On the bootstrap in cube root asymptotics
À propos du bootstrap pour des estimateurs convergeant à la vitesse racine cubique

Christian LÉGER* Brenda MACGIBBON†

CRM-3186

May 2005

*Département de mathématiques et de statistique, Université de Montréal, C.P. 6128, succ. Centre-ville, Montréal (Québec)H3C 3J7, Canada; leger@dms.umontreal.ca

†Département de mathématiques, Université du Québec à Montréal, C.P. 8888, succ. Centre-ville, Montréal (Québec) H3C 3P8, , Canada; brenda@math.uqam.ca

Abstract

The authors study the application of the bootstrap to a class of estimators which converge at a nonstandard rate to a nonstandard asymptotic distribution. They provide a theoretical framework to study its asymptotic behaviour. A simulation study clearly shows that in the case of an estimator such as Chernoff's estimator of the mode (1964) usually the basic bootstrap confidence intervals drastically undercover while the percentile bootstrap intervals overcover. This is one of the rare instances where basic and percentile confidence intervals, which have exactly the same length, behave in a very different way. In the case of Chernoff's estimator, if the distribution is symmetric, it is possible to bootstrap from a smooth symmetric estimator of the distribution for which the basic bootstrap confidence intervals will have the claimed coverage probability while the percentile bootstrap interval will have an asymptotic coverage of 1!

Key words and phrases: Bootstrap; confidence interval for the mode; counterexample; cube root asymptotics; least median of squares.

MSC 2000: Primary 62G09, 62G20.

Résumé

Les auteurs étudient l'application du bootstrap à une classe d'estimateurs qui convergent à une vitesse et à une distribution asymptotique non standard. Ils présentent un cadre théorique pour l'étude de son comportement asymptotique. Une simulation démontre clairement que dans le cas d'un estimateur du mode de Chernoff (1964), la probabilité de couverture de l'intervalle de confiance bootstrap de base est grandement inférieure au niveau prescrit alors que celle des intervalles de type percentile dépasse le niveau prescrit. C'est un des rares cas où les intervalles de confiance de base et percentile ont un comportement si différent malgré des longueurs identiques. Dans le cas de l'estimateur de Chernoff, si la distribution est symétrique, il est possible d'appliquer le bootstrap à partir d'un estimateur lisse et symmétrique de la distribution qui mènera à des intervalles bootstrap de base dont la probabilité de couverture asymptotique sera la bonne alors que celle de l'intervalle percentile convergera vers 1 !

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), Shao and Tu (1995), and Davison and Hinkley (1997) 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 behaviour 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 *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) do not have the right coverage.

In this paper, we use theory and simulations to study the behaviour of the bootstrap for such problems. Simulation results show that the basic bootstrap confidence intervals constructed from the empirical distribution function grossly undercover while *percentile intervals* overcover. Note that the two intervals have exactly the same length. A theoretical study shows that the bootstrap is unlikely to "work" in general for such problems. By that, we mean that the asymptotic distributions of the estimator and its bootstrapped version are not the same which implies that the coverage probability of basic bootstrap confidence intervals does not converge to the nominal level. The main reason for this failure is the rate of convergence of the estimator which causes the covariance kernel in the bootstrapped version to converge to a random one. 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 $n^{-1/3}$. For instance, if the distribution of the observations is symmetric and unimodal, we show that the bootstrap can be made to work in the case of Chernoff's estimator by resampling from a smooth symmetric estimate of the distribution. Although this results in an artificial type of estimator which is not designed to be used in practice, we can show that basic bootstrap confidence intervals based on inverting the bootstrap estimate of the distribution of $\hat{\theta} - \theta$ have asymptotically the right coverage, e.g., a 95% basic bootstrap confidence interval will asymptotically contain the true parameter 95% of the time. Interestingly, the corresponding 95% bootstrap percentile interval based on resampling from the same smooth symmetric estimate of the distribution of the observations will not only overcover, they will asymptotically cover the true parameter 100% of the time!

To our knowledge, this is the first example where the basic and percentile bootstrap confidence intervals behave in such a drastically different manner. 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 than $n^{-1/2}$ 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 behaviour 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 behaviour 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 basic bootstrap confidence intervals may asymptotically have the claimed coverage. We also indicate why, in

general, one of the crucial conditions, on the convergence of the bootstrap covariance structure, will not be met. Section 3 presents simulation results that investigate the small sample behaviour of bootstrap confidence intervals based on various estimates of the distribution of the observations when it is normal and Chernoff's modal estimator is used. We in no way endorse the estimators used in this illustrative simulation study as practical ones since, under the stringent conditions necessary for asymptotically consistent bootstrap intervals, better estimators such as the median with shorter bootstrap confidence intervals exist. These estimators here were chosen to illustrate bootstrap confidence interval behaviour, not for practical use. These results also illustrate the drastically different behaviour of the basic and percentile bootstrap confidence intervals. Since it is well known that the method of subsampling always works as long as there is an asymptotic distrivution (Politis, Romano, and Wolf, 2001), the simulation study also contains results using this method which indicate poor small sample results unless n is very large. In Section 4, we consider the modal estimator of Chernoff (1964) from a theoretical point of view and present a bootstrap version which works provided that the distribution of the original observations satisfies certain conditions, including symmetry, a rather strong condition for the practical use of this estimator. We show that this leads to asymptotically valid basic bootstrap confidence intervals, but the coverage probability of percentile intervals converges to 1 in this case so that these intervals (of exactly the same length) would be preferred. The theoretical results here support the simulation results in the previous section. Section 5 contains a short conclusion. An appendix contains the proofs of the results.

2. BOOTSTRAP FOR CUBE ROOT ESTIMATORS

In this section, we study the asymptotic behaviour 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, following the linear functional notation of Kim and Pollard (1990), $Pg(\,\cdot\,,\theta)$ means $E_P g(X,\theta)$ and where $\{g(\,\cdot\,,\theta) : \theta \in \Theta\}$ is a class of functions indexed by a subset $\Theta \subset I\!R^d$. An example of such a $\theta_0(P)$ is given by Chernoff (1964) where he 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.

Asymptotic results for the bootstrap can be obtained by considering a triangular array set-up, so let us consider sequences of fixed distribution functions $\{P_n\}$, where P_n will converge to P in some sense, and let $\hat{Q}_n$ be the e.d.f. of a sample of size n from P_n. In the bootstrap world, the parameter $\theta_0(P_n)$ the argmax of a functional, is often estimated by $\theta_0(\hat{Q}_n)$ which maximizes $\hat{Q}_n g(\,\cdot\,,\theta)$. We now consider a more general type of estimator $\theta_n(\hat{Q}_n)$ of $\theta_0(P_n)$ which essentially maximizes $\hat{Q}_n g(\,\cdot\,,\theta)$, that is, $\hat{Q}_n g(\,\cdot\,,\theta_n(\hat{Q}_n)) \geq \sup_{\theta\in\Theta} \hat{Q}_n g(\,\cdot\,,\theta) - o_{P_n}(n^{-2/3})$.

As in Kim and Pollard (1990), for each R and n, the class of functions $\mathcal{G}_{R,n}$ and its envelope $G_{R,n}$ are defined to make full use of the power of empirical process theory for maximal inequalities as follows:

$$\mathcal{G}_{R,n} = \{g(\,\cdot\,,\theta) : |\theta - \theta_0(P_n)| \leq R\}, \tag{1}$$

$$G_{R,n}(x) = \sup_{\mathcal{G}_{R,n}} |g(x,\theta)|. \tag{2}$$

The class $\mathcal{G}_{R,n}$ (for positive R near 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.

In order to find conditions under which the bootstrap behaves satisfactorily, more conditions will have to be imposed on the original distribution P and the sequence of fixed distributions $\{P_n\}$. We shall refer to these conditions as Condition 1, given below.

Condition 1. *Let $\{P_n\}$ be a sequence of fixed distributions with $\{\hat{Q}_n\}$ denoting the empirical distribution of a sample of size n from P_n and let P be another distribution. We say that $\{P_n\}$ and P satisfy Condition 1 if the following conditions are satisfied.*
Let $\{\theta_n(\hat{Q}_n)\}$ denote a sequence of estimators for which
(i) $\hat{Q}_n g(\,\cdot\,,\theta_n(\hat{Q}_n)) \geq \sup_{\theta\in\Theta} \hat{Q}_n g(\,\cdot\,,\theta) - o_{P_n}(n^{-2/3})$;
then we suppose that
(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 positive 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 neighbourhood of $\theta_0(P)$;

(v) letting $h_n(u,v) = P_n(g(\cdot\,,u)g(\cdot\,,v)) - P_n g(\cdot\,,u)P_n g(\cdot\,,v)$ *and* $h(u,v) = Pg(\cdot\,,u)g(\cdot\,,v) - Pg(\cdot\,,u)Pg(\cdot\,,v)$ *then*

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

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

exists for each s,t *in* $I\!R^d$ *and also for each* t *and each* $\epsilon > 0$

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

where $I\{\cdot\}$ *is the indicator function of the event;*
(vi) there is a $C_1 < \infty$ *such that* $P_n G_{R,n}^2 \le C_1 R$ *for all* R *in a neighbourhood of 0 and all* n *and for each* $\epsilon > 0$*, there is a* K *such that* $P_n G_{R,n}^2\{G_{R,n} > K\} < \epsilon R$ *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*.*

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, due to the particular nature of the problem, not because we are making stringent assumptions that could be relaxed, as will be explained at the end of this section.

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 for the bootstrap distribution if $\tilde{P}_n$ is not smooth. So, smooth estimators $\tilde{P}_n$ such as kernel estimators, will usually be considered.

The next theorem states that bootstrap resampling from $\tilde{P}_n$ will be asymptotically consistent provided the following conditions are satisfied. To simplify the notation, we assume that $g(\cdot\,,\theta_0(P)) \equiv 0$.

THEOREM 1. *If* $\{\tilde{P}_n\}$ *and* P *satisfy Condition 1 with probability 1, then*

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

This means that the distribution function of the statistic can uniformly be approximated by the bootstrap version. The main application of such a theorem is for the validity of confidence intervals. If $\Theta \subset I\!R^1$ 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) or Theorem 1.2.1 of Politis, Romano and Wolf (2001). On the other hand, the percentile confidence interval is given by

$$[L_n^{-1}(\alpha,\tilde{P}_n), L_n^{-1}(1-\alpha,\tilde{P}_n)], \tag{6}$$

where

$$L_n(x,\tilde{P}_n) = Prob_{\tilde{P}_n}(\theta_n(\hat{Q}_n) \le x), \tag{7}$$

and $\hat{Q}_n$ is the empirical distribution function of the bootstrap sample from the distribution $\tilde{P}_n$. The percentile interval can be rewritten in terms of the quantiles of K_n as follows:

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

Note that the percentile interval is centered at $\theta_0(\tilde{P}_n)$ rather than $\theta_n(\hat{P}_n)$ and that the left endpoint consists of adding the left quantile rather than subtracting the right quantile. In cases where the asymptotic distribution of $\theta_n(\hat{P}_n)$ is

symmetric (e.g., Chernoff's estimator, see Chernoff, 1964, and Groeneboom, 1989), *if* the asymptotic distribution of $\theta_0(\tilde{P}_n)$ is the same as that of $\theta_n(\hat{P}_n)$, then the validity of Theorem 1 would also imply that the asymptotic coverage probability of the percentile intervals would be the nominal one. In general, we have no information on the asymptotic distribution of $\theta_0(\tilde{P}_n)$ when $\tilde{P}_n$ is different from $\hat{P}_n$, but the next section will present a case where the asymptotic distributions differ considerably, even in the rate of convergence.

Unfortunately, Condition 1 is unlikely to be satisfied in general. The main problem is in part (v) with the convergence of $n^{1/3}h_n(\theta_0(P_n)+sn^{-1/3},\theta_0(P_n)+tn^{-1/3})$ to $H(s,t)$ (equation 4), the covariance kernel of the Gaussian process $Z(t)$ introduced previously. Consider a fixed sequence of distributions $\{P_n\}$. Heuristic arguments of Knight (1989) for bootstrapping the sample mean in the infinite variance case, which were made more precise and widely applicable later in Zarepour and Knight (1999), can be used here to show that $n^{1/3}h_n(\theta_0(P_n)+sn^{-1/3},\theta_0(P_n)+tn^{-1/3})$ converges to a random covariance kernel. Thus the naive bootstrap will fail. A method that always works as long as there is an asymptotic distribution is subsampling (Politis, Romano, and Wolf, 2001). Unfortunately, however, this is at the expense of poorly behaved small sample results unless the sample size is very large. This will be illustrated in the following section.

The following simple mathematical argument also illustrates these ideas clearly in the case of sufficiently smooth h_n. Let us consider :

$$\begin{aligned} n^{1/3}h_n(\theta_0(P_n)+sn^{-1/3},\theta_0(P_n)+tn^{-1/3}) &= n^{1/3}h_n(\theta_0(P)+sn^{-1/3},\theta_0(P)+tn^{-1/3}) \\ &+n^{1/3}(\theta_0(P_n)-\theta_0(P))\left[\frac{\partial h_n}{\partial x}(\theta_0(P)+sn^{-1/3},\theta_0(P)+tn^{-1/3})\right. \\ &\left.+\frac{\partial h_n}{\partial y}(\theta_0(P)+sn^{-1/3},\theta_0(P)+tn^{-1/3})\right]+O(n^{1/3}(\theta_0(P_n)-\theta_0(P))^2). \end{aligned} \tag{9}$$

Usually, in statistical estimation problems, $h_n(x,y)=h(x,y)+O(n^{-1/2})$. Provided that the extra terms in (9) 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 (9) will not be negligible leading to a random covariance kernel. For instance $\theta_0(\hat{P}_n)-\theta_0(P)$ 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 order that the extra terms in (9) be negligible.

In either case by using point process theory or the simpler mathematical argument above, there is clearly a problem for the basic bootstrap confidence interval. In order to understand these problems more fully an illustrative simulation study is performed for Chernoff's modal estimator. Simulations in the next section clearly show that the coverage probabilities of the basic bootstrap confidence intervals for Chernoff's estimator are relatively stable as n increases and very far from the nominal level supporting the thesis that Theorem 1 is not valid in that case. Therefore basic bootstrap confidence intervals will not asymptotically have the claimed coverage.

3. A SIMULATION STUDY OF 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

$$g(x,\theta)=I[\theta-\alpha\le x\le\theta+\alpha]-I[\theta_0(P)-\alpha\le x\le\theta_0(P)+\alpha]. \tag{10}$$

Chernoff's estimator is $\theta_0(\hat{P}_n)$ where $\theta_0(P)=\arg\max_{\theta\in\Theta} Pg(\,\cdot\,,\theta)$. Note that the second indicator function in (10) ensures that $g(\,\cdot\,,\theta_0(P))\equiv 0$ and by adding a constant independent of θ to $Pg(\,\cdot\,,\theta)$ this does not change the value θ which maximizes it.

Given the nature of this estimation problem, it would seem that symmetry might play a crucial role. Let us 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 this 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 $\theta_0(P)$ its 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.

Table 1: Coverage probabilities of one-sided and two-sided basic and percentile (perc) 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 1000 samples of size $n = 10$, 100, 200, and 1000. Each bootstrap confidence interval is based on 999 bootstrap samples. The samples are made up of standard normal observations.

		Sample size n							
		10		100		200		1000	
Bootstrap method	CI	perc	basic	perc	basic	perc	basic	perc	basic
Ordinary	left 5%	2.20	23.50	1.10	23.50	2.00	25.90	2.50	23.80
	left 95%	98.10	79.80	98.30	78.00	98.20	77.00	97.20	76.00
	2-sided 90%	95.90	56.30	97.20	54.50	96.20	51.10	94.70	52.20
Smooth,	left 5%	0.60	11.50	0.40	16.30	0.50	17.20	0.50	18.90
$\lambda = n^{-0.3}$	left 95%	99.60	87.80	99.70	85.80	99.90	84.10	99.10	82.10
	2-sided 90%	99.00	76.30	99.30	69.50	99.40	66.90	98.60	63.20
Smooth,	left 5%	1.10	23.30	0.60	22.90	1.80	25.80	2.00	23.10
$\lambda = 0.25n^{-0.3}$	left 95%	98.80	79.10	98.30	76.70	98.20	77.40	97.20	77.20
	2-sided 90%	97.70	55.80	97.70	53.80	96.40	51.60	95.20	54.10
Symmetric	left 5%	3.80	7.60	0.50	6.50	0.20	4.80	0.10	3.30
	left 95%	97.30	93.10	99.40	95.10	99.80	94.90	99.80	96.60
	2-sided 90%	93.50	85.50	98.90	88.60	99.60	90.10	99.70	93.30
Smooth symmetric,	left 5%	0.80	2.80	0.00	4.40	0.00	2.90	0.00	1.70
$\lambda = n^{-0.3}$	left 95%	99.30	96.90	100.0	97.00	100.0	97.60	100.0	97.50
	2-sided 90%	98.50	94.10	100.0	92.60	100.0	94.70	100.0	95.80
Smooth symmetric,	left 5%	3.00	6.80	0.40	6.40	0.20	3.70	0.00	2.70
$\lambda = 0.25n^{-0.3}$	left 95%	97.80	93.10	99.60	95.80	99.90	95.80	100.0	96.60
	2-sided 90%	94.80	86.30	99.20	89.40	99.70	92.10	100.0	93.90
Parametric	left 5%	0.20	4.70	0.00	5.50	0.00	6.10	0.00	5.20
	left 95%	99.80	95.20	100.0	95.30	100.0	95.40	100.0	95.00
	2-sided 90%	99.60	90.50	100.0	89.80	100.0	89.30	100.0	89.80

We also have some indications from the second section that smoothness can play an important role in the theory. Therefore let K be the distribution function of a symmetric density about 0, and let

$$\tilde{P}_{n,\lambda}(x) = \frac{1}{2n}\left[\sum_{i=1}^{n} K\left(\frac{x - X_i}{\lambda}\right) + \sum_{i=1}^{n} K\left(\frac{x + X_i - 2\hat{\theta}}{\lambda}\right)\right], \tag{11}$$

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 $n^{-1/2}$. Again arguments at the end of Section 2 seem to imply the consistency of the bootstrap in such a case.

As mentioned above, both symmetry and/or smoothness are expected to play a role in the small sample behaviour of basic and percentile bootstrap intervals based on Chernoff's modal estimator. The simulation study is designed with this in mind and we have tried to see to what extent such behaviour occurs for different sample sizes.

For each bootstrap method, we have generated 1000 samples of size $n = 10$, 20, 100, and 1000 from a standard Gaussian distribution. For each data set, we computed 90% basic and percentile bootstrap two-sided confidence intervals based on 999 bootstrap samples, as well as the corresponding 5% and 95% (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 smooth kernel estimate based on bandwidth $n^{-0.3}$ and $0.25n^{-0.3}$, 3) 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), 4) the smooth symmetric kernel estimate based on the previous 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 1000 simulated samples, the Monte Carlo standard errors in the estimates are 0.007 and 0.009 when we assume that the true probabilities are 0.05 and 0.9, respectively. Table 1 contains the results.

The basic confidence intervals of the ordinary bootstrap clearly do not work. The coverage of 90% two-sided intervals varies between 52.20% and 56.30%, even for samples of size 1000. So, the ordinary bootstrap does not work, i.e., the bootstrapped version of the estimator cannot have the same asymptotic distribution as the estimator itself. On the other hand, the coverage of the 90% percentile intervals is also quite stable varying from 94.10% and 97.30%. Therefore the percentile intervals significantly overcover. Smoothing the ordinary bootstrap does not really change

Table 2: Coverage probabilities of one-sided and two-sided confidence intervals and average length of the two-sided intervals for Chernoff's modal estimator when the half-window is 0.17. The intervals are the basic smooth symmetric bootstrap (smooth sym) with a bandwidth of $\lambda = 0.25n^{-0.3}$, the basic parametric bootstrap and the subsampling method using three subsample sizes for each sample size. The estimated probabilities are based on 1000 samples of size $n = 10$, 100, 200, and 1000. Each bootstrap and subsampling confidence interval is based on 999 resamples. The samples are made up of standard normal observations.

		Coverage probability			Average length of
Sample size	CI	left 5%	left 95%	2-sided 90%	the 2-sided interval
10	sub $b = 4$	24.90	78.30	53.40	1.777
	sub $b = 6$	28.50	73.80	45.30	1.387
	sub $b = 8$	33.40	68.20	34.80	0.926
	smooth sym	6.80	93.10	86.30	1.910
	parametric	4.70	95.20	90.50	1.931
100	sub $b = 4$	7.40	91.90	84.50	2.194
	sub $b = 6$	8.80	90.80	82.00	2.071
	sub $b = 8$	10.80	89.70	78.90	1.910
	smooth sym	6.40	95.80	89.40	1.235
	parametric	5.50	95.30	89.80	1.253
200	sub $b = 5$	5.60	95.80	90.20	2.195
	sub $b = 10$	8.50	92.10	83.60	1.896
	sub $b = 15$	10.50	90.90	80.40	1.744
	smooth sym	3.70	95.80	92.10	1.052
	parametric	6.10	95.40	89.30	1.079
1000	sub $b = 25$	3.60	95.80	92.20	1.630
	sub $b = 30$	4.20	95.30	91.10	1.573
	sub $b = 35$	5.20	95.00	89.80	1.526
	smooth sym	2.70	96.60	93.90	0.733
	parametric	5.20	95.00	89.80	0.751

the conclusions of both types of confidence intervals. Resampling from symmetric distribution functions has a very measurable effect. Coverage probabilities of basic confidence intervals are close to the claimed level supporting the theory. 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. Those of the percentile intervals converge to 1 as n increases. This phenomenon will be given a theoretical basis in the next section. It should be noted that for each sample size, the coverage probability of the basic intervals increases with the value of the smoothing parameter. Therefore for n large, the smooth symmetric basic intervals overcover. On the other hand, the parametric basic and percentile intervals behave as predicted from the theory.

Similar partial results were obtained for the shorth robust estimator. Also, the typical undercoverage and overcoverage of the basic and percentile confidence intervals from the ordinary bootstrap has been observed in the problem of age replacement policy found in Léger and Cléroux (1992).

We have also computed bootstrap estimates of the standard error of the modal estimator. For samples of size 10 and 20, the bootstrap estimates from the empirical distribution or a smoothed version were slightly biased upward whereas for samples of size 100 and 1000, the bias was not important. Resampling from symmetric distributions led to standard error estimates with basically no bias.

An attractive alternative for this class of problems could be subsampling. It consists of resampling *without* replacement subsamples of size b. Provided that the rate of convergence is known, that an asymptotic distribution exists, and that $b/n \to 0$ as $n \to \infty$, Theorem 2.2.1 of Politis, Romano, and Wolf (2001) shows that the coverage probability of confidence intervals based on subsampling will have the claimed level. These intervals are based on finding the quantiles of $b^{1/3}(\hat{\theta}_b - \hat{\theta}_n)$ and multiplying them by the correction factor $n^{-1/3}$, where $\hat{\theta}_b$ is the estimator computed from the subsample of b, and $\hat{\theta}_n$ is the estimator based on the original sample. The requirement that $b/n \to 0$ suggests that the subsample size b must be relatively small compared to the sample size n. Table 2 presents the results of a simulation study comparing the smooth symmetric bootstrap with $\lambda = 0.25n^{-0.3}$ and the parametric bootstrap to subsampling using three different subsampling sizes for each value of n. The subsample sizes were found by trial and error to include what seems like the best value of b plus two neighboring values. In terms of coverage probability, subsampling is not competitive until the sample size is 200, and the best value of b is then 5 and the coverage error quickly increases as b increases to 10 or 15. For a sample of size 1000, subsampling from

b between 25 and 35 will lead to very good coverage probabilities. On the other hand, the two-sided subsampling confidence intervals with good coverage probabilities are about twice as large as the bootstrap intervals. Consequently, subsampling is not a practical solution unless the sample size is very large, and even then, the confidence intervals are much larger than the bootstrap intervals. Moreover, subsampling is quite sensitive to the subsample size, unless n is very large. On the other hand, if the distribution of the observations is not symmetric, no bootstrap solution exists and one may be left with the imperfect subsampling solution.

4. THEORETICAL CONSIDERATIONS OF CHERNOFF'S MODAL ESTIMATOR

In this section we provide a firm theoretical basis for the behaviour of the different bootstrap confidence intervals in the simulation study of Section 3. As indicated before if P is symmetric, using estimates $\tilde{P}_n$ which are smooth and symmetric with respect to, e.g., the median, then $\theta_0(\tilde{P}_n) - \theta_0(P) = O_P(n^{-1/2})$. This condition is important to ensure that the covariance structure of the bootstrap process will be the same as the original process, (see part (iv) of Condition 1, and equation (9)). In this case, Theorem 1 is applicable and resampling from $\tilde{P}_n$ leads to a consistent bootstrap procedure. The next theorem gives sufficient conditions for the basic bootstrap confidence intervals based on Chernoff's estimator in the symmetric case to achieve the asymptotically correct coverage.

THEOREM 2. *We follow the definition and notations at the beginning of Section 3 and we assume that P satisfies 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 given in equation (11) 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 rth 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 satisfying the above conditions 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*

$$[\theta_0(\hat{P}_n) - n^{-1/3}K_n^{-1}(1-\alpha, \tilde{P}_{n,\lambda}), \theta_0(\hat{P}_n) - n^{-1/3}K_n^{-1}(\alpha, \tilde{P}_{n,\lambda})]. \tag{12}$$

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 1.2.1 of Politis, Romano and Wolf (2001).

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. However, as mentioned earlier, the percentile interval is given by the αth 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 $\tilde{P}_{n,\lambda}$ and the interval can also be written as

$$[\theta_0(\tilde{P}_{n,\lambda}) + n^{-1/3}K_n^{-1}(\alpha, \tilde{P}_{n,\lambda}), \theta_0(\tilde{P}_{n,\lambda}) + n^{-1/3}K_n^{-1}(1-\alpha, \tilde{P}_{n,\lambda})]. \tag{13}$$

Note that the interval is centered at $\theta_0(\tilde{P}_{n,\lambda})$ which is the median here, not $\theta_0(\hat{P}_n)$, Chernoff's modal estimator. The coverage probability of the percentile interval is

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

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. The length of the basic and percentile intervals is identical and is $O_P(n^{-1/3})$. The basic interval is centered at $\theta_0(\hat{P}_n)$ (Chernoff's modal estimator), which is within $O_P(n^{-1/3})$ of $\theta_0(P)$ and the asymptotic coverage is the nominal one. On the other hand, the percentile interval is centered at $\theta_0(\tilde{P}_{n,\lambda})$ (the median), which is within $O_P(n^{-1/2})$ of $\theta_0(P)$. The intervals are therefore an order of magnitude too large leading to an asymptotic coverage of 1, *despite the fact that they are exactly as large as the basic intervals.* Nevertheless, it is important to note that the basic interval of (12) and the percentile interval of (13) are both constructed as they should be when resampling from $\tilde{P}_{n,\lambda}$. Their drastically different asymptotic behaviours are an artifact of this particular problem, not the result of centering on inappropriate estimators.

Here we show, in the case of Chernoff's estimator how the limit covariance function of the bootstrap process (part (v) in Condition 1) turns out to be a problem unless we use a smooth symmetric distribution function estimator such as $\tilde{P}_{n,\lambda}$ from which to bootstrap. Since the function $g(x,\theta)$ of equation (10) is the difference of two indicator functions of intervals and the second one is always centered at $\theta_0(P)$ (to ensure that $g(\,\cdot\,,\theta_0(P)\equiv 0)$, the limit $H(s,t)$ of equation (3), based on the function $h(u,v)$, contains terms such as

$$\begin{aligned}\lim_{n\to\infty} \quad & n^{1/3}\{P(\theta_0(P)+sn^{-1/3}+\alpha)-P(\theta_0(P)+\alpha)\}\\ = \quad & \lim_{n\to\infty} n^{1/3}\{sn^{-1/3}p(\theta_0(P)+\alpha)+o(n^{-1/3})\}\\ = \quad & sp(\theta_0(P)+\alpha), \end{aligned} \tag{14}$$

where p is the density of P. In order for the bootstrap to work, the same limit must be obtained when P is replaced by P_n everywhere except in the argument of the second term which remains $\theta_0(P)+\alpha$, where P_n is a sequence of fixed distribution functions which can be viewed as estimating P. By using, e.g., a kernel estimator, we can make sure that $P_n(x)$ converges to $P(x)$ and its derivative $p_n(x)$ converges to $p(x)$. Then using a Taylor expansion on P_n,

$$\begin{aligned}\lim_{n\to\infty} \quad & n^{1/3}\{P_n(\theta_0(P_n)+sn^{-1/3}+\alpha)-P_n(\theta_0(P)+\alpha)\}\\ = \quad & \lim_{n\to\infty} n^{1/3}\{\theta_0(P_n)-\theta_0(P)+sn^{-1/3}\}\{p_n(\theta_0(P)+\alpha)+o(n^{-1/3})\}\\ = \quad & sp(\theta_0(P)+\alpha)+\lim_{n\to\infty} n^{1/3}\{\theta_0(P_n)-\theta_0(P)\}p_n(\theta_0(P)+\alpha),\end{aligned}$$

which is the same as (14) provided that $\theta_0(P_n)-\theta_0(P)=o(n^{-1/3})$. In the case of Chernoff's modal estimator using $\tilde{P}_{n,\lambda}$, this is the case (in probability), but in general, even if one resamples from a smooth distribution function estimator $\hat{P}_n$, one faces $\theta_0(\hat{P}_n)-\theta_0(P)=O_P(n^{-1/3})$ and the limiting covariance function of the bootstrap process will be random and therefore different from the covariance function of the original process. Thus, the bootstrap distribution of the cuberoot estimator will not asymptotically converge to the asymptotic distribution of the estimator.

To prove Theorem 2, one needs to 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 (10), $\{P_n\}$ and P satisfy Condition 1.*

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

Since the function $Pg(\,\cdot\,,\theta)$ has a 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

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

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 $0 < s < t$.

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

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

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

Hence

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

$$\left[tn^{-1/3}\big[p(\theta_0(P)+\alpha)-p(\theta_0(P)-\alpha)\big]+o(n^{-1/3})\right]\Bigg]$$
$$= \quad 0.$$

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

$$H(s,t) = \begin{cases} \min(|s|,|t|)\Big[p(\theta_0(P)+\alpha)+p(\theta_0(P)-\alpha)\Big], & \text{if } st > 0 \\ 0, & \text{if } st \le 0 \end{cases} \tag{15}$$

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.

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

The second case is 0, while the third case is

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

Now the second term in h_n is

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

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

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

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 $|g(\,\cdot\,,\theta_0(P)+tn^{-1/3})| \le 1$ 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

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

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

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

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

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

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.

5. CONCLUSION

In this paper, we have studied the application of 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 in order to study the behaviour of the bootstrapped version of the estimator. We have indicated the asymptotic distribution of the bootstrapped version is very different from that of the original estimator because of the convergence of the covariance kernel to a random one. Simulation results clearly show that the basic bootstrap confidence intervals drastically undercover while the percentile bootstrap intervals overcover. This is one of the rare instances where basic and percentile confidence intervals, which have exactly the same length, behave in a very different way. In the case of Chernoff's estimator, if the underlying distribution is symmetric, we constructed a bootstrap version which is asymptotically valid so that the basic confidence intervals will asymptotically have the right coverage. Interestingly, the coverage of the corresponding percentile intervals converged to 1! We consider this a very interesting phenomenon in its own right. However, we do emphasize that the bootstrapped version of the estimator considered here in the case of Chernoff's estimator are for illustrative purposes only. We do not recommend them for practical use. Although it is known that the subsampling method works if there is an asymptotic distribution, the simulation study here indicates that unless the sample size is very large (e.g., 200), the coverage error will be important; that unless the sample size is even larger (e.g., 1000), the coverage error is quite sensitive to the choice of subsample size around the optimal value; and that in either case, the subsampling confidence intervals are about twice as large as the basic bootstrap ones. Except for this special case, we do not know a bootstrap method that asymptotically has the claimed coverage level for estimators with cube root asymptotics. Subsampling is an alternative solution with some important practical limitations discussed above. However, be wary of using the ordinary bootstrap in nonstandard problems.

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 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)

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

(ii)

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

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 neighbourhood 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,

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

Using Markov's inequality, the jth summand is bounded by

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

By part (ii) of the maximal inequality, the assumption about $P_nG_{R,n}^2$, 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 $|\theta - \theta_0(P_n)| < C$

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

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

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

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

Let

$$Z_n(t) = \begin{cases} n^{2/3}\left[\hat{Q}_n g(\,\cdot\,,\theta_0(P_n) + tn^{-1/3}) - P_n g(\,\cdot\,,\theta_0(P_n))\right], & \text{if } \theta_0(P_n) + tn^{-1/3} \in \Theta \\ 0, & \text{otherwise} \end{cases} \tag{16}$$

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

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

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

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

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

$$\sqrt{\epsilon}J(1) + \sqrt{\mathbb{P}_n X_n}\sqrt{\mathbb{P}_n J^2(\min(1, n^{1/3}Y_n/\epsilon))}.$$

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:

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

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 $Kn^{-1/2}J(1)\sqrt{P_n F_n^2} = O(n^{-2/3})$ 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 (16) converge in distribution to the 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 continuous sample paths and covariance kernel

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

Proof: Lemma 5 established stochastic equicontinuity for the $\{W_n\}$ processes. Addition of the expected value $n^{2/3}[P_n g(\,\cdot\,,\theta_0(P_n) + tn^{-1/3}) - P_n g(\,\cdot\,,\theta_0(P_n))]$ 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 $n^{2/3}[\hat{Q}_n g(\,\cdot\,,\theta_0(P_n) + tn^{-1/3}) - P_n g(\,\cdot\,,\theta_0(P_n)]$ 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.

ACKNOWLEDGEMENTS

The authors would like to thank the editor and a referee for helpful and insightful comments which led to an improvement in the original version of this manuscript. Partial funding for this work has been supplied by the Natural Sciences and Engineering Research Council of Canada and MITACS. The authors also acknowledge the support of the Centre de Recherches Mathématiques of Université de Montréal as well as the GERAD research center.

REFERENCES

N. Altman, & C. Léger (1997). On the Optimality of Prediction-based Selection Criteria and the Convergence Rates of Estimators. *Journal of the Royal Statistical Society, Series B*, 59, 205–216.

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

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

R. Beran (1984). Bootstrap methods in statistics. *Jahresbericht der Deutschen Mathematiker-Vereinigung*, 86, 14–30.

P. J. Bickel & D. A. Freedman (1981). Some asymptotic theory for the bootstrap. *Annals of Statistics*, 9, 1196–1217.

H. Chernoff (1964). Estimation of the mode. *Annals of the Institute of Statistical Mathematics*, 16, 31–41.

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

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

P. Groeneboom (1989). Brownian motion with a parabolic drift and Airy functions. *Probability Theory and Related Fields*, 81, 79–109.

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

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

J. Kim & D. Pollard (1990). Cube root asymptotics. *Annals of Statistics*, 18, 191–219.

K. Knight (1989). On the bootstrap of the sample mean in the infinite variance case. *Annals of Statistics*, 17, 1168–1175.

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

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

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

D. N. Politis, J. P. Romano & M. Wolf (2001). *Subsampling.* Springer-Verlag, New York.

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

P. J. Rousseeuw (1984). Least median of squares regression. *Journal of the American Statistical Association*, 79, 871–880.

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

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

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

M. Zarepour & K. Knight (1999). Bootstrapping point processes with some applications, *Stochastic Processes and their Applications*, 84, 81-90.