tau_{n,i}|<1)\to 0,$$

and

(3.16) $$\sum_{i=1}^{n} E\tilde{Y}_{n,i}^2\to 1,$$

where $\tilde{Y}_{n,i}=Y_{n,i}\,\mathbf{1}(|Y_{n,i}^2-\tau_{n,i}|<1)$ and $\tau_{n,i}\in\operatorname{med}(\mathscr{L}(Y_{n,i}^2))$. Here, for a distribution Q we write $\tau\in\operatorname{med}(Q)$ if $0.5\leqq Q(X\leqq\tau)$ and $0.5\leqq Q(X\geqq\tau)$.

Suppose without loss of generality $\tau_{n,1} \to \tau$. Then $Y_{n,1}^2 \to \tau$ (in probability). With (3.10), ..., (3.13) one gets

$$Y_{n,1} \to m \qquad \text{(in probability)} \tag{3.17}$$

and

$$\sum_{i=2}^{n} Y_{n,i} \to N(-m, 1) \qquad \text{(in distribution)}, \tag{3.18}$$

i.e. $s^2 = 0$, $m^2 = \tau$. Note that $|m| > 0$ because of (3.9). With (3.14) and (3.16) one can follow

$$\sum_{i=2}^{n} \tilde{Y}_{n,i} \to N(-m, 1) \qquad \text{(in distribution)} \tag{3.19}$$

and

$$\sum_{i=2}^{n} E\,\tilde{Y}_{n,i}^2 \to 1 - m^2. \tag{3.20}$$

But (3.19) contradicts (3.20). This can be seen from the version of the central limit theorem given above. The central limit theorem implies for a $\tau > 0$ that $\sum_{i=2}^{n} E\,\tilde{Y}_{n,i}^2 \geqq \sum_{i=2}^{n} E\,\tilde{Y}_{n,i,\tau}^2 \geqq \sum_{i=2}^{n} E(\tilde{Y}_{n,i,\tau} - E\,\tilde{Y}_{n,i,\tau})^2 \to 1$, but this contradicts (3.20). This shows (2.1).

Proof of (2.2). We have to show

$$\sum_{i=1}^{n} (E Y_{n,i,\varepsilon})^2 \to 0, \tag{3.21}$$

where $Y_{n,i,\varepsilon} = Y_{n,i}\, 1(|Y_{n,i}| \leqq \varepsilon)$. Raikov's theorem implies

$$\sum_{i=1}^{n} (Y_{n,i} - E Y_{n,i,1})^2 \to 1 \qquad \text{(in probability)}.$$

Because of $\sum_{i=1}^{n} (Y_{n,i})^2 \to 1$ (in probability) one gets

$$-2 \sum_{i=1}^{n} Y_{n,i}\, E Y_{n,i,1} + \sum_{i=1}^{n} (E Y_{n,i,1})^2 \to 0 \qquad \text{(in probability)}. \tag{3.22}$$

Condition (3.5) of the central limit theorem implies that there exists a sequence $\varepsilon_n \to 0$ with

$$\sum_{i=1}^{n} P(Y_{n,i} \neq Y_{n,i,\varepsilon_n}) \to 0 \tag{3.23}$$

and

(3.24) $$\sum_{i=1}^{n} \operatorname{var}(Y_{n,i,\varepsilon_n}) \to 1.$$

This shows

(3.25) $$\sum_{i=1}^{n} Y_{n,i}\, EY_{n,i,1} = \sum_{i=1}^{n} Y_{n,i,\varepsilon_n}\, EY_{n,i,\varepsilon_n} + o_P(1)$$
$$= \sum_{i=1}^{n} (EY_{n,i,\varepsilon_n})^2 + o_P(1) = \sum_{i=1}^{n} (EY_{n,i,1})^2 + o_P(1),$$

where we have used

$$\sum_{i=1}^{n} (EY_{n,i} - Y_{n,i,\varepsilon_n})^2 \to 0,$$

$$\operatorname{var}\left(\sum_{i=1}^{n} Y_{n,i,\varepsilon_n}\, EY_{n,i,\varepsilon_n}\right) \leqq \sum_{i=1}^{n} \operatorname{var}(Y_{n,i,\varepsilon_n})\,(EY_{n,i,\varepsilon_n})^2 = O(\varepsilon_n^2)$$
$$= o(1),$$

and

$$\sum_{i=1}^{n} (EY_{n,i,\varepsilon_n})^2 - \sum_{i=1}^{n} (EY_{n,i,1})^2 \to 0.$$

(3.4) and (3.5) show $\sum_{i=1}^{n} (EY_{n,i,1})^2 \to 0$. This implies (3.21).

Proof of "(i)⇒(iii)". Put again $Y_{n,i} = (g_n(X_{n,i}) - t_n)/(n\sigma_n)$. From the proof of "(i)⇒(ii)" we know already

(3.26) $$S_n^2 = \sigma_n^2 \sum_{i=1}^{n} \left(Y_{n,i} - \frac{1}{n} \sum_j Y_{n,j}\right)^2 = \sigma_n^2 (1 + o_P(1)).$$

For (iii) we have to show conditions (3,3), ..., (3.5) of the central limit theorem for bootstrap samples. Put

$$Y_{n,i}^* = \frac{g_n(X_{n,i}^*) - \frac{1}{n} \sum_{j=1}^{n} g_n(X_{n,j})}{n\sigma_n}.$$

Now choose $\varepsilon_n \to 0$ such that $\sum_{i=1}^{n} P(|Y_{n,i}| > \varepsilon_n) \to 0$. Then $P(P^*(|Y_{n,i}^*| \leqq 2\varepsilon_n) = 1) \to 1$.
For (3.4) note that with probability tending to 1 for every $\tau > 0$:

$$nE^*(Y_{n,i,\tau}^* - E^* Y_{n,i,\tau}^*)^2 = nE^*(Y_{n,i}^* - E^* Y_{n,i}^*)^2$$
$$= S_n^2/\sigma_n^2 = 1 + o_p(1) \qquad \text{(see (3.26))}.$$

Proof of "(iii)⇒(ii)". Without loss of generality suppose $g_n = \mathrm{id}$. In disagreement with old notation we put $Y_{n,i} = X_{n,i}/n$ and $Y^*_{n,i} = X^*_{n,i}/n$. We suppose

(3.27) $$\operatorname{med}\left(1/n \sum_{i=1}^{n} P_{n,i}\right) \ni 0,$$

(3.28) $$P(|\hat{T}_n| \leqq 1) \geqq 1/2, \quad \text{and} \quad P(|\hat{T}_n| < 1) \leqq 1/2.$$

(3.27) and (3.28) can be achieved by replacing $X_{n,i}$ by $c_n X_{n,i} + b_n$ for appropriate constants c_n and b_n. The proof of (i) is given in several steps.

Step 1. First we show

(3.29) $$\limsup_n \sum_{i=1}^{n} P(|Y_{n,i}| > \delta) < +\infty \qquad \forall \delta > 0.$$

In particular, (3.29) implies that the bootstrap observations are asymptotically negligible in the following sense: there exist $\delta_n \to 0$ with

(3.30) $$P(P^*(|Y^*_{n,j}| \geqq \delta_n) \leqq \delta_n) \to 1.$$

Proof of (3.29). The sequence $\mathscr{L}(\hat{T}_n)$ is shift-compact, i.e. there exists a sequence s_n such that $\mathscr{L}(\hat{T}_n - s_n)$ is tight. This follows from (3.28) with Theorem 1 on p. 407 and Proposition 1 on p. 420 in Le Cam (1986). Because of (3.28) also $\mathscr{L}(\hat{T}_n)$ is tight. Therefore there exists a measure Q such that for a subsequence n_i

(3.31) $$\mathscr{L}(\hat{T}_{n_i}) \to Q.$$

We apply the following concentration inequality (see Theorem 1 on p. 407 in Le Cam (1986) and Esseen (1966)).

Theorem (Concentration inequality): *For n independent random variables $Z_1, \ldots, Z_n$ put $S = Z_1 + \ldots + Z_n$. Let μ_j be a median of Z_j. Then for $\tau > 0$*

(3.32) $$\sup_x P(x \leqq S \leqq x + \tau) \leqq \sqrt{8\pi} \left[\sum_{i=1}^{n} P(|Z_i - \mu_i| \geqq \tau)\right]^{-1/2}.$$

Because of (3.28) there exists a sequence x_n with

$$c = P(x_n \leqq \hat{T}_n \leqq x_n + \delta/2) > 0.$$

Then, because of (2.5),

(3.33) $$P^*(x_n \leqq \hat{T}^*_n - \hat{T}_n + t_n \leqq x_n + \delta/2) = c + o_p(1).$$

The concentration inequality (3.32) implies with $\hat{\mu}=\operatorname{med}(Y_{n,1}, \dots, Y_{n,n})$

$$\# \{i: |Y_{n,i}-\hat{\mu}| \geqq \delta/2\} = nP^*(|Y^*_{n,i}-\hat{\mu}| \geqq \delta/2) \leqq 8\pi c^{-2} + o_p(1). \tag{3.34}$$

(3.34) and (3.27) imply $|\hat{\mu}| \leqq \delta/2$ with probability tending to one. Therefore

$$\# \{i: |Y_{n,i}| \geqq \delta\} \leqq 8\pi c^{-2} + o_p(1) \tag{3.35}$$

and, furthermore,

$$\sup_n E \# \{i: |Y_{n,i}| \geqq \delta\} = \sup_n \sum_{i=1}^{n} P(|Y_{n,i}| \geqq \delta) < +\infty. \tag{3.36}$$

This shows (3.29).

Step 2. We show now that the $Y_{n,i}$'s centered around their medians are asymptotically negligible:

$$\sup_{1 \leqq i \leqq n} P(|Y_{n,i}-\mu_{n,i}| \geqq \varepsilon) \to 0 \quad \text{for every } \varepsilon > 0, \tag{3.37}$$

where $\mu_{n,i} \in \operatorname{med}(\mathscr{L}(Y_{n,i}))$.

Proof of (3.37). Suppose, that (3.37) does not hold and that

$$P(|Y_{n,n}-\mu_{n,n}| \geqq \varepsilon) \geqq \delta \tag{3.38}$$

for some ε, $\delta > 0$ and n large enough.

We consider now bootstrap samples with Poisson random sample size. For this we generate a Poisson variable N with $EN=n$ which is independent of $(X_{n,1}, \dots, X_{n,n})$. Given $N=k$ and $(X_{n,1}, \dots, X_{n,n})$ we generate k random variables $X^*_{n,1}, \dots, X^*_{n,k}$ with distribution $\hat{P}_n$. Let $m_n = \sum_{j=1}^{n} Y_{n,j} 1(|Y_{n,j}| \leqq \delta_n)/\# \{j: |Y_{n,j}| \leqq \delta_n\}$ where $\delta_n \to 0$ is chosen such that (3.30) is fulfilled. For $Y^*_{n,j} = X^*_{n,j}/n$ we write $\hat{T}^{*,P}_n = Y^*_{n,1} + \dots + Y^*_{n,k}$. According to a result of Kolmogorov (see Proposition 5, p. 431 in Le Cam (1986) (3.30) implies that

$$d_L(\mathscr{L}^*(\hat{T}^*_n - nm_n), \mathscr{L}^*(\hat{T}^{*,P}_n - Nm_n)) \to 0 \quad \text{(in probability)}, \tag{3.39}$$

where $d_L(Q_1, Q_2)$ denotes the Levy distance, i.e. the infimum of the numbers ε such that

$$-\varepsilon + Q_1\{(-\infty, x-\varepsilon]\} \leqq Q_2\{(-\infty, x]\} \leqq Q_1\{(-\infty, x+\varepsilon]\} + \varepsilon.$$

As in Sect. 2 we consider also another representation of $\hat{T}^{*,P}_n$. Suppose for simplicity that the $X_{n,1}, \dots, X_{n,n}$ are pairwise different (a.s.). Define as in Sect. 2

$$N_j = \# \{1 \leqq i \leqq N: X_{n,i} = X^*_{n,j}\}.$$

Then $\hat{T}_n^{*,P} = \sum_{j=1}^{n} N_j \cdot Y_{n,j}$. The N_j's are independent Poisson variables with $EN_j = 1$. Consider a new random variable $X'_{n,n}$ with distribution $\mathcal{L}(X_{n,n})$ which is independent of $(X_{n,1}, \ldots, X_{n,n})$. Denote

$$\hat{S}_n^{*,P} = \sum_{j=1}^{n-1} N_j Y_{n,j} + N_n Y'_{n,n},$$

where $Y'_{n,n} = X'_{n,n}/n$. Then, because of (3.31) and (2.5),

$$d_L(\mathcal{L}^+(\hat{S}_n^{*,P} - N m_n), \mathcal{L}^*(\hat{T}_n^{*,P} - N m_n)) \to 0 \qquad \text{(in probability)},$$

where $\mathcal{L}^+$ denotes the conditional distribution $\mathcal{L}(|X_{n,1}, \ldots, X_{n,n}, X'_{n,n})$. Using (3.31) and arguing with characteristic functions one gets:

$$Y_{n,n} - Y'_{n,n} \to 0 \qquad \text{(in probability)}.$$

This contradicts (3.38). Therefore (3.37) is proved.

Step 3. We strengthen now the result of Step 2 to

$$\lim_n \sum_{i=1}^{n} P(|Y_{n,i} - \mu_{n,i}| > \varepsilon) = 0 \qquad \forall \varepsilon > 0. \tag{3.40}$$

Proof of (3.40). Suppose that (3.40) does not hold. Then for a $\delta > 0$ and for a subsequence n_i (for simplicity suppose $(n_i) = (n)$)

$$\lim_n \sum_{i=1}^{n} P(|Y_{n,i} - \mu_{n,i}| > \delta) = C > 0. \tag{3.41}$$

Recall from Step 1 that $C < +\infty$. (3.37) implies that for n large enough

$$P\{|Y_{n,i} - \mu_{n,i}| > \delta\} \leqq P\{|Y_{n,i}| > \delta/2\} \qquad \text{for } 1 \leqq i \leqq n. \tag{3.42}$$

Therefore one gets

$$\begin{aligned} 0 &< \liminf_{n \to \infty} \sum_{i=1}^{n} P(|Y_{n,i}| > \delta/2) \\ &\leqq \limsup_{n \to \infty} \sum_{i=1}^{n} P(|Y_{n,i}| > \delta/2) < +\infty. \end{aligned} \tag{3.43}$$

For $k \geqq 0$ consider the event

$$A_{k,n} = \{\#\{i : |Y_{n,i}| > \delta/2\} = k, |\text{med}(Y_{n,1}, \ldots, Y_{n,n})| \leqq \delta/4\}.$$

We show now that for every $k_0 > 0$

$$\liminf_{n \to \infty} P(B_{k_0,n}) > 0, \tag{3.44}$$

where $B_{k_0,n} = \bigcup_{k \geqq k_0} A_{k,n}$.

For (3.44) note first that $M_n = \#\{i: |Y_{n,i} - \mu_{n,i}| > \delta\}$ is asymptotically Poisson distributed with parameter C: $|P(M_n = k) - \exp(-C)\, C^k/k| \to 0$. Because of (3.42) for n large enough the random variable $\tilde{M}_n = \#\{i: |Y_{n,i}| > \delta/2\}$ is stochastically larger than M_n. This implies

$$\liminf_{n\to\infty} P(\tilde{M}_n \geqq k_0) > 0.$$

Furthermore, $P(|\mathrm{med}(Y_{n,1}, \ldots, Y_{n,n})| \leqq \delta/4) \geqq P(\#\{i: |Y_{n,i}| > \delta/4\} < n/2) \geqq 1 - E\left(2n^{-1} \sum_{i=1}^{n} I(|Y_{n,i}| > \delta/4)\right) \geqq 1 - \sup_{1 \leqq i \leqq n} P(|Y_{n,i}| > \delta/4) \to 1$, because of (3.29). So for every k_0 the intersection $B_{k_0,n}$ of the two events $(\tilde{M}_n \geqq k_0)$ and $(|\mathrm{med}(Y_{n,1}, \ldots, Y_{n,n})| \leqq \delta/4)$ has eventually probability larger than a positive constant.

Consider now $\hat{T}_n^* = Y_{n,1}^* + \ldots + Y_{n,n}^*$. Then for every $k_0 > 0$ on the event $B_{k_0,n}$ one gets with the concentration inequality (3.32) for all x

$$\begin{aligned} & P^*(|\hat{T}_n^* - x| \leqq \delta/8)^2 \\ & \leqq 8\pi [n P^*(|Y_{n,1}^* - \mathrm{med}(Y_{n,1}, \ldots, Y_{n,n})| > \delta/4)]^{-1} \\ & \leqq 8\pi [n P^*(|Y_{n,1}^*| > \delta/2)]^{-1} \\ & \leqq 8\pi k_0^{-1}. \end{aligned}$$

This implies for every $k_0 > 0$ on $B_{k_0,n}$

$$P^*(|\hat{T}_n^* - \hat{T}_n + t_n| \leqq 1) \leqq (8/\delta + 1)\,(8\pi k_0^{-1})^{1/2}.$$

Because of (2.5) and (3.44) this shows

$$P(|\hat{T}_n| \leqq 1) \to 0.$$

But this would contradict (3.28). This shows (3.40).

Step 4. We show now that $\hat{T}_n$ is asymptotically normal: There exists a sequence σ_n, bounded away from 0 and $+\infty$, and a sequence u_n such that

(3.45) $$d_\infty(\mathscr{L}(\hat{T}_n), N(u_n, \sigma_n^2)) \to 0.$$

Note that then (3.28) implies that also u_n is bounded away from 0 and $+\infty$.

Proof of (3.45). Set $Z_{n,i} = Y_{n,i} - \mu_{n,i}$. For $\tau > 0$ we put $\lambda_{n,\tau} = E Z_{n,i,\tau}$, $\rho_{n,\tau}^2 = E(Z_{n,i,\tau} - \lambda_{n,\tau})^2$ and $V_{n,i}^\tau = (Z_{n,i,\tau} - \mu_{n,\tau})/(n\rho_{n,\tau})$, where $Z_{n,i,\tau} = Z_{n,i} 1(|Z_{n,i}| \leqq \tau)$. For the asymptotic normality of $\hat{T}_n$ it suffices to show that $V_{n,i} = V_{n,i}^1$ fulfill the conditions (3.3), …, (3.5) of the central limit theorem. We show first:

(3.46) $(n\rho_{n,\tau})$ is bounded away from 0 for every $\tau > 0$.

Proof of (3.46). Suppose that (3.46) does not hold and assume for simplicity that $n\rho_{n,\tau} \to 0$ for a $\tau > 0$. This would imply $\mathrm{var}\left(\sum_{i=1}^{n} Z_{n,i,\tau}\right) = n\rho_{n,\tau}^2 \leqq 2n\rho_{n,\tau}\tau \to 0$.

Therefore $\sum_{i=1}^{n} Z_{n,i,\tau} - n\lambda_{n,\tau} \to 0$ (in probability). Because of $P(Z_{n,i,\tau} = Z_{n,i}$ for $i=1, \ldots, n) \to 1$ (see (3.40)) this implies $\hat{T}_n - n\lambda_{n,\tau} - \sum_{i=1}^{n} \mu_{n,i} = \sum_{i=1}^{n} Z_{n,i} - n\lambda_{n,\tau} \to 0$ (in probability). But this would contradict (3.28).

(3.46) implies that $|V_{n,i}|$ is uniformly bounded. Therefore (3.3) and (3.4) hold for $V_{n,i}$ for τ large enough. It remains to show (3.5). But this follows from (3.40) and

$$\lambda_{n,1} \to 0. \tag{3.47}$$

Proof of (3.47). Choose ε with $0 < \varepsilon < 1$. Then $|\lambda_{n,1} - \lambda_{n,\varepsilon}| \leqq P(|Z_{n,i}| \geqq \varepsilon) \to 0$ because of (3.40). Therefore $|\lambda_{n,1}| = \varepsilon + o(1)$ because of $|\lambda_{n,\varepsilon}| \leqq \varepsilon$.

That σ_n is bounded away from $+\infty$ follows from (3.28).

Step 5. We show now that the maximal summand $\max_{1 \leqq i \leqq n} |Y_{n,i}|$ is of lower order than the sum $\hat{T}_n$:

$$\sum_{i=1}^{n} P(|Y_{n,i}| \geqq \varepsilon) \to 0 \tag{3.48}$$

for every $\varepsilon > 0$.

Proof of (3.48). We know $d_L(\mathscr{L}^*(\hat{T}_n^* - n m_n), \mathscr{L}^*(\hat{T}_n^{*,P} - N m_n)) \to 0$ (in probability). This implies that

$$d_L(\mathscr{L}^*(\hat{T}_n^{*,P} - (N-n)\, m_n - \hat{T}_n + t_n - u_n), \quad N(0, \sigma_n^2)) \to 0 \quad \text{(in probability).} \tag{3.49}$$

Remember that $\hat{T}_n^{*,P} - N m_n = \sum_{j=1}^{n} N_j \cdot (Y_{n,j} - m_n)$ is a sum of (conditionally, given $(X_{n,1}, \ldots, X_{n,n})$) independent variables. Using Levy's result that the convolution of two distributions Q_1 and Q_2 can only be normal if Q_1 and Q_2 are normal one gets

$$\sup_j |Y_{n,j} - m_n| \to 0 \quad \text{(in probability).}$$

Because of $m_n \to 0$ (in probability) this shows

$$\sup_{1 \leqq i \leqq n} |\mu_{n,i} - u_n/n| \to 0. \tag{3.50}$$

(3.50) and (3.40) imply (3.48).

Step 6. We show now

$$d_\infty(\mathscr{L}^*(\hat{T}_n^* - \hat{T}_n), N(0, S_n^2)) \to 0 \quad \text{(in probability).} \tag{3.51}$$

But this follows easily from the result of the last step by an application of the central limit theorem. (3.51) and (2.5) imply (ii).

Proof of Theorem 4. The proof can be carried out with the same approach as in the proof of Theorem 2: For the proof of "(i)⇒(ii)" one remarks first that consistency of bootstrap implies assertion (i) of Theorem 2 (asymptotic normality of $\hat{T}_n$) and one proceeds then similarly as in the proof of the inclusion "(i)⇒(iii)" of Theorem 2. The proof of "(ii)⇒(i)" can be done using the same steps as in the proof of the inclusion "(iii)⇒(ii)" of Theorem 2. Let us shortly mention two modifications.

In Step 1 choose $\mu \in \mathrm{med}(Q)$ and δ with

$$C(\delta)=Q(|Z-\mu|\geqq(\delta/2)^{1/2})>0.$$

Then one uses instead of (3.34)

$$\begin{aligned}&\#\{i:|Y_{n,i}-\hat{T}_n/n|\geqq(\delta/2)^{1/2}\}\\&\leqq C(\delta)^{-1}\sum_{i=1}^{n}P^*(|Y_{n,i}-\hat{T}_n/n|\,|Z_i^W-\mu|\geqq\delta/2)\\&=C(\delta)^{-1}\sum_{i=1}^{n}P^*(|Y_{n,i}^W-\mathrm{med}(\mathscr{L}^*(Y_{n,i}^W))|\geqq\delta/2)\\&\leqq 8\pi c^{-2}C(\delta)^{-1}+o_p(1).\end{aligned}$$

In Step 3 one does not need the Poissonisation because one can immediately apply the argument given at the end of this step.

References

Araujo, A., Giné, E.: The central limit theorem for real and banach valued random variables. New York: Wiley 1980

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

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

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

Csörgö, S., Mason, D.M.: Bootstrapping empirical functions. Ann. Stat. **17**, 1447–1471 (1989)

Esseen, C.G.: On the Kolmogorov-Rogozin inequality for the concentration function. Z. Wahrscheinlichkeitstheor. Verw. Geb. **5**, 210–216 (1966)

Feller, W.: An introduction to probability theory and its applications. Vol. II. New York: Wiley 1966

Gill, R.D.: Non- and semi-parametric maximum likelihood estimators and the von Mises method (part 1). Scand. J. Stat. **16**, 97–128 (1989)

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

Gnedenko, B.V., Kolmogorov, A.N.: Limit distributions for sums of independent random variables. Reading, Mass.: Addison Wesley 1954

Hall, P.: Central limit theorems for integrated square error of multivariate nonparametric density estimators. J. Multivariate Anal. **14**, 1–16 (1984)

Hall, P.: Asymptotic properties of the bootstrap for heavy-tailed distributions. Ann. Probab. **18**, 1342–1360 (1990)

Härdle, W., Mammen, E.: Comparing non parametric versus parametric regression fits. Preprint SFB 123, Universität Heidelberg, 1990

Jong, P. de: A central limit theorem for generalized quadratic forms. Probab. Theory Relat. Fields **75**, 261–277 (1987)

Knight, K.: On the bootstrap of the sample mean in the infinite variance case. Ann. Stat. **17**, 1168–1175 (1989)
Le Cam, L.: Asymptotic methods in statistical decision theory. Berlin Heidelberg New York: Springer 1986
Liu, R.: Bootstrap procedures under some non i.i.d. models. Ann. Stat. **16**, 1696–1708 (1988)
Liu, R., Singh, K.: Using i.i.d. bootstrap for general non-i.i.d. models. (Preprint, 1991)
Mammen, E.: Bootstrap and wild bootstrap for high-dimensional linear models. Ann. Stat. (to appear 1992a)
Mammen, E.: Higher-order accuracy of bootstrap for smooth functionals. Scand. J. Stat. (to appear 1992b)
Mammen, E.: When does bootstrap work: Asymptotic results and simulations. (Lect. Notes Stat. Berlin Heidelberg New York: Springer (to appear 1992c)
Raikov, D.A.: On a connection between the central limit theorem in the theory of probability and the law of large numbers. Izv. Akad. Nauk SSR, Ser. Mat. **2**, 323–338 (1938)
Wu, C.F.J.: Jackknife, bootstrap, and other resampling methods in regression analysis (with discussion). Ann. Stat. **14**, 1261–1295 (1986)