Ann. Inst. Statist. Math.
Vol. 40, No. 3, 565-586 (1988)

BOOTSTRAPPING THE MODE

JOSEPH P. ROMANO

Department of Statistics, Stanford University, Stanford, CA 94305, U.S.A.

(Received May 25, 1987; revised February 10, 1988)

Abstract. The problem of constructing bootstrap confidence intervals for the mode of a density is considered. Estimates of the mode are derived from kernel density estimates based on fixed and data-dependent bandwidths. The asymptotic validity of bootstrap techniques to estimate the sampling distribution of the estimates is investigated. In summary, the results are negative in the sense that a straightforward application of a naive bootstrap yields invalid inferences. In particular, the bootstrap fails if resampling is done from the kernel density estimate. On the other hand, if one resamples from a smoother kernel density estimate (which is necessarily different from the one which yields the original estimate of the mode), the bootstrap is consistent. The bootstrap also fails if resampling is done from the empirical distribution, unless the choice of bandwidth is suboptimal. Similar results hold when applying bootstrap techniques to other functionals of a density.

Key words and phrases: Bootstrap confidence intervals, mode, kernel density estimates.

1. Introduction

The bootstrap, first introduced by Efron (1979), is a general, powerful technique for constructing confidence intervals by approximating the sampling distribution of a pivot. Some asymptotic theory has been developed by Bickel and Freedman (1981) and Beran (1984), among others. The asymptotic validity of the bootstrap has been established for constructing confidence intervals for a wide variety of statistical functionals $T(F)$, when $T(F)$ is, in some sense, a smooth functional of the unknown distribution F. Relatively little is known about the performance of bootstrap confidence intervals for functionals of a density. In particular, the bootstrap simulates the distribution of an approximate pivot by resampling from an estimate of the underlying population. Often, the empirical distribution is a good choice of resampling distribution. However, when the population is known to be smooth and have a density, it makes sense to simulate observations from a continuous density. Moreover, if population parameters are estimated from an estimate $\hat{f}$ of the underlying density, an obvious choice is to resample from $\hat{f}$. In this paper, we study the performance of bootstrap confidence intervals for functionals of a density for various choices of resampling distributions. In short, naive choices of bootstrap procedures, such as resampling from the empirical distribution or certain density estimates, result in invalid inferences. This paper primarily focuses on a particular functional, the mode, though some general remarks are given in Subsection 2.4 in Discussion. A mode of a probability density $f(t)$ is a value θ which maximizes f.

We will consider estimates of the mode of a density via a kernel density estimate. That is, given a kernel K(a probability density on the real line), a bandwidth h_n, and a sample $X_1,\ldots, X_n$ from a c.d.f. F on R having a density f, the kernel density estimate is given by:

$$\hat{f}_{n,h_n} = \hat{f}_{n,h_n}(t;\, X_1,\ldots, X_n) = \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{t-X_i}{h_n}\right). \tag{1.1}$$

The bandwidth h_n may be data-dependent so that, in general, h_n is a measurable function of $X_1,\ldots, X_n$.

If K is bounded, continuous, and $\lim_{t\to\pm\infty} K(t) = 0$, then so is $\hat{f}_{n,h_n}$, so there will be a point $\hat{\theta}$ such that

$$\hat{f}_n(\hat{\theta}) = \sup_t \hat{f}_{n,h_n}(t)\,. \tag{1.2}$$

Because $\hat{\theta}$ may not be uniquely defined by this equation, consider the mode functional M defined by:

$$M(f) = \inf\{m \mid f(m) = \sup_t f(t)\}\,, \tag{1.3}$$

where f is a density on R. Then, the sample mode $\hat{\theta}_{n,h_n}$ is uniquely defined by:

$$\hat{\theta}_{n,h_n} = M(\hat{f}_{n,h_n})\,. \tag{1.4}$$

For ease of argument, we will use this definition throughout the paper, but all the results continue to hold if $\hat{\theta}_{n,h_n}$ is any random variable satisfying (1.2).

In Romano (1988), the limiting behavior of kernel density estimates of the mode is obtained under minimal assumptions on the underlying density for both fixed and data-dependent bandwidths. However, to apply these results, one must explicitly estimate $f^{(2)}(\theta)$ (and even higher derivatives of f at θ, depending on the choice of bandwidth), which is quite an intricate problem. In contrast, the bootstrap approach is automatic, and the asymptotic validity of this approach is the subject of this paper. As well as being applicable quite generally, the bootstrap is known to outperform conventional asymptotic approximations in some situations (see Beran (1987) for example). An interesting application (Silverman (1981)) of the bootstrap is testing whether a density is unimodal. In this case, no asymptotic approximation is known. However, the results here suggest that the bootstrap may not be valid either. In general, some asymptotic analysis is needed to justify the use of bootstrap methods, particularly in the context of bootstrapping functionals of a density.

A brief description of the bootstrap is now given in the context of this paper. Data $X_1,\ldots,X_n$ are sampled from an unknown probability distribution F, where F is assumed to belong to a collection $\boldsymbol{F}$ of distributions. The interest lies in estimating some parameter $T(F)$, and perhaps constructing a confidence interval for $T(F)$. We are thus led to considering a pivot $R_n(X_1,\ldots,X_n;\ F)$, which is just some functional depending on both $X_1,\ldots,X_n$ and F. For example, an estimator $\hat{T}_n = \hat{T}_n(X_1,\ldots,X_n)$ of $T(F)$ might be given, in which case a natural choice of R_n might be

$$R_n(X_1,\ldots,X_n;\ F) = \delta_n[\hat{T}_n(X_1,\ldots,X_n) - T(F)]\ ,$$

where δ_n is some normalizing sequence. Let $J_n(F)$ be the law of $R_n(X_1,\ldots,X_n;\ F)$ when $X_1,\ldots,X_n$ are i.i.d. F. In order to construct a confidence interval for $T(F)$, the sampling distribution, $J_n(F)$, of $R_n(X_1,\ldots,X_n;\ F)$ must be known or estimated. The bootstrap procedure is to estimate $J_n(F)$ by $J_n(\hat{G}_n)$, where $\hat{G}_n = \hat{G}_n(X_1,\ldots,X_n)$ is some estimate of F in $\boldsymbol{F}$. The (asymptotic) validity of the bootstrap follows by first showing

$$\rho(J_n(\hat{G}_n), J_n(F)) \to 0\ , \tag{1.5}$$

in probability or almost surely (under the law F), where ρ is any metric metrizing weak convergence. If, furthermore, $J_n(F)$ converges weakly to a strictly increasing continuous limiting distribution $J(F)$, then it follows that any upper α quantile of $J_n(\hat{G}_n)$ converges in probability to the upper α quantile of $J(F)$. It then follows that confidence intervals constructed from the appropriate quantiles of $J_n(\hat{G}_n)$ will have (asymptotically) the correct coverage probability. To prove (1.5), one typically first proves that $J_n(F)$ converges weakly to $J(F)$. Then, one shows $J_n(F_n)$ converges weakly to this same limiting distribution $J(F)$ whenever F_n is a sequence of distributions belonging to a certain set of sequences $C_F = \{F_n \in \boldsymbol{F}, n \geq 1\}$. The validity of (1.1) then follows if the sequence of estimates $\hat{G}_n$ falls in this set with probability one (see Theorem 1 of Beran (1984) for example).

Next, we consider the main problem of the present work, that of constructing a confidence interval for the location of the mode. Let $\boldsymbol{F}$ be the collection of all distributions F on the line having a density f and mode $M(f)$. Given a sample $X_1,\ldots, X_n$, let $\hat{f}_{n,h_n}$ be the kernel density estimate of f defined by (1.1). Here, we consider confidence intervals for $M(f)$ based on the pivot

$$R_{n,h_n}(X_1,\ldots, X_n; F) = (nh_n^3)^{1/2}\, [M(\hat{f}_{n,h_n}) - M(f)]\,,$$

where F has density f. Let $J_{n,h_n}(F)$ be the law of $R_{n,h_n}(X_1,\ldots, X_n; F)$ when $X_1,\ldots, X_n$ are i.i.d. F. As the notation suggests, the choice of pivot depends on the choice of bandwidth h_n. For now, suppose h_n is a fixed (nonrandom) sequence. The bootstrap procedure is to approximate $J_{n,h_n}(F)$ by $J_{n,h_n}(\hat{G}_n)$, where $\hat{G}_n$ is an estimate of F in $\boldsymbol{F}$. Let C_F be the set of sequences $\{F_n\}$ such that F_n has a density f_n such that, for $j = 0$, 2 and 3, $f_n^{(j)}$ converges to $f^{(j)}$ uniformly in some neighborhood of $M(f)$. Assume $nh_n^5/\log(n) \to \infty$ and $(nh_n^7)^{1/2} \to d$ for some $d < \infty$. In Section 2 we will see that, under weak assumptions on f and the kernel K, $J_{n,h_n}(F_n)$ converges weakly to a common continuous limit law $J(F)$ whenever $\{F_n\} \in C_F$. By the results given in Romano (1988), the assumptions on h_n cover the optimal rate at which h_n should tend to zero; that is, h_n should be of order $n^{-1/7}$. Now, let $\hat{G}_{n,b_n}$ be the distribution having density $\hat{f}_{n,b_n}$. If $nb_n^7/\log(n) \to \infty$ and $b_n \to 0$, then $\{\hat{G}_{n,b_n}\}$ lies in C_F with probability one. Hence, the bootstrap will be valid in this case if we resample from $\hat{G}_{n,b_n}$. But, note that this specifically rules out the case $h_n = b_n$. Moreover, in the case $h_n = b_n$, it is shown that the bootstrap is not valid as the convergence (1.5) fails. Hence, the naive bootstrap approach will fail, but by resampling from a kernel density estimate which is smoother than the one which yields the estimates of the mode, the bootstrap will be valid.

The essential reason why the bootstrap approach fails in the case $h_n = b_n$ is that the law of the pivot under F_n must converge to a limiting distribution, uniformly for F_n close to F. Uniformity here means F_n and its first three derivatives be close to those of F. However, it is the case that the third derivative of $\hat{f}_{n,h_n}$ is not a consistent estimate of the third derivative of f (even in a pointwise sense) under the assumption $(nh_n^7)^{1/2} \to d$ if $d < \infty$. Because the limiting distribution of the sample mode depends on the third derivative of f (in the bias term), failure of the bootstrap results. Of course, the optimal choice of bandwidth h_n to estimate the mode is chosen so that the bias term is of the same order as the variance, and so is not negligible in the limit. In fact, all bootstrap approaches considered actually estimate the variance of the kernel density estimate of the mode correctly, and it is specifically the bias of the kernel density estimates that causes all the trouble. In other words, the bootstrap estimate of bias of the kernel density estimate of the mode (properly normalized) is not consistent if one resamples from the original kernel density estimate.

In Section 2, the main weak convergence results on the bootstrap are presented. Fixed and random bandwidth sequences are considered. A modified pivot is also introduced so that one can resample from the empirical distribution, under certain conditions on the choice of bandwidth. However, it is seen that the bootstrap will not succeed for optimal choices of the bandwidth in this case. Some further remarks about bootstrapping functionals of a density are also given. Some simulation results are presented in Section 3. The proofs are given in Section 4.

2. Main results

If F is a c.d.f. on $\boldsymbol{R}$ with density f and mode $\theta = M(f)$, and $h_n = h_n(X_1, \ldots, X_n)$ is specified, consider the pivot

$$R_n(X_1, \ldots, X_n; F) = [(nh_n^3)^{1/2}(\hat{\theta}_{n,h_n} - \theta)] , \tag{2.1}$$

where $\hat{f}_{n,h_n}(t) = \hat{f}_{n,h_n}(t; X_1, \ldots, X_n)$ is given by (1.1) and $\hat{\theta}_{n,h_n} = M(\hat{f}_{n,h_n})$. Let $J_{n,h_n}(F)$ be the law of $R_n(X_1, \ldots, X_n)$ when $X_1, \ldots, X_n$ are i.i.d. F. Also, set $J_{n,h_n}(x; F)$ to be the c.d.f. under F of R_n.

To construct a confidence interval for θ, we need to estimate the sampling distribution $J_{n,h_n}(F)$. The bootstrap method is to estimate $J_{n,h_n}(F)$ by $J_{n,h_n}(\hat{G}_n)$, where $\hat{G}_n$ is a suitable estimate of F. Since the empirical c.d.f. of n observations $\hat{F}_n$ does not have a density (with respect to Lebesgue measure), $J_{n,h_n}(\hat{F}_n)$ does not make sense and we are led to considering another pivot defined as follows. Set

$$T_n(X_1, \ldots, X_n; F) = [(nh_n^3)^{1/2}(\hat{\theta}_{n,h_n} - \tilde{\theta}_n)] ,$$

where

$$\tilde{f}_n(t) = \frac{1}{h_n} \int_{-\infty}^{\infty} K\left(\frac{t-y}{h_n}\right) dF(y) ,$$

and $\tilde{\theta}_n = M(\tilde{f}_n)$. Let $L_{n,h_n}(F)$ be the law of $T_n(X_1, \ldots, X_n; F)$ when $X_1, \ldots, X_n$ are i.i.d. F. Another bootstrap procedure is to approximate $J_{n,h_n}(F)$ by $L_{n,h_n}(\hat{F}_n)$, where $\hat{F}_n$ is the empirical c.d.f. of n observations from F.

Before stating the main results, we need some weak assumptions. The assumption on the underlying density of the observations is stated below as Assumption (A), while the assumption on the kernel is given in (B).

(A) Assume f has a unique mode θ such that for every $\delta > 0$, $\sup_{\{t: |t-\theta| > \delta\}} f(t) < f(\theta)$. Also, f has a continuous third derivative in some neighborhood of θ with $f^{(2)}(\theta) < 0$.

(B) Assume the kernel K is symmetric and has a continuous second derivative of bounded variation. Also assume that, for some $p > 0$, K^{2+p}

and $|K^{(1)}|^{2+p}$ are integrable. For later use, define

$$I(K)=\int [K^{(1)}(z)]^2\,dz \qquad \text{and} \qquad H(K)=\int z^2K(z)dz\,.$$

For every $\delta>0$,

$$\frac{1}{h_n^3}\int_{\{z:|z|>\delta/h_n\}}|K^{(j)}(z)|dz\to 0 \qquad \text{for} \qquad j=0,1,2\,.$$

Also assume $|K|^3$ and $z^2|K^{(2)}(z)|^2$ are integrable.

Throughout this paper, ρ will denote any metric metrizing weak convergence of distributions.

THEOREM 2.1. *(Resampling from the empirical distribution-fixed bandwidths) Let $X_1, X_2,\ldots$ be i.i.d. F. Assume F has a density f with mode θ satisfying* (A), *and K satisfies* (B). *Let $\hat{F}_n$ be the empirical c.d.f. of $(X_1,\ldots,X_n)$. If $nh_n^5/\log(n)\to\infty$ and $h_n\to 0$, then*

(i) *$\rho(L_{n,h_n}(F), L_{n,h_n}(\hat{F}_n))\to 0$ for almost all sample sequences $X_1, X_2, X_3,\ldots$.*

(ii) *Moreover, $L_n(F)$ converges weakly to the law of Z, where Z is a Gaussian random variable having mean* 0 *and variance given by*:

$$\operatorname{Var}(Z)=\frac{f(\theta)}{[f^{(2)}(\theta)]^2}\cdot I(K)\,.$$

(iii) *If $nh_n^7\to 0$, then $\rho(L_{n,h_n}(\hat{F}_n), J_{n,h_n}(F))\to 0$ with probability one; otherwise, if $\liminf_{n\to\infty} nh_n^7>0$ and $f^{(3)}(\theta)\neq 0$, then $\liminf_{n\to\infty}\rho(L_{n,h_n}(\hat{F}_n), J_{n,h_n}(F))>0$ with probability one.*

Theorem 2.1 allows us to make confidence intervals for the mode θ, by approximating the quantiles of $J_{n,h_n}(F)$ by the corresponding quantiles of $L_{n,h_n}(\hat{F}_n)$, under the assumption $nh_n^7\to 0$. Unfortunately, this rules out optimal choices of the bandwidth since they must satisfy $nh_n^7\to d$ for $d>0$. Perhaps a more natural approach in the context of the mode is to estimate $J_{n,h_n}(F)$ by $J_{n,h_n}(\hat{G}_n)$, where $\hat{G}_n$ is some estimate of F with a density. In Theorem 2.2, we study the choice where $\hat{G}_n$ is the distribution having density $\hat{f}_n$.

THEOREM 2.2. *(Resampling from the kernel density estimate-fixed bandwidths) Let $X_1, X_2,\ldots$ be i.i.d. F. Assume F has a density f with mode θ satisfying* (A). *Assume K satisfies* (B) *and K has compact support. Let F^*_{n,h_n} be the (random) c.d.f. having density $\hat{f}_{n,h_n}$ given by* (1.1) *Assume*

$nh_n^5/\log(n) \to \infty$ and $(nh_n^7)^{1/2} \to d$.

(i) *If $b_n = h_n$, then $\liminf_{n\to\infty} \rho(J_{n,h_n}(F), J_{n,h_n}(F^*_{n,b_n})) > 0$ with probability one. Indeed,*

$$EJ_{n,h_n}(x; F^*_{n,b_n}) \to P(V + Z \le x),$$

where V and Z are independent Gaussian random variables, Z has the same Gaussian distribution as given in Theorem 2.1 and V has mean $-c$ and variance σ^2, where

$$c = \frac{d}{2} \cdot \frac{f^{(3)}(\theta)}{f^{(2)}(\theta)} \cdot H(K),$$

and

$$\sigma^2 = f(\theta) \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} K^{(1)}(w) K(y - w) dw \right]^2 dy .$$

(ii) *On the other hand, suppose $\hat{G}_n$ is an estimate of F based on $X_1, \dots, X_n$ such that $\hat{G}_n$ has a density $\hat{g}_n$ and, for $j = 0, 2$ and 3, $\hat{g}_n^{(j)}$ converges to $f^{(j)}$ uniformly in some neighborhood of $M(f)$ a.s. and $M(\hat{g}_n) \to M(f)$ a.s. Then,*

$$\rho(J_{n,h_n}(F), J_{n,h_n}(\hat{G}_n)) \to 0 \qquad a.s.$$

Thus, suppose $nb_n^7/\log(n) \to \infty$ and $b_n \to 0$ as $n \to \infty$. (Here, $J_n(F)$ is the law of the pivot given in (2.1) using bandwidth h_n, not b_n.) Then,

$$\rho(J_n(F), J_n(F^*_{n,b_n})) \to 0 \qquad a.s.$$

Remark 2.1.

(1) In words, statement (i) in Theorem 2.2 says that the unconditional distribution of $J_{n,1}(F^*_{n,b_n})$ converges to a Gaussian distribution with mean $-c$ and variance $\mathrm{Var}(Z) + \sigma^2$. The ordinary pivot $J_{n,h_n}(F)$ converges weakly to the law of Z. Roughly speaking, the random distribution $J_{n,h_n}(F^*_{n,b_n})$ converges weakly to a random Gaussian distribution having the same variance as Z_1, but a random mean depending on the sequence X_1, $X_2, \dots$. Hence, the unconditional distribution of $J_{n,h_n}(F^*_{n,b_n})$ converges weakly to a mixture of Gaussian distributions, which turns out to be Gaussian as well.

(2) The assumption in Theorem 2.2 that K has compact support can probably be weakened and is used only in the proof of (i). Since the naive bootstrap approach fails when resampling from the kernel density estimate, no attempt has been made to weaken this assumption on K.

(3) Part (ii) of the previous theorem says that we can bootstrap estimates of the mode based on kernel density estimates with bandwidths satisfying $nh_n^7 \to d$ (the optimal rate), but this is true as long as we resample from a kernel density estimate with a bandwidth b_n satisfying $nb_n^7/\log(n) \to \infty$ and $b_n \to 0$. In contrast, if $b_n = h_n$ satisfies $nh_n^7 \to d$, the bootstrap will fail.

Generalizations of the previous theorems are needed, allowing for the possibility of a data-dependent bandwidth.

THEOREM 2.3. *(Resampling from a kernel density estimate-random bandwidths) Let $X_1, X_2, \ldots$ be i.i.d. observations from a distribution F with density f and mode θ satisfying Assumption* (A) *and K satisfies Assumption* (B). *Let v_n be any fixed sequence of numbers such that $nv_n^5/\log(n) \to \infty$ and $(nv_n^7)^{1/2} \to d$. Given a statistic $S_n = S_n(Y_1, \ldots, Y_n) > 0$, let $Q_n(F)$ be the law of $S_n(Y_1, \ldots, Y_n)$ when $Y_1, \ldots, Y_n$ are i.i.d. F. Let the bandwidth $h_n = h_n(X_1, \ldots, X_n)$ be defined by $h_n(X_1, \ldots, X_n) = v_n \cdot S_n(X_1, \ldots, X_n)$. Let ω_n be any fixed sequence of numbers satisfying $n\omega_n^7/\log(n) \to \infty$ and $\omega_n \to 0$ as $n \to \infty$. Let $T_n = T_n(X_1, \ldots, X_n)$ satisfy $T_n \to t$ in probability for some $t > 0$. Let F^*_{n,b_n} be the (random) distribution having density*

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

where $b_n = \omega_n \cdot T_n$. Let δ_s denote the law placing mass 1 *at s. Assume, for almost all sample sequences $X_1, X_2, \ldots$*

$$\rho(Q_n(F^*_{n,b_n}), Q_n(F)) \to 0 , \tag{2.2}$$

and furthermore, for some $s > 0$,

$$\rho(Q_n(F), \delta_s) \to 0 .$$

Then,

$$\rho(J_{n,h_n}(F), J_{n,h_n}(F^*_{n,b_n})) \to 0 \qquad a.s.$$

Moreover, $J_{n,h_n}(F)$ converges weakly to the law of $Z - c \cdot s^{7/2}$.

By choosing S_n in Theorem 2.3 to be scale equivariant, the resulting estimate $\hat{\theta}_{n,h_n}$ is scale equivariant. The uniform consistency assumption (2.2) is weak and holds for typical estimators of scale such as interquartile range or standard deviation. For optimal choice of S_n, see Romano (1988).

Before further discussion of the above results and, in particular, the choice of bandwidth, it is helpful to understand the consistency properties of kernel density estimates and their derivatives. Silverman ((1978), Theorem C) proved that, under certain restrictions on the kernel K, if the underlying density f has a uniformly continuous j-th derivative, then for fixed bandwidth sequences h_n and $j \geq 1$, it is necessary and sufficient that $h_n \to 0$ and $n^{-1} h_n^{-2j-1} \cdot \log(h_n) \to 0$ as $n \to \infty$ in order for

$$\sup_t \left| \hat{f}^{(j)}_{n,h_n}(t) - f^{(j)}(t) \right| \to 0 ,$$

in probability and almost surely. In the context of this paper, uniform consistency may be too strong a requirement. For instance, in order to consistently estimate the bias of the kernel density estimate of the mode, it is necessary to consistently estimate $f^{(3)}(\theta)$. In this case, it is not necessary that $f^{(3)}(t)$ be consistently estimated uniformly in t (as we do not assume $f^{(3)}$ even exists everywhere); however, some uniformity is needed because θ is unknown. But, even if θ were known, the point here is that the conditions on the choice of bandwidth do not change much in order to yield convergence in probability of $\hat{f}^{(3)}_{n,h_n}(\theta)$ to $f^{(3)}(\theta)$ rather than uniform convergence of $\hat{f}^{(3)}_{n,h_n}$ to $f^{(3)}$. The following proposition substantiates this claim.

PROPOSITION 2.1. *Assume the kernel K has compact support and has a bounded, integrable j-th derivative. Also, assume f is a bounded density which is j-times continuously differentiable in a neighborhood of x. Let $X_1, X_2, \dots$ be i.i.d. observations having density f, and let*

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

Then, for fixed bandwidth sequences h_n, it is necessary and sufficient that $h_n \to 0$ and $nh_n^{2j+1} \to \infty$ as $n \to \infty$ in order for

$$\hat{f}^{(j)}_{n,h_n}(x) \to f^{(j)}(x) ,$$

in probability and in the mean-squared (L^2) sense.

DISCUSSION.

2.1 *Assumptions.* See Romano (1988) for a discussion of the assumptions on the kernel K and bandwidth h_n assumed in the previous theorems.

2.2 *Bootstrapping the location of the mode.* In the case of fixed bandwidths, Theorem 2.2 gives that bootstrapping $J_{n,h_n}(F)$ is not valid when resampling from the original kernel density estimate. Indeed, the asymptotic behavior of the bootstrap distribution is random and thus, with probability one, does not converge to the weak limit of $J_{n,h_n}(F)$ as $n \to \infty$. The problem is due to the bias of the location of the sample mode as an estimate of the mode θ. However, by resampling from F^*_{n,b_n}, where $nb_n^7/\log(n) \to \infty$, the bootstrap will be valid, and confidence intervals for θ based on $J_{n,h_n}(F^*_{n,b_n})$ give asymptotically correct coverage probabilities in this case. Alternatively, the bootstrap approximation $L_{n,h_n}(\hat{F}_n)$ may be used, but then one is forced to use a suboptimal bandwidth sequence.

2.3 *Choice of bandwidth.* The choice of a bandwidth which is scale equivariant results in a scale equivariant estimator of the location of the mode. The next issue is to determine the rate at which the bandwidth tends to zero. Section 2 of Romano (1988) shows that the optimal choice of bandwidth satisfies $nh_n^7 \to d$ for some $d > 0$. However, as discussed in 2.2 above, bootstrapping is not valid when resampling from either the empirical distribution or the kernel density estimate if $(nh_n^7)^{1/2} \to d$ and $d \neq 0$. In short, the problem is that, in order to estimate the bias of the sample mode, $f^{(3)}(\theta)$ must be estimated consistently, which apparently cannot be done under the assumption $(nh_n^7)^{1/2} \to d$. Indeed, Proposition 2.1 shows $f^{(3)}(\theta)$ cannot be estimated consistently (by $\hat{f}_n^{(3)}(\theta)$) under the assumption $(nh_n^7)^{1/2} \to d$ even if θ is known. Thus, one is forced to resample from a smoother kernel density estimate $\hat{f}_{n,b_n}$ with b_n satisfying $nb_n^7/\log(n) \to \infty$. In such case, one can bootstrap estimates based on optimal choices of the bandwidth satisfying $nh_n^7 \to d$ and h_n may be data-dependent as well (by Theorem 2.3).

Alternatively, the conventional asymptotic approach (i.e., the Gaussian approximation to the distribution of the sample mode) may be used, but then one must estimate θ by one choice of bandwidth and $f^{(3)}(\theta)$ by another. Another possibility occurs by changing one's point of view. By Theorem 2.1, $\rho(L_{n,h_n}(\hat{F}_n), L_{n,h_n}(F)) \to 0$ a.s.; thus, one can construct a confidence interval not for θ but for $\tilde{\theta}_n$, where

$$\tilde{f}_n(t) = \frac{1}{h_n}\int_{-\infty}^{\infty} K\left(\frac{t-y}{h_n}\right) f(y)dy\,,$$

and $\tilde{\theta}_n = M(\tilde{f}_n)$. In short, the results are extremely delicate and suggest that applications of these results may work only when the sample size is very large. In Section 3, simulations show the methods are not highly accurate for samples of size 100, and the results are quite sensitive to the choice of bandwidth.

2.4 *Bootstrap confidence bands for the unknown density.* Jhun (1985) proposed bootstrap confidence bands for the unknown density based on the pivot

$$\sup_t \left| \hat{f}_{n,h_n}(t) - f(t) \right| .$$

Under smoothness conditions on the kernel K and the underlying density f, he showed the bootstrap approach will be valid when resampling from the empirical distribution for nonrandom bandwidth sequences h_n satisfying $nh_n \to \infty$ and $nh_n^5 \to 0$ as $n \to \infty$. The case $nh_n^5 \to q$ $(0 < q < \infty)$ was not considered. However, the optimal rate at which h_n should tend to zero is precisely this rate $nh_n^5 \to q$. Unfortunately, the same phenomenon occurs here as in the case of the mode. The reason is that the optimal rate for the choice of bandwidth occurs when the asymptotic bias is comparable to the asymptotic variance, and so is not negligible in the limit. In this case, the asymptotic bias is

$$\lim_{n\to\infty} (nh_n)^{1/2}[E\hat{f}_{n,h_n}(t) - f(t)] \to f^{(2)}(t) \cdot H(K) .$$

But, by Proposition 2.1, $f^{(2)}(\theta)$ cannot be estimated consistently by $\hat{f}^{(2)}_{n,h_n}(t)$ if $nh_n^5 \to q$. Furthermore, for fixed t, the bootstrap approximation (when resampling from the empirical distribution) to the distribution of

$$(nh_n)^{1/2}[\hat{f}_{n,h_n}(t) - f(t)] ,$$

has a mean which is identically zero. As in the case of the location of the mode, one could correct for this bias effect by consistently estimating $f^{(2)}(t)$, but this can only be done by using a kernel density estimate with a bandwidth b_n such that $nb_n^5 \to \infty$. Even in the case $nh_n^5 \to 0$, the bootstrap approach will fail when resampling from the kernel density estimate (using bandwidth h_n) just as it did in the case of the location of the mode, but it will work if one resamples from a smoother kernel density estimate $\hat{f}_{n,b_n}$ with b_n satisfying $nb_n^5/\log(n) \to \infty$.

In fact, the same phenomenon persists when bootstrapping the most simple functional of a density, $f(t)$, the density evaluated at a fixed point t. Moreover, it should now be clear that a similar phenomenon will persist when trying to bootstrap quantities involving derivatives of densities, since the bias terms will depend on even higher derivatives which cannot be estimated consistently (based on kernel density estimates). Some results about joint bootstrap confidence intervals for θ and $f(\theta)$ are given in Romano (1986). Finally, it should be clear that similar results can be obtained for bootstrapping analogous functionals of a multivariate density.

3. Simulations

In this section, some simulations results are presented to see how well various methods perform for finite samples. In all the numerical results presented, estimates of the mode are obtained by using the standard Gaussian kernel.

First, consider confidence intervals for the mode based on Parzen's Gaussian approximation (1962), with the variance $V(f)$ estimated by replacing the unknown density f in $V(f)$ by the kernel density estimate $\hat{f}_n$. Suppose the data are 100 observations from the standard Gaussian distribution. The estimate of the mode is $\hat{\theta}_{n,h_n}$, where $h_n = h \cdot S_n$, S_n is the sample standard deviation, and h is some constant. Based on 1000 simulated data sets, the estimated coverage probabilities are reported in Table 1 for various choices of confidence levels and various choices of h. If the true coverage probability of the method is p, and $\hat{p}$ is the proportion in the 1000 simulations when the true value is covered, then the estimated standard error of $\hat{p}$ is the square root of $\hat{p}(1-\hat{p})/1000$. For example, for a normal 80% confidence interval with $h = 0.4$, the estimated true coverage probability of 0.0119; the usual large sample 95% confidence interval for p is 0.831 ± 0.023. As suggested by the asymptotics, the results are quite sensitive to the choice of h, and overall are unsatisfactory. Notice that, while a choice of $h = 0.4$ results in a good estimated coverage probability for the nominal 90% level, this choice of h is not good at other levels. To see how well the approximation works for data from a skewed distribution, the same experiment was run based on 100 observations from the chi-squared distribution with 4 degrees of freedom, and the results are reported in Table 2. Again, the results show the method is sensitive to the choice of bandwidth and can be quite inaccurate. Simulations were also done for smaller sample sizes and, as expected, the Gaussian approximation is even less useful.

Next, various bootstrap confidence intervals were simulated. Table 3 reports estimated coverage probabilities for bootstrap confidence intervals

Table 1. Estimated coverage probabilities based on normal approximation 1000 simulations: bandwidth $= h$-sample standard deviation, data are $N(0, 1)$, $n = 100$.

	Coverage Level			
	80%	90%	95%	99%
$h = 0.3$	0.695	0.795	0.849	0.933
$h = 0.4$	0.831	0.899	0.941	0.976
$h = 0.5$	0.881	0.951	0.980	0.996
$h = 0.6$	0.939	0.976	0.990	0.998
$h = 0.7$	0.956	0.989	0.997	0.999
$h = 0.8$	0.962	0.991	0.996	1.000

Table 2. Estimated coverage probabilities based on normal approximation 1000 simulations: bandwidth $= h$·sample standard deviation, data are $\chi^2(4)$, $n = 100$.

	Coverage Level			
	80%	90%	95%	99%
$h = 0.25$	0.759	0.834	0.890	0.944
$h = 0.28$	0.796	0.870	0.914	0.958
$h = 0.32$	0.817	0.899	0.940	0.975
$h = 0.35$	0.804	0.889	0.938	0.981
$h = 0.39$	0.782	0.889	0.841	0.986
$h = 0.42$	0.705	0.889	0.914	0.976
$h = 0.46$	0.634	0.826	0.918	0.983
$h = 0.5$	0.531	0.766	0.880	0.979

Table 3. Estimated coverage probabilities based on bootstrap 500 simulations, 200 bootstrap replications: estimating bandwidth $= h$·sample standard deviation, resampling bandwidth $= b$·sample standard deviation, data are $N(0, 1)$, $n = 100$.

	Coverage Level			
	80%	90%	95%	99%
$h = 0.2,\ b = 0.0$	0.710	0.887	0.962	1.000
$h = 0.2,\ b = 0.2$	0.682	0.786	0.898	0.956
$h = 0.2,\ b = 0.3$	0.672	0.868	0.942	0.976
$h = 0.2,\ b = 0.4$	0.724	0.882	0.952	0.984
$h = 0.4,\ b = 0.0$	0.718	0.880	0.942	0.980
$h = 0.4,\ b = 0.4$	0.646	0.852	0.924	0.978
$h = 0.4,\ b = 0.6$	0.728	0.876	0.938	0.982
$h = 0.4,\ b = 0.8$	0.764	0.898	0.960	0.988
$h = 0.6,\ b = 0.0$	0.660	0.848	0.962	0.982
$h = 0.6,\ b = 0.6$	0.628	0.846	0.918	0.982
$h = 0.6,\ b = 0.9$	0.718	0.852	0.916	0.980
$h = 0.6,\ b = 1.2$	0.758	0.902	0.944	0.984
$h = 0.8,\ b = 0.0$	0.754	0.830	0.906	0.980
$h = 0.8,\ b = 0.8$	0.704	0.842	0.914	0.980
$h = 0.8,\ b = 1.2$	0.786	0.894	0.946	0.992
$h = 0.8,\ b = 1.6$	0.776	0.888	0.960	0.998

for simulated Gaussian samples of size $n = 100$. The mode is estimated by using a bandwidth of $h \cdot S_n$, where S_n is the sample standard deviation. For each choice of $h = 0.2$, 0.4, 0.6 and 0.8, bootstrap confidence intervals were constructed by resampling from the kernel density estimate with bandwidths $b = 0.0$, h, $1.5h$ and $2h$. The choice of $b = 0.0$ corresponds to resampling from the empirical distribution. For each choice of h and b, 200 bootstrap data sets were generated to construct the bootstrap confidence interval, and this was repeated 500 times to estimate coverage probabilities at levels 80%, 90%, 95% and 99%. In practice, more than 200 bootstrap samples might be generated, but this was not done here in order to keep the total number of computations at a reasonable level. Also, the work of Hall (1986) suggests this number is adequate if one is mainly interested in coverage probability. Table 4 reports the corresponding results for simulated samples of size 100 from the chi-squared distribution with 4 degrees of freedom. Estimated standard errors can be obtained as described above for Table 1. While the results were not highly encouraging in that estimated coverage probabilities were not very close to the nominal levels, some interesting conclusions could be made. In particular, bootstrap confidence intervals constructed by resampling from the empirical distribution are more accurate for the simulated Gaussian data than for the simulated chi-squared data. This can be explained by the fact that the bias of the kernel density estimate of the mode is not negligible, except if the underlying density is symmetric (or locally symmetric) at the mode. As suggested by the asymptotics, resampling from the kernel density estimate with $h = b$ results in estimated coverage probabilities far from the nominal levels. Finally, increasing the resampling bandwidth from $b = h$ to $b = 1.5\,h$ and $b = 2h$ improves coverage accuracy, and results in estimated coverage probabilities superior to those presented in Tables 1 and 2 based on a Gaussian approximation. Unfortunately, larger samples are perhaps needed in order to obtain highly accurate results

Table 4. Estimated coverage probabilities based on bootstrap 500 simulations, 200 bootstrap replications: estimating bandwidth = h·sample standard deviation, resampling bandwidth = b·sample standard deviation, data are $\chi^2(4)$, $n = 100$.

	Coverage Level			
	80%	90%	95%	99%
$h = 0.25$, $b = 0.00$	0.758	0.872	0.928	0.986
$h = 0.25$, $b = 0.25$	0.790	0.910	0.966	1.000
$h = 0.25$, $b = 0.37$	0.822	0.924	0.972	0.996
$h = 0.25$, $b = 0.50$	0.818	0.926	0.968	0.992
$h = 0.32$, $b = 0.00$	0.644	0.786	0.854	0.964
$h = 0.32$, $b = 0.32$	0.824	0.940	0.980	1.000
$h = 0.32$, $b = 0.48$	0.816	0.922	0.956	0.992
$h = 0.32$, $b = 0.64$	0.820	0.920	0.962	0.994
$h = 0.39$, $b = 0.00$	0.582	0.764	0.862	0.970
$h = 0.39$, $b = 0.39$	0.770	0.876	0.950	0.996
$h = 0.39$, $b = 0.59$	0.792	0.916	0.954	0.990
$h = 0.39$, $b = 0.78$	0.804	0.908	0.956	0.992
$h = 0.46$, $b = 0.00$	0.538	0.688	0.786	0.918
$h = 0.46$, $b = 0.46$	0.744	0.856	0.894	0.980
$h = 0.46$, $b = 0.69$	0.784	0.924	0.964	0.992
$h = 0.46$, $b = 0.92$	0.792	0.918	0.954	0.990

based on bootstrapping. The simulations are encouraging for even larger sample sizes, but the problem of constructing a confidence interval for the mode for smaller sample sizes still remains a challenging one. In summary, the simulations reinforce the idea that generally automatic methods like the bootstrap need mathematical and numerical justification before their use can be recommended.

4. Proofs

In order to prove Theorem 2.1, we will need a *triangular* version of the asymptotic distribution of the sample mode and the size of the sample mode. That is, we will first study the asymptotic behavior of $L_n(F_n)$ for an appropriate choice of *nonrandom* F_n. The assumptions on F_n will be satisfied with probability one when $F_n = \hat{F}_n$ is the empirical of n observations sampled from F.

PROPOSITION 4.1. *Fix F with density f and mode θ satisfying* (A), *and K satisfies* (B). *Also, assume $nh_n^5/\log(n) \to \infty$ and $h_n \to 0$. Let F_n be a sequence of distributions on $\mathbf{R}$. Set*

$$\tilde{f}_n(t) = \frac{1}{h_n}\int_{-\infty}^{\infty} K\left(\frac{t-y}{h_n}\right) dF_n(y),$$

and $\tilde{\theta}_n = M(\tilde{f}_n)$. Assume $\{F_n\}$ satisfies the following:

(a) *$\tilde{f}_n$ and $\tilde{f}_n^{(2)}$ converge to f and $f^{(2)}$, respectively, uniformly in some neighborhood of θ.*

(b) *For every $\delta > 0$, $\limsup_{n\to\infty} \sup_{\{t:|t-\theta|>\delta\}} \tilde{f}_n(t) < f(\theta)$.*

(c) $(1/h_n)\boldsymbol{E}_{F_n}(|K^{(1)}((\tilde{\theta}_n - X)/h_n)|^m) \to f(\theta)\int_{-\infty}^{\infty} |K^{(1)}(y)|^m dy$ *for* $m = 1, 2, 3$.

(d) $(n/h_n)^{1/2}\boldsymbol{E}_{F_n}(K^{(1)}((\tilde{\theta}_n - X)/h_n)) \to 0$.

Then, $L_{n,h_n}(F_n)$ converges weakly to the law of Z, where Z is a Gaussian random variable with distribution as given in Theorem 2.1.

PROOF OF PROPOSITION 4.1. First note that assumptions (a) and (b) imply $\tilde{\theta}_n \to \theta$. For purposes of the proof, construct $X_{n,1},\ldots, X_{n,n}$ which are i.i.d. F_n. Set

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

and $\hat{\theta}_n = M(\hat{f}_n)$. The assumptions imply (by an argument similar to Theorem

1.1 of Romano (1988) or see Lemma 4.1 in Romano (1986)) that $\hat{\theta}_n - \tilde{\theta}_n \to 0$ in F_n^n probability. Since K has a continuous 2nd derivative, so does $\hat{f}_n$ and $\hat{f}_n^{(1)}(\hat{\theta}_n) = 0$. By Taylor's theorem,

$$0 = \hat{f}_n^{(1)}(\hat{\theta}_n) = \hat{f}_n^{(1)}(\tilde{\theta}_n) + (\hat{\theta}_n - \tilde{\theta}_n)\hat{f}_n^{(2)}(\theta_n^*) ,$$

for some random variable θ_n^* between $\hat{\theta}_n$ and $\tilde{\theta}_n$. Hence,

$$(nh_n^3)^{1/2}(\hat{\theta}_n - \tilde{\theta}_n) = -\frac{(nh_n^3)^{1/2}\hat{f}_n^{(1)}(\hat{\theta}_n)}{\hat{f}_n^{(2)}(\theta_n^*)} \quad \text{if} \quad \hat{f}_n^{(2)}(\theta_n^*) \neq 0 .$$

The result will follow by showing:

(1) The law of $S_n = \left(-\frac{(nh_n^3)^{1/2}\hat{f}_n^{(1)}(\hat{\theta}_n)}{f^{(2)}(\theta)} \right)$ converges weakly to the law of Z.

(2) $\hat{f}_n^{(2)}(\theta_n^*) \to f^{(2)}(\theta)$ in F_n^n probability for any sequence $\theta_n^* \to \theta$.

PROOF OF (1). Note that

$$S_n = \left(\frac{(nh_n^3)^{1/2}}{f^{(2)}(\theta)} \frac{1}{n} \sum_{j=1}^{n} V_{n,j} \right) ,$$

where the $V_{n,j}$ are independent and identically distributed as

$$V_{n,1} = \frac{-1}{h_n^2} K^{(1)}\left(\frac{\tilde{\theta}_n - X_{n,1}}{h_n} \right) .$$

By assumption (c) we have for $m = 1, 2, 2 + p$,

$$h_n^{2m-1} \boldsymbol{E}_{F_n} | V_{n,1} |^m \to f(\theta) \int_{-\infty}^{\infty} |K^{(1)}(y)|^m dy .$$

By assumption (d),

$$(nh_n^3)^{1/2} \boldsymbol{E}_{F_n} V_{n,1} \to 0 .$$

Hence,

$$(nh_n^3) \operatorname{Var}_{F_n}(\hat{f}_n^{(1)}(\tilde{\theta}_n)) \to f(\theta) \int_{-\infty}^{\infty} |K^{(1)}(y)|^2 \, dy ,$$

and so

$$\frac{E_{F_n}\left|V_{n,1}-E_{F_n}(V_{n,1})\right|^{2+p}}{n^{p/2}\sigma^{2+p}(V_{n,1})}=O(nh_n)^{-1/2}=o(1)\,.$$

Hence, by Lyapounov's central limit theorem, (1) is proved.

PROOF OF (2). This follows easily from assumption (a) and Corollary 5.3 of Romano (1986).

PROOF OF THEOREM 2.1. The proof consists of two steps. First, we show $L_{n,h_n}(\hat{F}_n)$ converges weakly to the law of Z with probability one, and then we show $L_{n,h_n}(F)$ converges weakly to the law of Z. The proof of (iii) then follows from (i), (ii) and Theorem 2.1 of Romano (1986). As usual, let

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

and $\hat{\theta}_n = M(\hat{f}_n)$.

Step 1. Apply Proposition 4.1. So, we must show assumptions (a)–(d) hold with probability one if we take $F_n=\hat{F}_n$. In this case, note that $\tilde{\theta}_n=\hat{\theta}_n$. Also, $\tilde{f}_n$ is the random density $\hat{f}_n$.

(a) By Lemma 5.2 and Corollary 5.2 of Romano (1986), $\hat{f}_n$ converges to f uniformly in some neighborhood of θ with probability one. The same is true for second derivatives by Proposition 5.1 and Corollary 5.3 of Romano (1986).

(b) Apply Lemma 5.2 and Corollary 5.2 of Romano (1986).

(c) Apply Corollary 5.5 of Romano (1986).

(d) Use the identity

$$E_{F_n}K^{(1)}\left(\frac{\tilde{\theta}_n-X}{h_n}\right)=\frac{1}{nh_n^2}\sum_{i=1}^{n}K^{(1)}\left(\frac{\hat{\theta}_n-X_i}{h_n}\right)=\frac{1}{h_n}\hat{f}_n^{(1)}(\hat{\theta}_n)=0\,.$$

Step 2. Again apply Proposition 4.1, but take $F_n=F$.

(a) Apply Lemma 5.1 and Proposition 5.1 of Romano (1986).

(b) Apply Lemma 5.2 of Romano (1986).

(c) Apply Proposition 5.1 of Romano (1986).

(d) This follows because

$$E_{F_n}K^{(1)}\left(\frac{\tilde{\theta}_n-X}{h_n}\right)=\int_{-\infty}^{\infty}K^{(1)}\left(\frac{\tilde{\theta}_n-X}{h_n}\right)f(x)dx=h_n^2\tilde{f}_n^{(1)}(\tilde{\theta}_n)=0\,,$$

and, by definition, $\tilde{\theta}_n=M(\tilde{f}_n)$.

PROOF OF THEOREM 2.2. By Theorem 2.1 of Romano (1986), $J_{n,h_n}(F)$ converges weakly to $J(F)$, the law of $Z-c$ given in the statement of Theorem 2.1. Proposition 4.1 of Romano (1986) studies the limiting behavior of $J_{n,h_n}(F_n)$ for nonrandom F_n. We apply the results there taking F_n to be the random distribution F^*_{n,b_n} having density $\hat{f}_{n,b_n}$. To verify condition (a), $\hat{f}_{n,b_n}$ converges uniformly to f in some neighborhood of θ with probability one by Lemma 5.2 and Corollary 5.2 of Romano (1986). The same is true for second derivatives by Proposition 5.1 and Corollary 5.3 of Romano (1986). Condition (b) holds with probability one by Lemma 5.2 and Corollary 5.2 of Romano (1986). Now, depending on the assumptions on b_n, the verification of condition (c) shows the cause of the difference in statements (i)–(ii) of the theorem. Let

$$\mu_n(F^*_{n,b_n}) = U_n = \left(\frac{n}{h_n}\right)^{1/2} \boldsymbol{E}_{F^*_{n,b_n}}\left(K^{(1)}\left(\frac{\theta_n - X_{n,1}}{h_n}\right)\right). \tag{4.1}$$

From Proposition 4.1 of Romano (1986), the asymptotic mean of $J_{n,h_n}(F^*_{n,b_n})$ is $f^{(2)}(\theta)\cdot U_n$. To prove (ii), $U_n \to c\cdot f^{(2)}(\theta)$ a.s., by an argument similar to the proof of Theorem 2.1 (ii) of Romano (1986). Thus, $J_{n,h_n}(\hat{G}_n)$ converges weakly to $J(F)$ with probability one, and the result follows. To prove (i), apply Lemma 4.2 (ii) (proved below), to get the law of U_n converges weakly to the law of U, where U has a Gaussian distribution with mean $d/2\cdot f^{(3)}(\theta)\cdot H(K)$ and variance σ^2 as given in the statement of the theorem. If we also had $U_n \to U$ a.s. as well, the rest of the argument would be easy. Indeed, $J_n(F^*_{n,b_n})$ would converge to the law of $Z-U$, where U is random and depends on the sequence $X_1, X_2, \ldots$. Hence, the asymptotic distribution of the bootstrap distribution of the location of the mode is Gaussian, but with a random mean U, instead of $-c$, and the result would follow. To get around the fact that the law of U_n does not converge almost surely to U, apply Skorohod's Almost Sure Representation Theorem (see Romano (1986) for details).

The following lemma is needed to prove Theorem 2.2. Note that not all the hypotheses of Theorem 2.2 are needed for the proof of (i) in Lemma 4.1.

LEMMA 4.1. *Assume the hypotheses of Theorem 2.2. Define the random process*

$$Z_n(w) = (nh_n)^{1/2}[\hat{f}_n(\theta - h_n w) - \boldsymbol{E}_F \hat{f}_n(\theta - h_n w)].$$

Regard $Z_n(\cdot)$ as a random variable taking values in $\boldsymbol{C(R)}$, the metric space of bounded continuous functions on $\boldsymbol{R}$ with a metric yielding the topology *of uniform convergence on compact sets. Then,*

(i) $Z_n(\cdot)$ *converges weakly to* $Z(\cdot)$, *a mean zero Gaussian process with covariance function*

$$EZ(w_1)Z(w_2) = f(\theta)\int_{-\infty}^{\infty} K(y - w_1)K(y - w_2)dy .$$

(ii) *Let* U_n *be defined as in* (4.1) *with* $h_n = b_n$. *Then, the law of* U_n *converges weakly to the law of* U, *where* U *is a Gaussian random variable with mean* $\mu = d/2 \cdot f^{(3)}(\theta) \cdot H(K)$ *and variance*

$$\text{Var}(U) = f(\theta)\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} K^{(1)}(w)K(y - w)dw\right]^2 dy .$$

PROOF OF LEMMA 4.1.

(i) This result is essentially contained in Bickel and Rosenblatt (1973), but different conditions are assumed. A direct proof, based on Theorem 12.3 of Billingsley (1968), is given in Romano (1986).

(ii) By a change of variables,

$$U_n = (nh_n)^{1/2}\int_{-\infty}^{\infty} K^{(1)}\left(\frac{\hat{\theta}_n - \theta}{h_n} + w\right)\hat{f}_{n,h_n}(\theta - h_n \cdot w)dw$$

$$= \int_{-\infty}^{\infty} K^{(1)}\left(\frac{\hat{\theta}_n - \theta}{h_n} + w\right)[\, Z_n(w) + (nh_n)^{1/2}E_F\hat{f}_n(\theta - h_n \cdot w)\,]dw .$$

We will show:

(1) $\int_{-\infty}^{\infty} K^{(1)}((\hat{\theta}_n - \theta)/h_n + w)Z_n(w)dw$ converges weakly to the law of $U - \mu$.

(2) $(nh_n)^{1/2}\int_{-\infty}^{\infty} K^{(1)}((\hat{\theta}_n - \theta)/h_n + w)E_F\hat{f}_{n,h_n}(\theta - h_n \cdot w)dw \to \mu$.

PROOF OF (1). Since $(nh_n^3)^{1/2}(\hat{\theta}_n - \theta)$ is tight and $nh_n^5 \to 0$, it follows that $(\hat{\theta}_n - \theta)/h_n \to 0$ in probability. By (i), we know the law of Z_n converges weakly to the law of Z. By Theorem 4.4 of Billingsley (1968), it follows that

$$\left(\frac{\hat{\theta}_n - \theta}{h_n}, Z_n\right) \to (0, Z) ,$$

weakly in the product space $R \times C(R)$. The distribution of U_n clearly depends only on the distribution of $((\hat{\theta}_n - \theta)/h_n, Z_n)$. Apply Skorohod's Almost Sure Representation Theorem to conclude there exist $\hat{\theta}_n^*$, Z_n^*, and Z^* (on some probability space) so that $(\hat{\theta}_n, Z_n)$ and $(\hat{\theta}_n^*, Z_n^*)$ have the same distribution, Z^* and Z have the same distribution, and

$$\left(\frac{\hat{\theta}_n^* - \theta}{h_n}, Z_n^* \right) \to (0, Z^*) \qquad \text{almost surely}.$$

Apply dominated convergence, using the fact that, for any $M > 0$,

$$\sup_{\{w: |w| \le M\}} |Z_n^*(w) - Z^*(w)| \to 0 \qquad \text{a.s.},$$

and the assumption that K has compact support to conclude

$$\int_{-\infty}^{\infty} K^{(1)} \left(\frac{\hat{\theta}_n^* - \theta}{h_n} + w \right) Z_n^*(w) dw \to \int_{-\infty}^{\infty} K^{(1)}(w) Z^*(w) dw \qquad \text{a.s.},$$

(and hence weakly). It follows that $\int_{-\infty}^{\infty} K^{(1)}((\hat{\theta}_n - \theta)/h_n + w) Z_n(w) dw$ converges weakly to $\int_{-\infty}^{\infty} K^{(1)}(w) Z(w) dw$. All we need now is the distribution of $\int_{-\infty}^{\infty} K^{(1)}(w) Z(w) dw$. But, argue as above to conclude that $\int_{-\infty}^{\infty} K^{(1)}(w) Z_n(w) dw$ also converges weakly to $\int_{-\infty}^{\infty} K^{(1)}(w) Z(w) dw$. We now show $\int_{-\infty}^{\infty} K^{(1)}(w) Z_n(w) dw$ converges weakly to $U - \mu$. Write

$$\int_{-\infty}^{\infty} K^{(1)}(w) Z_n(w) dw = (nh_n)^{-1/2} \sum_{i=1}^{n} S_{n,i}, \tag{4.2}$$

where the $S_{n,i}$ are independent and identically distributed as

$$\int_{-\infty}^{\infty} K^{(1)}(w) K\left(\frac{\theta - h_n \cdot w - X_n}{h_n} \right) dw$$
$$- \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} K^{(1)}(w) K\left(\frac{\theta - h_n \cdot w - x}{h_n} \right) dw f(x) dx.$$

Now, $\boldsymbol{E} S_{n,i} = 0$ and

$$\operatorname{Var}[S_{n,i}] = h_n \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} K^{(1)}(w) K(y - w) dw \right]^2 f(\theta - h_n \cdot y) dy$$
$$- \left[h_n \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} K^{(1)}(w) K(y - w) f(\theta - h_n \cdot y) dy \right]^2.$$

By dominated convergence, $h_n^{-1} \operatorname{Var}[S_{n,i}] \to \operatorname{Var}(U)$. Similarly, $h_n^{-1} \boldsymbol{E}|S_{n,i}|^3$ converges. Lyapounov's C.L.T. yields (4.2) converges weakly to the law of $U - \mu$.

PROOF OF (2). As in Step 1 above, apply Skorohod's Almost Sure Representation Theorem so that $(\hat{\theta}_n^* - \theta)/h_n \to 0$ almost surely. Also, if $\tilde{f}_n(x) = E_F \hat{f}_n(x)$, then by Proposition 5.1 of Romano (1986), for $j = 0, 1, 2$ and 3, we have $\tilde{f}_n^{(j)}(x)$ converges to $f^{(j)}(x)$ uniformly in some neighborhood of θ. Using these two facts, the argument is analogous to the proof of Theorem 2.1 (ii) of Romano (1986), where it was shown that $(nh_n)^{1/2} \int_{-\infty}^{\infty} K^{(1)}(w) f(\theta - h_n \cdot w)dw \to \mu$.

PROOF OF THEOREM 2.3. A brief sketch of the proof is given, since nothing new is involved. As in Proposition 4.1, first study the asymptotic behavior of $J_{n,1}(F_n)$ for fixed (nonrandom) sequences F_n, but this time as a stochastic process indexed by the bandwidth, just as was done in Theorem 2.2 of Romano (1986). A slight generalization of the argument of Proposition 4.1 will give the limiting finite dimensional distributions of such a process. The process will be tight by essentially the same argument given in the proof of Theorem 2.2 of Romano (1986). As in the proof of Theorem 2.2, argue that the assumptions on F_n are satisfied with probability one when F_n is taken to be $\hat{F}_n$.

PROOF OF PROPOSITION 2.1. To prove sufficiency, assume $nh_n^{2j+1} \to \infty$ and $h_n \to 0$. Proposition 5.1 of Romano (1986) yields

$$E\hat{f}^{(j)}_{n,h_n}(x) \to f^{(j)}(x)\,.$$

Also,

$$\begin{aligned}\operatorname{Var}[\hat{f}^{(j)}_{n,h_n}(x)] &= \frac{1}{nh_n^{2j+2}} \operatorname{Var}\left[K^{(j)}\left(\frac{x - X_1}{h_n}\right)\right] \\ &\le \frac{1}{nh_n^{2j+2}} E\left|K^{(j)}\left(\frac{x - X_1}{h_n}\right)\right|^2 .\end{aligned}$$

By Proposition 5.1 of Romano (1986),

$$\frac{1}{h_n} E\left|K^{(j)}\left(\frac{x - X_1}{h_n}\right)\right|^2 \to f(x)\int_{-\infty}^{\infty} |K^{(j)}(z)|^2 dz\,,$$

and sufficiency follows. Conversely, application of Theorem 1, p. 316 of Feller (1971), shows the conditions $nh_n^{2j+1} \to \infty$ and $h_n \to 0$ are necessary for $\hat{f}^{(j)}_{n,h_n}(x)$ to converge to $f^{(j)}(x)$ in probability as $n \to \infty$.

REFERENCES

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

Beran, R. (1987). Prepivoting to reduce level error of confidence sets, *Biometrika* (to appear).

Bickel, P. and Freedman, D. (1981). Some asymptotic theory of the bootstrap, *Ann. Statist.*, **9**, 1196–1217.

Bickel, P. J. and Rosenblatt, M. (1973). On some global measures of the deviations of density function estimates, *Ann. Statist.*, **1**, 1071–1095.

Billingsley, P. (1968). *Convergence of Probability Measures*, Wiley, New York.

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

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications*, 2nd ed., Vol. II, Wiley, New York.

Hall, P. (1986). On the number of bootstrap simulations required to construct a confidence interval, *Ann. Statist.*, **14**, 1431–1452.

Jhun, J. (1985). Bootstrap methods for density estimates and k-means clustering, Ph.D. dissertation, Department of Statistics, University of California, Berkeley.

Parzen, E. (1962). On estimation of a probability density function and mode, *Ann. Math. Statist.*, **33**, 1065–1076.

Romano, J. (1986). Bootstrapping the joint distribution of the location and size of the mode, Ph.D. dissertation, Department of Statistics, University of California, Berkeley.

Romano, J. (1988). On weak convergence and optimality of kernel density estimates of the mode, *Ann. Statist.*, **16**, 629–647.

Silverman, B. (1978). Weak and strong uniform consistency of the kernel estimate of a density and its derivatives, *Ann. Statist.*, **6**, 177–184.

Silverman, B. (1981). Using kernel density estimates to investigate multimodality, *J. Roy. Statist. Soc. Ser. B*, **43**, 97–99.

Whitt, W. (1970). Weak convergence of probability measures on the function space $C[0, \infty)$, *Ann. Math. Statist.*, **41**, 939–944.