On the Bootstrap in Cube Root Asymptotics

Christian Léger[*] [†] Brenda MacGibbon[‡] [§]

CRM-2561

August 1998

[*]Département de mathématiques et de statistique, Université de Montréal

[†]Supported by the Natural Sciences and Engineering Research Council of Canada. The author also acknowledges the support of the Centre de recherches mathématiques of Université de Montréal.

[‡]Département de mathématiques, Université du Québec à Montréal

[§]Supported by the Natural Sciences and Engineering Research Council of Canada. The author also acknowledges the support of the GERAD research center.

Abstract

In this paper, we attempt to study the bootstrap for a class of estimators of which Chernoff's estimator of the mode (1964) is a prototype. These estimators converge at the rate of the cube root of n, a rate different from the usual one, and their limit distributions can be expressed as functionals of Brownian motion with quadratic drift. We extend Kim and Pollard's results (1990) on functional central limit theorems for such an estimator to an analogous result for triangular arrays of estimators. Our result serves to illustrate the strong conditions (which are not necessarily true in general) in order for the bootstrap to work. Under such strong conditions we construct a bootstrap in Chernoff's problem. Our theoretical results are supported by a small simulation study.

Keywords: Bootstrap, counterexample, cube root asymptotics, least median of squares, confidence interval for the mode

Résumé

Dans cet article, nous étudions le bootstrap pour une classe d'estimateurs dont l'estimateur du mode de Chernoff (1964) est un prototype. Ces estimateurs convergent à la vitesse racine cubique de n, une vitesse différente de la vitesse habituelle, et la loi limite peut être exprimée comme une fonctionnelle d'un mouvement Brownien avec une tendance quadratique. Nous généralisons les résultats de Kim et Pollard (1990) sur les théorèmes limite centrale pour de tels estimateurs à un résultat similaire pour une suite triangulaire d'estimateurs. Ce résultat permet d'illustrer certaines conditions fortes (qui ne sont pas satisfaites en général) requises pour que le bootstrap fonctionne pour de tels estimateurs. Sous des conditions assez fortes sur la distribution des observations, nous construisons une version du bootstrap qui fonctionne pour le problème de Chernoff. Nos résultats théoriques sont supportés par une étude de simulation.

1 Introduction

In many practical problems in statistics, it is relatively simple to construct estimators but much more difficult to construct confidence intervals and regions, even approximate ones, because of the intractable nature of their finite sample and asymptotic distributions. The bootstrap has often been used with remarkable success to construct approximate confidence intervals which asymptotically achieve the claimed coverage probability. Theoretical accounts can be found in the books of Hall (1992), Efron and Tibshirani (1993) and Shao and Tu (1995), as well as in references therein.

A class of estimators with particularly intractable asymptotic distributions are those estimators defined as the value which minimizes (or maximizes) a given functional of the empirical distribution function. Kim and Pollard (1990) showed that such estimators converge at the rate $n^{-1/3}$ and their asymptotic distribution is the distribution of the argument which minimizes (maximizes) a Gaussian process. Examples include the shorth (Andrews et. al., 1972), an estimator of the location of a univariate distribution, a modal estimator introduced by Chernoff (1964), an estimator of the optimal age of replacement in a nonparametric age replacement policy of Arunkumar (1972) and the least median of squares of Rousseeuw (1984), a robust regression estimator with a high breakdown. With such an asymptotic distribution, it is difficult to make inferences based on these estimators. Efron and Tibshirani (1993) have used the bootstrap to estimate the standard error of the coefficients of the least median of squares estimator, but they only illustrated its use on a single data set. Léger and Cléroux (1992) have studied the behavior of the bootstrap for the estimator of the *cost* of a nonparametric replacement policy. They showed that the bootstrap works for this estimator of cost which has an asymptotic normal distribution and converges at the rate $n^{-1/2}$. The application of the bootstrap for the cost provides at the same time confidence intervals for the optimal age, whose estimator converges at the rate $n^{-1/3}$. In unpublished results for the research leading to that paper, they empirically found that the bootstrap did not work in this case.

In this paper, we will show that the bootstrap does not work in general for such problems. The main reason for this failure is the rate of convergence of the estimator. The estimator can be viewed as a functional of the distribution function. The ordinary bootstrap consists of resampling from the empirical distribution function, so that the value of the "parameter" in the bootstrap world is the functional of the empirical distribution function which converges at the rate $n^{-1/3}$. By using the method of proof of Beran (1984) for bootstrap results, we can investigate the behavior of the bootstrap for other estimates of the distribution of the observations. This allows us to show that the bootstrap can work if resampling is done from a smooth distribution function such that the functional evaluated at this smooth distribution function converges at a faster rate than $n^{-1/3}$. We will show that in the case of Chernoff's estimator, the bootstrap can be made to work if we assume that the distribution of the observations is symmetric and unimodal and that resampling is done from a smooth symmetric estimate of the distribution. This implies that basic bootstrap confidence intervals based on inverting the bootstrap estimate of the distribution of $\hat{\theta} - \theta$ have asymptotically the right coverage. Interestingly, bootstrap percentile intervals based on resampling from the same smooth symmetric estimate of the distribution of the observations are inconsistent. The practical implication of these results is of course limited: if we know that the observations come from a symmetric unimodal distribution, why should we ever want to construct a confidence interval for the mode by using Chernoff's modal estimator, which converges at the rate $n^{-1/3}$, when we could construct a confidence interval based on the sample median or the sample mean, which converge at the rate $n^{-1/2}$. Nevertheless, these results are interesting because they illustrate that the usual bootstrap cannot be used on such estimators and they show exactly wherein lies the problem. Previous negative bootstrap results can be found, for instance, in Shao and Tu (1995). They include statistics which are not sufficiently smooth, such as the absolute value of the mean when the true mean is 0, or which converge faster then $n^{-1/2}$ such as the maximum order statistic or a function of the mean $g(\bar{X})$ such that the derivative $g'(\mu)$ is 0. In the last two examples, the rate of convergence is n^{-1}. It is well known that the rate of convergence of estimators influences the behavior of resampling methods, see e.g., Altman and Léger (1997). Examples where the rate of convergence is slower than $n^{-1/2}$ include kernel density estimation where it is typically $n^{-2/5}$, but for which the asymptotic distribution is normal. The bootstrap can work in this case, see e.g., Léger and Romano (1990). Cube root estimators also converge slower than the typical estimators and their asymptotic distribution is different from the normal. It is therefore interesting to study the behavior of the bootstrap for this class of estimators.

In Section 2, we introduce the class of estimators studied by Kim and Pollard (1990) and give conditions under which a bootstrap confidence interval may asymptotically have the claimed coverage. We also indicate why, in general, the conditions will not be met. In Section 3, we consider the modal estimator of Chernoff (1964) and present a bootstrap version which works provided that the distribution of the original observations satisfy certain strong conditions, including symmetry. We show that this leads to asymptotically valid *basic bootstrap* confidence intervals, following the terminology of Davison and Hinkley (1997) (also known as the hybrid method in Hall, 1988, and Shao and Tu, 1995), but the coverage probability of *percentile* intervals converges to 1 in this case. Section 4 presents some simulation results that investigate the small sample behavior of bootstrap confidence intervals based on various estimates of the distribution of the observations when it is normal. We only consider the one-sample problem here. Preliminary simulation results for the least median of squares regression estimator seems to indicate a different behavior than in the one-sample problem and require further research which will be presented elsewhere. A better behavior for the bootstrap when applied to least squares regression estimators rather than the mean of a univariate distribution has already been noted, see e.g., Hall (1992). An appendix contains the proofs of the results.

2 Bootstrap for cube root estimators

In this section, we study the asymptotic behavior of bootstrap versions of a class of estimators studied by Kim and Pollard (1990) which converge at the rate $n^{-1/3}$. Let $X_1, X_2, \ldots, X_n$ be independently and identically distributed (i.i.d.) from the distribution P, with empirical distribution function (e.d.f.) $\hat{P}_n$. The parameter of interest is

$$\theta_0(P) = \arg\max_{\theta \in \Theta} Pg(\,\cdot\,, \theta),$$

where $\{g(\,\cdot\,, \theta) : \theta \in \Theta\}$ is a class of functions indexed by a subset $\Theta \in I\!R^d$. (We are following the linear functional notation of Kim and Pollard, 1990, so that $Pg(\,\cdot\,, \theta)$ means $E_P g(X, \theta)$.) For instance, Chernoff (1964) introduced an estimator of the mode of a unimodal distribution by letting $g(\,\cdot\,, \theta)$ be the indicator of an interval of length 2α centered at θ. The parameter $\theta_0(P)$ is estimated by $\theta_n(\hat{P}_n)$ which essentially maximizes $\hat{P}_n g(\,\cdot\,, \theta)$ as defined in part (i) of Condition 1. In many cases we use $\theta_n(\hat{P}_n) = \theta_0(\hat{P}_n)$.

Under conditions such as those of Condition 1 with $P_n \equiv P$, Kim and Pollard (1990) have shown that the process $n^{2/3}[\hat{P}_n g(\,\cdot\,, \theta_0(P) + tn^{-1/3}) - Pg(\,\cdot\,, \theta_0(P))]$ converges in distribution to a Gaussian process $Z(t)$ with continuous sample paths, expected value $-(1/2)t'Vt$ and covariance kernel H, where V and H are defined in parts (iv) and (v) of Condition 1. Moreover, $n^{1/3}(\theta_n(\hat{P}_n) - \theta_0(P))$ converges in distribution to the random vector that maximizes Z. To study bootstrap estimators, it is necessary to prove a result analogous to Kim and Pollard's for a triangular array of distributions rather than just one. This is accomplished in the main theorem of this paper under conditions on the problem that will unfortunately not often be met in practice.

Let $K_n(x, P)$ be the distribution function of $n^{1/3}[\theta_n(\hat{P}_n) - \theta_0(P)]$. To apply the bootstrap to approximate it, we need to estimate the unknown distribution P. Let $\tilde{P}_n$ be such an estimate based on the sample $X_1, \ldots, X_n$. Examples will follow. The bootstrap estimate of $K_n(x, P)$ is $K_n(x, \tilde{P}_n)$. To compute it, bootstrap samples $X_1^*, \ldots, X_n^*$ i.i.d. from $\tilde{P}_n$, with e.d.f. $\hat{Q}_n$, are generated. The bootstrap parameter being estimated is $\theta_0(\tilde{P}_n)$ and its estimate is $\theta_n(\hat{Q}_n)$. The distribution function $K_n(x, \tilde{P}_n)$ is approximated by the empirical distribution function of the values of $n^{1/3}(\theta_n(\hat{Q}_n) - \theta_0(\tilde{P}_n))$ obtained by a Monte Carlo simulation of B bootstrap samples.

The proof of the weak convergence of $\theta_n(\hat{P}_n)$ requires some smoothness of $Pg(\,\cdot\,, \theta)$ in θ. This may not be the case of the bootstrap distribution if $\tilde{P}_n$ is not smooth. Some simulations in Section 4 will support this claim. So, smooth estimators $\tilde{P}_n$ such as kernel estimators, will usually be considered.

To make full use of the power of empirical process theory for maximal inequalities we define, for each

R and n, the class of functions $\mathcal{G}_{R,n}$ and its envelope $G_{R,n}$ as follows:

$$\mathcal{G}_{R,n} = \{g(\cdot,\theta) : |\theta - \theta_0(P_n)| \le R\}, \tag{1}$$
$$G_{R,n}(x) = \sup_{\mathcal{G}_{R,n}} |g(x,\theta)|. \tag{2}$$

As in Kim and Pollard (1990), the class $\mathcal{G}_{R,n}$ (near $R > 0$) must be assumed to be "uniformly manageable" for the envelopes $G_{R,n}$, a term coined by Pollard (1989) to distinguish the regularity conditions that he uses from many similar ones in the empirical process literature. Consult these references for further details.

The next theorem states that bootstrap resampling from $\tilde{P}_n$ will be asymptotically consistent provided the following conditions are satisfied.

Condition 1 *Let $\{P_n\}$ be a sequence of fixed distributions and P be another distribution. Let $\hat{Q}_n$ be the empirical distribution function of a sample of size n from P_n. We say that $\{P_n\}$ satisfies Condition 1 if it satisfies the following conditions.*

Let $\{\theta_n(\hat{Q}_n)\}$ be a sequence of estimators for which
(i) $\hat{Q}_n g(\cdot,\theta_n(\hat{Q}_n)) \ge \sup_{\theta\in\Theta} \hat{Q}_n g(\cdot,\theta) - o_{P_n}(n^{-2/3})$.
Suppose
(ii) $\theta_n(\hat{Q}_n)$ converges in P_n-probability to the unique $\theta_0(P)$ that maximizes $Pg(\cdot,\theta)$;
(iii) $\theta_0(P)$ is an interior point of Θ.
The classes $\mathcal{G}_{R,n}$, for R near 0, must be uniformly manageable for the envelopes $G_{R,n}$ and satisfy
(iv) $P_n g(\cdot,\theta)$ is three times differentiable with second derivative matrix $-V_n(\theta)$ and third derivative array $R_n(\theta)$ such that $V_n(\theta_0(P_n)) \to V(\theta_0(P))$ where $-V(\theta)$ is the second derivative matrix of $Pg(\cdot,\theta)$ and $R_n(\theta)$ is uniformly bounded in a neighborhood of $\theta_0(P)$;
(v) Let $h_n(x,y) = n^{1/3}P_n g(\cdot,x)g(\cdot,y) - n^{1/3}P_n g(\cdot,x)P_n g(\cdot,y)$ and $h(x,y) = n^{1/3}Pg(\cdot,x)g(\cdot,y) - n^{1/3}Pg(\cdot,x)Pg(\cdot,y)$. Let

$$H(s,t) = \lim_{n\to\infty} h(\theta_0(P) + sn^{-1/3}, \theta_0(P) + tn^{-1/3}) \tag{3}$$
$$= \lim_{n\to\infty} h_n(\theta_0(P_n) + sn^{-1/3}, \theta_0(P_n) + tn^{-1/3}) \tag{4}$$

exist for each s,t in $I\!R^d$. Also for each t and each $\epsilon > 0$

$$\lim_{n\to\infty} n^{1/3} P_n g(\cdot,\theta_0(P_n) + tn^{-1/3})^2 \{|g(\cdot,\theta_0(P_n) + tn^{-1/3})| > \epsilon n^{1/3}\} = 0;$$

(vi) $\exists C_1 < \infty$ such that $P_n G^2_{R,n} \le C_1 R$ for all R in a neighborhood of 0 and all n and for each $\epsilon > 0$ there is a K such that $P_n G^2_{R,n}\{G_{R,n} > K\} < \epsilon R$ for R near 0 and all n;
(vii) There is a $C_2 < \infty$ such that for θ_1 and θ_2 near $\theta_0(P)$ $P_n|g(\cdot,\theta_1) - g(\cdot,\theta_2)| \le C_2|\theta_1 - \theta_2|$ for all n.

Theorem 1 *If $\{\tilde{P}_n\}$ satisfies Condition 1 with probability 1, then*

$$\sup_x |K_n(x,P) - K_n(x,\tilde{P}_n)| \to 0, \qquad \text{with probability 1.}$$

If $\Theta \in I\!R^1$ so that the problem is one-dimensional, a $1-2\alpha$ basic bootstrap confidence interval for $\theta_0(P)$ would be given by

$$[\theta_n(\hat{P}_n) - n^{-1/3}K_n^{-1}(1-\alpha,\tilde{P}_n), \theta_n(\hat{P}_n) - n^{-1/3}K_n^{-1}(\alpha,\tilde{P}_n)]. \tag{5}$$

If Theorem 1 is valid, then the coverage probability of this confidence interval converges to $1-2\alpha$ as $n\to\infty$ as a consequence of Theorem 1 of Beran (1984). Unfortunately, Condition 1 is unlikely to be satisfied in general. The main problem is in part (v) with the convergence of $h_n(\theta_0(P_n) + sn^{-1/3}, \theta_0(P_n) + tn^{-1/3})$ to $H(s,t)$ (equation 4), the covariance kernel of the Gaussian process $Z(t)$ introduced at the beginning of the section. Consider a fixed sequence of distributions $\{P_n\}$. Provided that h_n is sufficiently smooth in x and y,

$$n^{1/3}h_n(\theta_0(P_n)+sn^{-1/3},\theta_0(P_n)+tn^{-1/3}) = n^{1/3}h_n(\theta_0(P)+sn^{-1/3},\theta_0(P)+tn^{-1/3})$$
$$+n^{1/3}(\theta_0(P_n)-\theta_0(P))\left[\frac{\partial h_n}{\partial x}(\theta_0(P)+sn^{-1/3},\theta_0(P)+tn^{-1/3})\right.$$
$$\left.+\frac{\partial h_n}{\partial y}(\theta_0(P)+sn^{-1/3},\theta_0(P)+tn^{-1/3})\right]+O(n^{1/3}(\theta_0(P_n)-\theta_0(P))^2). \quad (6)$$

Usually, in statistical estimation problems, $h_n(x,y) = h(x,y) + O(n^{-1/2})$. Provided that the extra terms in (6) are negligible, it may be possible that

$$\lim_{n\to\infty} n^{1/3}h_n(\theta_0(P)+sn^{-1/3},\theta_0(P)+tn^{-1/3}) = \lim_{n\to\infty} n^{1/3}h(\theta_0(P)+sn^{-1/3},\theta_0(P)+tn^{-1/3}) = H(s,t),$$

so that (4) would be valid. Here this will not happen in general since for our problems $\theta_0(P_n) = \theta_0(P) + O(n^{-1/3})$. If this is the case, the extra terms in (6) will not be negligible leading to a random covariance kernel. For instance $\theta(\hat{P}_n) - \theta_0(P)$ is of order $O_P(n^{-1/3})$ when $\hat{P}_n$ is the empirical distribution function of a sample of size n from P. So it seems necessary to have $\theta_0(P_n) = \theta_0(P) + o(n^{-1/3})$.

In the next section, we present an estimator for which if P is symmetric, we can find estimates $\tilde{P}_n$ such that $\theta(\tilde{P}_n) - \theta_0(P) = O_P(n^{-1/2})$ so that Theorem 1 is applicable and resampling from $\tilde{P}_n$ leads to a consistent bootstrap procedure.

3 Chernoff's Modal Estimator

Chernoff (1964) introduced an estimator of the mode of a unidimensional distribution as follows. Let α be a fixed value and let $\hat{\theta}$ be the value of θ which maximizes $\hat{P}_n[\theta-\alpha,\theta+\alpha]$ where $\hat{P}_n$ is the e.d.f. of a sample of size n from P, and $\hat{P}_n[a,b]$ is the probability that an observation from $\hat{P}_n$ lies in the interval $[a,b]$. The parameter being estimated is the value θ which maximizes $P[\theta-\alpha,\theta+\alpha]$.

Using the notation of the previous section let

$$g(x,\theta) = I[\theta-\alpha \le x \le \theta+\alpha] - I[\theta_0(P)-\alpha \le x \le \theta_0(P)+\alpha]. \quad (7)$$

Chernoff's estimator is $\theta_n(\hat{P}_n) = \theta_0(\hat{P}_n)$ where $\theta_0(P) = \arg\max_{\theta\in\Theta} Pg(\,\cdot\,,\theta)$. Note that the second indicator function in (7) ensures that $g(\,\cdot\,,\theta_0(P)) \equiv 0$ and by adding a constant independent of θ to $Pg(\,\cdot\,,\theta)$ it does not change the value θ which maximizes it.

Suppose that P is symmetric with respect to θ_0. If θ_1 maximizes $Pg(\,\cdot\,,\theta)$, then by symmetry $2\theta_0-\theta_1$ also maximizes this criterion. It is easy to see that if P is symmetric and unimodal, then $\theta_0(P)$ is the point of symmetry. If more than one value maximizes the criterion, then the distribution may have many modes, each with the same value of the density or it may have a flat peak (such as the uniform distribution). In all of these cases, it may be reasonable to define $\theta_0(P)$ as the mean of all these values. This again leads to $\theta_0(P) = \theta_0$, the point of symmetry. This becomes crucial since if resampling is done from a symmetric estimate of P, say $\tilde{P}_n$, then $\theta_0(\tilde{P}_n)$ is its point of symmetry. By choosing to symmetrize $\tilde{P}_n$ with respect to an estimate such as the mean or the median, then $\theta_0(\tilde{P}_n) - \theta_0(P) = O_P(n^{-1/2})$ rather than $O_P(n^{-1/3})$.

Let K be a distribution function symmetric about 0, and let

$$\tilde{P}_{n,\lambda}(x) = \frac{1}{2n}\left[\sum_{i=1}^{n} K\left(\frac{x-X_i}{\lambda}\right) + \sum_{i=1}^{n} K\left(\frac{x+X_i-2\hat{\theta}}{\lambda}\right)\right],$$

where $\hat{\theta}$ is the median of the sample. Note that $\tilde{P}_{n,\lambda}$ is a smooth symmetric kernel estimate of P and so $\theta_0(\tilde{P}_{n,\lambda}) = \hat{\theta}$ which converges at the rate $n^{-1/2}$.

Theorem 2 *Let P satisfy the following conditions:*
(P1) P is symmetric with respect to $\theta_0(P)$;
(P2) $\sup\{Pg(\cdot,\theta) : |\theta - \theta_0(P)| > \delta\} < Pg(\cdot,\theta_0(P))$ for each $\delta > 0$;
(P3) P has a uniformly continuous second derivative and $P^{(3)}$ is uniformly bounded.

Suppose also that the following conditions on k, the derivative of K, and its bandwidth λ are satisfied:
(K1) k is symmetric with respect to 0 with $\int k(x)\,dx = 1$, $\int xk(x)\,dx = 0$, and $\int x^2k(x)\,dx < \infty$;
(K2) $k^{(r)}$ is uniformly continuous (with modulus of continuity $w_{k,r}$) and of bounded variation for $r = 0, 1$, where $k^{(r)}$ is the r^{th} derivative of k;
(K3) $\int k^{(r)}(x)\,dx < \infty$ and $k^{(r)}(x) \to 0$ as $|x| \to \infty$ for $r = 0, 1$;
(K4) $\int |x \log|x||^{1/2}\,|dk^{(r)}(x)| < \infty$, for $r = 0, 1$;
(K5) Letting $\xi_r(u) = \{w_{k,r}\}^{1/2}$, $\int_0^1 \{\log(1/u)\}^{1/2}\,d\xi_r(u) < \infty$ for $r = 0, 1$
(K6) $\lambda \to 0$, $(n\lambda)^{-1}\log n \to 0$ and $(n\lambda^3)^{-1}\log(1/\lambda) \to 0$ as $n \to \infty$.

Then

$$\sup_x |K_n(x,P) - K_n(x,\tilde{P}_{n,\lambda})| \to 0, \qquad \text{with probability 1},$$

where $K_n(x,P)$ is the distribution function of $n^{1/3}(\theta_0(\hat{P}_n) - \theta_0(P))$ and $\theta_0(\hat{P}_n)$ is Chernoff's modal estimator.

This theorem implies that basic bootstrap confidence intervals based on Chernoff's estimator will asymptotically have the claimed coverage probability as stated in the following corollary.

Corollary 1 *Consider the basic bootstrap $1-2\alpha$ two-sided confidence interval computed from bootstrap samples generated from $\tilde{P}_{n,\lambda}$ given by*

$$[\theta_0(\hat{P}_n) - n^{-1/3}K_n^{-1}(1-\alpha,\tilde{P}_{n,\lambda}), \theta_0(\hat{P}_n) - n^{-1/3}K_n^{-1}(\alpha,\tilde{P}_{n,\lambda})]. \tag{8}$$

Under the conditions of Theorem 2, the coverage probability of this interval converges to the nominal level $1-2\alpha$.

Proof: The proof follows immediately from Theorem 1 of Beran (1984).

Remark: In typical problems where the asymptotic distribution of the estimator is normal, whenever the basic bootstrap confidence interval is asymptotically valid, so is the percentile interval. In this problem, the interval is given by the α^{th} and $(1-\alpha)^{\text{th}}$ quantiles of the bootstrap distribution of the values $\theta_0(\hat{Q}_n)$, where $\hat{Q}_n$ is the empirical distribution function of the bootstrap sample distributed according to $\tilde{P}_{n,\lambda}$. This interval can also be written as

$$[\theta_0(\tilde{P}_{n,\lambda}) + n^{-1/3}K_n^{-1}(\alpha,\tilde{P}_{n,\lambda}), \theta_0(\tilde{P}_{n,\lambda}) + n^{-1/3}K_n^{-1}(1-\alpha,\tilde{P}_{n,\lambda})]. \tag{9}$$

Note that the interval is centered at $\theta_0(\tilde{P}_{n,\lambda})$, not $\theta_0(\hat{P}_n)$. The coverage probability of the percentile interval is

$$\begin{aligned} & Prob\{-K_n^{-1}(1-\alpha,\tilde{P}_{n,\lambda}) \le n^{1/3}(\theta_0(\tilde{P}_{n,\lambda}) - \theta_0(P)) \le -K_n^{-1}(\alpha,\tilde{P}_{n,\lambda})\} \\ = \quad & Prob\{-n^{1/6}K_n^{-1}(1-\alpha,\tilde{P}_{n,\lambda}) \le n^{1/2}(\theta_0(\tilde{P}_{n,\lambda}) - \theta_0(P)) \le -n^{1/6}K_n^{-1}(\alpha,\tilde{P}_{n,\lambda})\}, \end{aligned}$$

which converges to 1 since the middle term is $O_P(1)$, while the left and right terms in the probability statement converge to $-\infty$ and ∞, respectively. Note that $\tilde{P}_{n,\lambda}$ was used as the estimator of P precisely because $\theta_0(\tilde{P}_{n,\lambda})$ is $\sqrt{n}$-consistent.

To prove Theorem 2, we will consider P and a fixed sequence $\{P_n\}$ satisfying Condition 2, show that Condition 1 is then satisfied for $\{P_n\}$ and finally show that $\{\tilde{P}_{n,\lambda}\}$ satisfy Condition 2 with probability 1.

Condition 2 *We say that the fixed sequence* $\{P_n\}$ *and* P *satisfy Condition 2 if* P *satisfies the conditions of Theorem 2 and if:*
(i) $\sup|P_n^{(r)}(x) - P^{(r)}(x)| \to 0$ *for* $r = 0, 1, 2$*, where* $P_n^{(r)}$ *and* $P^{(r)}$ *are the* r^{th} *derivative of* P_n *and* P*, respectively;*
(ii) $\theta_0(P_n) = \theta_0(P) + o(n^{1/3})$

Lemma 1 *Let the sequence* $\{P_n\}$ *and* P *satisfy Condition 2. For* $g(\cdot,\theta)$ *as defined in (7),* $\{P_n\}$ *and* P *satisfy Condition 1.*

Proof: We verify that each part of Condition 1 is satisfied. By definition of $\theta_n(\hat{Q}_n) = \theta_0(\hat{Q}_n)$, so that part (i) is automatically satisfied.

Since the function $Pg(\cdot,\theta)$ has a clean maximum at $\theta_0(P)$ by condition (P2), to get the consistency of (ii), it is sufficient to show that $\hat{Q}_n g(\cdot,\theta)$ converges in probability to $Pg(\cdot,\theta)$, uniformly in θ. Because the function g is a difference of indicator functions, it is sufficient to show that

$$\sup_x |\hat{Q}_n(x) - P(x)| \to 0, \qquad \text{in probability,}$$

where $\hat{Q}_n(x)$ and $P(x)$ are the distribution functions of $\hat{Q}_n$ and P evaluated at x. Now

$$\begin{aligned}\sup_x |\hat{Q}_n(x) - P(x)| &\leq \sup_x |\hat{Q}_n(x) - P_n(x)| + \sup_x |P_n(x) - P(x)| \\ &\to 0, \qquad a.e.\end{aligned}$$

the first term because of the Maximal Inequality (Theorem 3), and the second by assumption (i) of Condition 2.

Part (iii) is clearly satisfied by the symmetry of P. The uniform manageability of the classes of functions $\mathcal{G}_{R,n}$ is immediate from Pollard (1989).

Assumption (i) of Condition 2, along with the assumption of a bounded third derivative for P (Assumption P3) ensures that part (iv) is satisfied.

The first half of part (v) requires more care since the products of indicator functions involved depend on the actual location of $\theta_0(P_n)$ and $\theta_0(P)$. We begin with equation (3). We treat one case in detail. Without loss of generality $s < t$ and suppose that $0 < s < t$.

$$\begin{aligned}
&\lim_{n\to\infty} n^{1/3} Pg(\cdot,\theta_0(P) + sn^{-1/3})g(\cdot,\theta_0(P) + tn^{-1/3}) \\
&= \lim_{n\to\infty} n^{1/3} P\Bigg[\Big[I\{\theta_0(P) + sn^{-1/3} - \alpha \leq X \leq \theta_0(P) + sn^{-1/3} + \alpha\} \\
&\quad -I\{\theta_0(P) - \alpha \leq X \leq \theta_0(P) + \alpha\}\Big]\Big[I\{\theta_0(P) + tn^{-1/3} - \alpha \leq X \leq \theta_0(P) + tn^{-1/3} + \alpha\} \\
&\quad -I\{\theta_0(P) - \alpha \leq X \leq \theta_0(P) + \alpha\}\Big]\Bigg] \\
&= \lim_{n\to\infty} n^{1/3} P\Bigg[I\{\theta_0(P) + tn^{-1/3} - \alpha \leq X \leq \theta_0(P) + sn^{-1/3} + \alpha\} \\
&\quad -I\{\theta_0(P) + sn^{-1/3} - \alpha \leq X \leq \theta_0(P) + \alpha\} - I\{\theta_0(P) + tn^{-1/3} - \alpha \leq X \leq \theta_0(P) + \alpha\} \\
&\quad +I\{\theta_0(P) - \alpha \leq X \leq \theta_0(P) + \alpha\}\Bigg] \\
&= \lim_{n\to\infty} n^{1/3}\Bigg[P(\theta_0(P) + sn^{-1/3} + \alpha) - P(\theta_0(P) + \alpha) \\
&\quad +P(\theta_0(P) + sn^{-1/3} - \alpha) - P(\theta_0(P) - \alpha)\Bigg]
\end{aligned}$$

$$
\begin{aligned}
&= \lim_{n\to\infty} n^{1/3}\Bigg[sn^{-1/3}p(\theta_0(P)+\alpha) + sn^{-1/3}p(\theta_0(P)-\alpha) + o(n^{-1/3})\Bigg] \\
&= s\Bigg[p(\theta_0(P)+\alpha) + p(\theta_0(P)-\alpha)\Bigg],
\end{aligned}
$$

by a Taylor series expansion of $P(x)$ where $p(x)$ is the density at x. The second term in h involves

$$
\begin{aligned}
Pg(\,\cdot\,,\theta_0(P)+sn^{-1/3}) &= P\Bigg[I\{\theta_0(P)+sn^{-1/3}-\alpha \le X \le \theta_0(P)+sn^{-1/3}+\alpha\} \\
&- I\{\theta_0(P)-\alpha \le X \le \theta_0(P)+\alpha\}\Bigg] \\
&= \Bigg[P(\theta_0(P)+sn^{-1/3}+\alpha) - P(\theta_0(P)+sn^{-1/3}-\alpha) \\
&- P(\theta_0(P)+\alpha) + P(\theta_0(P)-\alpha)\Bigg] \\
&= \Bigg[sn^{-1/3}p(\theta_0(P)+\alpha) + sn^{-1/3}p(\theta_0(P)-\alpha) + o(n^{-1/3})\Bigg]
\end{aligned}
$$

Hence

$$
\begin{aligned}
&\lim_{n\to\infty} n^{1/3}\Bigg[Pg(\,\cdot\,,\theta_0(P)+sn^{-1/3})Pg(\,\cdot\,,\theta_0(P)+tn^{-1/3})\Bigg] \\
&= \lim_{n\to\infty}\Bigg[\Big[s[p(\theta_0(P)+\alpha) - p(\theta_0(P)-\alpha)] + o(1)\Big] \\
&\qquad \Big[tn^{-1/3}[p(\theta_0(P)+\alpha) - p(\theta_0(P)-\alpha)] + o(n^{-1/3})\Big]\Bigg] \\
&= 0.
\end{aligned}
$$

Note that this is true irrespective of the value of s and t.

Similar simple computations for the cases $s < t < 0$ and $s < 0 < t$ leads to

$$
H(s,t) = \begin{cases} \min(|s|,|t|)\Bigg[p(\theta_0(P)+\alpha) + p(\theta_0(P)-\alpha)\Bigg], & \text{if } st > 0 \\ 0, & \text{if } st \le 0 \end{cases} \tag{10}
$$

Let us show that the limit in equation (4) is the same. There are three cases: $\theta_0(P) < \theta_0(P_n) + sn^{-1/3}$, $\theta_0(P_n) + sn^{-1/3} \le \theta_0(P) \le \theta_0(P_n) + tn^{-1/3}$, or $\theta_0(P_n) + tn^{-1/3} \le \theta_0(P)$. We treat the first case in detail.

$$
\begin{aligned}
&\lim_{n\to\infty} n^{1/3} P_n g(\,\cdot\,,\theta_0(P_n)+sn^{-1/3}) g(\,\cdot\,,\theta_0(P_n)+tn^{-1/3}) \\
&= \lim_{n\to\infty} n^{1/3} P_n\Bigg[I\{\theta_0(P_n)+tn^{-1/3}-\alpha \le X \le \theta_0(P_n)+sn^{-1/3}+\alpha\} \\
&\quad - I\{\theta_0(P_n)+sn^{-1/3}-\alpha \le X \le \theta_0(P)+\alpha\} - I\{\theta_0(P_n)+tn^{-1/3}-\alpha \le X \le \theta_0(P)+\alpha\} \\
&\quad + I\{\theta_0(P)-\alpha \le X \le \theta_0(P)+\alpha\}\Bigg]
\end{aligned}
$$

$$
\begin{aligned}
&= \lim_{n\to\infty} n^{1/3}\Bigg[P_n(\theta_0(P_n)+sn^{-1/3}+\alpha) - P_n(\theta_0(P_n)+tn^{-1/3}-\alpha) \\
&\quad -P_n(\theta_0(P)+\alpha) + P_n(\theta_0(P_n)+sn^{-1/3}-\alpha) - P_n(\theta_0(P)+\alpha) + P_n(\theta_0(P_n)+tn^{-1/3}-\alpha) \\
&\quad +P_n(\theta_0(P)+\alpha) - P_n(\theta_0(P)-\alpha)\Bigg] \\
&= \lim_{n\to\infty} n^{1/3}\Bigg[P_n(\theta_0(P_n)+sn^{-1/3}+\alpha) - P_n(\theta_0(P)+\alpha) \\
&\quad +P_n(\theta_0(P_n)+sn^{-1/3}-\alpha) - P_n(\theta_0(P)-\alpha)\Bigg].
\end{aligned}
$$

The second case is 0, while the third case is

$$
\begin{aligned}
&\lim_{n\to\infty} n^{1/3} P_n g(\,\cdot\,,\theta_0(P_n)+sn^{-1/3}) g(\,\cdot\,,\theta_0(P_n)+tn^{-1/3}) \\
&\quad = \lim_{n\to\infty} n^{1/3}\Bigg[P_n(\theta_0(P)+\alpha) - P_n(\theta_0(P_n)+tn^{-1/3}+\alpha) \\
&\quad\quad +P_n(\theta_0(P)-\alpha) - P_n(\theta_0(P_n)+tn^{-1/3}-\alpha)\Bigg]
\end{aligned}
$$

Now the second term in h_n is

$$
\begin{aligned}
P_n g(\,\cdot\,,\theta_0(P_n)+sn^{-1/3}) &= P_n\Bigg[I\{\theta_0(P_n)+sn^{-1/3}-\alpha \le X \le \theta_0(P_n)+sn^{-1/3}+\alpha\} \\
&- I\{\theta_0(P)-\alpha \le X \le \theta_0(P)+\alpha\}\Bigg] \\
&= \Bigg[P_n(\theta_0(P_n)+sn^{-1/3}+\alpha) - P_n(\theta_0(P_n)+sn^{-1/3}-\alpha) \\
&- P_n(\theta_0(P)+\alpha) + P_n(\theta_0(P)-\alpha)\Bigg].
\end{aligned}
$$

With assumptions (i) and (ii) of Condition 2, using a Taylor series expansion of $P_n(x)$, we have that $n^{1/3}(\theta_0(P_n)-\theta_0(P)) \to 0$ and so

$$
\begin{aligned}
&\lim_{n\to\infty} n^{1/3}\Bigg[P_n(\theta_0(P_n)+sn^{-1/3}+\alpha) - P_n(\theta_0(P_n)+sn^{-1/3}-\alpha) \\
&- P_n(\theta_0(P)+\alpha) + P_n(\theta_0(P)-\alpha)\Bigg]\Bigg[P_n(\theta_0(P_n)+tn^{-1/3}+\alpha) - P_n(\theta_0(P_n)+tn^{-1/3}-\alpha) \\
&- P_n(\theta_0(P)+\alpha) + P_n(\theta_0(P)-\alpha)\Bigg] \\
&= \lim_{n\to\infty} n^{1/3}\Bigg[\Big\{\theta_0(P_n)-\theta_0(P)+sn^{-1/3}\Big\}\Big\{p_n(\theta_0(P)+\alpha) - p_n(\theta_0(P)-\alpha) + o(n^{-1/3})\Big\}\Bigg] \\
&\qquad \Bigg[\Big\{\theta_0(P_n)-\theta_0(P)+tn^{-1/3}\Big\}\Big\{p_n(\theta_0(P)+\alpha) - p_n(\theta_0(P)-\alpha) + o(n^{-1/3})\Big\}\Bigg] \\
&= 0.
\end{aligned}
$$

So we only need to take care of the first term in h_n when we take the limit. Using the previous results, assumptions (i) and (ii) of Condition 2, and a Taylor series expansion of $P_n(x)$, we have the required convergence of $h_n(s,t)$ to $H(s,t)$.

The second half of part (v) is immediate since $|g(\,\cdot\,, \theta_0(P) + tn^{-1/3})| \leq 1$ so that the probability of the event that the absolute value being larger than $\epsilon n^{1/3}$ becomes 0 for n large enough.

Consider part (vi). Since we are interested in the case when R is small, we can assume that $R < (1/2)\alpha$. If $|\theta_0(P) - \theta_0(P_n)| < R$ then the envelope is

$$G_{R,n}(x) = \begin{cases} 1, & \text{if } \theta_0(P_n) - R - \alpha \leq x \leq \theta_0(P_n) + R - \alpha, \\ & \text{or } \theta_0(P_n) - R + \alpha \leq x \leq \theta_0(P_n) + R + \alpha, \\ 0, & \text{otherwise} \end{cases}$$

If $|\theta_0(P) - \theta(P_n)| > R$, and $\theta_0(P) > \theta(P_n)$ then

$$G_{R,n}(x) = \begin{cases} 1, & \text{if } \theta_0(P_n) - R - \alpha \leq x \leq \theta_0(P) - \alpha, \\ & \text{or } \theta_0(P_n) - R + \alpha \leq x \leq \theta_0(P) + \alpha, \\ 0, & \text{otherwise} \end{cases}$$

The other case is dealt similarly. In all cases,

$$P_n G_{R,n}^2 = O(R + |\theta_0(P) - \theta_0(P_n)|).$$

provided that the density of P_n is bounded above by a constant and that this bound holds for all n, which is the case by assumption (i) of Condition 2 and condition (P3), thus taking care of the first half of part (vi). For the second half, take any $K > 1$. Part (vii) is immediate by taking a Taylor series expansion of P_n using the bound on the densities mentioned above. This completes the proof of this lemma.

We now need to show that the sequence of random distribution functions $\{\tilde{P}_{n,\lambda}\}$ satisfies Condition 2 with probability 1.

Lemma 2 *Let P and K satisfy the conditions of Theorem 2. Then with probability 1, the sequence $\{\tilde{P}_{n,\lambda}\}$ satisfies Condition 2.*

Proof: Let

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

be the smooth (non symmetric) kernel distribution function estimate of P. Then using Theorems A and C of Silverman (1978), we have that $\sup|\hat{P}_{n,\lambda}^{(r)}(x) - P^{(r)}(x)| \to 0$ *a.s.* for $r = 1, 2$. The case $r = 0$ is an immediate consequence of Theorem 1 of Shorack and Wellner (1986, pp. 765). Now we need to show that these conditions are satisfied by the symmetric distribution functions $\tilde{P}_{n,\lambda}$. We will treat the case $r = 1$ in detail and the other two will follow using similar arguments based on the symmetry of the distribution. Let $\tilde{p}_{n,\lambda}(x)$ and $\hat{p}_{n,\lambda}(x)$ be the derivatives of $\tilde{P}_{n,\lambda}(x)$ and $\hat{P}_{n,\lambda}(x)$ respectively. Then, since k is symmetric about 0,

$$\tilde{p}_{n,\lambda}(x) = \frac{1}{2}[\hat{p}_{n,\lambda}(x) + \hat{p}_{n,\lambda}(2\hat{\theta}_n - x)],$$

where $\hat{\theta}_n$ is the center of symmetry of $\tilde{p}_{n,\lambda}$. Let P be symmetric about θ, hence

$$\begin{aligned} \sup_x |\tilde{p}_{n,\lambda}(x) - p(x)| &= \sup_x |\tilde{p}_{n,\lambda}(\theta + x) - p(\theta + x)| \\ &= \frac{1}{2}\sup_x |\hat{p}_{n,\lambda}(\theta + x) - p(\theta + x) + \hat{p}_{n,\lambda}(2\hat{\theta}_n - \theta - x) - p(\theta + x)| \\ &\leq \frac{1}{2}\sup_x |\hat{p}_{n,\lambda}(\theta + x) - p(\theta + x)| \\ &\quad + \frac{1}{2}\sup_x |\hat{p}_{n,\lambda}(2\hat{\theta}_n - \theta - x) - p(\theta - x)|, \end{aligned}$$

by symmetry of p with respect to θ.

The first term goes to 0 by Silverman (1978) whereas the second term is bounded by

$$\begin{aligned}\sup_x |\hat{p}_{n,\lambda}(2\hat{\theta}_n - \theta - x) - p(\theta - x)| &= \sup_x |\hat{p}_{n,\lambda}(\theta - x) + S_n(2\hat{\theta}_n - \theta - x, \theta - x) - p(\theta - x)| \\ &\leq \sup_x |\hat{p}_{n,\lambda}(\theta - x) - p(\theta - x)| \\ &\qquad + \sup_x |S_n(2\hat{\theta}_n - \theta - x, \theta - x)|,\end{aligned}$$

where $S_n(\cdot\,,\,\cdot)$ is the remainder in the Taylor series expansion. Both terms go to 0, the first by Silverman (1978), and the second since $\sup_x |\hat{p}'_{n,\lambda}(x) - p'(x)| \to 0$ a.e., and the sup is less than this sup times $|2\hat{\theta}_n - 2\theta|$ which also goes to 0 a.e.

Since $\tilde{P}_{n,\lambda}$ is symmetric with respect to the median of the sample, then $\theta_0(\tilde{P}_{n,\lambda})$ is the median and so $n^{1/2}(\theta_0(\tilde{P}_{n,\lambda}) - \theta_0(P))$ converges weakly to a normal distribution so that part (ii) is also satisfied with probability 1.

4 Simulation

The theory of the previous section shows that basic bootstrap confidence intervals based on Chernoff's modal estimator have a coverage probability that converges to the claimed level when the sample size increases provided that resampling is done from a smooth symmetric distribution. Moreover, the theory suggests that bootstrapping will not work if the resampling distribution is asymmetric. Also, the theoretical developments require some smoothness. In a small simulation study, we have tried to see to what extent are these findings valid for a small sample size.

For each bootstrap method, we have generated 1,000 samples of size 10 from a standard Gaussian distribution. For each data set, we computed an 80% basic bootstrap two-sided confidence interval based on 1,000 bootstrap samples, as well as the corresponding 10% and 90% (left) one-sided confidence intervals. Throughout the simulation, the size of the half-window, the parameter α, was arbitrarily set at 0.17. We considered five different bootstrap resampling schemes for a total of seven different bootstrap methods. The corresponding estimates of the distribution function are: 1) the empirical distribution function (ordinary bootstrap), 2) the symmetrized empirical distribution function (i.e., the empirical distribution function of $X_1, \ldots, X_n, 2\hat{\theta} - X_1, \ldots, 2\hat{\theta} - X_n$, where $\hat{\theta}$ is the median of the sample), 3) the smooth kernel estimate based on bandwidth 0.1 and 0.4, 4) the smooth symmetric kernel estimate based on the same bandwidths, and 5) a parametric bootstrap where the mean and the variance of the normal distribution are estimated by the mean and (unbiased) variance of the sample. With 1,000 simulated samples, the Monte Carlo standard errors in the estimates are 0.009 and 0.013 when we assume that the true probabilities are 0.1 and 0.8, respectively. Table 1 contains the results.

We see from these results that using a smooth symmetric distribution as an estimate of a symmetric distribution does indeed lead to confidence intervals with coverage probabilities close to the nominal level even with only 10 observations. The case of a smooth symmetric kernel estimate with bandwidth .4 is remarkable while the parametric bootstrap (which is also smooth and symmetric) and the smooth symmetric kernel estimate with bandwidth .1 also give estimators with reasonable coverage. We can also notice that symmetry alone is not sufficient, but seems much more important than smoothness, as was argued in the theoretical section. There is no doubt that the ordinary bootstrap does not work at all.

5 Appendix

In this appendix, we prove Theorem 1. The results in this section are basically a triangular version of the results in Kim and Pollard (1990). So we have tried to keep their notation and their structure of results as much as possible. In most cases, the generalization to a triangular array is straightforward, but some require suitable modifications. Thus for completeness we have stated all results and proofs. The

Table 1: Coverage probabilities one-sided and two-sided basic bootstrap confidence intervals for Chernoff's modal estimator for different bootstrap methods when the half-window is 0.17. The estimated probabilities are based on 1,000 samples of size 10. Each bootstrap confidence interval is based on 1,000 bootstrap observations. The samples are made up of 10 standard normal observations.

Bootstrap method	10% one-sided CI	90% one-sided CI	80% two-sided CI
Ordinary	0.250	0.716	0.466
Symmetric	0.147	0.859	0.712
Smooth, $\lambda = .1$	0.253	0.767	0.514
Smooth, $\lambda = .4$	0.181	0.820	0.639
Smooth symmetric, $\lambda = .1$	0.127	0.877	0.750
Smooth symmetric, $\lambda = .4$	0.092	0.907	0.815
Parametric	0.117	0.881	0.764

corresponding result of Kim and Pollard (1990) is mentioned in parentheses, e.g., KP 3.1 refers to their result 3.1. The first result is a maximal inequality over a class of functions.

Theorem 3 Maximal Inequality *(KP 3.1)*
Let $\mathcal{F}$ be a manageable class of functions with an envelope F, for which $P_nF^2 < \infty$, for all n. Suppose that $0 \in \mathcal{F}$. Then there exists a function J, not depending on n, such that
(i)

$$\begin{aligned}\sqrt{n}I\!P_n \sup_{\mathcal{F}} |\hat{Q}_n f - P_n f| &\leq I\!P_n \sqrt{\hat{Q}_n F^2} J(\sup_{\mathcal{F}} \hat{Q}_n f^2 / \hat{Q}_n F^2) \\ &\leq J(1)\sqrt{P_n F^2}\end{aligned}$$

(ii)

$$\begin{aligned}nI\!P_n \sup_{\mathcal{F}} |\hat{Q}_n f - P_n f|^2 &\leq I\!P_n \ \hat{Q}_n F^2 J^2(\sup_{\mathcal{F}} \hat{Q}_n f^2 / \hat{Q}_n F^2) \\ &\leq J^2(1)\sqrt{P_n F^2}\end{aligned}$$

The function J is continuous and increasing, with $J(0) = 0$ and $J(1) < \infty$.

This result is proved in Pollard (1989) for a fixed distribution P. But since the inequality is true for all n and all P satisfying the condition stated, we immediately have this generalization. The next lemma establishes an $O_{P_n}(n^{-1/3})$ rate of convergence for $\theta_n(\hat{Q}_n)$.

Lemma 3 *(KP Lemma 4.1)*
Suppose that the first half of part (vi) of Condition 1 is satisfied. Then for each $\epsilon > 0$, there exist random variables $\{M_n\}$ of order $O_{P_n}(1)$ such that

$$|\hat{Q}_n g(\,\cdot\,, \theta) - P_n g(\,\cdot\,, \theta)| \leq \epsilon |\theta - \theta_0(P_n)|^2 + n^{-2/3} M_n^2$$

for $|\theta - \theta_0(P_n)| \leq R_0$, where R_0 is the value defining the neighborhood in the above condition.

Proof: For ease of notation suppose that $R_0 = \infty$. Define $M_n(\omega)$ as the infimum (possibly $+\infty$) of these values for which the asserted inequality holds. Define $A(n, j)$ to be the set of those $\theta \in \Theta$ for which

$$(j-1)n^{-1/3} \leq |\theta - \theta_0(P_n)| < jn^{-1/3}.$$

Then for m constant,

$$\begin{aligned}
I\!P_n\{M_n > m\} &\leq I\!P_n\{\exists\theta : n^{2/3}|\hat{Q}_n g(\cdot,\theta) - P_n g(\cdot,\theta)| > \epsilon|\theta - \theta_0(P_n)|^2 + n^{-2/3}m^2\} \\
&\leq \sum_{j=1}^{\infty} I\!P_n\{\exists\theta \in A(n,j) : n^{2/3}|\hat{Q}_n g(\cdot,\theta) - P_n g(\cdot,\theta)| > \epsilon(j-1)^2 + m^2\}.
\end{aligned}$$

Using Markov's inequality, the j^{th} summand is bounded by

$$n^{4/3} I\!P_n \sup_{|\theta-\theta_0(P_n)|<jn^{-1/3}} \frac{|\hat{Q}_n g(\cdot,\theta) - P_n g(\cdot,\theta)|^2}{[\epsilon(j-1)^2 + m^2]^2}.$$

By part (ii) of the maximal inequality, the assumption about $P_n G^2_{R,n}$, and the assumption that $\{\mathcal{G}_{R,n} : R \leq R_0, n \geq 1\}$ is uniformly manageable, there is a constant C' such that the numerator of the last expression is less than $n^{4/3}(n^{-1}C'jn^{-1/3})$. Hence the sum is suitably small for all n by choosing m large enough.

Corollary 2 *(KP Corollary 4.2)*
Suppose that parts (i), (ii), (iv), and (vi) of Condition 1 are satisfied. Then

$$\theta_n(\hat{Q}_n) = \theta_0(P_n) + O_{P_n}(n^{-1/3}).$$

Proof: By part (i),

$$\hat{Q}_n g(\cdot,\theta_n(\hat{Q}_n)) > \hat{Q}_n g(\cdot,\theta_0(P_n)) - O_{P_n}(n^{-2/3}).$$

Using a Taylor series expansion of $P_n g(\cdot,\theta)$, the fact that $\theta_0(P_n)$ maximizes $P_n g(\cdot,\theta)$ so that its first derivative at $\theta_0(P_n)$ is 0 and its second derivative matrix $-V_n(\theta_0(P_n))$ is negative definite, the convergence of $V_n(\theta_0(P_n))$ to $V(\theta_0(P))$, and the fact that the third order derivatives are bounded, there exists $C > 0$ and $\epsilon > 0$ such that for $|\theta - \theta_0(P_n)| < C$

$$P_n g(\cdot,\theta) - P_n g(\cdot,\theta_0(P_n)) \leq -2\epsilon|\theta - \theta_0(P_n)|^2.$$

See e.g., the Corollary of Section 4.1 and Lemma 1 of Section 4.2 of Marsden and Tromba (1976). Then

$$\begin{aligned}
0 &= \hat{Q}_n g(\cdot,\theta_0(P_n)) - \hat{Q}_n g(\cdot,\theta_0(P_n)) \\
&< \hat{Q}_n g(\cdot,\theta_n(\hat{Q}_n)) - \hat{Q}_n g(\cdot,\theta_0(P_n)) + O_{P_n}(n^{-2/3}), \qquad \text{by part (i)} \\
&\leq P_n g(\cdot,\theta_n(\hat{Q}_n)) - P_n g(\cdot,\theta_0(P_n)) + \epsilon|\theta_n(\hat{Q}_n) - \theta_0(P_n)|^2 + 2n^{-2/3}M_n^2 + O_{P_n}(n^{-2/3}) \\
&\qquad \text{using Lemma 3 twice} \\
&\leq -2\epsilon|\theta_n(\hat{Q}_n) - \theta_0(P_n)|^2 + \epsilon|\theta_n(\hat{Q}_n) - \theta_0(P_n)|^2 + 2n^{-2/3}M_n^2 + O_{P_n}(n^{-2/3}) \\
&= -\epsilon|\theta_n(\hat{Q}_n) - \theta_0(P_n)|^2 + 2n^{-2/3}M_n^2 + O_{P_n}(n^{-2/3}).
\end{aligned}$$

This implies that $\theta_n(\hat{Q}_n) = \theta_0(P_n) + O_{P_n}(n^{-1/3})$ ending the proof.

Let

$$Z_n(t) = \begin{cases} n^{2/3}\left[\hat{Q}_n g(\cdot,\theta_0(P_n) + tn^{-1/3}) - P_n g(\cdot,\theta_0(P_n))\right], & \text{if } \theta_0(P_n) + tn^{-1/3} \in \Theta \\ 0, & \text{otherwise} \end{cases} \tag{11}$$

and the corresponding centered process

$$W_n(t) = \begin{cases} Z_n(t) - n^{2/3}\left[P_n g(\cdot,\theta_0(P_n) + tn^{-1/3}) - P_n g(\cdot,\theta_0(P_n))\right], & \text{if } \theta_0(P_n) + tn^{-1/3} \in \Theta \\ 0, & \text{otherwise} \end{cases}$$

Lemma 4 *(KP Lemma 4.5)*
Suppose that parts (iii), (iv), and (v) of Condition 1 are satisfied. Then the finite-dimensional projections of the process Z_n converge in distribution. The limit distributions correspond to the finite-dimensional projections of a process

$$Z(t) = -(1/2)t'Vt + W(t)$$

where $-V$ is the second derivative matrix of $Pg(\cdot, \theta)$ at $\theta_0(P)$ and W is a centered Gaussian process with covariance kernel H defined in part (v).

Proof: With fixed t, and by assuming that $\theta_0(P_n) \to \theta_0(P)$, part (iii) ensures that $\theta_0(P_n) + tn^{-1/3}$ belongs to Θ for n large enough. When that happens

$$W_n(t) = \sum_{i=1}^{n} n^{-1/3}[g(\xi_i, \theta_0(P_n) + tn^{-1/3}) - P_n g(\cdot, \theta_0(P_n) + tn^{-1/3})].$$

Part (iv) implies that

$$n^{2/3}[P_n g(\cdot, \theta_0(P_n) + tn^{-1/3}) - P_n g(\cdot, \theta_0(P_n)] \to -(1/2)t'Vt$$

as $n \to \infty$ which contributes the quadratic trend to the limit process for Z_n.

Now

$$\begin{aligned} cov(W_n(s), W_n(t)) &= n^{1/3} P_n g(\cdot, \theta_0(P_n) + sn^{-1/3}) g(\cdot, \theta_0(P_n) + tn^{-1/3}) \\ &\quad - n^{1/3} P_n g(\cdot, \theta_0(P_n) + sn^{-1/3}) P_n g(\cdot, \theta_0(P_n) + tn^{-1/3}) \\ &\to H(s,t) \end{aligned}$$

by part (v). The second half of part (v) implies the Lindeberg condition.

Lemma 5 *(KP Lemma 4.6)*
Suppose that the classes $\mathcal{G}_{R,n}$ are uniformly manageable for R near 0 for the envelopes $G_{R,n}$. Suppose also that parts (vi) and (vii) of Condition 1 are satisfied. Then the processes $\{W_n\}$ satisfy the stochastic equicontinuity condition (ii) of Theorem 2.3 of Kim and Pollard (1990).

Proof: Let $M > 0$ be fixed and let $\{\delta_n\}$ be a sequence of positive numbers converging to 0. Define $\mathcal{F}(n)$ to be the class of all differences $g(\cdot, \theta_0(P_n) + t_1 n^{-1/3}) - g(\cdot, \theta_0(P_n) + t_2 n^{-1/3})$ with $\max(|t_1|, |t_2|) \le M$ and $|t_1 - t_2| \le \delta_n$. The class has envelope $F_n = 2G_{R(n),n}$ where $R(n) = Mn^{-1/3}$. It is good enough to prove, for every such $\{\delta_n\}$ and M, that

$$n^{2/3} I\!P_n \sup_{\mathcal{F}(n)} |\hat{Q}_n f - P_n f| = o(1)$$

Define $X_n = n^{1/3}\hat{Q}_n F_n^2$ and $Y_n = \sup_{\mathcal{F}(n)} \hat{Q}_n f^2$. Then the uniform manageability of $\mathcal{G}_{R,n}$ and the Maximal Inequality provide a single increasing function $J(\cdot)$ such that

$$n^{2/3} I\!P_n \sup_{\mathcal{F}(n)} |\hat{Q}_n f - P_n f| \le I\!P_n \sqrt{X_n} J(n^{1/3} Y_n / X_n)$$

for n large enough. Notice how the $n^{2/3}$ splits into an $n^{1/2}$ required by the maximal inequality and an $n^{1/6}$, which is absorbed into the definition of $\sqrt{X_n}$. Split according to whether $X_n \le \epsilon$ or not, using the fact that $n^{1/3} Y_n \le X_n$ and invoking the Cauchy-Schwarz inequality for the contribution from $\{X_n > \epsilon\}$, to bound the last expected value by

$$\sqrt{\epsilon} J(1) + \sqrt{I\!P_n X_n} \sqrt{I\!P_n J^2(\min(1, n^{1/3} Y_n / \epsilon))}.$$

Part (vi) ensures that $\mathbb{P}_n X_n = n^{1/3} P_n F_n^2 = O(1)$. It therefore suffices to show that $Y_n = o_{P_n}(n^{-1/3})$. We will establish the stronger result, $\mathbb{P}_n Y_n = o(n^{-1/3})$ by splitting each f into two pieces, according to whether F_n is bigger or smaller than some constant K:

$$\begin{aligned} \mathbb{P}_n \sup_{\mathcal{F}(n)} \hat{Q}_n f^2 &\leq \mathbb{P}_n \sup_{\mathcal{F}(n)} \hat{Q}_n f^2 \{F_n > K\} + K \mathbb{P}_n \sup_{\mathcal{F}(n)} \hat{Q}_n |f| \\ &\leq \mathbb{P}_n \hat{Q}_n F_n^2 \{F_n > K\} + K \sup_{\mathcal{F}(n)} P_n |f| + K \mathbb{P}_n \sup_{\mathcal{F}(n)} \left| \hat{Q}_n |f| - P_n |f| \right|. \end{aligned}$$

Of these three bounding terms: The first can be made less than $\epsilon n^{-1/3}$ by choosing K large enough, according to (vi); with K fixed, the second is of order $O(n^{-1/3}\delta_n)$, by virtue of (vii) and the definition of $\mathcal{F}(n)$; the third is less than $K n^{-1/2} J(1) \sqrt{P_n F_n^2} = O(n^{-2/3})$ by virtue of the maximal inequality applied to the uniformly manageable classes $\{|f| : f \in \mathcal{F}(n)\}$ with envelopes F_n. The result follows.

Theorem 4 *(KP Theorem 4.7) Under the conditions of Lemmas 4 and 5, the processes $\{Z_n\}$ defined by (11) converge in distribution to the process*

$$Z(t) = -(1/2) t' V t + W(t),$$

where $-V$ is the second derivative matrix of $Pg(\,\cdot\,, \theta)$ at $\theta_0(P)$ and W is a centered Gaussian process with continuous sample paths and covariance kernel

$$H(s,t) = \lim_{\alpha \to \infty} \alpha P g(\,\cdot\,, \theta_0(P) + s/\alpha) g(\,\cdot\,, \theta_0(P) + t/\alpha).$$

Proof: Lemma 5 established stochastic equicontinuity for the $\{W_n\}$ processes. Addition of the expected value $n^{2/3}[P_n g(\,\cdot\,, \theta_0(P_n) + t n^{-1/3}) - P_n g(\,\cdot\,, \theta_0(P_n))]$ does not disturb this property. Thus $\{Z_n\}$ satisfies the two conditions of Theorem 2.3 of Kim and Pollard (1990) for convergence in distribution of stochastic processes with paths in $\mathrm{B}_{\mathrm{loc}}(\mathbb{R}^d)$; the process Z has the asserted limit distribution.

Theorem 5 *(KP Theorem 1.1)*
Let $\{P_n\}$ and P satisfy Condition 1. Then the process $n^{2/3}[\hat{Q}_n g(\,\cdot\,, \theta_0(P_n) + t n^{-1/3}) - P_n g(\,\cdot\,, \theta_0(P_n))]$ converges in distribution to a Gaussian process $Z(t)$ with continuous sample paths, expected value $-(1/2)t'Vt$ and covariance kernel H.

If V is positive definite and if Z has nondegenerate increments, then $n^{1/3}(\theta_n(\hat{Q}_n) - \theta_0(P_n))$ converges in distribution to the (almost surely unique) random vector that maximizes Z.

Proof: The conditions of Lemma 3 are satisfied; its Corollary 2, with parts (i) and (iii), give the $O_{P_n}(n^{-1/3})$ rate of convergence for $\theta_n(\hat{Q}_n)$. Parts (iii) to (vii) are the conditions of Lemmas 4 and 5, so Theorem 4 gives the convergence in distribution of Z_n to Z.

The kernel H necessarily has the rescaling property (2.4) of Kim and Pollard (1990). Together with the positive definiteness of V and the nondegeneracy of the increments of Z, this implies (Lemmas 2.5 and 2.6 of Kim and Pollard, 1990) that Z has all its samples paths in $\mathrm{C}_{\max}(\mathbb{R}^d)$. Theorem 2.7 of Kim and Pollard (1990), applied to $t_n = n^{1/3}(\theta_n(\hat{Q}_n) - \theta_0(P_n))$, completes the argument.

Proof of Theorem 1: If $\{P_n\}$ satisfies Condition 1, then Theorem 5 implies

$$\sup_x |K_n(x, P_n) - K_\infty(x, P)| \to 0,$$

where $K_\infty(x, P)$ is the distribution function of the asymptotic distribution of $n^{1/3}(\theta_n(\hat{Q}_n) - \theta_0(P))$ described in Theorem 5, where $\hat{Q}_n$ if the e.d.f. of a sample of size n from P. The continuity of the asymptotic distribution (e.g., Groeneboom, 1989) implies the uniform convergence. Likewise for $P_n \equiv P$, we have

$$\sup_x |K_n(x, P) - K_\infty(x, P)| \to 0.$$

Hence,

$$\sup_x |K_n(x, P_n) - K_n(x, P)| \to 0.$$

Given that $\{\tilde{P}_n\}$ satisfies Condition 1 with probability 1, the result is immediate.

REFERENCES

Altman, N. and Léger, C. (1997). On the Optimality of Prediction-based Selection Criteria and the Convergence Rates of Estimators. *J. R. Statist. Soc.* B, **59**, 205–216.

Andrews, D.F., Bickel, P.J., Hampel, F.R., Huber, P.J., Rogers, W.H., Tukey, J.W. (1972). *Robust Estimates of Location.* Princeton University Press, Princeton, New Jersey.

Arunkumar, S. (1972). Nonparametric age replacement policy. *Sankhyā*, Series A, **34**, 251–256.

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

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

Chernoff, H. (1964). Estimation of the mode. *Ann. Inst. Stat. Math.* **16** 31–41.

Davison, A.C. and Hinkley, D.V. (1997). *Bootstrap Methods and their Application.* Cambridge University Press, Cambridge.

Efron, B. and Tibshirani, R.J. (1993). *An Introduction to the Bootstrap.* Chapman and Hall, London.

Groeneboom, P. (1989). Brownian motion with a parabolic drift and Airy functions. *Probab. Theo. Rel. Fields* **81** 79–109.

Hall, P. (1988). Theoretical comparison of bootstrap confidence intervals. *Ann. Statist.* **16** 927–953.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion.* Springer-Verlag, New York.

Kim, J. and Pollard, D. (1990). Cube root asymptotics. *Ann. Statist.* **18** 191–219.

Léger, C. and Cléroux, R. (1992). Nonparametric age replacement: Bootstrap confidence intervals for the optimal cost. *Operations Research*, **40**, 1062–1073.

Léger, C. and Romano, J.P. (1990). Bootstrap choice of tuning parameters, *Annals of the Institute of Statistical Mathematics*, **4**, 709–735.

Marsden, J.E. and Tromba, A.J. (1976). *Vector Calculus.* W.H. Freeman and Company, San Francisco, California.

Pollard, D. (1989). Asymptotics via empirical processes. *Statist. Sci.* **4** 341–354.

Rousseeuw, P.J. (1984). Least median of squares regression. *J. Amer. Statist. Assoc.* **79** 871–880.

Shao, J. and Tu D. (1995). *The Jackknife and Bootstrap.* Springer, New York.

Shorack, G.R. and Wellner, J.A. (1986). *Empirical Processes With Applications to Statistics.* John Wiley & Sons, New York.

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