

Second-order properties of an extrapolated bootstrap without replacement under weak assumptions

PATRICE BERTAIL

INRA-CORELA, 65 Boulevard de Brandebourg, 94205 Ivry/Seine, France

This paper shows that a straightforward extrapolation of the bootstrap distribution obtained by resampling without replacement, as considered by Politis and Romano, leads to second-order correct confidence intervals, provided that the resampling size is chosen adequately. We assume only that the statistic of interest T_n , suitably renormalized by a regular sequence, is asymptotically pivotal and admits an Edgeworth expansion on some differentiable functions. The results are extended to a corrected version of the moving-block bootstrap without replacement introduced by Künsch for strong-mixing random fields. Moreover, we show that the generalized jackknife or the Richardson extrapolation of such bootstrap distributions, as considered by Bickel and Yahav, leads to better approximations.

Keywords: bootstrap; Edgeworth expansion; generalized jackknife; random fields; Richardson extrapolation; strong mixing; undersampling

1. Introduction

Politis and Romano (1994) give a general theory, under very weak assumptions, for the construction of first-order asymptotically valid approximations and confidence intervals based on a resampling scheme. The idea is to approximate the sampling distribution of a statistic T_n based on n observations by the empirical distribution of the suitably renormalized values of the statistic computed on all the subsets of size $b_n < n$. Although this undersampling scheme may be seen as a delete- d_n jackknife with $d_n = n - b_n$ (see Wu 1990; Shao and Wu 1989), for reasons that will appear more clearly later, we shall call this method the bootstrap without replacement (WoRB). Politis and Romano's (1994) work is based on two basic ideas – one being undersampling, the other the absence of replacement. Indeed, it is well known from previous work that failure of Efron's (1979) bootstrap with a resampling size equal to the original sample size may be solved in some cases by undersampling (see, for instance, Bretagnolle 1983; Bertail 1992). However, such a modification may not succeed in other cases. One of the reasons for this failure is the problem of ties induced by the bootstrap algorithm (consider, for instance, the cases of U -statistics in Bickel and Freedman 1981). The approach of Politis and Romano (1994), which goes back to Wu's (1990) jackknife histogram,

avoids these problems and leads to a kind of ‘general’ asymptotic validity of the WoRB. Extensions to strong-mixing random fields using the ideas of the moving-block bootstrap introduced by Künsch (1989) and Liu and Singh (1992) are also possible.

However, the main drawback of this method is that it does not provide second-order properties as does Efron’s bootstrap in regular cases. An explicit correction, as proposed in Tu (1992), may be used to overcome this deficiency; yet this requires some additional regularity conditions on T_n . Moreover, it is not automatic. The purpose of this paper, which extends Bertail (1993), is to give a modified version of the WoRB, in the independent identically distributed case and in a strong-mixing setting; this version yields a ‘general’ second-order asymptotic validity under the very weak assumption that the statistic T_n admits an Edgeworth expansion on some differentiable functions. The proposed transformation of the WoRB distribution has been independently investigated by Booth and Hall (1993). It is based on the observation that if the size used to construct the jackknife pseudo-values is b_n , then the error of the approximation of the jackknife histogram is of order $O(b_n^{-1/2})$ and thus cannot coincide with the true distribution up to the second order. However, if one mixes the normal approximation and the jackknife histogram in proportion $(b_n/n)^{1/2}$ and $(1 - (b_n/n)^{1/2})$, then one obtains second-order properties. In fact, this transformation is already contained in Bickel and Yahav’s (1988) paper on Richardson extrapolations, whose aim was to show how it is possible to construct an accurate approximation of the bootstrap distribution at a lower cost, using undersampling. The weighted scheme proposed along these lines may itself be seen as a generalized jackknife transformation (see Gray *et al.* 1972) of the asymptotic distribution and the jackknife histogram that yields a better approximation of the true distribution. In some sense, this is a transposition of bias reduction techniques to ‘order reduction’ techniques. We shall see that this idea may be applied to a very large category of statistics and may be generalized to obtain approximations correct up to $O(n^{-1+\epsilon})$, for any $\epsilon > 0$ in regular cases. We shall see that the present results may be useful in many applications and lead to new second-order results for many non-regular functionals. In particular, we obtain second-order correct confidence intervals for some econometric quantities that are asymptotically chi-squared distributed, for quantiles and U -quantiles, as well as for degenerate U statistics and von Mises functionals, for which the second-order properties of the usual bootstrap have never been examined or may even fail.

The layout of this paper is as follows. To motivate our main result and introduce the notation, we shall begin in Section 2 with general considerations on the case of the mean. Connections with other bootstrap methods are also pointed out. Section 3 gives the basic ideas and the main theorem on the second-order validity of the corrected bootstrap in the i.i.d. real case under very weak assumptions. The results are illustrated by many examples. As a particular illustration we show that it is possible to choose a resampling size which improves the results of Hall (1992a) in nonparametric density estimation and allows for the construction of more accurate confidence intervals. Some extensions to more general spaces are also considered. Section 4 extends the results of Section 3 to strong-mixing random fields using a modified version of Künsch’s (1989) moving-block bootstrap. The results are illustrated by some applications in robust estimation in autoregressive moving-average models and spectral density estimation. Section 5 exposes some refinements. In particular,

we shall show that computing the generalized jackknife of several WoRBs for different values of b_n improves the approximations and reduces the cost of computation. In Section 6 we give some simulation results, in the case of quantile estimation, for which the usual bootstrap fails to be second-order correct.

2. Bootstrapping without replacement

2.1. A METHOD OF GENERAL INTEREST

Let $\{X_1, X_2, \dots, X_n\}_{n \in \mathbb{N}}$ be sequences of i.i.d. random variables with distribution P and consider a statistic $T_n = T_n(X_1, X_2, \dots, X_n)$ estimating some real parameter $\theta(P)$ (we shall later see that the assumption that $\theta(P)$ is real may easily be relaxed). Consider also $S_n^2 \equiv S_n^2(X_1, X_2, \dots, X_n)$ a normalizing sequence converging in probability to a constant $\sigma^2 > 0$. This includes the case $S_n^2 = 1$. Assume that T_n converges to $\theta(P)$ at a rate τ_n . To fix some notation, define

$$K_n(x, P) \equiv P(\tilde{T}_n \leq x)$$

with

$$\tilde{T}_n \equiv \tilde{T}_n(X_1, X_2, \dots, X_n) \equiv H(\tau_n S_n^{-1}(T_n - \theta(P))),$$

where H is either the identity or the absolute value. The function H is introduced to consider simultaneously the construction of unilateral and bilateral confidence intervals. Consider now the following assumptions:

Assumption A1. \tilde{T}_n converges in distribution to a random variable having a non-degenerate continuous distribution function $K(x, P)$ with a uniformly bounded first derivative.

Assumption A1'. A1 holds under the weaker condition that $K(x, P)$ is continuously differentiable on a compact neighbourhood of the quantile of order $1 - \alpha$:

$$c(1 - \alpha) \equiv \inf \{x: K(x, P) \geq 1 - \alpha\}.$$

Let $b_n < n$ be the size of the samples taken without replacement from the observed population $\mathbf{X}_n = (X_1, X_2, \dots, X_n)$. These subsets may take only

$$N_n = \binom{n}{b_n}$$

possible values denoted by $\mathcal{X}_{b_n, i}$, $i = 1, \dots, N_n$. The WoRB distribution of the root \tilde{T}_n is defined by

$$\bar{K}_{b_n}(x|\mathbf{X}_n) \equiv N_n^{-1} \sum_{i=1}^{N_n} I\{H(\tau_{b_n}(T_{b_n}(\mathcal{X}_{b_n, i}) - T_n)/S_{b_n}(\mathcal{X}_{b_n, i})) \leq x\},$$

where in the following $I\{B\}$ is the indicator of the event B . The fact that $\bar{K}_{b_n}(x|\mathbf{X}_n)$ is almost

a U statistic with finite support leads to the result that if $\tau_{b_n}/\tau_n \rightarrow_{n \rightarrow \infty} 0$, $b_n/n \rightarrow_{n \rightarrow \infty} 0$ with $b_n \rightarrow_{n \rightarrow \infty} \infty$ then under Assumption A1

$$\sup_x |\bar{K}_{b_n}(x|\mathbf{X}_n) - K_n(x, P)| \rightarrow_{n \rightarrow \infty} 0 \text{ in probability}$$

without assuming any regularity condition on T_n (see Politis and Romano 1994). Direct application of Beran's (1984) results leads to first-order correct bootstrap confidence intervals under Assumption A1'. Extensions to general spaces to non-i.i.d. sequences (typically α -mixing sequences) are almost immediate using a version of the moving-block bootstrap without replacement introduced by Künsch (1989). Notice that, for the results to be useful in practice, the convergence rate has to be known, which may be a serious problem if the class of probability P under consideration is too big (consider, for instance, the case of $T_n = \min_{1 \leq i \leq n} X_i$ or the case of the mean when no assumptions are made on the variance). Using the 'right' convergence rate means that we have to adapt the WoRB to the model (but note that it is also the case with the usual bootstrap in the same situations). Bertail *et al.* (1995) recently showed that it is possible to circumvent this problem by first estimating consistently the convergence rate via undersampling and then using this estimator to construct the WoRB distribution. However, for many statistical applications the assumption concerning the convergence rate is rather weak. We shall consider that case now.

2.2. CONNECTIONS WITH OTHER BOOTSTRAPS

Connections with the generalized weighted bootstrap will probably make clearer the link between the WoRB, Efron's bootstrap and the Monte Carlo step. Mason and Newton (1992) have introduced a generalization of Efron's bootstrap that yields similar and sometimes better results in many situations (see Barbe and Bertail 1995). The idea is to weight each observation by stochastic exchangeable weights $(w_{n,i})_{1 \leq i \leq n}$ and to determine the distribution of the weighted statistic according to the law of the weights, conditioned on the original sample. Specific choices of weights lead to well-known resampling plans including the Bayesian bootstrap, Lo's (1991) Bayesian bootstrap clones and Efron's bootstrap. The WoRB actually belongs to this category since it is possible to choose weights defined for $i_k \in \{0, 1\}$, $k = 1, \dots, n$, such that $\sum_{k=1}^n i_k = b_n$ by

$$P_W(w_{n,1} = i_1, w_{n,2} = i_2, \dots, w_{n,n} = i_n) = N_n^{-1}. \quad (1)$$

In the case of the mean, $T_n \equiv \bar{X}_n = n^{-1} \sum_{i=1}^n X_i$ estimating $\theta(P) = E_P X$, under the assumption that $\sigma^2 = E_P(X - E_P X)^2 < \infty$, Theorem 2.1 of Mason and Newton (1992) applies (see their conditions 2.12 and 2.13 with weights defined by $Y_{i,n} = w_{n,i} \{b_n(n - b_n)\}^{-1/2}$). When $b_n/n \rightarrow_{n \rightarrow \infty} 0$, we obtain, for the mean, the validity of the WoRB as stated in Politis and Romano (1994). The asymptotic validity of the WoRB in some other regular cases (typically U or V statistics) follows directly from this representation. The approach of Politis and Romano (1994) is, of course, more general and more direct. However, conditions for second- and third-order validity developed in Barbe and Bertail (1995) fail for these weights essentially because of the 'undersampling' size $\sum_{k=1}^n w_{n,k} = b_n$.

2.3. SECOND-ORDER INVALIDITY

To motivate our results, consider the example of the mean studied in Bertail (1993) and Booth and Hall (1993). Assume that $E_P X^4 < \infty$ and choose $S_n^2 \equiv n^{-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$. Under regularity conditions (see Bhattacharya and Ghosh 1978), we have the traditional Edgeworth expansion:

$$P\{n^{1/2}S_n^{-1}(\bar{X}_n - \theta(P)) \leq x\} = \Phi(x) + n^{-1/2}k_3/6(2x^2 + 1)\phi(x) + o(n^{-1/2}), \quad (2)$$

where k_3 is the skewness of X . Babu and Singh's (1985) Edgeworth expansion for sampling without replacement from a finite population, with $b_n/n \rightarrow_{n \rightarrow \infty} 0$, yields the Edgeworth expansion for the WoRB studentized distribution

$$\bar{K}_{b_n}(x|\mathbf{X}_n) = \Phi(x(1 - b_n/n)^{-1/2}) + b_n^{-1/2}k_{3,n}/6(2x^2 + 1)\phi(x) + o(b_n^{-1/2}), \quad (3)$$

where $k_{3,n}$ is the empirical skewness. If $b_n/n = o(b_n^{-1/2})$ and since $k_{3,n} - k_3 = o_p(1)$, a Taylor expansion of (3) immediately leads to

$$\bar{K}_{b_n}(x|\mathbf{X}_n) = \Phi(x) + b_n^{-1/2}k_3/6(2x^2 + 1)\phi(x) + o_p(b_n^{-1/2}). \quad (4)$$

Notice that the condition $b_n/n = o(b_n^{-1/2})$ simply means that b_n should be small enough, typically of order $o(n^{2/3})$, to take into account the error induced by the undersampling scheme. Of course, the two Edgeworth expansions (2) and (4) fail to match as they do for the usual bootstrap. It is clear that $\bar{K}_{b_n}(x|\mathbf{X}_n)$ is not a good approximation: the order of the error $O(b_n^{-1/2})$ is worse than $O(n^{-1/2})$ obtained with the usual Gaussian approximation. However, a combination of these two different approximations leads to a better one. Indeed, putting

$$\bar{K}_n^*(x|\mathbf{X}_n) \equiv (b_n/n)^{1/2}\bar{K}_{b_n}(x|\mathbf{X}_n) + (1 - (b_n/n)^{1/2})\Phi(x),$$

which is actually a particular case of Richardson's extrapolations (see Isaacson and Keller 1966, p. 188) studied in Bickel and Yahav (1988), we immediately see that $\bar{K}_n^*(x|\mathbf{X}_n)$ is second-order correct.

Notice that $\bar{K}_n^*(x|\mathbf{X}_n)$ may also be interpreted as the generalized jackknife estimator of the two non-second-order correct approximations $\bar{K}_{b_n}(x|\mathbf{X}_n)$ and $\Phi(x)$ (see Gray *et al.* 1972) defined by:

$$\det \begin{pmatrix} \bar{K}_{b_n}(x|\mathbf{X}_n) & b_n^{-1/2} - n^{-1/2} \\ \Phi(x) & -n^{-1/2} \end{pmatrix} / \det \begin{pmatrix} 1 & b_n^{-1/2} - n^{-1/2} \\ 1 & -n^{-1/2} \end{pmatrix}.$$

The purpose of the next section is to prove that second-order correctness holds if the size of the resampling is chosen adequately, under the minimal assumption that the statistic under consideration, T_n , suitably renormalized, admits an Edgeworth expansion on an arbitrary continuously differentiable distribution.

3. Second-order properties of the extrapolated WoRB: the i.i.d. case

3.1. NOTATION AND ASSUMPTIONS

Let T_n , S_n^2 and \tilde{T}_n be as in Section 2. We shall assume that \tilde{T}_n has an Edgeworth expansion and, for ease of notation, we shall index the assumptions by the order of the Edgeworth expansion, I .

Assumption A2[I]. \tilde{T}_n admits an Edgeworth expansion of the form:

$$K_n(x, P) \equiv P(\tilde{T}_n \leq x) = E_n^{(I-1)}(x, P) + O(f_I(n)^{-1}), \quad (5)$$

uniformly in x , where

$$E_n^{(I-1)}(x, P) \equiv K(x, P) + \sum_{i=1}^{I-1} f_i(n)^{-1} p_i(x, P).$$

The $\{f_i(\cdot)\}_{1 \leq i \leq I-1}$ are an increasing sequence of positive non-decreasing functions converging to $+\infty$ at $+\infty$. The $\{p_i(x, P)\}_{1 \leq i \leq I-1}$ are uniformly continuous functions of x . For $I > 2$, we also assume that the $p_i(x, P)$ have uniformly bounded derivatives.

For the purpose of confidence intervals, only the following assumption is needed:

Assumption A2'[I]. Assumption A2[I] holds only on a neighbourhood of $c(1 - \alpha)$.

In some applications, we may not have precise control over the order of the Edgeworth expansion of the statistic T_n . In those cases it will not be possible to control more precisely the rate of the corrected WoRB approximation. For these reasons, we shall alternatively make the following assumptions:

Assumption B2[I - 1]. \tilde{T}_n admits an Edgeworth expansion of the form:

$$K_n(x, P) = E_n^{(I-1)}(x, P) + o(f_{I-1}(n)^{-1}), \quad (6)$$

where $f_i(\cdot)$ and $p_i(x, P)$, $i = 1, \dots, I - 2$ are as in Assumption A2[I], but $p_{I-1}(x, P)$ is only assumed to be uniformly continuous.

Assumption B2'[I - 1]. Assumption B2[I - 1] holds only on a compact neighbourhood of $c(1 - \alpha)$.

In many applications $K(x, P)$ is the cdf of the normal distribution or a chi-square distribution and the $p_i(x, P)$ are polynomials multiplied by the density of the normal or the chi-square distribution. In such cases, the functions $f_i(n)$ are known and conditions A2[I], A2'[I] or alternatively B2[I - 1] and B2'[I - 1] may be easily checked.

The following assumptions ensure that b_n is chosen adequately according to the functions f_i .

Assumption A3[i]. b_n is such that $b_n \rightarrow_{n \rightarrow \infty} \infty$ and $(b_n/n)f_i(b_n)^2 \log(f_i(b_n)) \rightarrow_{n \rightarrow \infty} 0$.

Assumption A4[i]. b_n is chosen such that $\tau_{b_n}/\tau_n \rightarrow_{n \rightarrow \infty} 0$ and $\tau_{b_n}f_i(b_n)|T_n - \theta(P)| \rightarrow_{n \rightarrow \infty} 0$, in probability.

Notice that Assumption A4[i] is satisfied if $\tau_{b_n}f_i(b_n)/\tau_n \rightarrow_{n \rightarrow \infty} 0$ under Assumption A1 (or A1').

Assumption A5[i]. b_n satisfies $\sum_{n>0} f_i(b_n) \exp[-nv_n\{b_n f_i(b_n)^2\}^{-1}] < \infty$, for some positive sequence v_n converging to 0.

As we shall later see (Remarks 1–3 and Examples 1–5), it is easy to find b_n in practical situations.

Finally, we need some control over the convergence of S_n^2 . The following assumption ensures that the tail of the distribution of S_n^2 is not too big and does not induce any perturbations on the WoRB distribution when one undersamples and standardizes by the variance estimated on the subsamples.

Assumption A6[i]. For some $\epsilon \in]0, 1/2[$, as n converges to ∞ ,

$$P(S_n^2/\sigma^2 \leq \epsilon) = o(f_i(n)^{-1}).$$

Assumption A6[i] is very weak since in many applications $f_i(n) = o(n^{-\beta i})$, for some $\beta > 0$ and $P\{S_n^2/\sigma^2 \leq \epsilon\} = o(n^{-r})$ for any $r > 0$ and $\epsilon < 1/2$. It is automatically satisfied if an exponential inequality holds for S_n^2 . If it does not hold, one may also use a truncated estimator, $\tilde{S}_n^2 = \max(S_n^2, \eta)$, $\eta > 0$. Notice that for the statement of Theorem 1 the standardization by S_n is of no real importance since we may choose $S_n = 1$ so that Assumption A6[i] is trivially satisfied. It will be important if we want to construct a pivotal quantity, i.e. such that $K(\cdot, P)$ is independent of P .

3.2. THE BASIC RESULT

The main results of this paper are based on the fact that, on the preceding assumptions, the WoRB distribution admits an Edgeworth expansion

$$\bar{K}_{b_n}(x|\mathbf{X}_n) = E_{b_n}^{(I-1)}(x, P) + O_P(f_I(b_n)^{-1}). \quad (7)$$

This result may look strange to those accustomed to Efron's bootstrap. Indeed, it is well known that the bootstrap distribution, say $K_n^E(x|\mathbf{X}_n)$, may be interpreted as an empirical Edgeworth expansion (see Bhattacharya and Qumsiyeh 1989). In contrast, (7) is equal to the original expansion developed on b_n instead of n . To explain this phenomenon, let Assumption

A2[3] hold and consider the case where $f_i(x) = x^{i/2}$. In many situations it is possible to show (see, for instance, Hall 1992b) that the bootstrap distribution is such that

$$K_n^E(x|\mathbf{X}_n) = E_n^{(2)}(x, P_n) + O_P(n^{-3/2}) = K(x) + \sum_{i=1}^2 n^{-i/2} p_i(x, P_n) + O_P(n^{-3/2}),$$

where the $p_i(x, P_n)$, $i = 1, 2$, are the empirical version of the $p_i(x, P)$ and the asymptotic distribution $K(x) = K(x, P)$ is independent of P . Since in most regular situations we have

$$p_i(x, P_n) - p_i(x, P) = O_P(n^{-1/2}), \tag{8}$$

the bootstrap distribution matches the true one up to $O_P(n^{-1})$. Considering the WoRB for general b_n , one may also obtain by direct methods – see, for instance, Babu and Singh (1985) for the mean – that the WoRB distribution admits an Edgeworth expansion of the type

$$\bar{K}_{b_n}(x|\mathbf{X}_n) = E_n^{(2)}(x, P_n) + O_P(b_n n^{-1} + b_n^{-3/2} + b_n^{-1/2} n^{-1/2}).$$

Assumptions A3[3] and A4[3] are satisfied if, for instance, we choose $b_n = n^{1/6}$ and we obtain for this choice and under (8):

$$\begin{aligned} \bar{K}_{b_n}(x|\mathbf{X}_n) &= E_{b_n}^{(2)}(x, P_n) + O(n^{-1/4}) \\ &= E_{b_n}^{(2)}(x, P) + O(n^{-1/4}). \end{aligned}$$

What is meant is that the order of the whole approximation is so bad that it makes no difference whether we use the empirical or the non-empirical version! Of course, this approximation is very poor. The scheme proposed in (18) and further in Section 5 overcomes this deficiency. We are now ready to state our basic result.

Theorem 1. (i) Under Assumptions A1, A2[I], A3[I], A4[I] and A6[I], the WoRB distribution admits an Edgeworth expansion

$$\bar{K}_{b_n}(x|\mathbf{X}_n) = E_{b_n}^{(I-1)}(x, P) + O_P(f_I(b_n)^{-1}), \tag{9}$$

uniformly in x .

(ii) Under Assumptions A1, B2[I – 1], A3[I – 1], A4[I – 1] and A6[I – 1] we have uniformly in x

$$\bar{K}_{b_n}(x|\mathbf{X}_n) = E_{b_n}^{(I-1)}(x, P) + o_P(f_{I-1}(b_n)^{-1}). \tag{10}$$

(iii) Under Assumptions A1', A2'[I] (B2'[I – 1]), A3[I] (A3[I – 1]) and A4[I] (A4[I – 1]), equation (9) (equation (10)) holds on a neighbourhood of $c(1 - \alpha)$.

(iv) If in addition we assume A5[I] (A5[I – 1]) and that A4[I] (A4[I – 1]) holds with probability one, then the results hold with probability one.

Proof. The proof is similar for (i) and (ii) (replace f_I by f_{I-1} and $O(\cdot)$ by $o(\cdot)$ in the proofs). Part (i) may be directly deduced from the following two lemmas.

Lemma 1. Under Assumptions A1 (A1'), A2[I] (A2'[I]) and A3[I],

$$U_{b_n}(x|\mathbf{X}_n) = N_n^{-1} \sum_{i=1}^{N_n} I_{\{\tau_{b_n}(T_{b_n}(X_{b_{n,i}}) - \theta(P))/S_{b_n}(X_{b_{n,i}}) \leq x\}}$$

admits with probability one an Edgeworth expansion

$$U_{b_n}(x|\mathbf{X}_n) = E_{b_n}^{(l-1)}(x, P) + O_P(f_I(b_n)^{-1}), \quad (11)$$

uniformly in x (uniformly on a neighbourhood of $c(1 - \alpha)$).

Proof. For any real-valued right-continuous distribution G , denote $G(x^-) = \lim_{y \rightarrow x, y < x} G(x)$ and define, for any $1 \leq k_n \leq n$ and $1 \leq j \leq k_n$,

$$x_{j,k_n} = \sup\{x, j/k_n \geq K_{b_n}(x^-, P)\}.$$

Then straightforward arguments show (see Chow and Teicher 1988, p. 265)

$$\begin{aligned} \sup_x |U_{b_n}(x|\mathbf{X}_n) - K_{b_n}(x, P)| &\leq \max_{0 \leq j \leq k_n} |U_{b_n}(x_{j,k_n}|\mathbf{X}_n) - K_{b_n}(x_{j,k_n}, P)| \\ &\quad + \max_{1 \leq j \leq k_n} |U_{b_n}(x_{j,k_n}^-|\mathbf{X}_n) - K_{b_n}(x_{j,k_n}^-, P)| + k_n^{-1}. \end{aligned}$$

But notice that $U_{b_n}(x|\mathbf{X}_n)$ is a U statistic with finite support and expectation $K_{b_n}(x, P)$. Hoeffding's inequality implies that, for any $t > 0$, for any x ,

$$P\{|U_{b_n}(x|\mathbf{X}_n) - K_{b_n}(x, P)| > t\} \leq 2 \exp(-2(n/b_n)t^2). \quad (12)$$

As a particular case, for any $\epsilon > 0$, if we choose $k_n = \lceil 2f_I(b_n)/\epsilon \rceil + 1$ and $t = \epsilon f_I(b_n)^{-1}$, then we have

$$\begin{aligned} &P\{\sup_x |U_{b_n}(x|\mathbf{X}_n) - K_{b_n}(x, P)| > \epsilon f_I(b_n)^{-1}\} \\ &\leq P\{\max_{0 \leq j \leq k_n} |U_{b_n}(x_{j,k_n}|\mathbf{X}_n) - K_{b_n}(x_{j,k_n}, P)| \\ &\quad + \max_{1 \leq j \leq k_n} |U_{b_n}(x_{j,k_n}^-|\mathbf{X}_n) - K_{b_n}(x_{j,k_n}^-, P)| > (\epsilon/2)f_I(b_n)^{-1}\} \\ &\leq 4(2f_I(b_n)\epsilon^{-1} + 2) \exp(-(n/b_n)\epsilon^2 f_I(b_n)^{-2}/2). \end{aligned} \quad (13)$$

Now, under A3[I], the right-hand side of (13) converges to 0. It follows that

$$U_{b_n}(x|\mathbf{X}_n) - K_{b_n}(x, P) = O_P(f_I(b_n)^{-1})$$

uniformly in x . But under A2[I] (A2'[I]), $K_{b_n}(x, P)$ admits an Edgeworth expansion, yielding the result of Lemma 1. \square

Lemma 2. Under Assumptions A1, A2[I] (A1' and A2'[I]), A3[I], A4[I] and A6[I],

$$\bar{K}_{b_n}(x|\mathbf{X}_n) - U_{b_n}(x|\mathbf{X}_n) = O_P(f_I(b_n)^{-1}),$$

uniformly on \mathbb{R} (uniformly on a neighbourhood of $c(1 - \alpha)$).

Proof. Define

$$F_{n,i} = \{\epsilon < \sigma^{-2} S_{b_n}(\mathcal{X}_{b_n,i})^2\}$$

and

$$E_n = \{\tau_{b_n} \sigma^{-1} |T_n - \theta(P)| \leq \epsilon^{1/2} \mu_n f_I(b_n)^{-1}\},$$

where ϵ is given in Assumption A6[I]. Under Assumption A4[I] we may choose μ_n to be a non-negative sequence converging to 0 such that

$$P(E_n) \rightarrow_{n \rightarrow \infty} 1. \quad (14)$$

In order to simplify the notation, write

$$I_{i,n} \equiv I_{\{H(\tau_{b_n}(T_{b_n}(\mathcal{X}_{b_n,i}) - T_n)/S_{b_n}(\mathcal{X}_{b_n,i})) \leq x\}}.$$

Then, we have, for any $\nu > 0$,

$$\begin{aligned} P \left\{ \left| \bar{K}_{b_n}(x|\mathbf{X}_n) - N_n^{-1} \sum_{i=1}^{N_n} I_{i,n} I_{\{E_n \cap F_{n,i}\}} \right| > \nu f_I(b_n)^{-1} \right\} \\ \leq P \left\{ I_{\{\bar{E}_n\}} + N_n^{-1} \sum_{i=1}^{N_n} I_{\{\bar{F}_{n,i}\}} > \nu f_I(b_n)^{-1} \right\}. \end{aligned} \quad (15)$$

But under Assumption A6[I], by the strong law of large numbers for U statistics, we obtain with probability one

$$N_n^{-1} \sum_{i=1}^{N_n} I_{\{\bar{F}_{n,i}\}} - P\{\epsilon \geq \sigma^{-2} S_{b_n}(\mathcal{X}_{b_n,1})^2\} \rightarrow_{n \rightarrow \infty} 0. \quad (16)$$

Since we have

$$P\{\epsilon \geq \sigma^{-2} S_{b_n}(\mathcal{X}_{b_n,1})^2\} = o(f_I(b_n)^{-1}), \quad (17)$$

it follows from (14), (16) and (17) that the right-hand side of inequality (15) converges to 0 yielding, uniformly in x ,

$$\left| \bar{K}_{b_n}(x|\mathbf{X}_n) - N_n^{-1} \sum_{i=1}^{N_n} I_{i,n} I_{\{E_n \cap F_{n,i}\}} \right| = o_P(f_I(b_n)^{-1}).$$

But we have on each $E_n \cap F_{n,i}$

$$\begin{aligned} |H(\tau_{b_n}(T_{b_n}(\mathcal{X}_{b_n,i}) - T_n)/S_{b_n}(\mathcal{X}_{b_n,i}))| &\leq |H(\tau_{b_n}(T_{b_n}(\mathcal{X}_{b_n,i}) - \theta(P))/S_{b_n}(\mathcal{X}_{b_n,i}))| \\ &\quad + \mu_n f_I(n)^{-1}. \end{aligned}$$

The preceding equations imply that, with probability one,

$$\begin{aligned} U_{b_n}(x - \mu_n f_I(b_n)^{-1}|\mathbf{X}_n) + o_P(f_I(n)^{-1}) &\leq \bar{K}_{b_n}(x|\mathbf{X}_n) \\ &\leq U_{b_n}(x + \mu_n f_I(b_n)^{-1}|\mathbf{X}_n) + o_P(f_I(n)^{-1}). \end{aligned}$$

Since, from Lemma 1, $U_{b_n}(x|\mathbf{X}_n)$ admits an Edgeworth expansion on differentiable functions,

a Taylor expansion of $K(x \pm \mu_n f_I(b_n)^{-1})$ and of each $p_i(x \pm \mu_n f_I(b_n)^{-1}, P)$, $i = 1, \dots, I - 2$, leads immediately to the fact that

$$U_{b_n}(x \pm \mu_n f_I(b_n)^{-1} | \mathbf{X}_n) = U_{b_n}(x | \mathbf{X}_n) + o(f_I(b_n)^{-1})$$

and Lemma 2 follows. The uniformity in x comes from the fact that the derivatives are uniformly bounded. Notice that under Assumption B2[$I - 1$] (B2' [$I - 1$]) uniform continuity of the last function $p_{I-1}(\cdot, P)$ is sufficient. \square

Combining Lemmas 1 and 2 leads to (i) and (ii). The proof of (iii) proceeds along similar lines: it suffices to apply the Borel–Cantelli lemma, using inequality (13), and to replace convergence in probability by convergence with probability one in Lemma 2. \square

3.3. SECOND-ORDER VALIDITY OF $\bar{K}_n^*(x | \mathbf{X}_n)$

In the following we shall assume that the root \tilde{T}_n is asymptotically pivotal in the sense that the asymptotic distribution of \tilde{T}_n is independent of P , $K(\cdot, P) \equiv K(\cdot)$, and the latter is known. This is the setting in which the second-order properties of the t -percentile method have been extensively studied (see, for instance, Hall 1992b). In many applications, it is a *sine qua non* for construction of asymptotic confidence intervals.

The generalized jackknife of $\bar{K}_{b_n}(x | \mathbf{X}_n)$ and $K(x)$ is then given by

$$\bar{K}_n^*(x | \mathbf{X}_n) \equiv f_1(b_n) f_1(n)^{-1} \bar{K}_{b_n}(x | \mathbf{X}_n) + (1 - f_1(b_n) f_1(n)^{-1}) K(x). \quad (18)$$

In the following, we shall also use the gap

$$\Delta_n \equiv \sup_x |\bar{K}_n^*(x | \mathbf{X}_n) - K_n(x, P)|.$$

We deduce from Theorem 1 the second-order validity of the corrected WoRB. If we can control precisely the remainder in the Edgeworth expansion, we are able to control precisely the rate of the extrapolation. The proof of the following corollary is straightforward.

Corollary 1. (i) Under Assumptions A1, A2[2], A3[2], A4[2] and A6[2] we have

$$\Delta_n = O_P(\max(f_1(b_n) f_2(b_n)^{-1} f_1(n)^{-1}, f_2(n)^{-1})). \quad (19)$$

Alternatively, under Assumptions A1, B2[1], A3[1], A4[1] and A6[1] we have

$$\Delta_n = o_P(f_1(n)^{-1}). \quad (20)$$

(ii) If, in addition to Assumption A3[2] (A3[1]), we have A5[2] (A5[1]) and if A4[2] (A4[1]) holds with probability one, equation (19) (equation (20)) holds almost surely.

(iii) Let $c_n^*(1 - \alpha) \equiv \inf \{x: \bar{K}_n^*(x | \mathbf{X}_n) \geq 1 - \alpha\}$ be the $(1 - \alpha)$ quantile of the extrapolated WoRB distribution; then under Assumptions A1', A2'[2], A3[2], A4[2] and A6[2] we have:

$$P(\tilde{T}_n \leq c_n^*(1 - \alpha)) = 1 - \alpha + O(\max(f_1(b_n) f_2(b_n)^{-1} f_1(n)^{-1}, f_2(f_2(n)^{-1})).$$

Alternatively, under Assumptions A1', B2'[1], A3[1] A4[1] and A6[1] we have:

$$P(\tilde{T}_n \leq c_n^*(1 - \alpha)) = 1 - \alpha + o(f_1(n)^{-1}). \quad (21)$$

Remark 1. The validity of the results whether H is equal to the identity or to the absolute value is important for the construction of one- and two-sided confidence intervals. Indeed, in many cases, the order of the first non-vanishing term in the Edgeworth expansion of $|\tilde{T}_n|$ is smaller than that of \tilde{T}_n . This is due to the fact that in many applications, the first polynomial $p_1(x, P)$ that appears in the expansion of \tilde{T}_n is even and vanishes in the Edgeworth expansion of $|\tilde{T}_n|$. If θ is not a real parameter, the function H used in the definition of \tilde{T}_n may also be replaced by any real-valued function satisfying a triangular inequality. This does not change the proof.

Remark 2. In many applications it is often possible to derive an Edgeworth expansion of type (6) with $f_1(x) = x^{1/2}$ and $\tau_n = n^{1/2}$. Assumption A3[1] holds for $b_n = o(n^{1/2} \log(n)^{-1/2})$. In that case Assumption A4[1] is automatically satisfied and we obtain $\Delta_n = o_P(n^{-1/2})$. Almost sure results are typically obtained with $b_n = O(n^{1/2}(\log n)^{-1})$. For two-sided confidence intervals, in standard applications, we have $f_1(x) = x$ and $\tau_n = n^{1/2}$. In that case Assumptions A3[1] and A4[1] are satisfied with $b_n = o(n^{1/3} \log(n)^{-1/3})$, yielding $\Delta_n = o_P(n^{-1})$.

Remark 3. With more work and regularity conditions on T_n it is often possible to obtain Edgeworth expansions of type (5) with $f_1(x) = x^{1/2}$, $f_2(x) = x$ and $\tau_n = n^{1/2}$. In that case one should take $b_n = o(n^{1/3} \log(n)^{-1/3})$ for unilateral confidence intervals. The greater the value of b_n , the better the corrected WoRB. If $b_n = n^{1/3} \log(n)^{-1}$, we obtain $\Delta_n = O_P(n^{-2/3} \log(n)^{1/2})$. As can be seen from the example of the mean in Section 2.3, the choice of b_n indicated by the theorem may not be optimal in the sense that admissible b_n may yield a better order of Δ_n in specific cases. According to Booth and Hall (1993) one may choose $b_n = O(n^{-2/3})$ and obtain $\Delta_n = O_P(n^{-5/6})$, if one accounts for the finite-population correction factor $(1 - b_n/n)$ which appears in the variance of (1) (see also Remark 4). In many regular applications (such as functionals of means) the usual bootstrap yields an error of order $O_P(n^{-1})$. With refinements proposed in Barbe and Bertail (1995) for the weighted bootstrap, one may obtain $O_P(n^{-3/2})$ if one chooses the random weights appropriately as a function of the statistic. This last result is, however, not automatic in the sense that we need some extra information on the structure of the statistic. Thus, the extrapolated distribution does not provide any improvements in regular situations. However, we shall see below that a refinement of the proposed method leads to the rate $O_P(n^{-1+\epsilon})$, $\epsilon > 0$, and thus may be used as an alternative to the usual bootstrap. The results of Theorem 1 and Corollary 1 are more interesting in non-regular situations including some non-Gaussian and lattice cases or situations in which the bootstrap has never been proved or is known to fail.

Remark 4. The interest of Corollary 1 lies mainly in the fact that the asymptotic distribution is known or tractable. However, if the asymptotic distribution is unknown or very difficult to handle, one may think of using another approximation, for instance a WoRB approximation, and define the generalized jackknife of these distributions which coincides exactly with the

Richardson extrapolation proposed by Bickel and Yahav (1988) when $f_1(x) = x^{1/2}$:

$$\bar{K}_n^{**}(x|\mathbf{X}_n) = \det \begin{pmatrix} \bar{K}_{b_{n,1}}(x|\mathbf{X}_n) & f_1(b_{n,1})^{-1} - f_1(n)^{-1} \\ \bar{K}_{b_{n,2}}(x|\mathbf{X}_n) & f_1(b_{n,2})^{-1} - f_1(n)^{-1} \end{pmatrix} / \det \begin{pmatrix} 1 & f_1(b_{n,1})^{-1} - f_1(n)^{-1} \\ 1 & f_1(b_{n,2})^{-1} - f_1(n)^{-1} \end{pmatrix}.$$

Such a solution has also been investigated recently in a paper by Bickel *et al.* (1994) made known to us during the revision of this work.

If $b_{n,1} \neq b_{n,2}$, say $b_{n,1} < b_{n,2}$, $\bar{K}_n^{**}(x|\mathbf{X}_n)$ improves over $\bar{K}_{b_{n,2}}(x|\mathbf{X}_n)$. However, it coincides with the true distribution up to $o(f_2(b_{n,2})^{-1})$ so that it is second-order correct if and only if (9) holds with $b_{n,2}$ such that

$$f_2(b_{n,2}) \gg f_1(n). \tag{22}$$

In most situations it is not possible to choose $b_{n,2}$ in such a way that both the Edgeworth expansion of the WoRB approximation and this condition hold. The scheme without replacement implies a loss in Edgeworth expansion typically of size b_n/n (see (5)) coming from the fact that the associated weights given by (2) have variance $(1 - b_n/n)$ instead of 1. It follows that if we want to obtain the Edgeworth expansion up to $o(f_2(b_{n,2})^{-1}) = o(b_{n,2}^{-1})$, we have to choose $b_{n,2}$ such that $b_{n,2}/n = o(b_{n,2}^{-1})$. This means that $b_{n,2}$ is $o(n^{1/2})$ and (22) cannot hold.

Notice, however, that, if we want to take into account the correction factor induced by the scheme without replacement, it makes more sense to consider $\bar{K}_{b_{n,i}}(x(1 - b_{n,i}/n)^{1/2}|\mathbf{X}_n)$, $i = 1, 2$, in the construction of $\bar{K}_n^{**}(x|\mathbf{X}_n)$, at least for non-degenerate linear statistics. This possibility has recently been investigated in Bertail and Politis (1996).

Remark 5. If $\theta(P)$ is a multidimensional parameter on \mathbb{R}^k endowed with a norm $\|\cdot\|$, it is easy to generalize the preceding results to construct second-order correct confidence regions based on the WoRB distribution of a root $\tilde{T}_n^* = \|\tau_n C_n(X_n)^{-1/2}(T_n - \theta(P))\|$, where $C_n(X_n)$ estimates the asymptotic covariance matrix of T_n . If Assumption A4[i] is replaced by $\tau_{b_n} f_i(b_n) \|T_n - \theta(P)\| \xrightarrow{n \rightarrow \infty} 0$ in probability, then it is easy to see that propositions (19) and (20) hold with

$$\bar{K}_{b_n}(x|\mathbf{X}_n) \equiv N_n^{-1} \sum_{i=1}^{N_n} I_{\{\tau_{b_n} \|C_{b_n}(X_{b_n,i})^{-1/2}(T_{b_n}(X_{b_n,i}) - T_n)\| \leq x\}}.$$

As a particular case, the results apply in linear (or nonlinear) models with homoscedastic errors.

If $\theta(P)$ takes its values in a separable Banach space \mathcal{B} of type 2, endowed with a norm $\|\cdot\|$, a straightforward generalization may also be obtained for the root $\tilde{T}_n \equiv H(n^{-1/2}\bar{X}_n - \theta(P))$, where H is a Fréchet differentiable function from \mathcal{B} to \mathbb{R} . Götze (1989) proves that under some regularity assumptions on H and the moments of X_i , \tilde{T}_n admits an Edgeworth expansion on a family of functions which are continuously differentiable (see his Theorem 1.3). This may be used, for instance, to construct confidence regions for likelihood ratios.

The preceding results apply also in many other situations, as shown in the following examples.

Example 1. If $\theta(P)$ is a second-order Fréchet or Hadamard differentiable functional whose first gradient (i.e. the influence function) is non-degenerate, then it can be shown that under some regularity conditions on the gradients of $\theta(P)$, $\theta(P_n)$ (where P_n is the empirical distribution function of the observations) admits an Edgeworth expansion on the normal distribution with $f_1(x) = x^{1/2}$ and $\tau_n = n^{1/2}$ (see Pfanzagl and Wefelmeier 1985, for instance) and the results apply with b_n as given in Remarks 2 and 3. Other examples (including the maximum likelihood estimator, rank statistics, non-degenerate U statistics, econometric estimators) and conditions for the validity of second- and third-order expansions may also be found, for instance, in Bhattacharya and Denker (1990) and Sargan (1976). It is easy to check conditions of Theorem 1 in those cases.

Example 2. Some asymptotic expansions have also been derived in the degenerate case (see Götze 1984). For instance, if we deal with degenerate bivariate von Mises functionals, then the assumptions of Theorem 1 are satisfied with $f_i(x) = x^i$ and $\tau_n = n$. As a particular case, we obtain the second-order validity of the extrapolated WoRB distribution of the Cramér–von Mises statistic. The asymptotic properties of the usual bootstrap of degenerate U statistics and functionals have also been studied by Bretagnolle (1983) and Bertail (1992). In that case, either the undersampling scheme or an adequate recentring of the statistic is fundamental for the validity of the bootstrap. However, it is difficult to obtain the second-order validity of the usual bootstrap of degenerate von Mises functionals because one has to check a highly technical smoothness condition that ensures that the bivariate influence function is not too degenerate with respect to the probability measures P and P_n (see Götze 1984). A similar problem appears when one wishes to construct two-sided confidence intervals for a third-order Fréchet differentiable functional (see Bertail 1992). Undersampling avoids these technical problems. Moreover, it is not necessary to make smoothness assumptions on the gradients of $\theta(P)$ or uniformity assumptions on the map $\theta(\cdot)$ in a neighbourhood of P as is usually the case – see the notion of strong differentiability in Tu (1992) and the Lipschitz conditions on the gradients of $\theta(\cdot)$ used in Bertail (1992).

Example 3. Many quantities in econometrics are asymptotically χ^2 distributed and admit an Edgeworth expansion on the family of χ^2 distributions (see Sargan 1979) with $\tau_n = n$ and $f_1(n) = n^{1/2}$. All the assumptions are satisfied in that case and the results hold with $b_n = o(n^{1/2} \log(n)^{-1/2})$ yielding an error of size $o(n^{-1/2})$ for confidence regions.

Example 4. The following example is on the same lines as Hall (1992a). Consider nonparametric density estimation from the point of view of coverage probability. To take into account the problem of bias in bootstrapping nonparametric density kernel estimators, Hall (1992a) shows that it is better to undersmooth the kernel estimator and oversmooth the bootstrap distribution than to estimate the bias and then to construct a bias-corrected bootstrap confidence interval. For a fourth-order kernel estimator, this leads to a two-sided bootstrap confidence interval coverage probability error of size $O(n^{-4/5})$ instead of $O(n^{-12/17})$ in the latter case. Undersampling may be an alternative solution to that problem. In the following, we give very crude arguments showing that an adequate choice of b_n may yield even better results. This example also illustrates the fact that, in some cases,

Assumption A3[i] should be replaced by a more adequate condition on b_n . This problem ought to be investigated in more detail, but is beyond the scope of this paper. Consider

$$\tilde{T}_n(x) \equiv (nh)^{1/2} \hat{\sigma}_n(x)^{-1} |\hat{f}_n(x) - f(x)|,$$

where $\hat{f}_n(x) \equiv (nh)^{-1} \sum_{i=1}^n K\{h^{-1}(x - X_i)\}$ is a kernel estimator of the density f , constructed with an r th-order bounded kernel $K(r \geq 4)$ and a bandwidth h . The estimated variance of $\hat{f}_n(x)$ is

$$\hat{\sigma}_n^2(x) \equiv (nh)^{-1} \sum_{i=1}^n K\{h^{-1}(x - X_i)\}^2 - h\hat{f}_n(x)^2.$$

Following Hall (1992a), assuming that the r th-derivative, say $f^{(r)}$, is continuous and bounded, the bias of $\hat{f}_n(x)$ is given by

$$b_n(x) \equiv k_r h^r f^{(r)}(x) + o(h^r),$$

where k_r is a constant. Now by straightforward arguments on Edgeworth expansion, if, for any x , for any $\epsilon > 0$,

$$P\{(nh)^{1/2} |b_n(x)| / \hat{\sigma}_n(x) > \epsilon((nh)^{-3/2} + n^{-1})\} = o((nh)^{-3/2} + n^{-1}) \tag{23}$$

then we have

$$P(\bar{T}_n(x) \leq y) = \Phi(y) - \Phi(-y) + 2(nh)^{-1} p_1(y, f, x) \phi(y) + O((nh)^{-3/2} + n^{-1}),$$

where p_1 is a polynomial whose form is given explicitly in Hall (1991a). But if we choose h of the form $Cn^{-\beta}$, $\beta > 0$, then (23) holds if β is such that

$$\beta > \max(2/(2+r), 3/2r+1). \tag{24}$$

Now if we choose b_n such that $n^{-1}b_n^{1+3(1-\beta)} \log(b_n) = o(1)$ and $n^{-1}b_n^3 \log(b_n) = o(1)$, then Assumptions A3[2] and A4[2] hold and we can check all the assumptions of Theorem 1 and Corollary 1. This means that if $b_n = \min(n^{1/(4-3\beta)}, n^{1/3}) \log(n)^{-1}$, then the corrected WoRB yields second-order correct bilateral confidence intervals with a coverage error of size $o(n^{-1+\beta}(n^{-(1-\beta)/(2(4-3\beta))} + n^{-\beta/3}) \log(n))$ which is minimized when we choose β near the frontier given by (24). In the particular case $r = 4$ studied in Hall (1992a), we can choose $b_n = o(n^{1/3} \log(n)^{-1})$ and we obtain an error of size $o(n^{-7/9+\nu})$, for any small $\nu > 0$. Again, this leads to a better rate than the bias-corrected bootstrap confidence interval. Nevertheless, it seems to be worse than the method proposed by Hall (1992a). This may be explained by the fact that, as in the case of the mean, Assumption A3[2] is too restrictive on the choice of b_n . Theorem 1 aims to give a general result, but this may be clearly improved in particular cases. Actually, using Hall's (1991a) Edgeworth expansions and the arguments of Section 2.3, it is possible to choose b_n such that $n^{-1}b_n = O(b_n^{-3/2(1-\beta)} + b_n^{-1})$, i.e. $b_n = O(\min(n^{2/(5-3\beta)}, n^{1/2}))$, and in the case $r = 4$ the corrected WoRB yields an error of size close to $o(n^{-5/6+\nu})$, for any small $\nu > 0$, when β is close to $1/3$, which is better than the rate $O(n^{-4/5})$. Of course, it would be interesting to use higher-order Edgeworth expansions in the spirit of Booth and Hall (1993) to investigate which is the best choice for b_n yielding the best rate when one takes into account the correction factor emphasized in Remark 4.

3.4. MONTE CARLO APPROXIMATION

As noted by Politis and Romano (1994), N_n may be very large and it may be difficult to compute the WoRB distribution exactly. It is, however, possible to construct a stochastic approximation by simulating the law of the ‘resampling plan’ described in (1). Let $\mathcal{X}_{b_n, i_1}, \dots, \mathcal{X}_{b_n, i_{B_n}}$ be B_n samples chosen with or without replacement among the N_n possible sets of b_n observations. Then $\bar{K}_{b_n}(x|\mathbf{X}_n)$ may be approximated by

$$\hat{K}_{b_n}^{(B_n)}(x|\mathbf{X}_n) \equiv B_n^{-1} \sum_{k=1}^{B_n} I\{\tilde{T}_{b_n}(\mathcal{X}_{b_n, i_k}) \leq x\},$$

leading to the Monte Carlo approximation

$$\hat{K}_n^{*(B_n)}(x|\mathbf{X}_n) \equiv f_1(b_n)f_1(n)^{-1}\hat{K}_{b_n}^{(B_n)}(x|\mathbf{X}_n) + (1 - f_1(b_n)f_1(n)^{-1})K(x).$$

Since the difference between $\hat{K}_{b_n}^{(B_n)}(x|\mathbf{X}_n)$ and $\bar{K}_{b_n}(x|\mathbf{X}_n)$ is of order $O(B_n^{-1/2})$ the difference between $\hat{K}_n^{*(B_n)}(x|\mathbf{X}_n)$ and $\bar{K}_n^*(x|\mathbf{X}_n)$ is of order $O(f_1(b_n)f_1(n)^{-1}B_n^{-1/2})$. B_n should be chosen so that it does not affect the second-order correction. Typically $B_n^{-1/2} = o(f_1(b_n)^{-1})$ is sufficient to obtain an approximation of order $o(f_1(n)^{-1})$.

4. Extensions to strong-mixing random fields

It is possible to generalize the results obtained for i.i.d. random variables to strong-mixing homogeneous random fields using the idea of the moving-block bootstrap as introduced by Künsch (1989) (see also work done independently by Liu and Singh 1992) and closely related to Carlstein’s (1986) proposal for the use of non-overlapping blocks to estimate the variance of a statistic (see also Sherman 1992; Politis and Romano 1993).

It is well known that the moving-block bootstrap does not provide second-order confidence intervals. More precisely, Lahiri (1992) shows that the moving-block bootstrapped mean has a random bias that actually determines the rate of the approximation, which appears to be worse than the asymptotic distribution. When standardizing the mean by the true variance, he shows that it is possible to get rid of this drawback by recentring the bootstrap mean by its expectation. However, if, in practical situations, one standardizes by an estimator of the variance, Götze and Künsch (1993) show in the case of the mean that it is essential for second-order correctness to hold that the moving-block bootstrap variance should have the same bias as the variance on the original sample. Generalizations to nonlinear functionals are not straightforward. One of the features of the corrected moving-block bootstrap without replacement that we will now study is that it is possible to use the bad recentring or standardization and to choose the resampling size so as to get (automatically) rid of these bias problems and then to rescale to obtain the second-order properties. As in the preceding section, results may be applied even in non-regular cases.

In what follows we shall mainly for the most part retain the notation of Politis and Romano (1994). Let $(X(t), t \in \mathbb{Z}^d)$ be a random field indexed by \mathbb{Z}^d , defined on a

probability space $(\Omega, \mathcal{A}, \mathcal{P})$, taking its values in \mathbb{R} . We shall assume that the random field satisfies some weak dependence condition. Define the strong-mixing coefficients, for $k \in \mathbb{N}^*$, $l_1 \in \mathbb{Z}^*$, $l_2 \in \mathbb{Z}^*$ by

$$\alpha_X(k, l_1, l_2) \equiv \sup_{\substack{E_1 \in \mathbb{Z}^d \\ E_2 \in \mathbb{Z}^d \\ \text{card}(E_i) \leq l_i, i = 1, 2, \bar{d}(E_1, E_2) \geq k}} \{P(A_1 A_2) - P(A_1)P(A_2), A_i \in F(E_i)\},$$

where $F(E_i)$ is the σ -algebra generated by $(X(t), t \in E_i)$ and the semi-distance \bar{d} is defined by

$$\bar{d}(E_1, E_2) \equiv \inf_{\substack{t \in E_1 \\ u \in E_2}} \{\Delta(t, u)\} \text{ with } \Delta(t, u) \equiv \sup_{1 \leq k \leq d} |u_k - t_k|.$$

We shall, in the following, make the following assumption:

Assumption α_0 . *The random field $(X(t), t \in \mathbb{Z}^d)$ is homogeneous and the strong-mixing coefficients $\alpha_X(k, l_1, l_2)$ converge to 0 as k converges to ∞ , for any l_1 and l_2 .*

Now let $T_n = T_n(X(t), t \in E_n)$ be a measurable function of the observations $\mathbf{X}_n = (X(t), t \in E_n)$ on the rectangle $E_n \subset \mathbb{Z}^d$, $n = (n_k)_{1 \leq k \leq d}$. T_n estimates a real parameter θ . As in the i.i.d. case, we shall also assume that there exists a normalizing sequence $S_n = S_n(X(t), t \in E_n)$ such that $\tilde{T}_n \equiv H(\tau_n S_n^{-1}(T_n - \theta))$ with distribution $K_n(\cdot, P)$, satisfy Assumption A1 (or A1'). As in Sections 2 and 3, H may be either the identity or the absolute value (or a norm if one wishes to handle multidimensional parameters).

Let $(j, b, h) \in (\mathbb{Z}^d)^3$ and define the blocks $E_{j,b,h}$ by

$$E_{j,b,h} \equiv \{i \in \mathbb{Z}^d, (j_k - 1)h_k + 1 \leq i_k \leq (j_k - 1)h_k + b_k, \text{ for } k = 1, \dots, d\}.$$

$b = (b_k)_{k=1, \dots, d}$ defines the shape of the blocks whereas $h = (h_k)_{k=1, \dots, d}$ is related to the amount of overlap in each direction of the space. The smaller the components of h , the greater the overlaps. Since we do not want a block to appear twice, we require that $h_k \geq 1$, $k = 1, \dots, d$: in that case if $j \neq j'$, $E_{j,b,h} \neq E_{j',b,h}$. The number of such blocks is given by

$$N_{n,h,b} \equiv \prod_{k=1}^d m_k, \quad m_k \equiv [(n_k - h_k)/b_k] + 1$$

and $j = (j_k)_{k=1, \dots, d}$ can take values only in

$$B_{n,h,b} \equiv \{j \in \mathbb{Z}^d, E_{j,b,h} \subset E_n\} = \prod_{k=1}^d \{1, \dots, m_k\}.$$

In the spirit of Sections 2 and 3, b is a function of the vector n that we shall denote by $b(n)$. Since $N_{n,h,b}$ is a decreasing function of the $(h_i)_{1 \leq i \leq d}$, it seems natural to choose $h = (1, 1, \dots, 1)$, which corresponds to maximum overlaps. However, we shall give our results for a general $h(n)$ possibly depending on n . Denote respectively by $\mathbf{B}(n)$ and $\mathbf{H}(n)$ the products of the $b_i(n)$ and of the $h_i(n)$.

The moving-block version of the bootstrap without replacement is then defined by (see Politis and Romano 1994)

$$\bar{K}_{n,h,b}(x|\mathbf{X}_n) \equiv N_{n,h,b}^{-1} \sum_{j \in B_{n,h,b}} I\{\hat{T}_{b(n),j} \leq x\}$$

with

$$\hat{T}_{b(n),j} \equiv H(\tau_{b(n)} S_{b(n),j}^{-1}(T_{b(n),j} - T_n)),$$

where $S_{b(n),j}$ and $T_{b(n),j}$ are the values of the statistic evaluated on the set of the $b(n)$ observations indexed by $E_{j,b,h}$, for $j \in B_{n,h,b}$.

Of course, the size of the blocks of observations should be chosen as a function of the dependence between the observations. Assumption A3[i] should be replaced by the following:

Assumption $\alpha 3[i]$. Define

$$\beta_n \equiv \sum_{k=1}^{\max_{j=1,\dots,d} n_j} k^{d-1} \alpha_X(k, \mathbf{B}(n), \mathbf{B}(n))$$

and

$$A_n \equiv [2^{-d} N_{n,h,b}] \mathbf{B}(n).$$

Then $b(n) = (b_j(n))_{j=1,\dots,d}$ and $h(n) = (h_j(n))_{j=1,\dots,d}$ are chosen such that there exists $(p_j)_{j=1,\dots,d}$ with $P_n = \prod_{j=1}^d p_j \leq 2^{-d} N_{n,h,b}$ converging to ∞ , satisfying the two conditions

$$N_{n,h,b} P_n^{-1} f_i(b(n))^{3/2} \alpha_X(\min_j p_j \min_j h_j - \max_j b_j, A_n, A_n) \rightarrow_{n \rightarrow \infty} 0, \tag{25}$$

$$f_i(b(n)) \ln f_i(b(n)) \max(N_{n,h,b}^{-1/2}, \mathbf{B}(n)^{1/2} \mathbf{H}(n)^{-1/2} N_{n,h,b}^{-1/2}, \beta_n^{1/2} N_{n,h,b}^{-1/2}, P_n N_{n,h,b}^{-1}) \rightarrow_{n \rightarrow \infty} 0. \tag{26}$$

It is easy to choose $b(n)$ in many practical situations (see Remark 6). Of course, if $h(n)$ is too big (and consequently $N_{n,h,b}$ too small), then Assumption $\alpha 3[i]$ may never hold. In many cases it seems more interesting to choose $h(n) = (1, \dots, 1)$. Notice also that at the first order (i.e for $i = 0$ and $f_i(x)$ equal to a constant) it is easy to obtain the conditions of Politis and Romano's (1994) Theorem 3.1.

In the same way, we replace Assumption A4[i] by Assumption $\alpha 4[i]$:

Assumption $\alpha 4[i]$. $b(n)$ is chosen such that $\tau_{b(n)}/\tau_n \rightarrow_{n \rightarrow \infty} 0$ and

$$\tau_{b(n)} f_i(b(n)) |T_n - \theta(P)| \rightarrow_{n \rightarrow \infty} 0.$$

The following theorem gives the analogue of Theorem 1 and Corollary 1 for asymptotically pivotal statistics of strong-mixing random fields.

Theorem 2. *Suppose that Assumption α_0 holds. If Assumptions A3[.] and A4[.] are replaced by $\alpha 3$ [.] and $\alpha 4$ [.] then the conclusions (i), (ii) and (iii) of Theorem 1, (i) and (iii) of Corollary 1 remain true with*

$$\bar{K}_n^*(x|\mathbf{X}_n) \equiv f_1(b(n))f_1(n)^{-1}\bar{K}_{n,h,b}(x|\mathbf{X}_n) + (1 - f_1(b(n))f_1(n)^{-1})K(x, P).$$

Proof. We follow the arguments of Theorem 1. Actually it suffices to obtain a similar version of Lemma 1 in a strong-mixing setting. The proof of Lemma 2 is similar: replace the U -statistics argument in (16) by the law of large numbers for strong-mixing random variables.

Similarly to the i.i.d. case, define

$$U_{b(n)}(x|\mathbf{X}_n) \equiv N_{n,h,b}^{-1} \sum_{i \in B_{n,h,b}} I\{\tilde{T}_{b(n),i} \leq x\},$$

where

$$\tilde{T}_{b(n),i} \equiv H(\tau_{b(n)}S_{b(n),i}^{-1}(T_{b(n),i} - \theta)).$$

We have by homogeneity of the random fields:

$$E\{U_{b(n)}(x|\mathbf{X}_n)\} = P(\tilde{T}_{b(n),1} \leq x) = K_{b(n)}(x, P).$$

The proof is similar to the proof in the i.i.d. case if we show that

$$P\{\sup_x |U_{b(n)}(x|\mathbf{X}_n) - K_{b(n)}(x, P)| > \epsilon f_1(b(n))^{-1}\} = o(1).$$

Since the main tool for proving this result is the existence of a Hoeffding-type inequality or a Bernstein inequality, we shall first discuss this point. A precise Bernstein inequality has recently been obtained by Bosq (1993) for strong-mixing times series (i.e. if $d = 1$) under a minorization condition that unfortunately does not hold in our situation. However, Rhomari (1993) shows that this condition may be removed to obtain a similar inequality. The result may be used to obtain the validity of our procedure for strong-mixing times series. A Bernstein inequality for β -mixing random fields may also be found in Doukhan (1994). Combining the approaches of Rhomari (1993) and Doukhan (1994), we give in Lemma 3 an extension of these inequalities to strong-mixing stationary random fields. In the following we use the notation $h_* \equiv \min_i (h_i(n))$, $b^* \equiv \max_i (b_i(n))$, $p_* \equiv \min (p_i(n))$, $n^* \equiv \max_i (n_i)$.

Lemma 3. *Let D_n be a sequence of rectangles of size $n = (n_i)_{i=1,\dots,d}$ with $\tilde{N}_n \equiv \prod_{i=1}^d n_i$ and consider $(Y_n(t), t \in D_n)$, an array of α -mixing stationary random fields on a probability space $(\Omega, \mathcal{A}, \mathcal{P})$ with strong-mixing coefficients $\alpha_Y(k, l_1, l_2)$. Assume, in addition, that*

- (i) $EY_n(t) = 0$;
- (ii) there exists $M_n > 0$, such that, for any $t \in \mathbb{Z}^d$, $|Y_n(t)| < M_n$;
- (iii) there exists $\sigma_n^2: \mathbb{N} \rightarrow \mathbb{R}^+$ such that, for any $A \subset \mathbb{Z}^d$,

$$E\left(\sum_{t \in A} Y(t)\right)^2 \leq \sigma_n^2(\text{card}(A)).$$

For any $\epsilon_n > 0$, for any integer $1 \leq p_i \leq [n_i/4]$, $i = 1, \dots, d$, put $P_n = \prod_{i=1}^d p_i$; then we have

$$P\left(\left|\sum_{t \in D_n} Y_n(t)\right| > \tilde{N}_n \epsilon_n\right) \leq 2^{d+2} \exp\{-\tilde{N}_n P_n \epsilon_n^2 2^{-d} / \{16\sigma_n^2(P_n) + 8M_n P_n^2 \epsilon_n\}\} + 33M_n^{1/2} \epsilon_n^{-1/2} \tilde{N}_n P_n^{-1} 2^d \alpha_Y(p_* - 1, [2^{-d} \tilde{N}_n], [2^{-d} \tilde{N}_n]). \quad (27)$$

Proof. Without any loss of generality we may assume that $M_n = 1$. We use blocking techniques as in Bosq (1993). For this purpose we shall cut the domain D_n into slices of equal size (except on the upper bounds of D_n). Let $1 \leq p_j \leq [n_j/2]$, $j = 1, \dots, d$, be the size of the slices. Consider $q_j = [n_j/(2p_j)]$, $j = 1, \dots, d$, and put $Q_n = \prod_{j=1}^d q_j$. For any $\delta = (\delta_1, \dots, \delta_d) \in \{0, 1\}^d$, for any $i = (i_j)_{j=1, \dots, d} \in \mathcal{E}_n = \prod_{j=1}^d \{1, \dots, q_j\}$, define

$$U_i(\delta) \equiv \sum_{k_1=a_{i_1}}^{c_{i_1}} \dots \sum_{k_2=a_{i_2}}^{c_{i_2}} \dots \sum_{k_d=a_{i_d}}^{c_{i_d}} Y(k),$$

where

$$a_{i_j} \equiv 2(i_j - 1)p_j + \delta_j p_j + 1$$

$$c_{i_j} \equiv \min(n_j, 2i_j p_j - (1 - \delta_j)p_j).$$

We use the convention that the sum is equal to 0 if $a_j > c_j$ for some $j = 1, \dots, d$. Now $U_i(\delta)$ is the sum of the $X(t)$ over the non-overlapping blocks defined by the a_{i_j} and c_{i_j} and by definition we have

$$S_n \equiv \sum_{t \in D_n} Y_n(t) = \sum_{\delta \in \{0,1\}^d} V_n(\delta),$$

where we define

$$V_n(\delta) \equiv \sum_{i \in \mathcal{E}_n} U_i(\delta).$$

Now notice that, for a fixed δ , the $U_i(\delta)$, for $i \in \mathcal{E}_n$, may be seen as a sequence of strong-mixing random variables indexed by \mathbb{Z} with strong-mixing coefficients satisfying

$$\alpha(k) \leq \alpha_Y(p_* - 1, [2^{-d} \tilde{N}_n], [2^{-d} \tilde{N}_n]). \quad (28)$$

We have under (i) and (ii), $E\{U_i(\delta)\} = 0$ and $|U_i(\delta)| \leq P_n$. Moreover, under (iii),

$$\tilde{\sigma}_n^2(k) \equiv \max_{\substack{A \subset \mathcal{E}_n \\ \text{card}(A)=k}} E\left(\sum_{i \in A} U_i(\delta)\right)^2 \leq \sigma_n^2(P_n k).$$

Now we may write

$$P(|S_n| > \tilde{N}_n \epsilon_n) \leq \sum_{\delta} P(|V_n(\delta)| > 2^{-d} \tilde{N}_n \epsilon_n)$$

and by applying Rhomari's (1993) Bernstein inequality on each $V_n(\delta)$ and using (28) we obtain that, for any $1 \leq \tilde{R}_n \leq [Q_n/2]$, we have

$$\begin{aligned} P(|V_n(\delta)| > 2^{-d} \tilde{N}_n \epsilon_n) &= P(|V_n(\delta)| > Q_n(2^{-d} \tilde{N}_n \epsilon_n Q_n^{-1})) \\ &\leq 4 \exp \left\{ -\tilde{N}_n^2 \tilde{R}_n \epsilon_n^2 2^{-2d} / \{16\tilde{\sigma}_n^2(\tilde{R}_n) Q_n + 4P_n \tilde{N}_n \tilde{R}_n (\tilde{R}_n + 1) \epsilon_n 2^{-d}\} \right\} \\ &\quad + 33\epsilon_n^{-1/2} Q_n \tilde{R}_n^{-1} \alpha_Y(p_* - 1, [2^{-d} \tilde{N}_n], [2^{-d} \tilde{N}_n]). \end{aligned}$$

Since we have $Q_n \leq 2^{-d} P_n^{-1} \tilde{N}_n$, it follows that

$$\begin{aligned} P(|S_n| > \tilde{N}_n \epsilon_n) &\leq 2^{d+2} \exp \left\{ -\tilde{N}_n \tilde{R}_n \epsilon_n^2 2^{-d} / \{16\sigma_n^2(P_n \tilde{R}_n) P_n^{-1} + 4P_n \tilde{R}_n (\tilde{R}_n + 1) \epsilon_n\} \right\} \\ &\quad + 2^d 33\epsilon_n^{-1/2} \tilde{N}_n P_n^{-1} \tilde{R}_n^{-1} \alpha_Y(p_* - 1, [2^{-d} \tilde{N}_n], [2^{-d} \tilde{N}_n]). \end{aligned}$$

Now if we choose $\tilde{R}_n = 1$, we obtain the result of Lemma 3. □

We now return to the proof of Theorem 3. The random variables $I\{\tilde{T}_{b(n),i} \leq x\}_{i \in B_{n,h,b}}$, $n \in \mathbb{N}$, define an array of stationary strong-mixing random fields with strong-mixing coefficients which satisfy, for any $(l_1, l_2) \in \mathbb{N}$ and for $k > k_0 = b^*/h_*$,

$$\alpha^*(k, l_1, l_2) \leq \alpha_X(kh_* - b^*, l_1 \mathbf{B}(n), l_2 \mathbf{B}(n)).$$

Now, it is easy to check the assumptions of Lemma 3 with

$$Y_n(i) = I\{\tilde{T}_{b(n),i} \leq x\} - EI\{\tilde{T}_{b(n),i} \leq x\}.$$

In that case, we have $M_n = 2$ and

$$V\left(\sum_{i \in A} Y_n(i)\right) \leq \text{card}(A) + \sum_{i \in A} \sum_{\substack{j \in A \\ i \neq j}} \text{cov}(Y_n(i), Y_n(j)). \tag{29}$$

But, for $\Delta(i, j) = \max_k (|i_k - j_k|) \geq k_0$, the covariance inequality gives

$$\text{cov}(Y_n(i), Y_n(j)) \leq 4\alpha^*(\Delta(i, j)h_* - b^*, \mathbf{B}(n), \mathbf{B}(n)).$$

Let c_0, c_1, c_2, \dots be some non-negative constants. Now, following the arguments of Politis and Romano (1994) the right-hand side of inequality (29) is bounded by

$$\begin{aligned} \sigma_n^2(A) &= \text{card}(A) + c_0 \text{card}(A) \mathbf{B}(n) \mathbf{H}(n)^{-1} \\ &\quad + 4 \text{card}(A) \sum_{k=k_0+1}^{n^*} k^{d-1} \alpha^*(kh_* - b^*, \mathbf{B}(n), \mathbf{B}(n)). \end{aligned}$$

Now, apply inequality (27) to obtain that, for any $\epsilon > 0$, for any $P_n \leq 2^{-d} N_{n,h,b}$,

$$P\{|U_{b(n)}(x|\mathbf{X}_n) - K_{b(n)}(x, P)| > \epsilon f_I(b(n))^{-1}\} \leq E_n(\alpha)$$

with

$$\begin{aligned} E_n(\alpha) &= c_1 \exp \left\{ -c_5 N_{n,h,b} \epsilon^2 f_I(b(n))^{-2} \{1 + c_2 \mathbf{B}(n) \mathbf{H}(n)^{-1} + c_3 \beta_n + c_4 P_n \epsilon f_I(b(n))^{-1}\}^{-1} \right\} \\ &\quad + c_6 \epsilon^{-1/2} f_I(b(n))^{1/2} N_{n,h,b} P_n^{-1} \alpha_X(p_* h_* - b^*, [2^{-d} N_{n,h,b}] \mathbf{B}(n), [2^{-d} N_{n,h,b}] \mathbf{B}(n)). \end{aligned}$$

Now apply the arguments of Lemma 1 to obtain the uniform bound

$$P\{\sup_x |U_{b(n)}(x|\mathbf{X}_n) - K_{b(n)}(x, P)| > \epsilon f_I(b(n))^{-1}\} \leq c_7 f_I(b(n)) E_n(\alpha). \quad (30)$$

Then, it is easy to see that under Assumption $\alpha 3[I]$ the right-hand side of inequality (30) converges to 0 as n converges to ∞ . \square

Remark 6. Assume that α_X satisfies

$$\alpha_X(k, B(n), B(n)) \leq c^{-1} e^{-ck}, \quad (31)$$

for some constant $c > 0$. This inequality holds, for instance, for finite stationary ARMA models, some strictly stationary homogeneous Markov chain and m -dependent random fields. A similar condition may be found in Götze and Hipp (1983) and is used to obtain asymptotic expansions for sums of weakly dependent processes.

Consider the case with a maximum overlap: $\mathbf{H}(n) = 1$ and all the dimensions n_i going to ∞ . In this case β_n is bounded, and if we put $p_j = (n_j - b_j)^{1/2}$, $j = 1, \dots, d$, it is easy to check that (26) and (25) hold, for $i = 1$, if we choose $b_j(n) = o(n_j^{1/2} \log(n_j)^{-1/2})$.

Example 5. In the real case, validity of Edgeworth expansions for functionals of stationary ARMA processes may be derived from the conditions of Götze and Hipp (1983). The validity of the bootstrap in estimating the parameters of a stationary autoregressive model has been proved by Bose (1988). The result mainly relies on the fact that the residuals are assumed to be i.i.d. This allows resampling in the estimated residuals. Our results include the more general case when the residuals are stationary MA(q) (where q is unknown). Moreover, it is possible to check that M -estimators in time series based on a criterion ψ (see Künsch 1984) satisfy the conditions of Götze and Hipp (1983) provided that ψ is sufficiently differentiable. Since in that case the explicit Edgeworth expansion may be difficult to obtain, the corrected moving-block bootstrap without replacement may be a useful tool for constructing second-order correct confidence regions for the parameters.

Example 6. Politis and Romano (1992; 1994) proposed different methods based on the moving-block bootstrap to construct confidence intervals for the spectral density of a strong-mixing process based on a kernel spectral density estimate. Our results may be used to improve confidence intervals for pointwise estimation. However, asymptotic confidence bands for the spectral density using a supremum norm may yield a very inadequate level. Indeed, following Hall (1991b), since (under some regularity assumptions) the spectral density kernel estimator may be approximated by a Gaussian process, the error induced uniformly on the level by using the asymptotic approximation is of order $f_1(n)^{-1} = \log(n)^{-1}$. It is a very poor approximation that may be improved by the corrected WoRB. A more complete study of this case will be given in a forthcoming paper.

Example 7. Asymptotic expansions up to $o(N^{-d/2})$, when $n = (N, \dots, N)$, for a potential function of m -dependent random fields generated by i.i.d. variables may be found in Götze and Hipp (1989). It is easy to see that the first term in the Edgeworth expansion (of size

$N^{-d/2}$) satisfies the regularity conditions of Theorem 2. Since in that case, the strong-mixing coefficients are of type (31), if we choose $h = (1, 1, \dots, 1)$, the adequate size for $b(n)$ is given by $b_i(n) = o(N^{1/2} \log(N)^{-1/2})$ (see Remark 6).

5. Further refinements: generalized jackknife of WoRB distributions

5.1. IMPROVING THE APPROXIMATIONS

In the preceding sections we have seen that a linear combination of two bad approximations of the same distribution may lead to a better approximation. As already noted, we can see this as a transposition of jackknife techniques to get rid of some bad orders. More generally, compute $I - 1$ different WoRB distributions $\bar{K}_{b_{n,k}}$, $k = 1, \dots, I - 1$. They are different in the sense that they are constructed for different values of $b_{n,k}$ satisfying Assumptions A3[I - 1], A4[I - 1]. From Theorem 1 and 2 under their respective hypotheses, such distributions admit Edgeworth expansions

$$\bar{K}_{b_{n,k}}(x|\mathbf{X}_n) = E_{b_{n,k}}^{(I-1)}(x, P) + o_P(f_{I-1}(b_{n,k})^{-1}), \quad k = 1, \dots, I - 1, \quad (32)$$

so that we obtain a Cramér system, $k = 1, \dots, I - 1$,

$$\begin{aligned} \bar{K}_{b_{n,k}}(x|\mathbf{X}_n) - K_n(x, P) &= \sum_{i=1}^{I-1} (f_i(b_{n,k})^{-1} - f_i(n)^{-1}) p_i(x, P) + o_P(f_{I-1}(b_{n,k})^{-1}) \\ &\quad + o_P(f_{I-1}(n)^{-1}). \end{aligned}$$

This leads to a new approximation of $K_n(x, P)$, linear in Φ and the $\bar{K}_{b_{n,1}}(x|\mathbf{X}_n)$, by inverting the system. Classical linear algebra leads to the generalized jackknife estimator (see Gray *et al.* 1972)

$$G_n^{(I-1)}(x) \equiv a^{-1} \det A(K(x), \bar{K}_{b_{n,1}}(x|\mathbf{X}_n), \dots, \bar{K}_{b_{n,I-1}}(x|\mathbf{X}_n)),$$

with

$$a = \det A(1, 1, \dots, 1),$$

where we define

$$A(x_1, x_2, \dots, x_I) \equiv \begin{pmatrix} x_1 & -f_1(n)^{-1} & \dots & -f_{I-1}(n)^{-1} \\ x_2 & f_1(b_{n,1})^{-1} - f_1(n)^{-1} & \vdots & f_{I-1}(b_{n,1})^{-1} - f_{I-1}(n)^{-1} \\ x_3 & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ x_I & f_1(b_{n,I-1})^{-1} - f_1(n)^{-1} & \dots & f_{I-1}(b_{n,I-1})^{-1} - f_{I-1}(n)^{-1} \end{pmatrix}.$$

In the case $f_i(x) = x^{i/2}$, $G_n^{(I-1)}$ may also be seen as the Richardson extrapolation of the asymptotic distributions and the WoRB distributions studied by Bickel and Yahav (1988). We clearly have $G_n^{(1)}(\cdot) = \bar{K}_n^*(\cdot | \mathbf{X}_n)$. The fact that $G_n^{(I-1)}$ may improve the preceding approximations is stated in the following theorem.

Theorem 4. (i) Assume that (32) holds uniformly and that the $(b_{n,k})_{1 \leq k \leq I-1}$ are subject to the conditions

$$0 < \prod_{j=1}^{I-1} \max_{1 \leq k \leq I-1} f_j(b_{n,k}) / \min_{1 \leq k \leq I-1} f_j(b_{n,k}) < \infty \tag{33}$$

and, for some positive constant C ,

$$|a| \prod_{j=1}^{I-1} \min_{1 \leq k \leq I-1} (f_j(b_{n,k})^{-1}) \geq C. \tag{34}$$

Denote

$$v_n = \max_{1 \leq i \leq I-1} f_i(n)^{-1} (\min_{1 \leq k \leq n} f_{I-1}(b_{n,k}))^{-1} \min_{1 \leq k \leq n} f_i(b_{n,k}).$$

Then we have

$$\Delta_n^{(I-1)} \equiv \sup_x |G_n^{(I-1)}(x) - K_n(x, P)| = o_P(v_n) = o_P(f_1(n)^{-1}).$$

(ii) Assume that (32) holds uniformly in a neighbourhood of $c(1 - \alpha)$ in addition to conditions (33) and (34). If we define

$$c_n^{(I-1)}(1 - \alpha) \equiv \inf \{x: G_n^{(I-1)}(x) \geq 1 - \alpha\},$$

the $(1 - \alpha)$ quantile of the generalized jackknife of the WoRB distributions, then we have

$$P(\tilde{T}_n \leq c_n^{(I-1)}(1 - \alpha)) - 1 + \alpha = o(v_n) = o(f_1(n)^{-1}).$$

Proof. Under (34), we have $a \neq 0$ and classical linear algebra calculation leads to

$$G_n^{(I-1)}(x) - K_n(x, P) = O_P(\delta_n),$$

with

$$\delta_n \equiv a^{-1} \det A(\mu_{n,0} f_{I-1}(n)^{-1}, \mu_{n,1} f_{I-1}(b_{n,1})^{-1}, \dots, \mu_{n,I-1} f_{I-1}(b_{n,I-1})^{-1})$$

for some stochastic sequences $(\mu_{n,i})_{0 \leq i \leq I-1}$ converging to 0. The results follow by straightforward algebra. □

Remark 7. Expression (33) means that the range of the $(b_{n,k})_{1 \leq k \leq I-1}$ satisfying Assumptions A3[$I - 1$] and A4[$I - 1$] should not be too big and is fixed by the order of the functions f_i . However, (34) means that the $b_{n,k}$ should not be too close. In many situations, this condition is not satisfied if the differences between each $b_{n,i}$ do not converge to ∞ . Nevertheless, in many applications the f_i are of each type $f_i(x) = x^{\beta i}$, $\beta > 0$. In that case one may choose $b_{n,1}$

according to Assumptions A3[$I - 1$] and A4[$I - 1$] and take $b_{n,i} = K_i b_{n,1}$, where the $\{K_i\}_{1 \leq i \leq I-1}$ are distinct non-negative constants. In that case, a is no more than the determinant of a Vandermonde matrix. Since the K_i are distinct, (34) holds.

If $f_i(x) = x^{i/2}$, Assumptions A3[$I - 1$] and A4[$I - 1$] are satisfied for instance if $b_{n,i} = K_i n^{1/I} \log(n)^{-1}$. In that case, under Assumptions A1 and B2[$I - 1$], we obtain

$$\Delta_n^{(I-1)} = o_p(n^{-1+1/I} \log(n)^{(I-2)/2}).$$

This means that when I increases, the order of the extrapolated approximation becomes better and approaches the rate of the usual bootstrap $O_p(n^{-1})$. However, this holds only if we have sufficient regularity conditions on the statistic T_n to have Edgeworth expansion up to $o(n^{-(I-1)/2})$ (for conditions, see Bhattacharya and Ghosh 1978; Sargan 1976; 1979; Götze 1984).

Remark 8. The idea of using two different resampling plans to get a better estimator of the distribution may also be found in Barbe and Bertail (1995). In their case, however, the aim was to obtain some estimators of the terms involved in the Edgeworth expansion in order to invert it. It is noteworthy that the system may also be used to obtain simultaneous estimators of the functions involved in the Edgeworth expansion.

Remark 9. In practice the $I - 1$ WoRB distributions are also approximated by Monte Carlo simulations as in Section 3.4. Let $\{\hat{K}_{b_{n,i}}^{(B_{n,i})}\}_{i=1, \dots, I-1}$ be such approximations respectively constructed with $B_{n,i}$, $i = 1, \dots, I - 1$. Then the Monte Carlo approximation of $G_n^{(I-1)}(x)$ is given by

$$\hat{G}_n^{(I-1)}(x) \equiv a^{-1} \det A(\Phi(x), \hat{K}_{b_{1,1}}^{(B_{n,1})}(x, P), \dots, \hat{K}_{b_{n,I-1}}^{(B_{n,I-1})}(x, P)).$$

Since the error induced by using $\hat{K}_{b_{n,i}}^{(B_{n,i})}$ instead of $\bar{K}_{b_{n,i}}$ is of order $O(B_{n,i}^{-1/2})$, if we choose $B_{n,i}^{-1/2} = o(f_i(b_{n,i})^{-1})$ then the total error induced by the simulation does not destroy the higher-order correction.

Take, for instance, $\tau_n = n^{1/2}$, $f_i(x) = x^{i/2}$ and $I = 4$. If Assumption B2[3] holds, then we may choose $b_{n,i} = i n^{1/4} \log(n)^{-1}$ and $B_{n,i} = n^{1/2}$, $i = 1, \dots, I - 1$. In that case we need only $3n^{1/2}$ simulations to obtain an approximation correct up to $o_p(n^{-3/4})$ whereas, in general, we need at least $B_n = n^{3/2}$ simulations to construct the usual bootstrap distribution to obtain an error of the same order. This method thus seems computationally less expensive. However, for small and moderate sample sizes, the $b_{n,i}$ that are supposed to converge to ∞ may be rather small and a simulation study is required to validate these results.

6. Some simulation results

Some simulation results on the encouraging performance of extrapolations even for small sample sizes are given in Bickel and Yahav (1988) for the mean case. We shall show in this section that the extrapolated WoRB leads to some interesting improvements over the

bootstrap and the asymptotics, even if the sample size is small (which implies that b_n may be very small). In regular cases, this may be explained heuristically by the fact that the WoRB distribution gives, up to the correcting factor $(b_n/n)^{1/2}$, the correct skewness of the distribution. We mainly study a non-regular case, confidence intervals for quantiles, a problem which has its own statistical interest (see Hall and Martin 1991).

Let X_1, X_2, \dots, X_n be i.i.d. real random variables with distribution function F . Let $X_{1,n} \leq X_{2,n} \leq \dots \leq X_{n,n}$ be the ordered sample. Denote, for $\alpha \in]0, 1[$,

$$\theta = F^{-1}(\alpha) = \inf \{t, F(t) \geq \alpha\},$$

the quantile of order α , estimated by $X_{r,n} \equiv F_n^{-1}(\alpha)$, where $F_n = n^{-1} \sum_{i=1}^n 1_{\{X_i \leq x\}}$. It is known that the bootstrap does not provide second-order correction for the distribution of $F_n^{-1}(\alpha)$. The asymptotic distribution typically has an error of size $n^{-1/2}$, whereas the bootstrap yields an error of size $n^{-1/4}$ (see, for instance, Falk and Reiss 1989). Smoothing F_n and using a smooth bootstrap improves this result but we do not obtain second-order results even for two-sided confidence intervals. The coverage probability of bootstrap confidence intervals turns out to be of order $O(n^{-1/2})$ in both cases and does not reach second-order correctness. Moreover, for a fixed given level, even distribution-free methods based on order statistics are known not to be second-order correct (except for very special levels and sizes of the sample): indeed, in that case the two-sided confidence interval is of the form $I_{l,u} = [X_{l,n}, X_{u,n}]$. The coverage probability is then given by $P(l \leq B(n, \alpha) \leq u)$ where $B(n, \alpha)$ is a binomial random variable, which has jumps of size $n^{-1/2}$. To standardize the distribution and apply our results, we use Maritz and Jarrett's (1978) variance estimator,

$$\hat{\sigma}_n^2 = n \sum_{i=1}^n (X_{i,n} - X_{r,n})^2 w_i$$

with

$$w_i = r \binom{n}{r} \int_{(i-1)/n}^{i/n} x^{r-1} (1-x)^{n-r} dx,$$

which turns out to be the exact computation of the bootstrap estimator of the variance (Efron 1979).

Hall and Martin (1991) prove that, under $E|X|^\epsilon < \infty$ and, if $f = F'$ (with $f(\theta) > 0$) has two bounded derivatives in the neighbourhood of θ , the following Edgeworth expansion holds:

$$P(n^{1/2}(X_{r,n} - \theta)/\hat{\sigma}_n \leq x) = \Phi(x) + n^{-1/2} Q(x, f)\phi(x) + O(n^{-3/4}),$$

where $Q(x, f)$ is a polynomial in x depending on f and its derivatives. As noted by Hall and Martin (1991), for unilateral confidence intervals, an explicit Edgeworth correction by inversion may be difficult and illusory, because of the bad performance of kernel estimators of the density and its derivatives for small sample sizes. However, it is easy to construct the WoRB distribution of the root $\hat{t} = n^{1/2}(X_{r,n} - \theta)/\hat{\sigma}_n$ and use the extrapolation.

The following simulations were performed on the exponential distribution. We chose respectively $n = 25$, $n = 50$ and $n = 100$. We chose an undersampling size $b_{n,1} =$

$[3n^{1/2} \log(n)^{-1}]$, that is, the undersampling sizes become 4, 5 and 6 respectively for the three values of n . We also chose an undersampling size $b_{n,2} = 4b_{n,1}$. For the theoretical reasons discussed in Section 3, we expect to obtain less accurate results with this undersampling size. In the following the $E_i(\beta)$ denote bounds for \hat{t} at level β , where i indexes the different methods used. $i = 0$ corresponds to the true bound (estimated with 10 000 simulations). $i = 1$ corresponds to the asymptotic bound, $i = 2$ to the extrapolation of the WoRB distribution with $b_{n,1}$, $i = 3$ to the same method with $b_{n,2}$, $i = 4$ to the extrapolation of the two undersampling distributions as done in Bickel *et al.* (1994), $i = 5$ to the usual bootstrap.

Table 1 gives the mean of the estimated bounds over 2000 repetitions of the procedure, for $\alpha = 75\%$ at level $\beta = 5\%$; the empirical standard deviations (Std) of the bounds are given in parentheses. The estimated coverage probability C (as a percentage), over the 2000 repetitions of the procedure (except for E_0 which gives the true one) is also given.

Among all the methods studied, the extrapolation of the undersampling distribution with a small undersampling size, i.e. E_2 , gives the best results in terms of coverage probability. On average, the estimated bound is closest to the true one and its standard deviation is rather small compared to others. Even if we take into account this variability, it performs better than the asymptotic E_1 . The gain is not negligible for the construction of confidence intervals for θ because the variance of the estimated fractile may be rather big. To give some idea, on average, over the repetitions of the procedure, for $n = 25$, we have $F_n^{-1}(0.75) = 1.416$ with standard deviation 0.372. If the undersampling size is too big (corresponding to E_3), as we expect, we do not obtain an accurate approximation. On average, the usual bootstrap E_5 gives satisfactory results for the bound. It is even better than the asymptotic, but the standard deviations are so big that it seems quite dangerous to rely on this method for small sample sizes. The extrapolation of the two undersampling distributions performs very badly in these simulations. The adequate choice of the undersampling sizes in that case is an open problem. Similar results have been obtained for different levels β and different values of $\alpha \in [25\%, 75\%]$. However, for $\alpha > 75\%$, the estimator of the quantile in the subsamples tends to coincide with the maximum. Since we centre at $F_n^{-1}(\alpha)$, we expect the undersampling distribution to have very long tails, so that we may wonder if, for large quantiles and a smaller level, it may be possible to keep the

Table 1. Bounds for \hat{t} , with $\alpha = 75\%$ at level $\beta = 5\%$

	$n = 25$			$n = 50$			$n = 100$		
	Mean	(Std)	C (%)	Mean	(Std)	C (%)	Mean	(Std)	C (%)
$E_0(\beta)$	-2.074	(-)	5.0	-1.842	(-)	5.0	-1.701	(-)	5.0
$E_1(\beta)$	-1.645	(-)	9.1	-1.645	(-)	7.9	-1.645	(-)	5.3
$E_2(\beta)$	-1.886	(0.048)	6.3	-1.774	(0.043)	5.5	-1.671	(0.029)	5.2
$E_3(\beta)$	-1.485	(0.354)	11.9	-1.584	(0.207)	10.6	-1.628	(0.150)	5.6
$E_4(\beta)$	-1.127	(0.525)	18.1	-1.280	(0.366)	14.6	-1.428	(0.396)	10.7
$E_5(\beta)$	-1.791	(0.652)	15.8	-1.743	(0.462)	12.3	-1.639	(0.416)	10.2

Table 2. Comparison of the performance of the extrapolations with $\alpha = 80\%$ and $\alpha = 95\%$ at level $\beta = 1\%$

	$\alpha = 80\%$						$\alpha = 95\%$					
	$n = 25$			$n = 50$			$n = 25$			$n = 50$		
	Mean	(Std)	C (%)	Mean	(Std)	C (%)	Mean	(Std)	C (%)	Mean	(Std)	C (%)
$E_0(\beta)$	-2.831	(-)	1.0	-2.616	(-)	1.0	-7.666	(-)	1.0	-4.203	(-)	1.0
$E_1(\beta)$	-2.326	(-)	2.5	-2.326	(-)	1.7	-2.326	(-)	11.9	-2.326	(-)	5.6
$E_2(\beta)$	-2.835	(0.098)	1.4	-2.660	(0.058)	1.3	-3.146	(0.133)	6.8	-2.938	(0.091)	2.4
$E_3(\beta)$	-2.639	(0.326)	2.4	-2.672	(0.225)	1.6	-3.182	(0.330)	6.6	-3.012	(0.241)	2.9
$E_4(\beta)$	-2.130	(0.643)	4.2	-2.228	(0.664)	4.7	-10.155	(17.384)	14.2	-2.735	(2.015)	9.0
$E_5(\beta)$	-3.280	(1.311)	2.4	-3.001	(0.871)	2.4	-9.277	(8.388)	6.1	-4.636	(2.576)	3.1

Table 3. Confidence bounds of the extrapolated WoRB compared to the asymptotics, at level $\beta = 5\%$: $n = 50$, $b_n = 5$

	$\chi^2(1)$						$LN(0, 1)$					
	Mean	(Std)	C (%)	Mean	(Std)	C (%)	Mean	(Std)	C (%)	Mean	(Std)	C (%)
$E_0(\beta)$	-1.769	(-)	5.0	-1.836	(-)	5.0	-1.641	(-)	5.0	-1.868	(-)	5.0
$E_1(\beta)$	-1.645	(-)	6.3	-1.645	(-)	6.4	-1.645	(-)	5.0	-1.645	(-)	7.2
$E_2(\beta)$	-1.744	(0.039)	5.1	-1.804	(0.036)	5.2	-1.640	(0.041)	4.8	-1.798	(0.042)	5.5

($b_n = 5$)

second-order properties of the extrapolation. Moreover, for two different fractiles, the WoRB differs only by the recentring factor. So we may wonder whether the WoRB works better for some α than for others. Table 2 illustrates what happens for $\alpha = 80\%$ and $\alpha = 90\%$ at level $\beta = 1\%$. It is constructed, like Table 1, over 2000 repetitions of the procedure.

In every case the extrapolation with a small undersampling size improves the coverage probability. $E_3(\beta)$ gives also good results but with a greater variability. It is clear that the results are better for $\alpha = 80\%$ than for $\alpha = 95\%$. This may be explained by the fact that the recentring factor for $\alpha = 95\%$ is not estimated accurately for small sampling size. Curiously, the bootstrap seems to perform relatively well for a large quantile compared to the asymptotics.

Finally, to study the robustness of these empirical results for quantiles we give in Table 3 the performance of E_2 against the asymptotics, for $n = 50$ and $b_n = 5$ and 2000 repetitions of the procedure. We study the Gaussian $N(0, 1)$, the chi-square $\chi^2(1)$, the negative chi-square $-\chi^2(1)$ (with opposite skewness) and the lognormal $\text{LN}(0, 1)$ distributions. The results confirm what was observed for the exponential case. The standard deviations are relatively stable for the different distributions.

Our conclusion (limited, of course, by the case studied and the range of the simulations) is that extrapolating an undersampling distribution with the asymptotics may yield some substantial improvements, even if the undersampling distribution is constructed with a very small resampling size. We would also like to emphasize the importance of the finite population factor b_n/n which may have strong effects on the accuracy of the results if it is too big. Further investigations, as started in Bickel *et al.* (1994) and Bertail and Politis (1996), may be needed to validate the use of the extrapolation of undersampling distributions when the asymptotic distribution is unknown.

Acknowledgements

Many thanks are due to Martin Bruegel, Lancelot James, David Mason and Dimitris Politis for their suggestions and careful reading of the manuscript. This work was revised while visiting Albert Y. Lo at University of Science and Technology, Hong Kong; I would like to thank him for his kind hospitality and stimulating discussions. I am also indebted to Peter J. Bickel who pointed out Bickel and Yahav's (1988) paper at the Institute of Mathematical Statistics meeting in Chapel Hill.

References

- Babu, G. and Singh, K. (1985) Edgeworth expansions for sampling without replacement from finite populations. *J. Multivariate Anal.* **17**, 261–278.
- Barbe, P. and Bertail, P. (1995) *The Weighted Bootstrap*, Lecture Notes in Statist. 98. New York: Springer-Verlag.
- Beran, R. (1984) Bootstrap methods in statistics. *Jahresber. Deutsch. Math.-Verein.*, **86**, 24–30.

- Bertail, P. (1992) La méthode du bootstrap, quelques applications et résultats théoriques. Doctoral thesis, Université de Paris IX.
- Bertail, P. (1993) Second-order properties of a corrected bootstrap without replacement. Technical Report 9306, INRA-CORELA, Ivry, France.
- Bertail, P. and Politis, D.N. (1996) Extrapolation of subsampling distribution estimators in i.i.d. and strong-mixing cases. Technical Report 9604, INRA-CORELA, Ivry, France.
- Bertail, P., Politis, D.N. and Romano, J.P. (1995) On subsampling estimators with unknown rate of convergence. Technical Report 9501, INRA-CORELA, Ivry, France.
- Bhattacharya, R.N. and Denker, M. (1990) *Asymptotic Statistics*. Boston: Birkhäuser Verlag.
- Bhattacharya, R.N. and Ghosh, J. (1978) On the validity of Edgeworth expansion. *Ann. Statist.*, **6**, 434–451.
- Bhattacharya, R.N. and Qumsiyeh, M. (1989) Second-order comparisons between the bootstrap and empirical Edgeworth expansions. *Ann. Statist.*, **17**, 160–169.
- Bickel, P.J. and Freedman, D.A. (1981) Some asymptotic theory for the bootstrap. *Ann. Statist.*, **9**, 1196–1217.
- Bickel, P.J. and Yahav, J.A. (1988) Richardson extrapolation and the bootstrap. *J. Amer. Statist. Assoc.*, **83**, 387–393.
- Bickel, P.J., Götze, F. and van Zwet, W.R. (1994) Resampling fewer than n observations: gains, losses and remedies for losses. Technical report no. 419, University of California.
- Booth, J.G. and Hall, P. (1993) An improvement of the jackknife distribution function estimator. *Ann. Statist.*, **21**, 1476–1485.
- Bose, A. (1988) Edgeworth corrected by bootstrap in autoregressions. *Ann. Statist.*, **16**, 1709–1722.
- Bosq, D. (1993) Bernstein type large deviation inequalities for partial sums of strong-mixing processes. *Statistics*, **24**, 59–70.
- Bretagnolle, J. (1983) Lois limites du bootstrap de certaines fonctionelles. *Ann. Inst. H. Poincaré Probab. Statist.*, **19**, 281–296.
- Carlstein, E. (1986) The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *Ann. Statist.*, **14**, 1171–1179.
- Chow, Y.S. and Teicher, H. (1988) *Probability Theory, Independence, Interchangeability, Martingales*, 2nd edn. New York: Springer-Verlag.
- Doukhan, P. (1994) *Mixing: Properties and Examples*, Lecture Notes in Statist. 85. New York: Springer-Verlag.
- Efron, B. (1979) Bootstrap methods: another look at the jackknife. *Ann. Statist.*, **7**, 1–26.
- Falk, M. and Reiss, R.D. (1989) Weak convergence of smoothed and non-smoothed bootstrap quantile estimates. *Ann. Probab.*, **17**, 362–371.
- Götze, F. (1984) Expansions for von mises functionals. *Z. Wahrscheinlichkeitstheorie Verw. Geb.*, **65**, 599–625.
- Götze, F. (1989) Edgeworth expansions in functional limit theorems. *Ann. Probab.*, **17**, 1602–1634.
- Götze, F. and Hipp, C. (1983) Asymptotic expansions for sums of weakly dependent random vectors. *Z. Wahrscheinlichkeitstheorie Verw. Geb.*, **64**, 211–239.
- Götze, F. and Hipp, C. (1989) Asymptotic expansions for potential functions of i.i.d. random fields. *Probab. Theory Related Fields*, **82**, 349–370.
- Götze, F. and Künsch, H.R. (1993) Second-order correctness of the blockwise bootstrap for stationary observations. Preprint 93-061, Sonderforschungsbereich 343 Bielefeld.
- Gray, H., Schucany, W. and Watkins, T. (1972) *The Generalized Jackknife Statistic*. New York: Marcel Dekker.
- Hall, P. (1991a) Edgeworth expansions for nonparametric density estimators, with applications. *Statistics*, **22**, 215–232.

- Hall, P. (1991b) On convergence rates of suprema. *Probab. Theory Related Fields*, **89**, 447–455.
- Hall, P. (1992a) Effect of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability density. *Ann. Statist.*, **20**, 675–694.
- Hall, P. (1992b) *The Bootstrap and Edgeworth Expansion*. New York: Springer-Verlag.
- Hall, P. and Martin, M. (1991) On the error incurred using the bootstrap variance estimate when constructing confidence intervals. *J. Multivariate Anal.*, **38**, 70–81.
- Isaacson, E. and Keller, H.B. (1966) *Analysis of Numerical Methods*. New York: Wiley.
- Künsch, H.R. (1984) Infinitesimal robustness for autoregressive processes. *Ann. Statist.*, **12**, 843–863.
- Künsch, H.R. (1989) The jackknife and the bootstrap for general stationary observations. *Ann. Statist.*, **17**, 1217–1241.
- Lahiri, S.N. (1992) Edgeworth corrected by ‘moving-block’ bootstrap for stationary and non stationary data. In R. Le Page and L. Billard (eds), *Exploring the Limits of the Bootstrap*. New York: Wiley.
- Lo, A.Y. (1991) Bayesian bootstrap clones and a biometry function. *Sankhyā A*, **53**, 320–333.
- Liu, R. and Singh, K. (1992) Moving blocks jackknife and bootstrap capture weak dependence. In R. Le Page and L. Billard (eds), *Exploring the Limits of the Bootstrap*. New York: Wiley.
- Maritz, J.S. and Jarrett, R.G. (1978) A note on estimating the variance of the sample median. *J. Amer. Statist. Assoc.*, **82**, 155–162.
- Mason, D. and Newton, M.A. (1992) A rank statistics approach to the consistency of a general bootstrap. *Ann. Statist.* **20**, 1611–1624.
- Pfanzagl, J. and Wefelmeier, W. (1985) *Asymptotic Expansions for General Statistical Models*. New York: Springer-Verlag.
- Politis, D.N. and Romano, J.P. (1992) A general resampling scheme for triangular arrays of α -mixing random variables with applications to the problem of spectral density estimation. *Ann. Statist.*, **20**, 1985–2007.
- Politis, D.N. and Romano, J.P. (1993) Nonparametric resampling for homogeneous strong-mixing random fields. *J. Multivariate Anal.*, **47**, 301–328.
- Politis, D.N. and Romano, J.P. (1994) A general theory for large sample confidence regions based on subsamples under minimal assumptions. *Ann. Statist.*, **22**, 2031–2050.
- Rhomari, N. (1993) Remarque sur l’inégalité de type exponentielle pour des sommes partielles d’un processus fortement mélangeant. Preprint, LSTA Paris VI and CREST-ENSAE.
- Sargan, D. (1976) Econometric estimators and the Edgeworth expansion. *Econometrica*, **44**, 421–448.
- Sargan, D. (1979) Some approximations to the distribution of econometric criteria asymptotically distributed as chi-squared. *Econometrica*, **49**, 1107–1128.
- Sherman, M. (1992) Subsampling and asymptotic normality for a general statistic from a random field. Ph.D. thesis, Dept. of Statistics, University of North Carolina, Chapel Hill.
- Shao, J. and Wu, C.F.J. (1989) A general theory for jackknife variance estimation. *Ann. Statist.*, **17**, 1176–1197.
- Tu, D. (1992) Approximating the distribution of a general standardized functional statistic with that of jackknife pseudo-values. In R. Le Page and L. Billard (eds), *Exploring the Limits of the Bootstrap*. New York: Wiley.
- Wu, C.F.J. (1990) On the asymptotic properties of the jackknife histogram. *Ann. Statist.*, **18**, 1438–1452.

Received August 1994 and revised October 1996