ON SUBSAMPLING ESTIMATORS WITH UNKNOWN RATE OF CONVERGENCE

by

Patrice Bertail and Dimitris N. Politis
INRA-Conso Purdue University
Joseph P. Romano
Stanford University

Technical Report #95-20

Department of Statistics
Purdue University

April 1995

◇

On subsampling estimators with unknown rate of convergence *

Patrice Bertail
INRA-Conso, 65 Bd. de Brandebourg, 94205 Ivry/Seine, France

Dimitris N. Politis
Department of Statistics, Purdue University, W. Lafayette, IN 47907, USA

Joseph P. Romano
Department of Statistics, Stanford University, Stanford, CA 94305, USA

Abstract

In Politis and Romano (1992, 1993, 1995) a general subsampling methodology has been put forth for the construction of large-sample confidence regions for a general unknown parameter θ associated with the probability distribution generating the stationary sequence $X_1, ..., X_n$. The subsampling methodology hinges on approximating the large-sample distribution of a statistic $T_n = T_n(X_1, ..., X_n)$ that is consistent for θ at some *known* rate τ_n.

Although subsampling has been shown to yield confidence regions for θ of asymptotically correct coverage under very weak assumptions, the applicability of the methodology as it has been presented so far is limited if the rate of convergence τ_n happens to be unknown or intractable in a particular setting. In this report we show how it is possible to circumvent this limitation by (a) using the subsampling methodology to derive a consistent estimator of the rate τ_n, and (b) employing the estimated rate to construct asymptotically correct confidence regions for θ based on subsampling.

Keywords. Bootstrap, confidence regions, jackknife, rate of convergence, strong mixing, subsampling.
AMS 1991 subject classifications: Primary 62G09; secondary 62M10.

*Research partially supported by NSF grants DMS94-04329 and DMS94-03826.

1 Introduction

Let $\underline{X}_n = (X_1, ..., X_n)$ be an observed stretch of a (strictly) stationary, strong mixing sequence of random variables $\{X_t, t \in Z\}$ taking values in an arbitrary sample space S; the probability measure generating the observations is denoted by P. The strong mixing condition essentially means that the sequence $\alpha_X(k) = \sup_{A,B} |P(A \cap B) - P(A)P(B)|$ tends to zero as k tends to infinity, where A and B are events in the σ-algebras generated by $\{X_t, t < 0\}$ and $\{X_t, t \geq k\}$ respectively; the case where $X_1, ..., X_n$ are independent, identically distributed (i.i.d.) is an important special case of the general scenario.

In Politis and Romano (1992, 1993, 1995) a general subsampling methodology has been put forth for the construction of large-sample confidence regions for a general unknown parameter $\theta = \theta(P)$ under very minimal conditions; see also Bickel, Goetze and van Zwet (1994), Bertail (1994). The subsampling methodology hinges on approximating the sampling distribution (under P) of a statistic $T_n = T_n(\underline{X}_n)$ that is consistent for θ at some *known* rate τ_n. Under the assumption that there is a nondegenerate asymptotic distribution for the centered 'dilated' statistic $\tau_n(T_n - \theta)$, i.e., if there is a $K(x, P)$, continuous in x, such that

$$K_n(x, P) \equiv \Pr{}_P\{\tau_n(T_n - \theta) \leq x\} \underset{n\to\infty}{\longrightarrow} K(x, P) \tag{1}$$

for any real number x, then the subsampling methodology was shown to 'work'provided that

$$b_n \underset{n\to\infty}{\longrightarrow} \infty \tag{2}$$

and

$$\frac{b_n}{n} \underset{n\to\infty}{\longrightarrow} 0. \tag{3}$$

As a matter of fact, the additional assumption $\frac{\tau_{b_n}}{\tau_n} \underset{n\to\infty}{\longrightarrow} 0$ must be satisfied as well, but this is trivially satisfied in the case where $\tau_n = n^\alpha$ for some $\alpha > 0$, or even if $\tau_n = n^\alpha L(n)$ where L is a normalized slowly varying function; see Section 2.2 for a definition of the notion of slowly varying function. In other words, subsampling yields confidence regions for θ of asymptotically correct coverage under the very weak assumption (1), provided care is taken so that (2) and (3) are satisfied as well.

In the case of i.i.d. data, subsampling may be seen as a delete-d (with $d = n - b_n$) jackknife (cf. Shao and Wu (1989) and Wu (1990)), but also as resampling (bootstrap) without replacement with a resampling size b_n

smaller than the original sample size n; since the difference between sampling with replacement, i.e., Efron's (1979) classical bootstrap with resample size b_n, and sampling without replacement, i.e., subsampling with subsample size b_n, is negligible if $b_n^2/n = o_P(1)$, the general validity of subsampling implies general validity of the bootstrap with appropriately small resample size (cf. Politis and Romano (1992, 1993)) even in situations where the usual bootstrap (with resample size n) fails. In the case of stationary data (time series or random fields), subsampling is closely related to the blocking methods introduced by Carlstein (1986), Künsch (1989), and Liu and Singh (1992); see also Sherman and Carlstein (1994).

Although existence of an asymptotic distribution is almost a *sine que non* condition for the purposes of approximating the large-sample sampling distribution of the statistic T_n, and some form of weak dependence condition (e.g., mixing) is required for consistency of estimation as the sample size increases, it may be possible that the convergence rate $\tau_{\cdot}$ (as a function of n) is unknown (or impossible to calculate), in which case it is difficult to see how subsampling could be used for construction of confidence regions; however, subsampling would readily give some information on the shape of $K(x, P)$ -as a function of x- which can be a helpful diagnostic tool (cf. Sherman (1992)).

The aim of this article is to actually show that it is possible to drop the hypothesis concerning the explicit knowledge of the convergence rate in order to use subsampling for the construction of nonparametric confidence regions. In particular, we will first use the subsampling methodology to derive a consistent estimator of the rate $\tau_{\cdot}$, and we will then use the estimated rate to get asymptotically correct confidence regions based on subsampling. The underlying idea is that it is possible to construct the subsampling distribution of T_n itself (rather than that of $\tau_n T_n$); the speed by which it degenerates to a Dirac measure as $n \to \infty$ is closely related to τ_{b_n}. Constructing several subsampling distributions for different choices of b_n gives valuable information on the shape of τ_n as a function of n.

The following examples illustrate why it may be interesting to estimate the convergence rate before subsampling.

Example 1: Nonnormal limit distribution. Consider the functional $\theta(P) = (E_P X)^2$ and its empirical counterpart $T_n = \left(n^{-1} \sum_{i=1}^n X_i\right)^2$, where the X_i's are i.i.d. with variance σ^2. If $E_P X \neq 0$ then we have $n^{1/2}(T_n - \theta(P)) \underset{n\to\infty}{\longrightarrow} N(0, 4\sigma^2)$ and the usual bootstrap works. If $E_P X = 0$

then we have $n(T_n - \theta(P)) \underset{n\to\infty}{\longrightarrow} \sigma^2\chi_1^2$ and the usual bootstrap fails; see, for example, Datta (1995) where a modified bootstrap is proposed to remedy the situation. Note that subsampling (or bootstraping with smaller resampling size) would work in this case here as well *provided* the rate of convergence is known, i.e., provided it is known whether or not $E_P X = 0$.

Example 2: The extreme of n i.i.d. observations. Another interesting example is the case of the extreme order statistic. It is well known that the convergence rate of the extreme depends on the domain of attraction of the underlying distribution and that the usual bootstrap fails because of a certain lack of uniformity (Bickel and Freedman (1981)). Moreover since the underlying distribution is in general unknown, the convergence rate is unknown, and we can not construct the subsampling distribution unless we have some extra information on P.

Example 3: The sample mean of a time series with moderate dependence. Let $\bar{X}_n = n^{-1}\sum_{t=1}^n X_t$ be the sample mean, and $\theta = EX_0$ be the mean. Suppose that, although the sequence $\{X_t\}$ is strong mixing, the mixing coefficients decrease to zero slowly enough so that $\sum_{k=1}^\infty |Cov(X_1, X_k)| = \infty$; then the variance of $\bar{X}_n$ is not of order n^{-1}. Suppose that actually $\sum_{k=1}^n Cov(X_1, X_k) \sim n^{2\beta}$, and therefore $Var(\bar{X}_n) \sim \sigma_\infty^2 n^{2\beta-1}$, for some $0 < \beta < 1/2$, and $\sigma_\infty^2 > 0$. Assuming $E|X_0|^{2+\delta} < \infty$, for some $\delta > 0$, it follows from Rosenblatt (1984) that $n^{\frac{1}{2}-\beta}(\bar{X}_n - \theta)$ has an asymptotic normal $N(0, \sigma_\infty^2)$ distribution. To use either the normal limit distribution or its subsampling approximation, the exponent β must be estimated from the data; see, e.g., Künsch (1989).

Example 4: Nonparametric regression. Suppose bivariate data $(Y_i, X_i), i = 1, ..., n$, are available and we are interested in estimating the conditional expectation $E(Y|X = x)$, for some x, using some nonparametric regression technique. Many such smoothing methods are currently available, e.g., lowess, AVAS, and ACE, and are included as ready-to-use functions in statistical software such as S+; see, e.g., Becker, Chambers, and Wilks (1988). One can employ the classical bootstrap with resample size n to get a measure of accuracy of the estimated $E(Y|X = x)$; as a matter of fact, it is hard to avoid using the bootstrap in this case (Efron (1994)). Nevertheless, for robustness purposes, one might want to use subsampling (or bootstrap with resample size $b_n < n$) instead; to do this, it is unavoidable that the

rate of convergence of the estimated $E(Y|X = x)$ to its true value should be calculated or estimated.

The structure of the remainder of the paper is as follows: Section 2 presents the basic idea of rate estimation using subsampling, and the proposed estimators are shown to be asymptotically consistent at an appropriate rate. Section 3 shows how the estimated rate can be employed in the construction of asymptotically correct confidence regions based on another round of subsampling. A small finite-sample experiment is conducted and its outcomes are presented in Section 4. Finally, Section 5 contains the technical proofs of our results.

2 Estimation of the rate

2.1 Some heuristic ideas

Until otherwise stated, let us make the simplifying assumption that T_n and θ are real-valued, and that $K(x, P)$ is continuous in x on the whole real line. Although the case of i.i.d. data is a special case of the strong mixing case, the construction of the subsampling distribution can take advantage of the i.i.d. structure, when such a structure exists; of course, if one is unsure regarding the independence assumption, it is safer (and more robust) to operate under the general strong mixing assumption.

- **General case (strong mixing data).** Define Y_i to be the subsequence $(X_i, X_{i+1}, ..., X_{i+b_n-1})$, for $i = 1, ..., q$, and $q = n - b_n + 1$; note that Y_i consists of b_n consecutive observations from the $X_1, \ldots, X_n$ sequence, and the order of the observations is preserved.
- **Special case (i.i.d. data).** Let $Y_1, \ldots, Y_q$ be equal to the $q = \binom{n}{b_n}$ subsets of size b_n chosen from $\{X_1, \ldots, X_n\}$, and then ordered in any fashion; here the subsets Y_i consist of unordered observations.

In either case, let $T_{b_n,i}$ be the value of the statistic T_b applied to the subsample Y_i. The subsampling distribution of the root $\tau_n(T_n - \theta)$, based on a subsample size b_n, is defined by

$$K_{b_n}(x \mid \underline{X}_n, \tau_\cdot) \equiv q^{-1} \sum_{i=1}^{q} 1\{\tau_{b_n}(T_{b_n,i} - T_n) \leq x\}. \tag{4}$$

In both the strong mixing and i.i.d. cases, Politis and Romano (1992, 1995) showed that if (1), (2), (3) hold, then

$$\sup_x |K_{b_n}(x \mid \underline{X}_n, \tau_\cdot) - K(x, P)| = o_P(1) \tag{5}$$

as n tends to infinity.

Denote by $K_{b_n}(x \mid \underline{X}_n) \equiv K_{b_n}(x \mid \underline{X}_n, 1)$ the subsampling distribution of the root $(T_n - \theta)$. Let $\hat{V}_{b_n}$ be the variance of a random variable with distribution $K_{b_n}(x \mid \underline{X}_n)$; in other words, $\hat{V}_{b_n}$ is the subsampling estimator of the variance of T_{b_n}. Under strong enough moment and mixing conditions (cf. Carlstein (1986), Künsch (1989), Liu and Singh (1992), Politis and Romano (1992, 1995)), the subsampling variance estimator is consistent, i.e., (i) there is a positive constant V such that $\tau_n^2 Var(T_n) \to V$, and (ii) $\tau_{b_n}^2 \hat{V}_{b_n} \to V$ in probability, as n tends to infinity.

From (ii) above it follows that

$$\log \hat{V}_{b_n} = \log V - 2 \log \tau_{b_n} + o_P(1). \tag{6}$$

Thus, if we construct several (at least two) subsampling distributions for different sizes b_n, we may be able to estimate the convergence rate. So let $b_{1,n} \neq b_{2,n}$; then we have

$$\frac{1}{2} \log \hat{V}_{b_{2,n}} - \frac{1}{2} \log \hat{V}_{b_{1,n}} = \log(\frac{\tau_{b_{1,n}}}{\tau_{b_{2,n}}}) + o_P(1).$$

Therefore, having constructed two subsampling distributions, it is possible to estimate the shape of function $\tau_\cdot$ up to some slowly varying function in Karamata's sense (see for instance Bingham, Goldie and Teugels (1987)). To see this, assume that τ_n is of the form $\tau_n = n^\alpha$ where α is an unknown positive constant; then we have

$$\frac{1}{2} \log(\tfrac{b_{1,n}}{b_{2n}})^{-1} \left(\log \hat{V}_{b_{2,n}} - \log \hat{V}_{b_{1,n}} \right) = \alpha + o_P(\log(\tfrac{b_{1,n}}{b_{2,n}})^{-1}). \tag{7}$$

In other words, if we ensure that $\log(\frac{b_{1,n}}{b_{2,n}}) \underset{n \to \infty}{\longrightarrow} \infty$, which can be done by letting $b_{i,n} = n^{\beta_i}$, where $1 > \beta_1 > \beta_2 > 0$ are some constants, equation (7) suggests using $\frac{1}{2} \log(\frac{b_{1,n}}{b_{2n}})^{-1} \left(\log \hat{V}_{b_{2,n}} - \log \hat{V}_{b_{1,n}} \right)$ as a consistent estimator of α.

In general, we could construct I subsampling distributions based on subsample sizes $b_{i,n}$, $i = 1, .., I$, and we would then have the corresponding

variance estimators $\hat{V}_{b_{i,n}}$, $i = 1, .., I$. Still under the assumption that $\tau_n = n^\alpha$, equation (6) yields the system of equations

$$y_i = \frac{-\log V}{2} + \alpha x_i + u_i \tag{8}$$

for $i = 1, .., I$, where $y_i = -\log \hat{V}_{b_{i,n}}/2$, $x_i = \log b_{i,n}$, and $u_i = o_P(1)$. Note that (8) can be interpreted as a straight line regression of the y_i's on the x_i's, where the slope of the fitted straight line is given by Ordinary Least Squares to be

$$\alpha_I \equiv \frac{\sum_{i=1}^I (y_i - \overline{y})(\log(b_{i,n}) - \overline{\log})}{\sum_{i=1}^I (\log(b_{i,n}) - \overline{\log})^2} \tag{9}$$

where $\overline{y} = I^{-1} \sum_{i=1}^I y_i$, and $\overline{\log} = I^{-1} \sum_{i=1}^I \log(b_{i,n})$. It is easy to see that α_2 is just the estimator suggested by (7).

The following theorem asymptotically validates the use of the estimator α_I as an estimator of α.

Theorem 1 *Let $b_{i,n} = n^{\beta_i}$, $1 > \beta_1 > \cdots > \beta_I > 0$, and assume that for some $\tau_n = n^\alpha$*
(i) there is a positive constant V such that $\tau_n^2 Var(T_n) \underset{n\to\infty}{\longrightarrow} V$, and
(ii) $\tau_{b_{i,n}}^2 \hat{V}_{b_{i,n}} \underset{n\to\infty}{\longrightarrow} V$, in probability, for $i = 1, ..., I$.
Then, $\alpha_I = \alpha + o_P((\log n)^{-1})$.

Remark 1a Note that, although (5) is valid in great generality (cf. Politis and Romano (1992, 1993, 1995)), conditions *(i)* and *(ii)* of Theorem 1 may easily fail to be satisfied. Since the spirit of the present paper is to obtain valid results under the weakest assumptions possible, having to assume conditions (i) and (ii) is somewhat unsatisfactory; for example, it might be the case that T_n does not even possess a finite 2nd moment. A way out is proposed in the following section, where it is seen that a consistent estimator of α can be constructed without assuming conditions (i) and (ii) by looking at quantiles.

Remark 1b Note that in the regression equation (8), the 'errors' u_i are neither of mean zero, nor are they uncorrelated. The reason Ordinary Least Squares (OLS) works asymptotically is that the u_i's become asymptotically negligible as n increases. In special cases, e.g., the sample mean of a strong mixing sequence, the asymptotic order of the bias of $\log \hat{V}_{b_{i,n}}$ as an estimator of $\log V$, as well as the asymptotic covariance between $\log \hat{V}_{b_{i,n}}$ and $\log \hat{V}_{b_{j,n}}$

can be calculated, and thus could be taken into account in order to conduct a Generalized Least Squares (GLS) regression; nevertheless, the OLS will be asymptotically equally as efficient as the GLS, so the extra complications can be avoided.

Remark 1c Choosing the β_i's properly is another important consideration. Note that to reduce the magnitude of the error in estimating α by α_I, the β_i's should be taken to be spread out over the interval $(0,1)$, far apart from one another; see equation (7). In the i.i.d. case, this choice of spread-out β_i's should cause no problem; however, in the strong mixing case, we generally have a bias-variance trade-off: if $b_{i,n}$ is small, $\log \hat{V}_{b_{i,n}}$ is quite biased but is not very variable, whereas if $b_{i,n}$ is large, $\log \hat{V}_{b_{i,n}}$ has small bias but is quite variable. In the particular case of the sample mean of a strong mixing sequence, to minimize the Mean Squared Error of $\log \hat{V}_{b_{i,n}}$ as an estimator of $\log V$ we would need to take $b_{i,n}$ proportional to $n^{1/3}$; see, for example, Künsch (1989) or Politis and Romano (1994). Thus, in order to render the 'errors' u_i asymptotically negligible as fast as possible, the β_i's should only be taken to span the interval $(1/3-\delta, 1/3+\delta)$, for some appropriate δ. Alas, it is not possible to prescribe an optimal choice of the β_i's in the general set-up considered in this paper; the difficult problem of choosing the β_i's requires more work on a case-by-case basis.

2.2 Rate estimation: scalar parameter case

Given a distribution G on the real line and a number $t \in (0,1)$, we will let $G^{-1}(t)$ denote the quantile transformation, i.e., $G^{-1}(t) = \inf\{x : G(x) \geq t\}$, which reduces to the regular inverse of the function G if G happens to be invertible. Note that we have

$$K_{b_n}(x\ \tau_{b_n}^{-1} \mid \underline{X}_n) = K_{b_n}(x \mid \underline{X}_n, \tau_\cdot) \tag{10}$$

ans thus it is easy to see that

$$K_{b_n}^{-1}(t \mid \underline{X}_n, \tau_\cdot) = \tau_{b_n}\ K_{b_n}^{-1}(t \mid \underline{X}_n). \tag{11}$$

The following lemma (see for instance de Haan (1970)) will be useful to us later on.

Lemma 1 *Let $k_0 = \sup\{x : K(x,P) = 0\}$ and $k_1 = \inf\{x : K(x,P) = 1\}$, and assume that $K(x,P)$ is continuous and strictly increasing on (k_0, k_1) as a function of x. If (5) is true as n tends to infinity, then*

$$K_{b_n}^{-1}(t \mid \underline{X}_n, \tau_\cdot) = K^{-1}(t,P) + o_P(1)$$

for any $t \in (0,1)$.

As a consequence of Lemma 1 and equation (11) we have that

$$\tau_{b_n} K_{b_n}^{-1}(t \mid \underline{X}_n) = K^{-1}(t,P) + o_P(1)$$

or equivalently

$$K_{b_n}^{-1}(t \mid \underline{X}_n) = \tau_{b_n}^{-1}\, K^{-1}(t,P) + o_P(\tau_{b_n}^{-1}). \tag{12}$$

It is observed that $K_{b_n}^{-1}(t \mid \underline{X}_n)$ is approximately proportional to $\tau_{b_n}^{-1}$; thus, as before, if we construct several (at least two) subsampling distributions for different sizes b_n, we may be able to estimate the convergence rate.

More precisely, for any point $t > K(0,P)$ we have

$$\log\left(K_{b_n}^{-1}(t \mid \underline{X}_n)\right) = \log\left(K^{-1}(t,P)\right) - \log(\tau_{b_n}) + o_P(1). \tag{13}$$

It follows that if we choose $b_{1,n} \neq b_{2,n}$, then we have

$$\log\left(K_{b_{2,n}}^{-1}(t \mid \underline{X}_n)\right) - \log\left(K_{b_{1,n}}^{-1}(t \mid \underline{X}_n)\right) = \log(\frac{\tau_{b_{1,n}}}{\tau_{b_{2,n}}}) + o_P(1);$$

as before, it follows that with two subsampling distributions, it is possible to estimate the shape of function $\tau_{\cdot}$ up to some slowly varying function. If τ_n is of the form $\tau_n = n^{\alpha}$ where α is an unknown constant, then

$$\log(\frac{b_{1,n}}{b_{2n}})^{-1}\left(\log\left(K_{b_{2,n}}^{-1}(t \mid \underline{X}_n)\right) - \log\left(K_{b_{1,n}}^{-1}(t \mid \underline{X}_n)\right)\right) = \alpha + o_P(\log(\frac{b_{1,n}}{b_{2,n}})^{-1}). \tag{14}$$

It is easy to see that the mean over several points t_j, $j = 1, ..., B$ of the left hand side of (14) is a consistent estimator of α, provided that we choose $b_{1,n}$ and $b_{2,n}$ such that $\log(\frac{b_{1,n}}{b_{2,n}}) \underset{n\to\infty}{\longrightarrow} \infty$; for instance take $b_{i,n} = n^{\beta_i}$, where $1 > \beta_1 > \beta_2 > 0$ are some constants. More generally, we can construct I subsampling distributions based on subsample sizes $b_{i,n}$, $i = 1, .., I$; then equation (13) evaluated at some points $t_j > K(0,P)$, $j = 1, .., J$, yields

$$y_{i,j} \equiv \log\left(K_{b_{i,n}}^{-1}(t_j \mid \underline{X}_n)\right) = a_j - \alpha \log(b_{i,n}) + u_{i,j} \tag{15}$$

where $a_j \equiv \log\left(K^{-1}(t_j,P)\right)$ and $u_{i,j} = o_P(1)$, $i = 1, .., I$ and $j = 1, .., J$. The above equation may be interpreted as an ANOVA (analysis of variance) set-up, in which case the following estimator of α suggests itself, namely

$$\alpha_{I,J} \equiv -\frac{\sum_{i=1}^{I}(y_{i,\cdot} - \overline{y})(\log(b_{i,n}) - \overline{\log})}{\sum_{i=1}^{I}(\log(b_{i,n}) - \overline{\log})^2} \tag{16}$$

where $y_{i,.} = J^{-1}\sum_{j=1}^{J} y_{i,j} = J^{-1}\sum_{j=1}^{J}\log\left(K^{-1}_{b_{i,n}}(t_j \mid \underline{X}_n)\right)$, $\overline{y} = (IJ)^{-1}\sum_{i=1}^{I}\sum_{j=1}^{J} y_{i,j}$ and $\overline{\log} = I^{-1}\sum_{i=1}^{I}\log(b_{i,n})$. Notice that for $I = 2$ and $J = 1$, then

$$\alpha_{2,1} = \log(\tfrac{b_{1,n}}{b_{2n}})^{-1}\left(\log\left(K^{-1}_{b_{2,n}}(t \mid \underline{X}_n)\right) - \log\left(K^{-1}_{b_{1,n}}(t \mid \underline{X}_n)\right)\right)$$

is the estimator suggested by (14).

The following theorem asymptotically validates the use of the estimator $\alpha_{I,J}$ as an estimator of α.

Theorem 2 *Assume that (1) holds for some $\tau_n = n^{\alpha}$, and some $K(x,P)$ continuous and strictly increasing on (k_0,k_1) as a function of x, where $k_0 = \sup\{x : K(x,P) = 0\}$ and $k_1 = \inf\{x : K(x,P) = 1\}$. Let $b_{i,n} = n^{\beta_i}$, $1 > \beta_1 > \cdots > \beta_I > 0$, and consider some points $t_j > K(0,P)$, $j = 1,...J$. Then $\alpha_{I,J} = \alpha + o_P((\log n)^{-1})$.*

Remark 2a In the i.i.d. case, it may be the case that the classical bootstrap with resample size equal to b_n provides an asymptotically consistent estimator of the distribution K under conditions (2) and (3); this can be checked (cf. Bickel *et al.* (1994)), or enforced by letting $0.5 > \beta_1 > \cdots > \beta_I > 0$ in Theorem 2 (cf. Politis and Romano (1992, 1993)). In such a case, this bootstrap distribution estimator could equally be used in place of the subsampling distribution estimator K_{b_n} for the purposes of rate estimation; however, there does not seem to be an advantage in doing this.

Remark 2b Under stronger conditions, the rate of convergence of $\alpha_{I,J}$ to α may be faster than $(\log n)^{-1}$. For example, it may be assumed not only that (1) holds, but that the convergence to the limit law $K(x,P)$ occurs at some rate, e.g., $K_n(x,P) = K(x,P) + O_P(n^{-1/2})$, or some other asymptotic expansion; see, for example, Bertail (1994), or Barbe and Bertail (1995).

Remark 2c If τ_n is a more complicated function, equation (13) suggests to estimate the shape of $h(n) = log(\tau_n)$, under the constraint that $h(.)$ is an increasing function equal to $+\infty$ at $+\infty$, using some nonparametric technique. In many interesting situations, τ_n is a regular varying function of index α, say $\tau_n = n^{\alpha}L(n)$ where L is a normalized slowly varying function that is is such that $L(1) = 0$ and for any $\lambda > 0$, $\lim_{x\to\infty}\frac{L(\lambda x)}{L(x)} = 1$ (see Bingham, Goldie, and Teugels (1987)). By the Karamata representation theorem, there exists $\epsilon(.), \epsilon(u) \underset{u\to\infty}{\longrightarrow} 0$ such that $L(n) = \exp\int_1^n u^{-1}\epsilon(u)du$,

and (13) may be written

$$\log\left(K_{b_n}^{-1}(t \mid \underline{X}_n)\right) = \log\left(K^{-1}(t,P)\right) - \alpha\log(b_n) + \int_1^{b_n} u^{-1}\epsilon(u)du + o(1)$$

which may be seen for a fixed t, as a partial spline regression model (see Engle *et al.* (1986)). However, in that case we necessarily need more than two subsampling distributions to estimate the slowly varying function part, here $h(b_n) = \int_1^{b_n} u^{-1}\epsilon(u)du$. Since we only want to obtain a convergent estimator of the convergence rate, we only need a convergent estimator of $h(.)$ and an estimator $\hat{\alpha}$ such that $\hat{\alpha} = \alpha + o_P((\log n)^{-1})$. This may be achieved, for instance, by constructing $I = \log(n)$ subsampling distibutions with $b_{i,n} = n^{\beta_i}, i = 1, ..., I,$, where $\beta_1 < ... < \beta_I$. However, it is likely that a huge sample size is needed to obtain an accurate estimator of the slowly varying function; therefore, we focus on the more feasible task of estimating α under the assumption of $\tau_n = n^\alpha$ from now on.

Note that $K_{b_n}(0 \mid \underline{X}_n, \tau_\cdot) = K_{b_n}(0 \mid \underline{X}_n) = K(0,P) + o_P(1)$, i.e., $K(0,P)$ may be estimated without knowing the rate τ_n; thus, choosing the t_j in order to apply Theorem 2 will not be a problem in practice. Nevertheless, the requirement $t_j > K(0,P)$ seems a bit cumbersome. We now show that, by looking at differences of quantiles, this requirement could be circumvented.

So let t_{2j}, for $j = 1, ..., J$, be some points in $(0.5, 1)$, and let t_{2j-1}, for $j = 1, ..., J$, be some points in $(0, 0.5)$; as before, let $b_{i,n}$ be some different subsample sizes, for $i = 1, ..., I$. From equation (12) it follows that

$$y_{i,j} \equiv \log\left(K_{b_{i,n}}^{-1}(t_{2j} \mid \underline{X}_n) - K_{b_{i,n}}^{-1}(t_{2j-1} \mid \underline{X}_n)\right) = a_j - \alpha\log(b_{i,n}) + u_{i,j} \quad (17)$$

where $a_j \equiv \log\left(K^{-1}(t_{2j},P) - K^{-1}(t_{2j-1},P)\right)$ and $u_{i,j} = o_P(1)$, $i = 1, .., I$ and $j = 1, .., J$. Here as well we can use the estimator $\alpha_{I,J}$ as defined in (16), where now

$$y_{i,\cdot} = J^{-1}\sum_{j=1}^{J} y_{i,j} = J^{-1}\sum_{j=1}^{J}\log\left(K_{b_{i,n}}^{-1}(t_{2j} \mid \underline{X}_n) - K_{b_{i,n}}^{-1}(t_{2j-1} \mid \underline{X}_n)\right),$$

$\overline{y} = (IJ)^{-1}\sum_{i=1}^{I}\sum_{j=1}^{J} y_{i,j}$ and $\overline{\log} = I^{-1}\sum_{i=1}^{I}\log(b_{i,n})$.

The following theorem asymptotically validates the use of the new estimator $\alpha_{I,J}$ as an estimator of α.

Theorem 3 *Assume that (1) holds for some $\tau_n = n^\alpha$, and some $K(x,P)$ continuous and strictly increasing on (k_0, k_1) as a function of x, where $k_0 =$*

$\sup\{x : K(x,P) = 0\}$ *and* $k_1 = \inf\{x : K(x,P) = 1\}$. *Let* $b_{i,n} = n^{\beta_i}$, $1 > \beta_1 > \cdots > \beta_I > 0$, *and consider some points* $t_{2j} \in (0.5, 1)$, *and* $t_{2j-1} \in (0, 0.5)$, *for* $j = 1, ...J$. *Then* $\alpha_{I,J} = \alpha + o_P((\log n)^{-1})$.

For example, suppose that $J = 1$ and that $t_1 = 0.25$ and $t_2 = 0.75$; then $y_{i,j}$ as given in (17) is the inter-quartile range of the subsampling distribution $K_{b_{i,n}}(x|\ \underline{X}_n)$, and a_j as given in (17) is the inter-quantile range of the limit distribution $K(x,P)$. Obviously, the inter-quantile range is a more robust estimate of scale as compared to the variance which was used in Theorem 1; so it is not surprising that Theorem 3 is true in greater generality. Nevertheless, it is recommended to take $J > 1$, i.e., to look at many differences of quantiles, in order that $\alpha_{I,J}$ becomes more accurate.

In the following section, a different way to side-step the cumbersome requirement $t_j > K(0,P)$ of Theorem 2 is proposed in the more general setting of a vector valued parameter θ.

2.3 Rate estimation: vector parameter case

Now assume that θ takes values in a normed linear space Θ with norm $||\cdot||$, and that $T_n = T_n(\underline{X}_n)$ is an estimator consistent for θ at rate $\tau_n = n^\alpha$, where α may be unknown. A confidence region for θ can be constructed if an approximation to the sampling distribution of $\tau_n||T_n - \theta||$ is available; for example, Θ might be a function space, and $||\cdot||$ might be the sup-norm in which case the confidence region has the form of a uniform confidence band for the unknown function θ (cf. Politis, Romano, and You (1993)).

Consequently, let us denote

$$\tilde{K}_n(x,P) \equiv \Pr{}_P(\tau_n||T_n - \theta|| \le x) \tag{18}$$

and we will assume as before that

$$\tilde{K}_n(x,P) \underset{n\to\infty}{\longrightarrow} \tilde{K}(x,P) \tag{19}$$

holds for some $\tilde{K}(x,P)$ continuous in x.

The subsample values $T_{b_n,i}$ are defined in a fashion identical to the construction leading to equation (4), and the subsampling distribution of the root $\tau_n||T_n - \theta||$ is given by

$$\tilde{K}_{b_n}(x \mid \underline{X}_n, \tau_\cdot) \equiv q^{-1} \sum_{i=1}^{q} 1\{\tau_{b_n}||T_{b_n,i} - T_n|| \le x\}; \tag{20}$$

Politis, Romano, and You (1993) showed that if (19), (2), (3) hold, then

$$\sup_x |\tilde{K}_{b_n}(x \mid \underline{X}_n, \tau_.) - \tilde{K}(x, P)| = o_P(1) \tag{21}$$

as n tends to infinity. The following analog of Lemma 1 will be useful here.

Lemma 2 *Let* $\tilde{k}_0 = \sup\{x : \tilde{K}(x, P) = 0\}$ *and* $\tilde{k}_1 = \inf\{x : \tilde{K}(x, P) = 1\}$, *and assume that* $\tilde{K}(x, P)$ *is continuous and strictly increasing on* $(\tilde{k}_0, \tilde{k}_1)$ *as a function of* x. *If (21) is true as* n *tends to infinity, then*

$$\tilde{K}_{b_n}^{-1}(t \mid \underline{X}_n, \tau_.) = \tilde{K}^{-1}(t, P) + o_P(1)$$

for any $t \in (0, 1)$.

So as before, denote by $\tilde{K}_{b_n}(x \mid \underline{X}_n) \equiv \tilde{K}_{b_n}(x \mid \underline{X}_n, 1)$, but let

$$y_{i,j} \equiv \log\left(\tilde{K}_{b_{i,n}}^{-1}(t_j \mid \underline{X}_n)\right) = a_j - \alpha \log(b_{i,n}) + u_{i,j} \tag{22}$$

where $a_j \equiv \log\left(\tilde{K}^{-1}(t_j, P)\right)$ and $u_{i,j} = o_P(1)$, $i = 1, .., I$ and $j = 1, .., J$. Now we can use $\alpha_{I,J}$ exactly as given in definition (16) –but now using the $y_{i,j}$s as given above in (22)– to get a consistent estimator of α.

Theorem 4 *Assume that (19) holds for some* $\tau_n = n^{\alpha}$, *and some* $\tilde{K}(x, P)$ *continuous and strictly increasing on* $(\tilde{k}_0, \tilde{k}_1)$ *as a function of* x, *where* $\tilde{k}_0 = \sup\{x : \tilde{K}(x, P) = 0\}$ *and* $\tilde{k}_1 = \inf\{x : \tilde{K}(x, P) = 1\}$. *Let* $b_{i,n} = n^{\beta_i}$, $1 > \beta_1 > \cdots > \beta_I > 0$, *and consider some points* $t_j > 0$, $j = 1, ...J$. *Then* $\alpha_{I,J} = \alpha + o_P((\log n)^{-1})$.

Although Theorem 4 is quite similar to Theorem 2, note that the cumbersome condition $t_j > K(0, P)$ has been replaced by the more natural condition $t_j > 0$; the reason is that $\tilde{K}(0, P) \equiv 0$ because of the continuity of $\tilde{K}(x, P)$. Theorem 4 might of course be used even for real-valued parameters; see also Theorem 6 in what follows.

3 Subsampling with estimated rate of convergence

Having shown that it is possible to consistently estimate the rate of convergence $\tau_.$ in general situations, we will now focus on our original goal, namely the construction of confidence regions for the unknown parameter θ.

The obvious strategy in order to use subsampling in the case of unknown rate of convergence $\tau_.$ is to estimate $\tau_.$, and used the estimated $\tau_.$ in constructing the subsampling distribution. As Theorems 5 and 6 below demonstrate, this plug-in strategy gives valid results, i.e., subsampling with an estimated rate of convergence yields confidence regions for θ of asymptotically correct coverage.

The following theorem establishes the asymptotic validity of subsampling with an estimated rate of convergence in the case of real-valued T_n and θ.

Theorem 5 *Assume that (1) holds for some $\tau_n = n^\alpha$, and some $K(x, P)$ continuous and strictly increasing on (k_0, k_1) as a function of x, where $k_0 = \sup\{x : K(x, P) = 0\}$ and $k_1 = \inf\{x : K(x, P) = 1\}$; also assume (2) and (3). Let $\hat{\alpha} = \alpha + o_P((\log n)^{-1})$, and put $\hat{\tau}_n = n^{\hat{\alpha}}$. Then*

$$\sup_x |K_{b_n}(x \mid \underline{X}_n, \hat{\tau}_.) - K(x, P)| = o_P(1). \tag{23}$$

Let $t \in (0, 1)$, and let $c_n(t) = K_{b_n}^{-1}(t \mid \underline{X}_n, \hat{\tau}_.)$ be the t^{th} quantile of the subsampling distribution $K_{b_n}(x \mid \underline{X}_n, \hat{\tau}_.)$. Then

$$\Pr_P\{\hat{\tau}_n(T_n - \theta) \leq c_n(t)\} \underset{n\to\infty}{\longrightarrow} t. \tag{24}$$

Thus the asymptotic coverage probability of the interval $[T_n - (\hat{\tau}_n)^{-1}c_n(t), \infty)$ is the nominal level t.

Note that the estimator $\hat{\alpha}$ in Theorem 5 can be obtained by *any* method (and from the same dataset $\underline{X}_n$), as long as $\hat{\alpha} = \alpha + o_P((\log n)^{-1})$; for example, any of the rate estimation methods discussed so far in Theorems 2, 3, or 4 could be used.

If a particular model can be assumed, then it might be more effective to use a model-specific estimator of the convergence rate. The model-specific rate estimators should be more accurate (if the model assumed is indeed true), and thus the subsampling distribution with estimated rate will be more accurate as a result; however, it is reassuring to know that there are many model-free estimators (given by our Theorems 2, 3, and 4) that are accurate enough to be used in conjunction with confidence intervals based on subsampling. For instance in our Example 3, we only need to estimate β which is linked to the strong mixing coefficients of the model; see, for example, Beran (1994) regarding estimation of the 'long-memory' parameter β.

The next theorem establishes the asymptotic validity of subsampling with an estimated rate of convergence in the general case where θ (and therefore T_n as well) takes values in a normed linear space Θ with norm $||\cdot||$. In the case Θ is the real line, Theorem 6 allows for the construction of symmetric confidence intervals for θ.

Theorem 6 *Assume that (19) holds for some $\tau_n = n^\alpha$, and some $\tilde{K}(x, P)$ continuous and strictly increasing on $(\tilde{k}_0, \tilde{k}_1)$ as a function of x, where $\tilde{k}_0 = \sup\{x : \tilde{K}(x, P) = 0\}$ and $\tilde{k}_1 = \inf\{x : \tilde{K}(x, P) = 1\}$. Let $\hat{\alpha} = \alpha + o_P((\log n)^{-1})$, and put $\hat{\tau}_n = n^{\hat{\alpha}}$. Then*

$$\sup_x |\tilde{K}_{b_n}(x \mid \underline{X}_n, \hat{\tau}_\cdot) - \tilde{K}(x, P)| = o_P(1). \tag{25}$$

Let $t \in (0,1)$, and let $c_n(t) = \tilde{K}_{b_n}^{-1}(t \mid \underline{X}_n, \hat{\tau}_\cdot)$ be the t^{th} quantile of the subsampling distribution $\tilde{K}_{b_n}(x \mid \underline{X}_n, \hat{\tau}_\cdot)$. Then

$$\Pr_P\{\hat{\tau}_n ||T_n - \theta|| \leq c_n(t)\} \underset{n\to\infty}{\longrightarrow} t. \tag{26}$$

Thus the asymptotic coverage probability of the confidence region $\{\theta : \hat{\tau}_n ||T_n - \theta|| \leq c_n(t)\}$ is the nominal level t.

Theorems 5 and 6 show that the construction of large-sample confidence regions based on subsampling can be performed in a fairly automatic and totally model-free fashion, in the sense that we do not have to adapt the resampling methodology to the model as it is usually the case for the bootstrap. Nonetheless, the subsampling methodology may yield poor approximations of the true distribution for finite sample sizes. Some ways of improving the accuracy of the subsampling distributions (for instance through linear combinations of subsampling distribution) are proposed in Bertail (1994) and Bickel *et al.* (1994). In the following section the subsampling methodology with estimated rate is put to the test by considering some finite-sample simulation results.

4 Some simulation results

In the previous subsections it was demonstrated that subsampling with an estimated convergence rate gives asymptotically valid results. However, due to the potentially very slow rate of convergence of $\hat{\alpha}$ to α, it may well be the case that a very large sample size n is required for the methodology

to be applicable in practice. To see how subsampling with an estimated convergence rate performs in finite samples, a small simulation experiment was performed. We give some empirical evidence on the performance of the estimators of the convergence rate and provide some indications on the choice of the estimator.

We mainly consider the example of the square of the mean of i.i.d. data, i.e., $T_n = \left(n^{-1}\sum_{i=1}^{n} X_i\right)^2$, where the X_i's are i.i.d. with mean $E_P X$, and variance σ^2. Note that this should be regarded as a toy-example since there are other (and better) methods for dealing with this particular problem; see the discussion regarding Example 1 of the Introduction. Nonetheless, this example will serve us well for the purposes of evaluating the finite-sample performance and the reliability of the proposed ideas. In what follows the $X_i's$ are i.i.d. generated according to a $N(E_P X, 1)$ law; other distributions, e.g., exponential and chi-square with different degrees of freedom were also considered but the results were qualitatively the same. Each subsampling distribution is approximated by a stochastic approximation Monte-Carlo procedure, with a $B = 3000$. In other words, since the number $q = \binom{n}{b_n}$ of subsets of size b_n can be huge, a random selection of B of those subsets is sufficient for the construction of subsampling distributions; see Politis and Romano (1992, 1995) for more details. All calculations were performed on a Sun-workstation with the software Gauss.

We first begin with the less expensive –in terms of computer time– method based on only two subsampling distributions.

The following Table 1 presents various estimations of α, under the assumption that $\tau_n = n^{\alpha}$. The estimators $\alpha_{2,1}(t_i)$ based on quantiles (referred to as the 'quantile method' in the tables) were constructed using Theorem 2, for a single fixed point t_i (see equation (14)); here t_1, t_2, t_3 were chosen to be 99%, 95% and 90% respectively. The estimators $\alpha_{2,15}([75\%, 95\%])$, $\alpha_{2,30}([75\%, 95\%])$ and $\alpha_{2,600}([75\%, 95\%])$ (also based on Theorem 2) were constructed using 15, 30 and 600 points respectively which were chosen to be between 75% and 95%. The estimators $\alpha_{2,1}(t_1 - t_2)$, based on difference of quantiles (referred to as the 'range method'), were built using Theorem 3 (see equation (17) with $I = 2$ and $J = 1$). The estimators $\alpha_{I,J}([1\%, 25\%] - [99\%, 75\%])$, with $I = 2$ and $J = 10$ or 600 were also based on the 'range method' of Theorem 3. Here J points are taken in the interval $[99\%, 75\%]$, and J in the interval $[1\%, 25\%]$; their differences defined the ranges to be considered.

The sample sizes used were $n = 100, 1000$ and 5000. In Table 1, the two subsampling distributions have been built respectively with $b_{1,n} = 0.9 * n$ and

$b_{2,n} = 0.4 * n$. We give the mean and the standard error over 100 repetitions of the procedure. Note that, as mentioned in the Introduction, $\alpha = 1/2$, if $E_P X \neq 0$, whereas $\alpha = 1$, if $E_P X = 0$. The simulations show that the estimators perform relatively differently according to the value of α.

$E_P X = 2$						
$\alpha_{I,J}$	$n = 100$		$n = 1000$		$n = 5000$	
Quantile method	Mean	S.E.	Mean	S.E.	Mean	S.E.
$\alpha_{2,1}(99\%)$	0.531	0.071	0.508	0.063	0.5000	0.0302
$\alpha_{2,1}(95\%)$	0.530	0.072	0.498	0.056	0.5039	0.0251
$\alpha_{2,1}(90\%)$	0.536	0.077	0.496	0.072	0.5045	0.0255
$\alpha_{2,15}([75\%, 95\%])$	0.535	0.069	0.496	0.066	0.5041	0.0400
$\alpha_{2,30}([75\%, 95\%])$	0.522	0.067	0.504	0.051	0.505	0.0252
$\alpha_{2,600}([75\%, 95\%])$	0.514	0.043	0.502	0.031	0.500	0.0246
Range method						
$\alpha_{2,1}(1\% - 99\%)$	0.506	0.048	0.497	0.037	0.4978	0.0241
$\alpha_{2,1}(5\% - 95\%)$	0.513	0.046	0.495	0.033	0.4999	0.0181
$\alpha_{2,1}(25\% - 75\%)$	0.488	0.046	0.494	0.034	0.5006	0.0201
$\alpha_{2,10}([1\%, 25\%] - [99\%, 75\%])$	0.509	0.040	0.498	0.030	0.4999	0.0149
$\alpha_{2,600}([1\%, 25\%] - [99\%, 75\%])$	0.506	0.037	0.496	0.028	0.5003	0.0152

Table 1: Comparison of the estimators build with two subsampling distributions, for different sample sizes; case $\alpha = 1/2$ and $E_P X = 2$.

$E_P X = 0$						
$\alpha_{I,J}$	$n = 100$		$n = 1000$		$n = 5000$	
Quantile method	Mean	S.E.	Mean	S.E.	Mean	S.E.
$\alpha_{2,1}(99\%)$	0.843	0.134	0.872	0.124	0.893	0.120
$\alpha_{2,1}(95\%)$	0.837	0.128	0.871	0.122	0.896	0.115
$\alpha_{2,1}(90\%)$	0.843	0.135	0.858	0.137	0.874	0.124
$\alpha_{2,15}([75\%, 95\%])$	0.842	0.129	0.872	0.125	0.892	0.121
$\alpha_{2,30}([75\%, 95\%])$	0.844	0.127	0.872	0.122	0.886	0.114
$\alpha_{2,600}([75\%, 95\%])$	0.865	0.119	0.891	0.112	0.897	0.109
Range method						
$\alpha_{2,1}(1\% - 99\%)$	0.753	0.174	0.814	0.161	0.852	0.151
$\alpha_{2,1}(5\% - 95\%)$	0.723	0.200	0.812	0.167	0.839	0.151
$\alpha_{2,1}(25\% - 75\%)$	0.771	0.221	0.798	0.201	0.844	0.150
$\alpha_{2,10}([1\%, 25\%] - [99\%, 75\%])$	0.787	0.183	0.813	0.183	0.845	0.149
$\alpha_{2,600}([1\%, 25\%] - [99\%, 75\%])$	0.785	0.176	0.811	0.185	0.836	0.152

Table 2: Comparison of the estimators build with two subsampling distributions, for different sample sizes; case $\alpha = 1$ and $E_P X = 0$.

The variance of the estimators is found to be relatively large for $\alpha = 1$, in comparison to the case $\alpha = 1/2$; in the latter case, very accurate estimation of the convergence rate is possible even for very small sample sizes ($n = 30$ gives similar results!). Also, the variance seems to decrease very slowly with increasing n, probably due to the slow $\log(n)$ convergence rate obtained in Theorems 2, 3, and 4); this is particularly evident in the $\alpha = 1$ case. However, the estimates give the right order for the true α even for a relatively small sample size, and using them it is easy to discriminate between the $\sqrt{n}$ and the n case. In some situations the convergence rate (for instance for U or V statistics) is typically of the form n^{β/β_0} where β_0 is known and fixed and β is integer : in that case we may use $[[\beta_{I,J} * \beta_0]]/\beta_0$, where $[[.]]$ denotes the nearest integer, as a better estimator for $\alpha = \beta/\beta_0$. The simulations suggest that we can get the right convergence rate and obtain accurate approximations in that case.

Note that the estimations based one one t-point are relatively stable whatever the choice of the points. Clearly the estimator based on all the points beetween $[0.75, 0.95]$ allowed by the Monte-Carlo approximation gives the better results. The estimator seems to be biased downward in the $\alpha = 1$ case (and probably biased upward in the case $\alpha = 1/2$, although the bias is smaller in the latter case).

We can also attempt a comparison between the 'quantile' and the 'range' methods of estimating α. As the entries of Tables 1 and 2 indicate, the 'range' method is better (both in terms of bias, as well as variance) than the 'quantile' method in the case $\alpha = 1/2$, while the 'quantile' method is better in the case $\alpha = 1$. This phenomenon was to be expected since the case $\alpha = 1/2$ yields a symmetric (normal) distribution whose variability is well-captured by a single number, be it an inter-quartile range, a standard deviation, etc. On the other hand, the case $\alpha = 1$ yields a skewed (χ^2) distribution whose variability is best described by means of individual quantiles, and *not* ranges.

Consider now the case where we base our estimator on more than 2 subsampling distributions. Of course this may be more computer-time expensive but may be justified if more accurate results are obtainable this way. The following simulations strongly support the fact that by using a large I we gain in precision over the case $I = 2$. Moreover, it may be interesting to build several subsampling distribution and consider a linear extrapolation of these distributions as done in Bertail(1994) and Bickel *et al.* (1994) to

obtain better approximations.

Tables 3 and 4 are constructed in the same fashion as Tables 1 and 2, although now we choose to build $I = 20$ subsampling distributions with $b_{i,n} = n * i/21, i = 1, ..., 20$.

$E_P X = 2$						
$\alpha_{I,J}$	$n = 100$		$n = 1000$		$n = 5000$	
Quantile method	Mean	S.E.	Mean	S.E.	Mean	S.E.
$\alpha_{20,1}(99\%)$	0.5525	0.0185	0.5186	0.0122	0.5084	0.0066
$\alpha_{20,1}(95\%)$	0.5386	0.0167	0.5137	0.0096	0.5057	0.0056
$\alpha_{20,1}(90\%)$	0.5323	0.0189	0.5114	0.0103	0.5044	0.0064
$\alpha_{20,1}(75\%)$	0.5249	0.0344	0.5054	0.0167	0.5026	0.0096
$\alpha_{20,600}([75\%, 95\%])$	0.5302	0.0228	0.5093	0.0113	0.5037	0.0065
Range method						
$\alpha_{20,1}(1\% - 99\%)$	0.4984	0.0084	0.4996	0.0076	0.5000	0.0045
$\alpha_{20,1}(5\% - 95\%)$	0.4992	0.0058	0.5000	0.0052	0.5002	0.0035
$\alpha_{20,1}(25\% - 75\%)$	0.5015	0.0081	0.5004	0.0069	0.5006	0.0045
$\alpha_{20,10}([1\%, 25\%] - [99\%, 75\%])$	0.5003	0.0052	0.5003	0.0047	0.5003	0.0031
$\alpha_{20,600}([1\%, 25\%] - [99\%, 75\%])$	0.5009	0.0054	0.5004	0.0048	0.5003	0.0033

Table 3: Comparison of the estimators build with $I = 20$ subsampling distributions, for different sample sizes; case $\alpha = 1/2$ and $E_P X = 2$.

$E_P X = 0$						
$\alpha_{I,J}$	n=100		n=1000		n=5000	
Quantile method	Mean	S.E.	Mean	S.E.	Mean	S.E.
$\alpha_{20,1}(99\%)$	0.9006	0.0761	0.9086	0.0779	0.9204	0.0758
$\alpha_{20,1}(95\%)$	0.9146	0.0854	0.9012	0.0892	0.9098	0.0884
$\alpha_{20,1}(90\%)$	0.8941	0.0923	0.8986	0.0934	0.9071	0.0956
$\alpha_{20,1}(75\%)$	0.9128	0.0987	0.9143	0.0996	0.9248	0.1002
$\alpha_{20,600}([75\%, 95\%])$	0.8991	0.0940	0.9025	0.0960	0.9121	0.0977
Range method						
$\alpha_{20,1}(1\% - 99\%)$	0.8764	0.0984	0.8845	0.1029	0.8940	0.1007
$\alpha_{20,1}(5\% - 95\%)$	0.8574	0.1169	0.8647	0.1238	0.8698	0.1232
$\alpha_{20,1}(25\% - 75\%)$	0.8244	0.1516	0.8315	0.1550	0.8341	0.1535
$\alpha_{20,10}([1\%, 25\%] - [99\%, 75\%])$	0.8416	0.1334	0.8484	0.1383	0.8528	0.1402
$\alpha_{20,600}([1\%, 25\%] - [99\%, 75\%])$	0.8380	0.1336	0.8441	0.1420	0.8483	0.1430

Table 4: Comparison of the estimators build with $I = 20$ subsampling distributions, for different sample sizes; case $\alpha = 1$ and $E_P X = 0$.

The estimators reported in Tables 3 and 4 seem to be quite more accurate than those in Tables 1 and 2; the extra computational effort seems to be well worth it. In particular for the $\alpha = 1/2$, the results are quite precise. In the $\alpha = 1$ case the estimators based on the quantile method with $I = 20$ have half the variance in comparison to those based on $I = 2$.

Note that the (empirical) bias of the estimators of α decreases as n increases in Tables 3 and 4 as it is expected from our consistency theorems. However, it is quite apparent that in the case $\alpha = 1$ (see Table 4), the (empirical) standard errors of the estimators of α do not seem to decrease as n increases, whereas in the case $\alpha = 1/2$ (see Table 3) the standard errors do decrease. This phenomenon can be attributed to the fact that in the simulations that generated Tables 3 and 4 the subsample sizes used were proportional to n (i.e., $b_{i,n} = n * i/21, i = 1, ..., 20$), and equation (3) was not satisfied. Intuitively, to control the bias of the subsampling distribution estimator a large subsample size b_n is required, while to control its variance a small ratio b/n is needed. Thus, without letting $b_n/n \to 0$, there is no guarantee –unless of course stronger model assumptions are made– that the variance of the subsampling distribution estimator will tend to zero as n increases.

Nevertheless, in very regular cases (e.g., i.i.d. data with normal limit distributions and $\sqrt{n}$ rate of convergence) it is possible to get a consistent estimate of scale using subsampling foregoing equation (3) and replacing it with the condition that the subsample size b_n is proportional to n; see Shao and Wu (1989). Therefore, it is not surprising, in the regular case $\alpha = 1/2$ of Table 3, that the standard errors do decrease as n increases.

To empirically verify the consistency of our estimators of α under the necessary condition $b_n/n \to 0$, the simulations that resulted into Tables 3 and 4 were repeated (with different subsample sizes) and presented in Tables 5 and 6 that follow; here subsample sizes $b_{i,n}$ were were chosen such that the ratio $\frac{b_{i,n}}{n}$ is decreasing with n. In particular, for $n = 100$, the $b_{i,n}$s are the same as those used in Tables 3 and 4; thus the first two columns of Table 5 are identical to the first two columns of Table 3, and the first two columns of Table 6 are identical to the first two columns of Table 4. However, for $n = 1000$, we chose $b_{i,n} = i * n/42$, $i = 1, .., 20$, i.e, half the subsample sizes used in the middle two columns of Tables 3 and 4, and for $n = 5000$, we chose $b_{i,n} = i * n/168$, $i = 1, .., 20$, i.e, one-eighth of the subsample sizes used in the last two columns of Tables 3 and 4.

$E_P X = 2$						
$\alpha_{I,J}$	$n = 100$		$n = 1000$		$n = 5000$	
Quantile method	Mean	S.E.	Mean	S.E.	Mean	S.E.
$\alpha_{20,1}(99\%)$	0.5525	0.0185	0.5249	0.0116	0.5228	0.0107
$\alpha_{20,1}(95\%)$	0.5386	0.0167	0.5182	0.0090	0.5161	0.0085
$\alpha_{20,1}(90\%)$	0.5323	0.0189	0.5145	0.0094	0.5128	0.0086
$\alpha_{20,1}(75\%)$	0.5249	0.0344	0.5084	0.0153	0.5076	0.0125
$\alpha_{2,600}([75\%, 95\%])$	0.5302	0.0228	0.5122	0.0100	0.5112	0.0085
Range method						
$\alpha_{20,1}(1\% - 99\%)$	0.4984	0.0084	0.4992	0.0075	0.4994	0.0070
$\alpha_{20,1}(5\% - 95\%)$	0.4992	0.0058	0.4999	0.0052	0.4998	0.0050
$\alpha_{20,1}(25\% - 75\%)$	0.5015	0.0081	0.5002	0.0070	0.5003	0.0068
$\alpha_{20,10}([1\%, 25\%] - [99\%, 75\%])$	0.5003	0.0052	0.4999	0.0046	0.5000	0.0044
$\alpha_{20,600}([1\%, 25\%] - [99\%, 75\%])$	0.5009	0.0054	0.5001	0.0049	0.5000	0.0046

Table 5: Comparison of the estimators build with $I = 20$ subsampling distributions, for different sample sizes; case $\alpha = 1/2$ and $E_P X = 2$ (using decreasing ratios $\frac{b_{i,n}}{n}$).

$E_P X = 0$						
$\alpha_{I,J}$	$n = 100$		$n = 1000$		$n = 5000$	
Quantile method	Mean	S.E.	Mean	S.E.	Mean	S.E.
$\alpha_{20,1}(99\%)$	0.9006	0.0761	0.9389	0.0527	0.9719	0.0341
$\alpha_{20,1}(95\%)$	0.9146	0.0854	0.9344	0.0606	0.9732	0.0344
$\alpha_{20,1}(90\%)$	0.8941	0.0923	0.9362	0.0628	0.9752	0.0327
$\alpha_{20,15}(75\%)$	0.9128	0.0917	0.9556	0.0575	0.9864	0.0239
$\alpha_{20,600}([75\%, 95\%])$	0.8991	0.0919	0.9419	0.0614	0.9787	0.0294
Range method						
$\alpha_{20,1}(1\% - 99\%)$	0.8764	0.0984	0.9254	0.0661	0.9666	0.0397
$\alpha_{20,1}(5\% - 95\%)$	0.8574	0.1169	0.9123	0.0818	0.9642	0.0443
$\alpha_{20,1}(25\% - 75\%)$	0.8244	0.1516	0.8952	0.1076	0.9601	0.0520
$\alpha_{20,10}([1\%, 25\%] - [99\%, 75\%])$	0.841	0.1334	0.9044	0.0937	0.9623	0.0471
$\alpha_{20,600}([1\%, 25\%] - [99\%, 75\%])$	0.8386	0.1336	0.9021	0.0967	0.9618	0.0482

Table 6: Comparison of the estimators build with I =20 subsampling distributions, for different sample sizes; case $\alpha = 1$ and $E_P X = 0$ (using decreasing ratios $\frac{b_{i,n}}{n}$).

It is quite evident from the entries of Table 6 that ensuring $\frac{b_{i,n}}{n} \to 0$ fixes up the problem of nondecreasing standard errors in the $\alpha = 1$ case of Table 4. In particular, it is apparent that both bias and standard error of our estimators of α are evidently decreasing as n increases, and at a reasonably fast rate. Nevertheless, the empirical results of Table 5 are worse than those of Table 3, i.e., bias and standard error of our estimators of α decrease faster with n in Table 3 as compared to those in Table 5; this seems to indicate that in very regular cases (as in the Shao and Wu (1989) setting) the $b_{i,n}$ could (and actually should) be are taken to be proportional to n. However, in the paper at hand it is assumed that the convergence rate τ_n is unknown, and hence it is also unknown *a priori* that a regular set-up holds. Equation (3) provides a general condition that ensures consistency of our estimators of α without special model assumptions.

It is also noteworthy that the comparison between the range and the quantile methods, which was attempted on the basis of the entries of Tables 1 and 2, seems to still hold after examining all presented simulations: the range method seems to be best in the regular case (normal limit distribution and $\sqrt{n}$ rate of convergence), whereas in the not-regular case (skewed limit distribution and not $\sqrt{n}$ rate of convergence) the quantile method is preferable. Of course, more simulations involving other perhaps more complicated examples may be needed to confirm the reported empirical results. It is clear however even from our limited simulation experiment that the estimation of α using subsampling is not infeasible in finite samples, and may prove to be an important practical tool.

5 Technical proofs

Proof of Theorem 1 First note that under conditions (i) and (ii) it follows that $u_i = o_P(1)$ in equation (8). Therefore (9) yields

$$\alpha_I = \alpha + \frac{\sum_{i=1}^{I}(u_i - \overline{u})(\log(b_{i,n}) - \overline{\log})}{\sum_{i=1}^{I}(\log(b_{i,n}) - \overline{\log})^2},$$

where $\overline{u} = I^{-1}\sum_{i=1}^{I} u_i$. But since $b_{i,n} = n^{\beta_i}$, $i = 1, .., I$, then $\overline{\log} = I^{-1}\sum_i \beta_i \log n = A \log n$, and it is easy to see that

$\sum_{i=1}^{I}(\log(b_{i,n}) - \overline{\log})^2 = (\sum_i(\beta_i - I^{-1}\sum_k \beta_k)^2)(\log n)^2 = B(\log n)^2$, for some A, B constants; thus, for fixed I, we have $\alpha_I = \alpha + o_P((\log n)^{-1})$. QED

Proof of Lemma 1 Let $\epsilon > 0$; by (5), we have that $\Pr_P\{|K_{b_n}(x \mid \underline{X}_n, \tau_\cdot) - K(x,P)| < \epsilon\} \to 1$, uniformly in x.

So let $z = K_{b_n}^{-1}(t-\epsilon \mid \underline{X}_n, \tau_\cdot)$, thus with probability tending to one

$$K_{b_n}(z \mid \underline{X}_n, \tau_\cdot) \geq t - \epsilon \Rightarrow K(z,P) \geq t - 2\epsilon \Rightarrow z \geq K^{-1}(t-2\epsilon, P).$$

Similarly, let $y = K^{-1}(t,P)$, thus with probability tending to one

$$K(y,P) \geq t \Rightarrow K_{b_n}(y \mid \underline{X}_n, \tau_\cdot) \geq t - \epsilon \Rightarrow y \geq K_{b_n}^{-1}(t-\epsilon; \mid \underline{X}_n, \tau_\cdot).$$

Hence, for any t and any $\epsilon > 0$, we have that

$$K^{-1}(t-2\epsilon, P) \leq K_{b_n}^{-1}(t-\epsilon \mid \underline{X}_n, \tau_\cdot) \leq K^{-1}(t,P)$$

with probability tending to one. Now let $\epsilon \to 0^+$ to conclude that $K_{b_n}^{-1}(t \mid \underline{X}_n, \tau_\cdot) = K^{-1}(t,P) + o_P(1)$. QED

Proof of Theorem 2 First note that equations (2) and (3) follow from our assumption on the $b_{i,n}$'s, hence equation (5) –with any of the $b_{i,n}$'s used as subsample size– is seen to hold as well from the results in Politis and Romano (1992, 1995).

Now note that, under (5), Lemma 1 validates the ANOVA equation (15), and in particular the fact that $u_{i,j} = o_P(1)$; therefore (16) yields

$$\alpha_{I,J} = \alpha + \frac{\sum_{i=1}^{I}(u_{i,\cdot} - \overline{u})(\log(b_{i,n}) - \overline{\log})}{\sum_{i=1}^{I}(\log(b_{i,n}) - \overline{\log})^2},$$

where $\overline{u} = (IJ)^{-1}\sum_{i=1}^{I}\sum_{j=1}^{J} u_{i,j}$. As in the proof of Theorem 1, note that $\overline{\log} = A\log n$, and $\sum_{i=1}^{I}(\log(b_{i,n}) - \overline{\log})^2 = B(\log n)^2$, for some A, B constants; thus, for fixed I and J, we have $\alpha_{I,J} = \alpha + o_P((\log n)^{-1})$. QED

Proof of Theorem 3 Similar to the proof of Theorem 2.

Proof of Lemma 2 Similar to the proof of Lemma 1.

Proof of Theorem 4 Similar to the proof of Theorem 2 using Lemma

2 and the results of Politis, Romano, and You (1993). QED

Proof of Theorem 5 Let x be a real number and note that

$$K_{b_n}(x \mid \underline{X}_n, \hat{\tau}_\cdot) \equiv q^{-1} \sum_{i=1}^{q} 1\{b_n^{\hat{\alpha}}(T_{b_n,i} - T_n) \leq x\}$$

$$= q^{-1} \sum_{i=1}^{q} 1\{b_n^{\hat{\alpha}}(T_{b_n,i} - \theta) + b_n^{\hat{\alpha}}(T_n - \theta) \leq x\}.$$

Put $U_n(x) = q^{-1} \sum_{i=1}^{q} 1\{b_n^{\hat{\alpha}}(T_{b_n,i} - \theta) \leq x\}$ and $E_n = \{b_n^{\hat{\alpha}} | T_n - \theta| \leq \epsilon\}$. Since $\hat{\alpha} = \alpha + o_P((\log n)^{-1})$, it follows that

$$b_n^{\hat{\alpha}} = b_n^{\alpha} \exp(\log b_n(\hat{\alpha} - \alpha)) = b_n^{\alpha}(1 + o_P(1)).$$

As in Politis and Romano (1992, 1995), it follows that (2) and (3) imply that $P(E_n) \underset{n \to \infty}{\longrightarrow} 1$; hence, with probability tending to one,

$$U_n(x - \epsilon) \leq K_{b_n}(x \mid \underline{X}_n, \hat{\tau}_\cdot) \leq U_n(x + \epsilon).$$

Now it suffices to show that $U_n(x)$ converges to $K(x, P)$ in probability; for in that case we could let $\epsilon \to 0$ in the above inequality to conclude that $K_{b_n}(x \mid \underline{X}_n, \hat{\tau}_\cdot) \to K(x, P)$ in probability. The uniform convergence (23) would then be a consequence of Polya's theorem, since $K(x, P)$ is continuous.

To show $U_n(x)$ converges to $K(x, P)$ in probability, first note that

$$EU_n(x) = E(1\{b_n^{\hat{\alpha}}(T_{b_n,1} - \theta) \leq x\}) = \Pr\{b_n^{\hat{\alpha}}(T_{b_n,1} - \theta) \leq x\}\};$$

but $b_n^{\hat{\alpha}}(T_{b_n,i} - \theta) = b_n^{\alpha}(T_{b_n} - \theta)(1 + o(1))$ and it follows that $EU_n(x) \to K(x, P)$.

Now let us look at the two cases that correspond to the two different constructions of the subsampling distribution (4).

General case (strong mixing data). Now, similarly to the proof of Theorem 3.1 in Politis and Romano (1992, 1995), it follows that

$$Var(U_n(x)) \leq \frac{2b_n + 1}{n} + \frac{8}{n} \sum_{k=1}^{n} \alpha_X(k);$$

under the assumption of strong mixing it is now apparent that the Cesaro-sum $\frac{1}{n} \sum_{k=1}^{n} \alpha_X(k) \to 0$, and using (2) we conclude that $Var(U_n(x)) \to 0$,

and therefore $U_n(x)$ converges to $K(x,P)$ in probability.

Special case (i.i.d. data). As in Theorem 2.1 in Politis and Romano (1992, 1995), Hoeffding's inequality gives

$$\Pr{}_P\{U_n(x) - EU_n(x) \geq \epsilon\} \leq \exp(-2nb_n^{-1}\epsilon^2)$$

which, in combination with $EU_n(x) \to K(x,P)$, shows that $U_n(x)$ converges to $K(x,P)$ in probability.

Finally note that the proof of (24) is very similar to the proof of Theorem 1 of Beran (1984) given result (23) and noting that $\hat{\tau}_n/\tau_n \to 1$ in probability, since $\hat{\alpha} = \alpha + o_P((\log n)^{-1})$. QED

Proof of Theorem 6 Similar to the proof of Theorem 5 using Theorem 4 in conjuction with the results of Politis, Romano, and You (1993). QED

Acknowledgment

Many thanks are due to Prof. E. Carlstein (Univ. of North Carolina) for his interesting question after a talk by one of the authors at the IMS Meeting in Chapel Hill (June 1994), and to Prof. Peter Bickel (Univ. of California, Berkeley) for his astute reply that foreshadowed our result on the inter-quartile range in Theorem 3.

References

[1] Barbe, Ph. and Bertail, P.(1995). *The Weighted Bootstrap*, Lecture Notes in Statistics #98, Springer Verlag, New York.

[2] Becker, R.A., Chambers, J.M., Wilks, A.R. (1988) *The New S Language: A Programming Environment for Data Analysis and Graphics*, Wadsworth & Brooks/Cole.

[3] Beran, J. (1994). *Statistical Methods for Long Memory Processes*, Chapman and Hall, New York.

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

[5] Bertail P. (1994). Second order properties of an extrapolated bootstrap without replacement: the i.i.d. and the strong mixing cases, CORELA-INRA working paper n°9407, submitted.

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

[7] Bickel, P., Götze, F. and van Zwet, W.R. (1994). Resampling fewer than n observations: Gains, losses and remedies for losses. Technical Report n^0 419, University of California, Berkeley.

[8] Bingham, N.H., Goldie, C.M. and Teugels, J.L. (1987). *Regular Variation*. Cambridge University Press.

[9] Carlstein, E.(1986). The use of subseries values for estimating the variance of a general statistics from a stationary sequence. *Ann. Statist.* **14** 1171-1179.

[10] Datta, S. (1995). On a modified bootstrap for certain asymptotically nonormal statistics. *Statist. Prob. Letters* (to appear).

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

[12] Efron, B. (1994). Comment on the paper "Bootstrap: More than a stab in the dark?" by G. A. Young, *Statist. Science*, vol. 9, no. 3, 396-398.

[13] Engle, R.F., Granger, C.W.J., Rice, J., and Weiss, A. (1986). Semiparametric estimates of the relation between weather and electricity sales. *J.Amer.Statist.Assoc.*, 81, 310-320.

[14] de Haan, L. (1970). *On Regular Variation and its Application to the Weak Convergence of Sample Extremes*. Math. Centre Tract 32, Amsterdam.

[15] Künsch, H.R.(1989). The jackknife and the bootstrap for general stationary observations. *Ann. Statist.* **17** 1217-1241.

[16] Liu R. and Singh K.(1992). Moving blocks jackknife and bootstrap capture weak dependence. In *Exploring The Limits of The Bootstrap.* Ed. Le Page R and Billard, L., John Wiley, New York.

[17] Politis, D.N. and Romano, J.P.(1992). A general theory for large sample confidence regions based on subsamples under minimal assumptions, Technical Report No. 399, Department of Statistics, Stanford University.

[18] Politis, D.N. and Romano, J.P.(1993). 'Estimating the Distribution of a Studentized Statistic by Subsampling', *Bulletin of the International Statistical Institute*, 49th Session, Firenze, August 25 - September 2, 1993, Book 2, pp. 315-316.

[19] Politis, D.N. and Romano, J.P.(1994). 'The Stationary Bootstrap', *J. Amer. Statist. Assoc.*, vol. 89, No. 428, 1303-1313.

[20] Politis, D.N. and Romano, J.P.(1995). 'Large Sample Confidence Regions Based on Subsamples under Minimal Assumptions', to appear in *Ann. Statist.* in 1995.

[21] Politis, D.N., J.P.Romano, L.You (1993). 'Uniform confidence bands for the spectrum based on subsamples', in *Computing Science and Statistics, Proceedings of the 25th Symposium on the Interface*, San Diego, California, April 14-17, 1993, (M. Tarter and M. Lock, eds.), The Interface Foundation of North America, pp. 346-351.

[22] Rosenblatt, M. (1984). Asymptotic normality, strong mixing and spectral density estimates, *Ann. Prob.*, 12, 1167-1180.

[23] Shao, J. and Wu, C.F.J. (1989). A general theory for jackknife variance estimation. *Ann. Statist.* **17**, 1176-1197.

[24] Sherman, M. (1992). Subsampling and asymptotic normality for a general statistic from a random field, Ph. D. Thesis, Dept. of Statistics, Univ. of North Carolina, Chapel Hill (Report # 2081, Mimeo Series).

[25] Sherman, M. and Carlstein, E. (1994), Nonparametric estimation of the moments of a general statistic computed from spatial data, *J. Amer. Statist. Assoc.*, **89**, No. 426, pp.496-500.

[26] Wu, C.F.J. (1990). On the asymptotic properties of the jackknife histogram. *Ann. Statist.* **18**, 1438-1452.