Statistica Sinica **18**(2008), 967-985

ON THE CHOICE OF m IN THE m OUT OF n BOOTSTRAP AND CONFIDENCE BOUNDS FOR EXTREMA

Peter J. Bickel and Anat Sakov

University of California, Berkeley and Tanor Group

Abstract: For i.i.d. samples of size n, the ordinary bootstrap (Efron (1979)) is known to be consistent in many situations, but it may fail in important examples (Bickel, Götze and van Zwet (1997)). Using bootstrap samples of size m, where $m \to \infty$ and $m/n \to 0$, typically resolves the problem (Bickel et al. (1997), Politis and Romano (1994)). The choice of m is a key issue. In this paper, we consider an adaptive rule, proposed by Bickel, Götze, and van Zwet (personal communication), to pick m. We give general sufficient conditions for first order validity of the rule, and consider its higher order behavior when the ordinary bootstrap fails, and when it works. We then examine the behavior of the rule in the context of setting confidence bounds on high percentiles, such as the asymptotic expected maximum.

Key words and phrases: Adaptive choice, bootstrap, choice of m, data-dependent rule, extrema, m out of n bootstrap.

1. Introduction

The non-parametric bootstrap with sample size n (Efron (1979)) is a powerful inferential tool, but in important situations it can fail (see Mammen (1992) for a review). Bickel et al. (1997), Götze (1993), and Politis and Romano (1994) revived a discussion of resampling smaller bootstrap samples, namely, instead of resampling bootstrap samples of size n (referred to here as the n-bootstrap), take bootstrap samples of size m, where $m \to \infty$ and $m/n \to 0$, with or without replacement. Earlier papers, discussing the idea of smaller sample size, include Bickel and Freedman (1981), Bretagnolle (1983), Swanepoel (1986) and Athreya and Fukuchi (1993).

The choice of m can be crucial, and two issues are involved. The first is that the user does not know, a-priori, whether the bootstrap works or not, in his case. The second is the choice of m, in case of n-bootstrap failure. Earlier papers, discussing the choice of m, include Datta and McCormick (1995), Hall, Horowitz and Jing (1995) and Politis, Romano and Wolf (1999).

In the independent case, Bickel, Götze, and van Zwet (personal communication) proposed a data dependent rule for selecting m in estimating the limit law of statistics. The study of this rule was independently begun by Sakov (1998) and Götze and Račkauskas (2001). Both sets of authors gave conditions for the rule to select m, such that $m/n \to 0$ and $m \to \infty$, when the bootstrap is inconsistent.

Götze and Račkauskas, using delicate analysis of the behavior of U-statistics with kernels growing in n, focused on conditions under which the rule gives "optimal" choices of m, that is, resulting in an approximation of the same order as the one that could be obtained by an oracle, knowing the underlying distribution. They verified their conditions in a number of classical examples, such as pivots based on the minimum of $U(0,\theta)$ variables, and quadratic statistics, both finite and infinite dimensional. Sakov (1998) also focused on quadratic statistics, but under a different set of conditions. Sakov's work was less general, but on the other hand, she showed how the m out of n bootstrap could, in some situations, be modified using extrapolation to give an approximation as good as the best other available approximation, an impossibility even for the oracle (Bickel and Sakov (2002a)).

In this paper, we formulate what we mean by failure of the bootstrap, in terms of convergence of measure-valued random elements. This formulation is more abstract than that of Götze and Račkauskas but, we believe, it clarifies the rationale of the rule. We then state and prove some elementary results on the behavior of the rule when the n-bootstrap is consistent, and when it is not. This is done under conditions that are plausible, but whose validation can be non-trivial, as might be expected from the verifications in Götze and Račkauskas.

Our major result is, in fact, the validation of these conditions for a class of pivotal statistics, based on $\max(X_1,\ldots,X_n)$, for setting a confidence bounds on $F^{-1}(1-1/n)$. This is an example which, except for specific cases such as the uniform distribution on $(0,\theta)$, does not fall within the scope of Götze and Račkauskas.

The paper is organized as follows. In Section 2, we motivate and describe a rule for choosing m. The properties of the rule are discussed in Section 3. For the case that the n-bootstrap is inconsistent, Subsection 3.1 gives rather general criteria under which the chosen m, denoted by $\hat{m}$, behaves properly, i.e., $\hat{m}/n \to 0$ and $\hat{m} \to \infty$. For the case that the n-bootstrap works and is optimal in the sense of Beran (1982), Subsection 3.2, shows that if the Edgeworth expansions exists, then $\hat{m}/n \to 1$, as it should. For the case that n-bootstrap is inconsistent, but Edgeworth or similar expansions are available, Subsection 3.3 shows that $\hat{m}$ not only behaves properly, but in fact gives essentially the best rates possible for estimation of the limiting distribution of the statistics. In some cases, the rule gives the best possible minimax rates for estimation of the limit in the sense given in Bickel et al. (1997). The application to extrema is presented in Section 4. Section 5 presents some simulations supporting the asymptotics. An online supplement to the paper contains technical arguments, and may be found at http://www3.stat.sinica.edu.tw/statistica/.

2. The Rule

Assume $X_1, \ldots, X_n$ are i.i.d. from a distribution $F \in \mathcal{F}$, $X_1 \in R^d$. Let $T_n = T_n(X_1, \ldots, X_n; F)$ be a random variable (rv) with cdf $L_n(x) = P(T_n \leq x)$. We assume a known rate of convergence of T_n to a non-degenerate limiting distribution L, i.e., $L_n \overset{\mathcal{L}}{\Rightarrow} L$. For simplicity, we suppress the dependence of L_n and L on F. The goal is to estimate, or construct a confidence interval for, $\theta_n = \gamma(L_n)$ for some functional γ. The bootstrap estimates L_n, and this estimate, in turn, is plugged into γ to estimate θ_n.

For any positive integer m, let the bootstrap sample, $X_1^*, \ldots, X_m^*$, be a sample drawn from the empirical cdf, $\hat{F}_n$, and denote the m-bootstrap version of T_n by $T_m^* = T_m(X_1^*, \ldots, X_m^*; \hat{F}_n)$, with bootstrap distribution $L_{m,n}^*(x) \equiv P^*(T_m^* \leq x) = P(T_m^* \leq x|\hat{F}_n)$.

We say that the bootstrap 'works', if $L_{m,n}^*$ converges weakly to L in probability for all $m, n \to \infty$ and, in particular, for $m = n$.

When the bootstrap does not 'work' then, under minimal conditions, using a smaller bootstrap sample size rectifies the problem, i.e., although $L_{n,n}^*$ does not have the correct limiting distribution, $L_{m,n}^*$, with 'small', but "not too small", m does (Bickel et al. (1997), Politis and Romano (1994)).

For any $m < n$, bootstrap samples may be drawn with or without replacement. If it is done without replacement (known as subsampling), then $L_{m,n}^* \overset{\mathcal{L}}{\Rightarrow} L$ under minimal conditions, as shown in Politis and Romano (1994) and Götze (1993). If the resampling is done with replacement (what we call the m-bootstrap), then this limit holds if, in addition to the minimal conditions, T_m is not affected much by the order of $\sqrt{m}$ ties (Bickel et al. (1997)). Thus, subsampling is more general than the m-bootstrap since fewer assumptions are required. However, the m-bootstrap has the advantage that it allows for the choice of $m = n$. In particular, if the n-bootstrap works and is known to be second order correct for some pivotal roots, the selection rule for m includes the particular case $m/n \to 1$ (Section 3.2). In that case, unlike subsampling, the m-bootstrap enjoys the second order properties of the n-bootstrap. Since in all situations of interest, so far, the conditions for consistency of the m-bootstrap are satisfied, we consider only the sampling with replacement case.

Using a smaller bootstrap sample requires a choice of m. To motivate the rule, we use the following example (Bickel et al. (1997), Sakov (1998)). Let X_i be i.i.d. with mean μ, and variance σ^2. Consider the null hypothesis that $\mu = 0$ with the usual test statistic $T_n = \sqrt{n}\bar{X}_n$ (the subscript on $\bar{X}_n$ indicates the sample size). An alternative for using the usual normal quantile in setting a critical value is to use the bootstrap distribution of $\sqrt{n}(\bar{X}_n^* - \bar{X}_n)$, i.e., resample the residuals $X_i - \bar{X}_n$. This is a special case of resampling the residuals in regression through the origin (Freedman (1981)). Suppose that, instead, we use the bootstrap distribution of $T_n^* = \sqrt{n}\bar{X}_n^*$, the direct, and naive, bootstrap version of T_n. Then, using a quantile of this distribution is inconsistent, as we discuss below. Consider the bootstrap distribution of $T_m^* = \sqrt{m}\bar{X}_m^*$. When m is fixed, say $m = k$, the bootstrap distribution is

$$L_{k,n}^*(x) = \frac{1}{n^k}\sum_{i_1=1}^{n}\cdots\sum_{i_k=1}^{n} 1\left(\sqrt{k}\frac{\sum_{l=1}^{k} X_{i_l}}{k} \leq x\right),$$

where $1(A)$ denotes the indicator function of A. $L_{k,n}^*$ is a function of the data only, and is a V-statistic. As $n \to \infty$, $L_{k,n}^* \to L_k$, and the limit depends on k (Serfling (1980)). The only distribution for which the limit is the same for all k is the normal distribution (Kagan, Linnik and Rao (1973)), and this case is discussed in Sakov (1998).

On the other hand, when $m, n \to \infty$, $\sqrt{m}(\bar{X}_m^* - \bar{X}_n) \stackrel{\mathcal{L}}{\Rightarrow} N(0, \sigma^2)$ (Bickel and Freedman (1981)). Hence, $\sqrt{m}\bar{X}_m^*$ behaves like $N(\sqrt{m}\bar{X}_n, \sigma^2)$. As $m/n \to \lambda \geq 0$, $\sqrt{m}\bar{X}_n = \sqrt{m/n}\sqrt{n}\bar{X}_n \stackrel{\mathcal{L}}{\Rightarrow} N(0, \lambda\sigma^2)$. Thus, the limiting distribution of $\sqrt{m}\bar{X}_m^*$ can be thought of as a random probability distribution, with randomness coming from the marginal distribution of the original sample. Denote the limit by $\mathcal{L}_\lambda \equiv N(\sqrt{\lambda}\sigma Z, \sigma^2)$, where $Z \sim N(0, 1)$. The limit is degenerate and equal to the desired $N(0, \sigma^2)$ iff $\lambda = 0$ (i.e., when $m/n \to 0$). For $\lambda > 0$, the limit is non-degenerate and depends on λ. Furthermore, $\lambda \mapsto \mathcal{L}_\lambda$ is $1 - 1$.

Loosely speaking, it follows that when m is in the "right range" of values, the bootstrap distributions for different possible samples are 'close' to each other; when m is "too large" or fixed, the bootstrap distributions (or processes) are different. This suggests looking at a sequence of values of m and their corresponding bootstrap distributions. A measure of discrepancy, between these distributions, should show large discrepancies when m is "too large" or fixed. The discrepancies should be small when m is of the "right order".

In essentially all examples considered so far, the failure of the n-bootstrap is of the following type: $L_{n,n}^*$, viewed as a probability distribution on the space of all probability distributions, does not converge to a point mass at the correct limit L, but rather converges (in a sense to be made precise in Appendix A) to a nondegenerate distribution, call it $\mathcal{L}_1$, on that space. If $m \to \infty$, $m/n \to \lambda$, $0 < \lambda < 1$, one gets convergence to a non-degenerate distribution, $\mathcal{L}_\lambda$, which is typically different from $\mathcal{L}_1$. We expect that $\mathcal{L}_0 = L$. This behavior suggests the following adaptive rule for choosing m.

1. Consider a sequence of m's of the form

$$m_j = \left[q^j n\right], \qquad \text{for } \ j = 0, 1, 2, \ldots, \qquad 0 < q < 1, \tag{1}$$

   where $[\alpha]$ denotes the smallest integer $\geq \alpha$.
2. For each m_j, find $L^*_{m_j,n}$ (in practice this is done by Monte-Carlo).
3. Let ρ be some metric consistent with convergence in law, and set

$$\hat{m} = \underset{m_j}{\operatorname{argmin}}\, \rho\Big(L^*_{m_j,n}, L^*_{m_{j+1},n}\Big).$$

   If the difference is minimized for a few values of m_j, then pick the largest among them. Denote the j corresponding to $\hat{m}$ by $\hat{\jmath}$.
4. The estimator of L is $\hat{L} = L^*_{\hat{m},n}$. Estimate θ by $\hat{\theta}_n = \gamma(\hat{L})$, or use the quantiles of $\hat{L}$ to construct confidence interval for θ.

Our discussion, and proofs, are for the case $\rho(F, G) = \sup_x |F(x) - G(X)|$, the Kolmogorov sup distance (KS). Götze and Račkauskas (2001) consider more general metrics of the form $\rho(P, Q) = \sup\{|P(h) - Q(h)| : h \in \mathcal{H}\}$, where $\mathcal{H}$ is a Donsker class of functions and $P(h) \equiv E_P(h(X))$. The results of Sections 3.1 and 3.2 are valid for this generalization also, but are not formally pursued since simulations using other ρ, such as the Wasserstein metrics, for our application to extrema did not show substantial differences. However, Götze and Račkauskas (2001) obtained better results for $T_n = \sqrt{n}\bar{X}_n$. One of us, Sakov (1998), successfully used metrics based on quantiles of P and Q in applications to the bump test for multimodality (Silverman (1981)).

Throughout the paper we assume the data are i.i.d. The idea of subsampling can be extended to time series. However, in order to maintain the dependence structure, the time series is first divided into blocks of size b, and then, instead of sampling individual observations, blocks are sampled. The block size, b, should satisfies the same conditions as m, i.e., $b \to \infty$ and $b/n \to 0$. This suggests that a similar data-dependent rule to choose b may work, although we have not pursued this. For more information see Politis et al. (1999).

3. General Behavior of the Rule

3.1. Order of $\hat{m}$ when the n-bootstrap is inconsistent

Recall that T_n is the rv of interest with exact cdf L_n, and T^*_m is its bootstrap version with bootstrap distribution $L^*_{m,n}$. For a given m, the bootstrap distribution is a stochastic process whose distribution depends on $\hat{F}_n$. To study the behavior of such objects carefully, we introduce the following framework. On a single large probability space $(\Omega, \mathcal{A}, P)$ we suppose that we can define:

(a) $X_1, \ldots, X_n, \ldots$ i.i.d. F on R^d;

(b) X^*_{jk} $j \geq 1$, $k \geq 1$ such that the conditional distribution of $X^*_{1n}, X^*_{2n}, \ldots$, given $(X_1, \ldots, X_n)$, is that of i.i.d. variables with common distribution $\hat{F}_n$.

We represent the laws of rv by their distribution functions, viewed as elements of the Skorokhod space $D(\bar{R})$, where $\bar{R} = [-\infty, +\infty]$. Thus, $L^*_{m,n}$ is a measurable map from Ω to $D(\bar{R})$. In Appendix A, we define what we mean when saying that $L^*_{m,n}$ *converges in law to* (a *random*) L *in probability.*

If m is fixed, say $m = k$, and T_n does not depend on F, then

$$V_k \equiv L^*_{k,n}(x) = \frac{1}{n^k} \sum_{i_1=1}^{n} \cdots \sum_{i_k=1}^{n} 1\Big(T_k(X_{i_1}, \ldots, X_{i_k}; F) \leq x\Big).$$

V_k is a V-statistic whose kernel has finite moments of all orders, and hence a limiting distribution that is the expected value of the kernel (Serfling (1980)). In general, it is reasonable to expect that $T_k(X_1, \ldots X_k; \hat{F}_n)$ will behave like $T_k(X_1, \ldots X_k; F)$, where F can now be treated as fixed. In Theorem 1, we build this into the assumptions.

For $m \to \infty$ and $m/n \to \lambda$ $(0 < \lambda \leq 1)$, define

$$U_n(\lambda) = \begin{cases} L^*_{[n\lambda]+1,n}, & 0 < \lambda \leq 1 - \frac{1}{n} \\ L^*_{n,n}, & 1 - \frac{1}{n} < \lambda \leq 1 \end{cases}, \tag{2}$$

which can be viewed as a stochastic process $U_n : (0, 1] \times \Omega \to D(\bar{R})$. With these notations, here are our assumptions.

(A.1) If $m = k$ fixed, then $L^*_{k,n} \overset{\mathcal{L}}{\Rightarrow} L_k$ as $n \to \infty$, where $L_k(x) = P(T_k(X_1, \ldots, X_k; F) \leq x)$.

(A.2) $L_n \overset{\mathcal{L}}{\Rightarrow} L$ as $n \to \infty$, where L is a continuous cdf. Viewed as a random distribution function, L is fixed with probability 1 and belongs to $C(\bar{R})$, the continuous functions on $\bar{R}$.

(A.3) For $m \to \infty$ and $m/n \to 0$, $L^*_{m,n} \overset{\mathcal{L}}{\Rightarrow} L$ in probability (see Bickel et al. (1997) for conditions, and Appendix A for meaning).

(A.4) For L_k defined in assumption (A.1), the map $k \mapsto L_k$ is $1 - 1$.

(A.5) For all $(\lambda_1, \ldots, \lambda_l)$, where $l \geq 1$, $(U_n(\lambda_1), \ldots, U_n(\lambda_l)) \overset{\mathcal{L}}{\Rightarrow} (U(\lambda_1), \ldots, U(\lambda_l))$ in probability, where $U : [0, 1) \times \Omega \to C(\bar{R})$, U is continuous from the right at $\lambda = 0$, and $U_n(\lambda_n) \overset{\mathcal{L}}{\Rightarrow} U(0+)$ as $\lambda_n \to 0$. That is, if ρ is the product Skorokhod metric on $D(\bar{R}) \times \cdots \times D(\bar{R})$, then for suitable $\tilde{U}_n, U$ (as in (a), (a') of Appendix A)

$$\rho\bigg(\Big(\tilde{U}_n(\lambda_1), \ldots, \tilde{U}_n(\lambda_l)\Big), \Big(U(\lambda_1), \ldots, U(\lambda_l)\Big)\bigg) \overset{p}{\to} 0.$$

(A.6) $P(\lambda \mapsto U(\lambda) \text{ is } 1 - 1) = 1$ for λ on $[0, 1]$.

If assumptions (A.5) and (A.6) hold, then the n-bootstrap is inconsistent. To see that, note that by assumption (A.5), $U(0) = L$ with probability 1. However, if the n-bootstrap works then $U(0) = U(1) = L$, which contradicts (A.6).

Theorem 1. *Let $\hat{m}$ be the m chosen by the rule. If assumptions* (A.1)−(A.6) *hold, then $\hat{m} \xrightarrow{p} \infty$, and $\hat{m}/n \xrightarrow{p} 0$.*

The proof is given in Appendix 2.1 of the online supplement.

3.2. Choice of m when the n-bootstrap works

This section is motivated by examples like the t statistic, and $\sqrt{n}(\bar{X} - \mu)$, for which the n-bootstrap works (Bickel and Freedman (1981), Singh (1981)). In such cases, using bootstrap samples of size $m < n$, in general, causes a loss in efficiency (Bickel et al. (1997)). The conditions under which the rule picks m, such that there is no loss in efficiency, involve Edgeworth expansions. Suppose that

$$L_n(x) = P(T_n \leq x) = A_0(x; F) + n^{-\frac{1}{2}} A_1(x; F) + o\Big(n^{-\frac{1}{2}}\Big), \tag{3}$$

$$L^*_{m,n}(x) = P^*(T^*_m \leq x) = A_0(x; \hat{F}_n) + m^{-\frac{1}{2}} A_1(x; \hat{F}_n) + o_p\Big(m^{-\frac{1}{2}}\Big). \tag{4}$$

We write $X_n = \Omega_p(Y_n)$ iff $X_n = O_p(Y_n)$, and $Y_n = O_p(X_n)$. Assume that

$$\|A_0(\cdot; \hat{F}_n) - A_0(\cdot; F)\| = \Omega_p\Big(n^{-\frac{1}{2}}\Big) \tag{5}$$

$$\|A_1(\cdot; \hat{F}_n) - A_1(\cdot; F)\| = o_p(1), \tag{6}$$

where $A_1(F) \neq 0$, and $\| \cdot \|$ is the norm used in our rule. Define

$$\kappa^*_n(F) = \inf_j \|L^*_{m_j,n}(\cdot) - L_n(\cdot)\|, \quad \hat{\kappa}^*_n(F) = \|L^*_{\hat{m},n}(\cdot) - L_n(\cdot)\|. \tag{7}$$

Our focus, in this subsection, is on the behavior of m when the n-bootstrap is consistent (which follows from equations (3)−(5)).

Theorem 2.

(a) *If assumptions* (A.1)−(A.4), *and* (3)−(6) *are fulfilled, then*

$$\hat{\kappa}^*_n(F) = \Omega_p\Big(n^{-\frac{1}{2}}\Big) \quad \text{and} \quad \kappa^*_n(F) = \Omega_p\Big(n^{-\frac{1}{2}}\Big). \tag{8}$$

As a consequence, $\hat{\kappa}^*_n(F)/\kappa^*_n(F) = \Omega_p(1)$, for all $F \in \mathcal{F}$.

(b) *Further assume that $A_0(\cdot; F)$ does not depend on F, that $o_p(1)$ in* (6) *is $\Omega_p(n^{-1/2})$, and that $o_p(n^{-1/2})$ in* (3) *is $O_p(n^{-1})$. Then,*

$$\frac{\kappa^*_n(F)}{\hat{\kappa}^*_n(F)} \xrightarrow{p} 1 \quad \text{and} \quad \hat{\kappa}^*_n(F) = O_p(n^{-1}). \tag{9}$$

Part (a) of the theorem corresponds to statistics like $\sqrt{n}(\bar{X}-\mu)$ where the m-bootstrap, with m chosen by the rule, behaves equivalently to the n-bootstrap, and both commit errors of order $n^{-1/2}$. Part (b) corresponds to statistics like the t, where the n-bootstrap produces coverage probabilities which differ by $O(n^{-1})$ from the nominal ones, and our conclusion (for $\|\cdot\|_\infty$) is that the $\hat{m}$-bootstrap achieves the same performance.

The proof is given in Appendix 2.2 of the online supplement.

Corollar 1. *Under the conditions of Theorem* 2, $\hat{m}/n \xrightarrow{p} 1$.

3.3. Optimality of $\hat{m}$ when the n-bootstrap is inconsistent

When the n-bootstrap is inconsistent, it is natural to ask the following.

(i) Does $\hat{m}$ achieve the optimal rate for the m-bootstrap? Using (7), the question becomes: do we have $\hat{\kappa}_n^*(F)/\kappa_n^*(F) = \Omega_p(1)$ for all $F \in \mathcal{F}$?

(ii) If so, does $\hat{m}$ also achieve the minimax rate for estimating L_n, for a suitable norm? This question can be framed as follows. Assume $\rho_n = \inf_\delta \max_{\mathcal{F}}(E_F(\|\delta(X_1,\ldots,X_n) - L_n\|)) \to 0$, where δ ranges over all possible estimates of L_n. Then, does $\mathrm{E}_F(\|L_{\hat{m}}^*(\cdot) - L_n(\cdot)\|)/\rho_n = \Omega(1)$?

Götze (1993) addresses the first question, for a class of smooth functionals $T(\hat{F}_n)$ and certain norms, obtaining an affirmative answer. We briefly address this question here. Both questions will be addressed in Section 4.2, in the context of extrema.

The basic requirement for addressing (i) is the existence of an Edgeworth type expansion for L_n and $L_{m,n}^*$. Assume (3) holds, $A_0(x;\hat{F}_n) = A_0(x;F) + O_p(n^{-\gamma})$, and $A_1(x;\hat{F}_n) = A_1(x;F) + \Omega_p(1)$. The expansion for $L_{m,n}^*$ is assumed to satisfy

$$L_{m,n}^*(x) = A_0(x;\hat{F}_n) + m^{-\frac{1}{2}} A_1(x;\hat{F}_n) + \Omega_p(m^\beta n^{-\gamma}) + o_p\Big(m^{-\frac{1}{2}} + m^\beta n^{-\gamma}\Big). \quad (10)$$

The difference between (10) and (4) is the term of order $m^\beta n^{-\gamma}$. If $\beta \geq \gamma$, then the n-bootstrap is inconsistent. Here is an example where such a term arises. The Edgeworth expansion of $\sqrt{n}\bar{X}_n$, if $\mu = 0$, has the form (3), but the bootstrap distribution of $\sqrt{m}\bar{X}_m^*$ has an expansion with $\beta = \gamma = 1/2$. This term arises from the difference $\Phi(x) - \Phi(-\sqrt{m}\bar{X}_n + x)$, and is the bias of the bootstrap distribution; it is of order $\Omega_p(\sqrt{m}\bar{X}_n) = \Omega_p(\sqrt{m/n})$.

Theorem 3.

1. *Assume* (3) *and* (10) *hold and* $\beta \geq \gamma$, *then*

$$\inf_m \|L_{m,n}^*(\cdot) - L_n(\cdot)\| = \Omega_p\Big(n^{-\frac{1}{2}\gamma(\frac{1}{2}+\beta)^{-1}}\Big). \quad (11)$$

2. *If* $\hat{m}$ *is chosen by the rule, then*

$$\|L_{\hat{m},n}^*(\cdot) - L_n(\cdot)\| = \Omega_p\Big(n^{-\frac{1}{2}\gamma(\frac{1}{2}+\beta)^{-1}}\Big). \quad (12)$$

3. *If* m_{opt} *is the* m *which minimizes* (11), *then* $m_{opt}/\hat{m} = \Omega_p(1)$.

The proof is given in Appendix 2.3 of the online supplement.

4. Confidence Bounds for Extrema

Our main example is setting confidence bounds for extrema. Assume $X_1, \ldots,$ X_n is an i.i.d. sample from a distribution F, with density f in the domain of attraction of G for the maximum. That is, if $X_{(n)} = \max(X_1, \ldots, X_n)$, there exist normalizing constants $a_n > 0$, and $b_n \in R$ (depending on F) such that,

$$P(a_n(X_{(n)} - b_n) \leq x) = F^n\Big(\frac{x}{a_n} + b_n\Big) \to G(x), \tag{13}$$

for all x in the support of G. In that case, G must be one of the three types (for example, David (1981) and Reiss (1989)):

$$\begin{aligned}
&G(x) = \exp(-x^{-\gamma}), && x \geq 0, \quad \gamma > 0 && \text{(I)},\\
&G(x) = \exp(-(-x)^{\gamma}), && x \leq 0, \quad \gamma > 0 && \text{(II)},\\
&G(x) = \exp(-e^{-x}) && && \text{(III)}.
\end{aligned}$$

By type of G, we refer to the location-scale family generated by G. Denote by $\omega(F)$ the supremum of the support of F. From (5.1.8)–(5.1.11) and P.5.3 in Reiss (1989), it follows that a_n and b_n can be chosen to be:

$$a_n^{-1}(F) = \begin{cases} b_n, & \text{for type I} \\ \omega(F) - b_n, & \text{for type II,} \\ (nf(b_n))^{-1}, & \text{for type III} \end{cases} \quad b_n(F) = F^{-1}\Big(1 - \tfrac{1}{n}\Big). \tag{14}$$

von Mises gave simple sufficient conditions on F to belong to the domain of attraction of G. We consider a slightly specialized form of these conditions (Reiss (1989, p.159)): If $F'' = f'$ exists, then for $\gamma > 0$,

$$\lim_{x \uparrow \omega(F)} \left[\frac{1-F}{f}\right]'(x) = \begin{cases} \gamma^{-1}, & \text{(I)} \\ -\gamma^{-1}, & \text{(II)} \\ 0, & \text{(III)} \end{cases}. \tag{15}$$

We refer to these conditions as (vM)(I)$-$(III), respectively. The original von Mises conditions (which are more tedious to work with, and harder to check directly) are implied by (15). We prove our results under the specialized form, since all common examples satisfy them.

Our goal is to set an upper confidence bound on b_n defined in (14). If G and a_n are known, we could use $X_{(n)} - a_n^{-1}G^{-1}(\alpha)$. If G is unknown, a natural alternative is to bootstrap the distribution of $a_n(X_{(n)} - F^{-1}(1-1/n))$, and replace

$G^{-1}(\alpha)$ by the α-th quantile of the bootstrap distribution of $a_n(X^*_{(n)} - \hat{F}_n^{-1}(1 - 1/n))$. Unfortunately, the n-bootstrap fails (Athreya and Fukuchi (1993)). Here is the m-bootstrap derivation. Denote by $X_{(k,n)}$ the k-th order statistic of a sample of size n, and by $X^*_{(k,m)}$ the k-th order statistic of a bootstrap sample of size m. If we apply our paradigm to

$$T_n = a_n\Big(X_{(n,n)} - F^{-1}\Big(1 - \frac{1}{n}\Big)\Big), \tag{16}$$

we are led to consider the bootstrap distribution of

$$T^*_m = a_m\Big(X^*_{(m,m)} - \hat{F}_n^{-1}\Big(1 - \frac{1}{m}\Big)\Big) = a_m\Big(X^*_{(m,m)} - X_{([n-\frac{n}{m}],n)}\Big), \tag{17}$$

where $m \to \infty$, and $m/n \to 0$. Denote the α-th quantile of this distribution by $\hat{G}_m^{*-1}(\alpha)$, and use $X_{(n)} - a_n^{-1}\hat{G}_m^{*-1}(\alpha)$ as an upper confidence bound for $b_n(F)$. We propose to use $\hat{m}$ as the choice of m. The success of this strategy, in the sense of Theorem 1, is discussed in the following section.

4.1. Order of $\hat{m}$

Theorem 4. *Suppose F satisfies* (15), *and the search for $\hat{m}$ is restricted to j such that $m_j = [q^j n]$ and $[n/m_j]$ are distinct integers, as j varies. Then $L^*_{\hat{m},n}(x) \to G(x)$, where G is the limit law of T_n (convergence as defined in Appendix* A).

Proof. We show that assumptions (A.1)−(A.6) of Section 3.1 hold. The theorem then follows from Theorem 1.

Condition (A.1) is immediate by continuity of $x \to T^*_m$.

Condition (A.2) is the classic result of von Mises (Reiss (1989)).

Condition (A.3): This condition is established if, for $m \to \infty$ and $m/n \to 0$, the bootstrap distribution of T^*_m is shown to converge weakly in probability to the same limiting distribution as T_n. Write,

$$T^*_m = a_m(X^*_{(m,m)} - b_m) - a_m\Big(X_{([n-\frac{n}{m}],n)} - b_m\Big).$$

Athreya and Fukuchi (1993) showed that $a_m(X^*_{(m,m)} - b_m)$ converges weakly to the desired limit. In Lemma 1 (Appendix 3.1 in the online supplement), we show that the right hand side is $o_p(1)$, and (A.3) follows.

Condition (A.4): Bickel and Sakov (2002b) showed that if F satisfies (vM)(I)−(III), then (A.4) is valid unless F is an extreme value distribution, in which case $L_1(F) = L_2(F) = \cdots$. The condition follows from this result. In the exceptional case that F is itself one of the three types, we need only show $\hat{m}/n \xrightarrow{p} 0$, which requires conditions (A.5) and (A.6), since $\hat{m}$ staying bounded poses no problems.

Condition (A.5): The proof of this condition uses the Poisson approximation to the Binomial distribution, and is given in Appendix 3.2 of the online supplement.

Condition (A.6) follows from Lemma 5 (Appendix 3.3 of the online supplement).

4.2. Optimality of $\hat{m}$

Here is a class of situations where (3) and (10) hold for $a_n(X_{(n,n)} - b_n)$. Suppose F satisfies the conditions of Theorem 5.2.11 (p. 176) of Reiss (1989). Then (3) holds with $\beta = 1/2$, and α specified in Reiss (the power in (3) is now α rather than 1/2) for the Hellinger norm and, hence, also the smaller L_∞ norm. To see that (10) holds, consider $m = \Omega(n^r)$ for some $0 < r < 1$. Set, $C(\epsilon) = \{x : F(x/a_m + b_m) \geq \epsilon\}$ for ϵ arbitrarily small, and $x^{(m)} = x/a_m + b_m$. Then,

$$\begin{aligned}
&\|L^*_{m,n}(\cdot) - L_m(\cdot)\| \\
&\quad = \sup_x \left\{\left|\hat{F}^m_n(x^{(m)}) - F^m(x^{(m)})\right| : x \in C(\epsilon)\right\} + o\Big(mn^{-\frac{1}{2}}\Big) = \Omega_p\Big(mn^{-\frac{1}{2}}\Big). \qquad (18)
\end{aligned}$$

The argument for (18) is as follows. First,

$$\begin{aligned}
&\sup_x \left\{m \cdot \left|\log(\hat{F}_n(x^{(m)})) - \log(F(x^{(m)}))\right| : x \in C(\epsilon)\right\} \\
&\quad = \Omega_p\left(m \sup_x \left\{|\hat{F}_n(x) - F(x)| : x \in C(\epsilon)\right\}\right) = \Omega_p\Big(mn^{-\frac{1}{2}}\Big),
\end{aligned}$$

and $\sup\{|\hat{F}^m_n(x^{(m)}) - F^m(x^{(m)})| : x \notin C(\epsilon)\} = o_p(\epsilon^m) = o_p(mn^{-1/2})$.

From (18) and (3):

$$\begin{aligned}
L^*_{m,n}(x) - L_n(x) &= (L^*_{m,n}(x) - L_m(x)) + (L_m(x) - L_n(x)) \\
&= A_1(x; F)(m^{-\alpha} - n^{-\alpha}) + o_p(m^{-\alpha}) + \Omega_p\Big(mn^{-\frac{1}{2}}\Big) \\
&= O_p(m^{-\alpha}) + \Omega_p\Big(mn^{-\frac{1}{2}}\Big).
\end{aligned}$$

Hence (10) holds, with $\beta = 1$, $\gamma = 1/2$, and α as above, for $\|\cdot\| = \|\cdot\|_\infty$. Götze and Račkauskas (2001) note this result for the standard uniform distribution.

The answer to question (ii) of Section 3.3 is more complicated. Here is what follows in a special case. Suppose we are dealing with a type II limit with $\omega(F) < \infty$ unknown. Suppose that F has support on $[0, 1]$, and is such that

$0 < f(1-) < \infty$ and $|f'(1-)| \leq M < \infty$. Then it is well known that the minimax rate for estimating $f(1)$, using root mean square error, is $n^{-1/3}$. However, since $\beta = 1$, $\gamma = 1/2$, and $\alpha = 1$, the m-bootstrap, with $\hat{m}$ chosen as above, can only yield a rate of $n^{-1/4}$.

This is a consequence of using $\|\cdot\|_\infty$ as our measure of departure. If, instead, we use the norm $\|g\| \equiv \sup\{|g(x)| : |x| \leq M\}$ for $M < \infty$, it may be shown that we obtain a rate of $n^{-1/3}$ for $\|L_{\tilde{m},n}(\cdot) - L_n(\cdot)\|$, where $\tilde{m}$ is selected using this norm. This implies that a better estimate of L_n, even for the original $\|\cdot\|_\infty$, is to estimate $f(1-)$ using $L_{\tilde{m},n}$, and then plug the resulting estimate into the limiting exponential distribution.

From a statistical point of view, there is an unfortunate general conclusion. We can design a pivot $a_n(\hat{F}_n)(X_{(n,n)} - F^{-1}(1 - 1/n))$ that has a limiting distribution, independent of F, in the domain of attraction of a particular type of extremal law. However, unlike what happens with pivots such as the t statistic, the $\hat{m}$-bootstrap distribution of this pivot is not, theoretically, a better approximation to the law of the pivot than the limit is.

4.3. Comments

Throughout the paper we assume that the rate of convergence is known. Consider the case when it is unknown, but is of the form $a_n \sim cn^\alpha$ for some $\alpha \geq 0$, $c > 0$. Bertail, Politis and Romano (1999) and Politis et al. (1999) estimate α using the m out of n bootstrap, and we applied it to the current problem of extrema (as a preliminary stage, before applying our rule; see Section 5). In the context of extrema, this form covers many important cases, e.g., the exponential, the uniform, the generalized Pareto, etc. However, it does rule out cases like the Gaussian, where $a_n = \sqrt{2\log n}$. In principle, the approach can be extended to such a functional form for a_n, but in practice this is difficult (Bertail et al. (2004)).

Subsampling, or the m out of n bootstrap without replacement, is also an option here (Politis et al. (1999)). For the extrema problem we expect that typically $\hat{m}/\sqrt{n} \to 0$, in which case sampling with and without replacement are the same with probability tending to 1.

5. Simulations

5.1. Choice of m

We begin by illustrating, qualitatively, that $\hat{m}$ behaves as we expect. Recall that $\hat{m}$ minimizes the *successive differences* $\hat{\Delta}_j = \|L^*_{m_j,n}(\cdot) - L^*_{m_{j+1},n}(\cdot)\|$. Ideally, we would like to use m which minimizes the *exact differences* $\|L^*_{m_j,n}(\cdot) - L_n(\cdot)\|$. In the simulation, we replace L_n by its limit, L, and plot $\Delta_j = \|L^*_{m_j,n}(\cdot) - L(\cdot)\|$, as well as $\hat{\Delta}_j$, as functions of m_j. Three examples of n-bootstrap failure, and one example when it works, are shown in Figure 1:

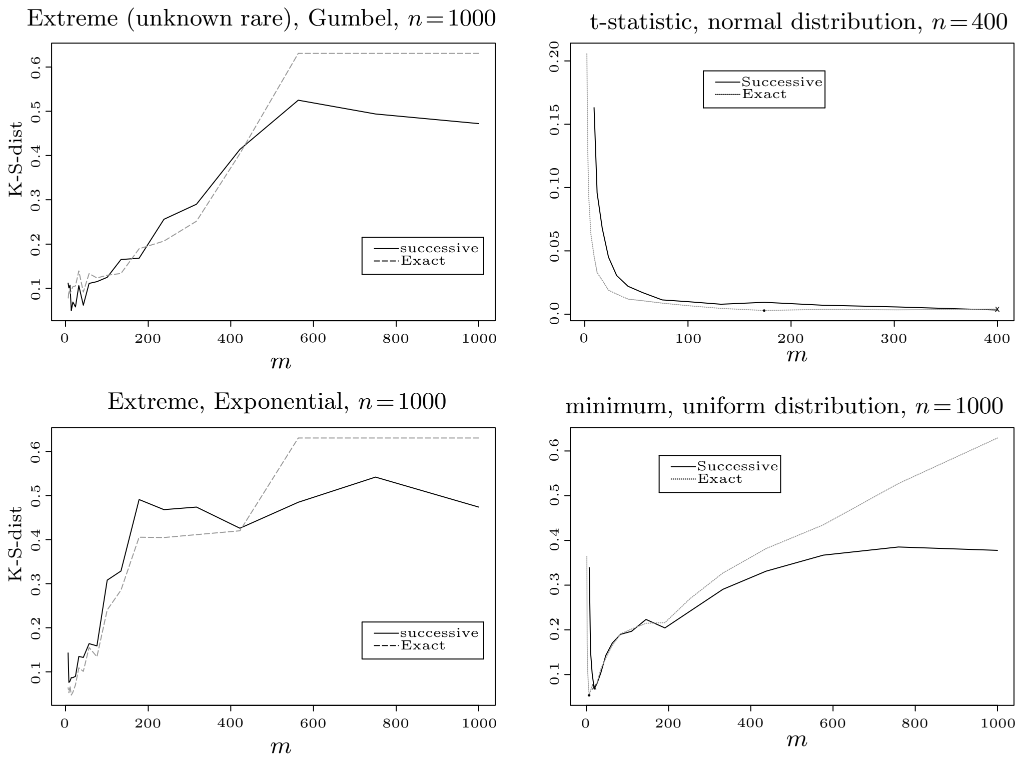

Figure 1. Choice of m using the rule.

1. The extrema problem when F is exponential and a_n is known.
2. The extrema problem when F is the Gumbel distribution. Here a_n is unknown; it is estimated using the approach of Bertail et al. (1999) (for details see Section 5.2).
3. If F is the Uniform$(0, \theta)$ distribution, the n-bootstrap fails to estimate the distribution of the maximum (Bickel and Freedman (1981)).
4. For the t-statistic, the n-bootstrap works (Bickel and Freedman (1981), Singh (1981)).

The sample size is indicated above each plot. Here and later, the number of bootstrap samples is 1,000, and $q = 0.75$.

As can be seen, there are two ('typical') types of curves: when the n-bootstrap fails, and when it works. When it fails, the successive differences go down drastically and then increase. On the other hand, when it works, the differences drop down and then remain pretty constant. This reflects the fact that, when the n-bootstrap works, once m is large enough there will not be large differences, to first order, between bootstrap distributions at increasing sample sizes. Thus, the rule also provides a diagnostic for n-bootstrap failure. Note that the two curves, in each plot, are not always close to each other, but both achieve their minimum for about the same m, as our analysis in Section 3.3 concludes. For more plots see Sakov (1998) and Götze and Račkauskas (2001).

5.2. Coverage when m is chosen by the rule

We evaluate the coverage of a confidence bound of a parameter, by constructing 1,000 bounds, and checking how many of them cover the true parameter. The desired level is 95%.

Extrema - known rate

In Table 1, an upper bound for $b_n = F^{-1}(1 - 1/n)$ is shown, following the approach of Section 4. Four distributions were considered: exponential, normal and Gumbel (domain of attraction is type III), and uniform (domain of attraction is type II). The a_n are taken from Embrechts, Klüppelberg and Mikosch (1997). For all but the exponential distribution, $F^{-1}(1 - 1/n)$ is different from the b_n given in Embrechts et al. (1997). However, this is only a location change of the limiting distribution. The coverage, mean, SD, and median of the chosen m over the 1,000 repetitions are given for sample sizes of 500 and 1,000. The coverages, for sample size 10,000, were 0.935, 0.926, 0.933 and 0.945, for the exponential, normal, Gumbel, and uniform, respectively.

Extrema - unknown rate

Table 2 is similar, but when a_n is unknown and of the form $a_n = n^\alpha$, for some α. In each repetition, α is estimated using the approach of Bertail et al. (1999), and then a_n and a_m in (16) and (17), are replaced by $n^{\hat{\alpha}}$ and $m^{\hat{\alpha}}$, respectively. To estimate α, the method which is based on ranges is used (see Bertail et al. (1999) for details. We used $I = 15$, $J = 50$, and quantiles between $(0.75, 0.95)$ and $(0.05, 0.25)$). Note that a_n, for the normal distribution, does not have the desired form. The mean and SD of the α's over 1,000 repetitions are given in Table 2.

Table 1. Coverage for b_n, known a_n

	Distribution	Exponential		Normal		Gumbel		Uniform	
	n	500	1000	500	1000	500	1000	500	1000
	Coverage	0.92	0.93	0.92	0.94	0.92	0.93	0.92	0.92
	$\Delta_n(\bar{b}_n)$	0.24	0.22	0.16	0.14	0.22	0.21	0.005	0.002
$\hat{m}$	Mean	17	19	27	32	17	17	25	22
	SD	11	12	17	21	10	9	48	30
	Median	15	17	21	31	15	13	15	17

Table 2. Coverage for b_n, unknown a_n

	Distribution	Exponential			Normal			Gumbel		
	n	500	1000	10,000	500	1000	10,000	500	1000	10,000
	Coverage	0.84	0.85	0.88	0.82	0.86	0.88	0.8	0.84	0.88
$\hat{m}$	Mean	17	18	40	17	19	40	16	17	33
	SD	10	12	28	10	12	31	9	11	26
	Median	16	14	32	16	14	32	12	14	24
$\hat{\alpha}$	Mean	-0.02	-0.01	0	0.18	0.16	0.13	0.02	0.01	0
	SD	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1

	Distribution	Uniform		
	n	500	1000	10,000
	Coverage	0.89	0.92	0.92
$\hat{m}$	Mean	17	19	39
	SD	12	12	28
	Median	16	18	32
$\hat{\alpha}$	Mean	0.94	0.94	0.98
	SD	0.1	0.1	0.1

Precision of the upper bound

We briefly explore precision when a_n is known, for the cases in Table 1. Our precision measure is $\Delta_n(\bar{b}_n) \equiv (\mathrm{Med}(\bar{b}_n) - b_n)/b_n$.

The exponential and the normal, with sample size 500, may be roughly compared to the results in Breiman, Stone and Kooperberg (1990). The figures are of the same order of magnitude. Breiman et al. (1990) has the only previous careful analysis of a nonparametric method for setting upper confidence bounds, to the best of our knowledge. Their method involves the fit of a power transformation family, followed by parametric fitting of an extreme value family to the tail of the data. There are two tunable parameters: the fraction of data used in estimating the tail parameter, and a parameter called tail heaviness. They carried out extensive simulations and, to some extent, tuned the parameter to these. For $n = 500$, they obtained Δ_n ranging roughly from 0.05 to 0.1 for the normal, and between 0.2 to 0.4 for the exponential. Unfortunately, they considered a coverage of 0.9, while we considered a coverage of 0.95. However, actual coverage is about the same. Moreover, in the normal case they are estimating the 0.999 percentile, while we estimate the 0.998 percentile. In one respect their task is easier than ours, estimating at the 90% rather than 95% level, where we expect Δ_n to be larger. On the other hand, for the Gaussian, their task is a bit harder because of the higher percentile they are estimating. That the results are qualitatively similar is encouraging. As a final comment on this comparison, their method has one significant advantage over our present method. By specifying parametric tails they are able to get bounds for extreme $F^{-1}(1 - c/n)$,

Table 3. Three other cases

	Case	μ, known SD				μ, unknown SD			θ		
	Distribution	Exponential		Normal		Normal			Uniform		
	n	100	500	100	500	100	500	1000	100	500	1000
	Coverage	0.922	0.93	0.953	0.961	0.952	0.951	0.947	0.912	0.926	0.936
	Mean	54	195	37	132	58	188	315	10	14	15
$\hat{m}$	SD	29	156	29	149	27	154	298	6	8	8
	Median	57	159	24	67	57	119	178	8	12	11

for $c < 1$. Our method could be coupled with extrapolation, as was done in Bickel and Sakov (2002a), but we leave that to future work.

Other parameters

In Table 3, estimated coverage is reported in three more cases.

Case 1: An upper bound for μ, when $\sigma = 1$, is $\bar{X}_n - z_\alpha/\sqrt{n}$; the n-bootstrap works (Singh (1981), Bickel and Freedman (1981)). We use the rule to choose m, and then replace the normal quantile by the bootstrap quantile of $\sqrt{m}(\bar{X}^*_m - \bar{X}_n)$. The distributions considered were the normal, and the exponential (shifted to have a 0 mean).

Case 2: An upper bound for μ, when σ is unknown, is $\bar{X}_n - t_{n-1,\alpha} s_n/\sqrt{n}$; the n-bootstrap works (Singh (1981), Bickel and Freedman (1981)), and the t-quantile is replaced by the quantile of the bootstrap distribution of $\sqrt{m}(\bar{X}^*_m - \bar{X}_n)/s^*_m$.

Case 3: An upper bound for θ, when the data follows the uniform distribution on $(0, \theta)$, is $X_{(n)} - \log(\alpha)/n$ (the bound is based on $n(\theta - X_{(n)})$). This is a classical example of n-bootstrap failure (Bickel and Freedman (1981)). Using the bootstrap, the exponential quantile is replaced by the quantile of the bootstrap distribution of $m(X_{(n)} - X^*_{(m)})$.

5.3. Choices of q, metric, and smoothing

In the simulations presented, $q = 0.75$ at (1). In Sakov (1998), experiments with $q = 0.5$ gave, qualitatively, the same answers. In the set-up of Table 1, for the exponential and normal with $n = 1,000$, coverage and precision for b_n are about the same for $q = 0.75$, 0.65, 0.6, 0.5. For selected cases in the table, we made comparison on the 95% as well as the 90% bound, and again the effect of the choice of q was negligible.

It may be argued that, based on theoretical grounds, a metric based on comparison of the appropriate quantiles of the bootstrap distributions is preferable to the KS distance. We noted, in Section 3.3, that a restricted KS distance should also give better results for the upper confidence bound of the upper end of the support in the case of bounded Lipschitz densities. In the simulation presented, we have concentrated on the KS distance. In addition, we studied the Wasserstein (W) metric, with $p = 1$ and 2, as well as a quantile-based metric. In simulation the cdf was evaluated on a finite grid over a bounded interval, so the KS distance is actually a restricted KS distance, as is the W metric. When using the W metric, with $p = 1$ and 2, coverage was about $1-2\%$ smaller than the coverage achieved using the KS distance for the exponential distribution. For the normal distribution, the coverage using the W metric was about $3-4\%$ lower. We also picked m to minimize the distance between the $5th$ quantile of the bootstrap distributions. Here the coverage was about $1-1.5\%$ lower than the coverage using the KS distance (for the exponential, normal and uniform distributions). The claim of Section 3,3, that an order-grounds-restricted KS distance should do better, was confirmed, but only for sample sizes of at least 1,000.

We also explored the effect of smoothing the curve whose minima we find, before locating them. We did this for the KS distance, and found that differences in coverage were less than 1%.

All in all, these variants did not make any substantial differences in performance for the sample sizes considered.

Acknowledgement

Supported by NSF Grant DMS-98-02960.

Appendix

A. Convergence in law in probability

We say that $L^*_{m,n}$ *converges in law to* (a *random*) L *in probability*, provided the following hold.

(i) There are maps $\tilde{L}^*_{m,n} : \Omega \to D(\bar{R})$, $m, n \geq 1$.

(ii) There is a map $L : \Omega \to C(\bar{R})$ ($C(\bar{R})$ are the continuous functions on $\bar{R}$ endowed with the sup norm) such that

(a) the distributions of $\{L^*_{m,n}\}_{m\geq 1}$, and $\{\tilde{L}^*_{m,n}\}_{m\geq 1}$ agree for all n, i.e., for any $k, j_1, \ldots, j_k$,

$$P\Big(\Big(L^*_{j_1,n}, \ldots, L^*_{j_k,n}\Big)^{-1}\Big)(\cdot) = P\Big(\Big(\tilde{L}^*_{j_1,n}, \ldots, \tilde{L}^*_{j_k,n}\Big)^{-1}\Big)(\cdot),$$

or

(a′) $\|L^*_{m,n}(\cdot) - \tilde{L}^*_{m,n}(\cdot)\|_\infty = o_p(1)$ as $m, n \to \infty$, and

(b) if ρ is the Skorokhod (or Prohorov) metric,

$$\rho\Big(\tilde{L}^*_{m,n}, L\Big) \xrightarrow{p} 0, \quad \text{as} \quad m, n \to \infty. \tag{19}$$

It is possible to combine (a) and (a') into a single condition but with no gain in simplicity.

Since L is continuous with probability 1, (19) implies that

$$\|\tilde{L}^*_{m,n}(\cdot) - L(\cdot)\|_\infty \xrightarrow{p} 0. \tag{20}$$

We use an extension of these notions by considering $T_n(\cdot)$ whose values themselves lie in R^k, or generally a separable metric function space $(\mathcal{F}, d)$. For instance, consider $T_n(\hat{F}_n, F) \equiv \sqrt{n}(\hat{F}_n - F)$. Then by the law $\mathcal{L}^*(T_m(\hat{F}^*_m, \hat{F}_n))$, we mean a measurable map from Ω to the space of probability distributions on $D(\bar{R})$, endowed with the Prohorov metric. Similarly, L is a measurable map from Ω to the space of all probabilities on $C(\bar{R})$. Definition (19) for convergence in law of $\mathcal{L}^*(T_m(\hat{F}^*_m, \hat{F}_n))$ in probability carries over, save that ρ is replaced by the Prohorov metric, and (20) is no longer relevant. In principle, we can consider $T_n(\cdot,\cdot)$ taking values in $l^\infty(\mathcal{F})$, where $\mathcal{F}$ is a set of functions on R^d, and formulate results, as in van der Vaart and Wellner (1996), dropping the measurability requirements, but we do not pursue this.

References

Athreya, K. and Fukuchi, J. (1993). Bootstrapping extremes of i.i.d. random variables. In *Proceedings of the conference on extreme value theory and applications* **3**, 23-29.

Barbour, A. (1988). Stein's method and Poisson process convergence. *J. Appl. Probab.* **25**, 175-184.

Beran, R. (1982). Estimated sampling distributions: the bootstrap and competitors. *Ann. Statist.* **10**, 212-225.

Bertail, P., Politis, D. and Romano, J. (1999). On subsampling estimators with unknown rate of convergence. *J. Amer. Statist. Assoc.* **94**, 569-579.

Bertail, P., Haefke, C., Politis, D. N. and White, H. (2004). Subsampling the distribution of diverging statistics with applications to finance. *J. Econometrics* **120**, 295-326.

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

Bickel, P. and Sakov, A. (2002a). Extrapolation and the bootstrap. *Sankhyā D* **64**, 640-652.

Bickel, P. and Sakov, A. (2002b). Equality of types for the distribution of the maximum for two values of n implies extreme value type. *Extremes* **5**, 45-53.

Bickel, P., Götze, F. and van Zwet, W. (1997). Resampling fewer than n observations: gains, losses and remedies for losses. *Statist. Sinica* **7**, 1-31.

Breiman, L., Stone, C. and Kooperberg, C. (1990). Robust confidence bounds for extreme upper quantiles. *J. Statist. Comput. Simulation* **37**, 127-149.

Bretagnolle, J. (1983). Lois limites du bootstrap de certaines fonctioneles. *Ann. Inst. H. Poincaré Probab. Statist.* **19**, 281-296.

Datta, S. and McCormick, W. P. (1995). Bootstrap inference for a first-order autoregression with positive innovations. *J. Amer. Statist. Assoc.* **90**, 1289-1300.

David, H. (1981). *Order Statistics.* 2nd edition. Wiley, Efron, B. (1979). Bootstrap methods: another look at the Jackknife. *Ann. Statist.* **7**, 1-26.

Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events*. Springer.

Freedman, D. (1981). Bootstrapping regression models. *Ann. Statist.* **9**, 1218-1228.

Götze, F. (1993). Asymptotic approximation and the bootstrap. *IMS Bulletin*, page 305.

Götze, F. and Račkauskas, A. (2001). Adaptive choice of bootstrap sample sizes. In *State of the Art in Probability and Statistics*, 286-309. I.M.S publications.

Hall, P., Horowitz, J. L. and Jing, B. (1995). On blocking rules for the bootstrap with dependent data. *Biometrika* **82**, 561-574.

Hodges, J. and LeCam, L. (1960). The Poisson approximation to the Poisson Binomial distribution. *Ann. Math. Statist.* **31**, 737-740.

Kagan, A., Linnik, Y. and Rao, C. (1973). *Characterization problems in Mathematical Statistics*. Wiley.

Mammen, E. (1992). *When Does the Bootstrap Work ?*. Springer-Verlag.

Politis, D. and Romano, J. (1994). Large sample confidence regions on subsamples under minimal assumptions. *Ann. Statist.* **22**, 2031-2050.

Politis, D., Romano, J. and Wolf, M. (1999). *Subsampling*. Springer.

Reiss, R. (1989). *Approximate Distributions of Order*. Springer.

Sakov, A. (1998). Using the m out of n bootstrap in hypothesis testing. Ph.D. thesis, University of California, Berkeley.

Serfling, R. (1980). *Approximation Theorems to Mathematical Statistics*. Wiley.

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

Singh, K. (1981). On the asymptotic accuracy of Efron's bootstrap. *Ann. Statist.* **9**, 1187-1195.

SintesBlanc, A. (1991). On the Poisson approximation for some multinomial distributions. *Statist. Probab. Lett.* **11**, 1-6.

Swanepoel, J. (1986). A note on proving that the (modified) bootstrap works. *Comm. Statist., Part A* **15**, 3193-3203.

van der Vaart, A. and Wellner, J. (1996). *Weak Convergence and Empirical Processes. With Applications to Statistics*, Springer-Verlag.

Wang, Y. (1986). Coupling methods in Statistics. *Canad. J. Statist.* **14**, 69-74.

Department of Statistics, University of California, Berkeley, CA, 94720-3860, U.S.A.

E-mail: bickel@stat.berkeley.edu

Tanor Group, 4 Simtat Hashezif, Zichron Yaacov 30900, Israel.

E-mail: anat@sakov.com

(Received September 2006; accepted April 2007)