EXTRAPOLATION OF SUBSAMPLING DISTRIBUTION ESTIMATORS

PATRICE BERTAIL AND DIMITRIS N. POLITIS

Abstract. In Politis and Romano (1994) a general subsampling methodology was put forth for the construction of large-sample confidence regions for a general unknown parameter $\mu = \mu(P)$ under very minimal conditions. Nevertheless, in some specific cases –e.g. in the case of the sample mean of i.i.d. data– it has been noted that the subsampling distribution estimators underperform as compared to alternative estimators such as the bootstrap and/or the asymptotic normal distribution (with estimated variance). In the present report we investigate the extent to which the performance of subsampling distribution estimators can be improved by an extrapolation technique, while at the same time retaining the robustness property of consistent distribution estimation even in nonregular cases; both i.i.d. and weakly dependent (mixing) observations are considered.

Abstract. Politis et Romano (1994) ont introduit une méthode de sous-échantillonnage générale permettant de construire des régions de confiance asymptotiques pour un paramètre $\mu(P)$; sous des hypothèses minimales sur les statistiques et les données en jeu. Néanmoins, il a été montré que les distributions de sous-échantillonnage sont inefficaces par rapport à d'autres méthodes d'approximations tels le Bootstrap ou la distribution asymptotique gaussienne, dans les cas spécifiques où ces dernières sont utilisables (par exemple la moyenne dans le cas i.i.d.). Dans cet article, nous étudions dans quelle mesure les méthodes d'interpolation et extrapolation permettent d'améliorer les méthodes de sous-échantillonange tout en conservant leur aspect robuste dans les cas non-réguliers. Les cas i.i.d. et dépendants sous des hypothèses de mélange sont étudiés et donnent lieu à des résultats différents.

Key words and phrases: Confidence limits, extrapolation, interpolation, large sample inference, subsampling, bootstrap, stationarity, ARMA.
AMS 1991 Subject Classification: Primary 62M20; Secondary 62G05.

1. Introduction

Let $\underline{X}_n = (X_1, \ldots, X_n)$ be an observed stretch of a (strictly) stationary, strong mixing sequence of real-valued random variables $\{X_t, t \in \mathbb{Z}\}$; the probability measure generating the observations is denoted by P. The strong mixing condition 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 $X_1, \ldots, X_n$ are independent, identically distributed (i.i.d.) is an important special case where $\alpha_X(k) = 0$ for all $k > 0$.

In Politis & Romano (1994) a general subsampling methodology has been put forth for the construction of large-sample confidence regions based on a statistics $T_n = T_n(\underline{X}_n)$ estimating a general unknown parameter $\theta = \theta(P)$; under very minimal conditions. In the case of stationary data (time series or random fields), subsampling is closely related to the blocking methods introduced by Hall (1985), Carlstein (1986), Künsch (1989), and Liu & Singh (1992) (see chapter 9 of Shao and Tu (1995)); see also Wu (1990), Sherman (1992) and Sherman & Carlstein (1994) for related ideas.

Let us now make the simplifying assumption that T_n and θ are real-valued. To obtain asymptotically pivotal (or at least, scale-free) statistics, a standardization (also known as 'studentization' when the norming is data-based and random) is often required. Since we will later discuss the influence of the studentization, we introduce a statistic $S_n = S_n(\underline{X}_n) > 0$ converging in probability to some constant $\sigma > 0$; heuristically, σ^2 may stand for the asymptotic variance of $\tau_n(T_n - \theta)$, but this is not necessarily always the case. Without loss of generality, the unstudentized case corresponds to $S_n = 1$.

Although i.i.d. data can be seen as a special case of stationary strong mixing data, the construction of the subsampling distribution can take advantage of the i.i.d. structure when such a structure exists:

- **General case (strong mixing data).** Define Y_i to be the subsequence $(X_i, X_{i+1}, \ldots, X_{i+b-1})$, for $i = 1, \ldots, q$, and $q = n - b + 1$; note that Y_i consists of b 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 $q = \frac{n!}{b!(n-b)!}$ subsets of size b 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 $T_{b,i}$ and $S_{b,i}$ be the values of statistics T_b and S_b as calculated from just subsample Y_i. The subsampling distribution of the root $\tau_n S_n^{-1}(T_n - \theta)$, based on a subsample size b, is defined by

$$K_b(x) \equiv q^{-1} \sum_{i=1}^{q} \mathbf{1}\{\tau_b S_{b,i}^{-1}(T_{b,i} - T_n) \leq x\}. \tag{1}$$

If there is a non-degenerate distribution $K(x, P)$, continuous in x, such that

$$K_n(x, P) \equiv \Pr_P\{\tau_n S_n^{-1}(T_n - \theta) \leq x\} \to K(x, P) \tag{2}$$

as $n \to \infty$, for any real number x, the subsampling methodology was shown to 'work' asymptotically provided that the integer "subsample size" b satisfies $b \to \infty$

and, as $n \to \infty$,

$$\max\left(\frac{b}{n}, \frac{\tau_b}{\tau_n}\right) \to 0$$

The subsampling distribution turns out to be a relatively low-accuracy approximation to the true sampling distribution $K_n(x; P)$, and is actually worse than the asymptotic normal distribution (with estimated variance). Indeed in Bertail (1997) it was proved that the subsampling distribution admits, for suitable b, the same Edgeworth expansion as $K_n(x; P)$ –when such an expansion exists– but in powers of b instead of n. This result has a straighforward consequence when there exists a standardization S_n such that the asymptotic distribution is pivotal and known, i.e. if $K(x; P) = K(x)$ not depending on P. If the rate of the first term in the Edgeworth expansion $f_1(n)$ is known (typically $f_1(n) = n^{1/2}$ in the regular case) then it is possible to improve the subsampling distribution by considering a linear combination of that distribution with the asymptotic distribution:

$$K_n^{int}(x) = \left(1 - \frac{f_1(b_n)}{f_1(n)}\right) K(x) + \frac{f_1(b_n)}{f_1(n)} K_b(x)$$

This type of linear (convex) combination with positive coefficients may be seen as an interpolation in that $K_n^{int}(x)$ is an intermediate point on the straight line segment joining $K_b(x)$ to the asymptotic $K(x)$, in the same way that sample size n is an intermediate point between sample sizes b and ∞; note the ordering $b < n < \infty$ and recall that we are interested in obtaining an estimate of the ordinate (sampling distribution) at sample size n (based on the ordinates at sample sizes b and ∞). This interpolation idea was first considered in Bickel & Yahav(1988) and generalized in Bertail (1997).

Nevertheless the generality of the subsampling methodology lies in the fact that $K(x; P)$ does not have to be known in order for subsampling to work. Therefore, it is of interest to seek improvements upon the subsampling distribution estimators that do not explicitly involve $K(x; P)$. In the present paper, we explore the asymptotic performance of extrapolation similar to the notion of Richardson extrapolation considered by Bickel & Yahav (1988), Bertail(1997) and Bickel *& al.* (1997) to seek the desired improvement. In the present paper we show that the extrapolation of two undersampling distribution $K_{b_1}(x)$ and $K_{b_2}(x)$ can be used to provide us with the linear combination effecting the aforementioned extrapolation of subsampling distribution estimators, and we quantify the improvement achieved by the extrapolation.

In Section 2 we focus on the i.i.d. case and show that since in this case subsampling amounts to sampling without replacement from a finite population, the finite population correction $1 - f$; with $f = \frac{b}{n}$, should necessarily be taken into account to build an accurate approximation of the true distribution. Then extrapolation of subsampling distributions for the sample mean or non degenerate U-V statistics of i.i.d. data achieves second order accuracy.

The strong mixing case studied in section 3 is more complicated because of inaccurate variance estimation. Thus, we include an Appendix where a simple variance estimator is proposed based on the ideas of Politis & Romano (1995) with the property of being almost $\sqrt{n}$ consistent; this accurate variance estimator can then be used in the construction of confidence intervals for the mean using the normal approximation, the subsampling approximation, or other studentized methods (e.g. the block-bootstrap). Then interpolation of subsampling distributions for the sample mean of strong mixing observations achieves second order accuracy; this finding extends the i.i.d. result of Booth & Hall (1993). In particular this results apply to many econometric models with stationary data. Section 4 presents some finite-sample simulations in the context of an ARMA model.

2. Independent identically distributed (i.i.d.) data

2.1. Finite population correction. Bertail (1997) noticed that subsampling may be seen as a particular case of the weighted bootstrap considered in Barbe & Bertail (1995) with some exchangeable weights $W_n = (w_{i,n})_{1\le i\le n}$. The proper normalizing factor for these weights is none other than the finite population correction factor $1-f$ with $f=\frac{b}{n}$; a result foreshadowed by Shao & Wu (1989) in the case of variance estimation (see also Booth & Hall(1993)). This suggests that the adequate renormalization factor in the subsampling distribution is τ_r (instead of τ_b) where r is defined by

$$r = b(1-f)^{-1}$$

and to define more generally the corrected subsampling distribution as

$$\tilde{K}_b(x) = q^{-1}\sum_{i=1}^{q} 1\{\tau_r S_{b,i}^{-1}(T_{b,i}-T_n) \le x\}$$

Clearly, the factor $(1-f)$ has no first-order asymptotic effect on the subsampling distribution which remains consistent under very weak assumptions provided that $f \to 0$. However, the $(1-f)$ factor is of great importance for second order properties as shown in the following paragraphs.

2.2. The studentized sample mean. Consider the problem of estimating the mean $\mu(P) = E_P X_1$. In the following we assume that $E_P X_1^4 < \infty$; and we take $T_n = \bar{X}_n = n^{-1}\sum_{i=1}^n X_i$, and $S_n^2 = n^{-1}\sum_{i=1}^n (X_i-\bar{X}_n)^2$ as usual. From Bhattacharya & Ghosh (1978), under the usual Cramér condition, we have the Edgeworth expansion :

(3)
$$K_n(x;P) = P\left\{ n^{1/2}S_n^{-1}(\bar{X}_n-\mu(P)) \le x \right\} = \Phi(x) + n^{-1/2}p_1(x;P)\phi(x) + O(n^{-1})$$

$$p_1(x;P) = \frac{k_3}{6}(2x^2+1)$$

where k_3 is the skewness.

Following Booth & Hall (1993), under the Cramér condition and assuming that $E|X_i|^{8+\delta} < \infty$, for some $\delta > 0$ we have from Babu & Singh (1985) an Edgeworth expansion for sampling without replacement from a finite population, with $b/n \xrightarrow[n\to\infty]{} 0$; so for any $\varepsilon > 0$;

$$\text{(4)}\quad \tilde{K}_b(x) = \Phi(x) + b^{-1/2}p_1(x;P)\phi(x) + b^{-1}p_2(x;P)\phi(x) - b^{1/2}n^{-1}\frac{1}{4}k_3\phi(x) + O_P(b^{-1/2}n^{-1/2+\varepsilon} + b^{3/2}n^{-2}) + o(b^{-1});$$

where

$$p_2(x;P) = 12^{-1}k_4(x^3-x) - 18^{-1}k_3^2(x^5+2x^3-3x) - 4^{-1}(x^3+3x)$$

and k_4 is the kurtosis. Notice first that for b such that $b/n \to 0$ the right hand side of (4) can not be made smaller than $O_P(n^{-1/2})$: Thus even with the finite population correction the subsampling distribution will not be second order correct.

The extrapolation of two subsampling distributions (one with subsample size b_1 and the other with b_2) is given by

$$\hat{K}_n^{(2)}(x) = \lambda_1 \hat{K}_{b_1}(x) + \lambda_2 \hat{K}_{b_2}(x) \tag{5}$$

where λ_1 and λ_2 are chosen to solve

$$\begin{aligned} \lambda_1 + \lambda_2 &= 1 \\ \lambda_1 b_1^{-1/2} + \lambda_2 b_2^{-1/2} &= n^{-1/2}. \end{aligned}$$

in order to match the second order term in (3) and get a second order valid approximation. We then get

Proposition 1: *Let* b_1 *and* b_2 *such that* $\frac{b_i}{n} \underset{n\to\infty}{\to} 0$ *(for* $i = 1;2$*), and* $\frac{b_2}{b_1} \underset{n\to\infty}{\to} C \in [0;1)$*, then the extrapolation of the two finite-population-corrected subsampling distributions of the studentized mean is actually second order correct with the best choice given by* $b_1 = C_1 n^{2/3}$ *and* $b_2 = C_2 n^{2/3}$ *with* $C_2 < C_1$*. And we have*

$$\sup_x |\hat{K}_n^{(2)}(x) - K_n(x;P)| = O_P(n^{-2/3}); \tag{6}$$

Proof :

From 4 and the definition of the b_i's, we get

$$\begin{aligned} \hat{K}_n^{(2)}(x) - K_n(x;P) = &-b_1^{-1/2} b_2^{-1/2} p_2(x;P)\phi(x) \\ &+ b_1^{1/2}(1 + C^{1/2})\, n^{-1}\frac{1}{4}k_3\phi(x) + O_P(b_1^{-1/2} n^{-1/2+\epsilon} + b_1^{3/2} n^{-2}) + o(b_1^{1/2} n^{-1}): \end{aligned} \tag{7}$$

Minimizing the order of the right hand side of (7) leads to choose b_1 and b_2 such that $b_1 b_2^{1/2}$ is proportional to and yields the result.

Remark 1 : Notice that the order of the whole approximation is worse than the one that we obtain by considering the interpolation of one subsampling distribution with its asymptotic distribution when the latter is known; see Booth & Hall (1993), Bertail (1997) who obtain an error of size $n^{-5/6}$.

Remark 2 : It is important to point out that if we do not take into account the finite population correction factor then the second order validity of the extrapolated version of the two distributions fails; in the case of interpolation, the second order property still holds but with a loss in term of coverage probability; see for instance Bertail (1997) and Bickel *& al.* (1997, p. 17) in which the importance of this correction factor was recognized but not really exploited. Indeed in the absence of the finite population correction, if we let $K_n^{(2)}(x)$ be the extrapolation of $K_{b_1}(x)$ and $K_{b_2}(x)$, then we have

$$\begin{aligned} K_n^{(2)}(x) &= \Phi(x) + n^{-1/2} p_1(x;P)\phi(x) + O_P\left(\frac{b_1}{n}\right) \\ &\quad + O_P(b_1^{-1/2} n^{-1/2+\epsilon} + b_1^{3/2} n^{-2} + b_1^{-1/2} b_2^{-1/2} + b_1^{1/2} n^{-1}); \end{aligned}$$

once again to obtain the second order correctness we would have to choose $n^{1/2} = o(b_1)$ and the loss induced by the sampling-without-replacement scheme (typically of order b_1/n) implies that second order correctness can not be attained.

2.3. General extrapolation result in the i.i.d. case. Of course, in the case of the mean, interpolation and/or extrapolation can be thought to give no advantage since we already know that the usual bootstrap (Efron(1979)) gives an approximation up to $O_P(n^{-1})$ (see Hall (1992)). It is quite obvious that the same holds for smooth functions of means. Even though the usual bootstrap works in these mean-like cases, subsampling may still be useful for computational reasons, since constructing two subsampling distributions requires less simulations than constructing the usual bootstrap distribution. Moreover, in contrast to the interpolation schemes studied in Booth & Hall (1993) and Bertail (1997), the extrapolation does not depend on the asymptotic approximation and is thus more robust. Indeed, if the original assumptions break down, e.g. the assumptions leading to the Edgeworth expansion, or even if the statistic T_n is not asymptotically Gaussian in which case interpolation is not applicable and the bootstrap fails, the extrapolation of the (corrected or uncorrected) subsampling distributions *remains* a consistent distribution estimator.

We conjecture however that in a great number of situations, extrapolation together with a finite population correction will yield second order accuracy, provided that one takes τ_r instead of τ_b in the definition of $\hat{K}_b$, for b such that $f_2^{-1}(b) = o(f_1(n)^{-1})$ by analogy to our Sections 2.1 and 2.2. Using recent Edgeworth expansions results in finite population by Bloznelis & Götze(2000), it is easy to see that this conjecture holds for U statistics of degree 2 with non degenerate first gradient (influence function) and as a consequence for any smooth statistical functional, differentiable according to some nice metric (see Barbe & Bertail(1995)). In that case, $\tau_n = n^{1/2}$ and $\tau_r = b^{1/2}(1-f)^{-1/2}$ is the adequate normalization which makes the extrapolation second order correct. In any case, even if second order accuracy is not achieved, the extrapolated distribution will always improve over each individual subsampling distribution, i.e., an extrapolated distribution will always improve over a single distribution.

2.4. When the convergence rate to the asymptotic approximation is unknown. A case of interest in some practical applications occurs when the order of the difference between the asymptotic and the true distribution is unknown. Indeed the knowledge of the rate f_1 of the asymptotic approximation is implicit in the construction of both the interpolation and extrapolation. Consider for instance the simple case of estimating the mean. If $E_P(X_i - E_P X_i)^3 \neq 0$ then (5) yields a second order correct approximation and improves over the asymptotic. But if $E_P(X_i - E_P X_i)^3 = 0$, then the asymptotic distribution is already second order correct and (5) is less accurate; this follows from the fact that when the skewness is zero, the correct extrapolation (which will indeed improve upon the asymptotic distribution) should be built with $f_1(n) = n$ and not $n^{1/2}$. Other problematic situations occur when the distribution of the X_is is lattice. These examples suggest that we either have to make a preliminary test on some parameter appearing in the Edgeworth expansion (depending on the statistic and the underlying distribution), or we have to directly construct and employ an accurate estimator of f_1 which may then be used in forming the extrapolation. Nevertheless, the first suggestion is not very satisfactory because it is highly problem-dependent, and the Edgeworth expansion may be quite complicated.

Under very general conditions on the statistic and under some conditions on b, the order of the subsampling distribution is $f_1(b)$. Thus if we study a collection of subsampling distributions, when b varies in its domain, we should be able to observe their convergence to the asymptotic distribution in connection to f_1: When $f_1(n) = n^{-\alpha}$, where α is unknown, the following proposition shows that it is possible to estimate the accuracy rate f_1 by a simple regression.

Proposition 2

Let $b_1^{(i)}; b_2^{(i)}$, for $i = 1; \ldots I$, (for simplicity $b_2^{(i)} = b_2$ may be chosen to be the same for all i) be several pairs of different subsampling sizes, satisfying the assumptions of Bertail (1997) A1-A5, with $b_1^{(i)} = n^{-\beta_i}$; $1/2 > \beta_1 > \ldots > \beta_I$, such that $\frac{b_1^{(i)}}{b_2^{(i)}} \underset{n\to\infty}{\longrightarrow} 0$. Assume in addition that $f_1(n) = n^{-\alpha}$, where α is unknown. Then we have uniformly in x

$$\log\left(|K_{b_1^{(i)}}(x) - K_{b_2^{(i)}}(x)|\right) = -\alpha\log(b_1^{(i)}) + \log(|p(x;P)|) + o_P(1); \quad i = 1; \ldots I: \tag{8}$$

Let $\hat\alpha$ be the least square estimator obtained by regressing $\log\left(|K_{b_1^{(i)}}(x) - K_{b_2^{(i)}}(x)|\right)$ on -$\log b_1^{(i)}$. Then we have

$$\hat\alpha = \alpha + o_P(\log(n)^{-1}); \tag{9}$$

As a consequence, interpolation and extrapolation of the finite-population-corrected subsampling distributions with estimated rate $\hat f_1(n) = n^{-\hat\alpha}$ are second order correct.

Proof : Under our assumptions, $K_b(x)$ has the same Edgeworth expansion as $K_n(x;P)$ but on functions of b instead of n. If we choose b_1 and b_2 such that $\frac{b_1}{b_2} \underset{n\to\infty}{\longrightarrow} 0$ then it is easy to see that

$$|K_{b_1}(x) - K_{b_2}(x)| = f_1(b_1)^{-1}(1 + o(1))|p(x;P)|; \tag{10}$$

and (8) follows by taking the log. Now, since $\sum_{i=1}^{I}\left(\log(b_1^{(i)}) - I^{-1}\sum_{i=1}^{I}\log(b_1^{(i)})\right)^2 = C_0(\log n)^2$ for some constant C_0; we thus have (9). Finally, it is easy to see that if we now use $\hat f_1(n) = n^{-\hat\alpha} = f_1(n)(1+o_P(1))$ in place of $f_1(n)$ then the extrapolation and the interpolation remain second order correct.

Remark : Notice that this idea is different but in the same spirit as Bertail & al(1999) who were trying to estimate the rate of the statistics T_n itself. In the more general case, when the functional form of f_1 is unknown, we may consider (10), or rather, its logarithm, as a nonparametric regression for $f_1(\cdot)$. Under a monotonicity constraint, and assuming that $f_1(b) \to \infty$ as $b \to \infty$, consistent estimation of $f_1(\cdot)$ may still be possible albeit more complicated and slower. We will not pursue this approach here.

3. Strong mixing data

The case of strong mixing data is complicated by the fact that an adequate standardization is needed (see Götze & Künsch (1996)). In Hall & Jing (1996), interpolation was used in the context of stationary data, however their results are weakened by the fact that their hypothesis on the Edgeworth expansion with a remainder of size $O(n^{-1})$ only hold in very particular circumstances (typically i.i.d. data with an adequate standardization). Indeed, for dependant data, the bias of the variance estimator may be so important that the second order validity of the interpolation may not even hold, a fact which explains the bad results of their simulation. Thus an accurate simple variance estimator is proposed in our appendix.

Even with this estimator, a close study of the improvements of interpolations and extrapolations is needed.

3.1. The studentized sample mean. Let $T_n = \bar{X}_n = n^{-1}\sum_{t=1}^{n} X_t$ be the sample mean, and $\mu = EX_0$ be the mean. Also let $R(s) = E(X_t - \mu)(X_{t+|s|} - \mu)$, for $s = 0, \pm 1, \pm 2, \dots$ be the autocovariance sequence. Both μ and $R(\cdot)$ are generally unknown, and the objective is to obtain interval estimates for μ based on the data in a nonparametric fashion. Following Götze & Künsch (1996) we assume that

$$\alpha_X(k) \le d^{-1}e^{-dk} \tag{11}$$

for some $d > 0$, and that

$$E|X_0|^s < \infty \tag{12}$$

for some $s \ge 5$. Also assume the Cramér-type regularity conditions A3, A5, and A6 of Götze & Künsch (1996).

Let s_n^2 be an estimator of σ_∞^2 based on $X_1, \dots, X_n$, and accurate enough so that $s_n^2/\sigma_\infty^2 + O_P(\sqrt{\log n/n})$ under conditions (11) and (12); e.g. we can let $s_n^2 = \hat{\sigma}^2_{0.5M,M,n}$ with $M = A\log n$, where the estimator $\hat{\sigma}^2_{m,M,n}$ is defined in the Appendix.

Consider now the subsampling distribution of the studentized sample mean which is defined by

$$L_b(x) \equiv q^{-1}\sum_{i=1}^{q} 1\{\sqrt{b}\,\frac{T_{b,i} - T_n}{s_{b,i}} \le x\}, \tag{13}$$

where $T_{b,i} = b^{-1}\sum_{k=1}^{i+b-1} X_k$ and $s_{b,i}$ is the statistic s_b computed on block $\{X_1, \dots, X_{i+b-1}\}$.

The following proposition states that interpolation of an undersampling distribution with the adequate standardization is second order correct. We give some rate of convergence. Interpolation is not second order correct but improves over only one subsampling distribution.

Proposition 3 *Under the preceding assumptions, we have :*

If $b = \Theta(n^{\frac{s}{3s-4}}\log n^{-1})$, *then the interpolation of* L_b *with* Φ *is second order correct with an error rate* $O_P(\frac{\log n}{n^{(2s-4)/(3s-4)}})$ *close to* $O_P(n^{-2/3})$ *when* s *is large.*

Let $b_i = c_i^2 b$, $i = 1, 2$ *with* $b = \Theta((n\log^2 n)^{s/(3s-4)})$ *then the extrapolation of two undersampling distributions satisfy* $L_2(x) = \Phi(x) + O_P(n^{(2-s)/(3s-4)}(\log n)^{s/(3s-4)})$, *which for large* s *becomes close to* $n^{-1/3}$ *thus improving upon the* $n^{-1/4}$ *rate of* $L_b(x)$, *but not achieving second order correctness.*

Proof :

In the following we use the notation $b = \Theta(m)$ to mean that $b/m \to$ const. $\neq 0$. Define the Edgeworth expansion

$$Q(x) = \Phi(x) + n^{-1/2}k_1\sigma_\infty^{-3/2}[\frac{1}{6}\phi^{(2)}(x) - \frac{1}{2}\phi(x)];$$

where $k_1 = \sum_{i,j} E[(X_0 - \mu)(X_i - \mu)(X_j - \mu)]$ is finite because of (11) and (12); here $\Phi(x), \phi(x), \phi^{(k)}(x)$ denote the standard normal distribution, density, and kth derivative of its density respectively. Now under the assumed conditions, we can employ the results of Götze & Künsch (1996) to infer that

$$\sup_x |P(\sqrt{n}\,\frac{\bar{X}_n - \mu}{s_n} \le x) - Q(x)| = O(\frac{M}{n^{1-2/s}}) + O(\beta_n) = O(\frac{\log n}{n^{1-2/s}}) \tag{14}$$

where

$$\beta_n = E s_n^2 - \sigma_n^2;$$

and where M is the equivalent width of the autocovariance window; see our Appendix for more details. Using the choice $M = A \log n$ for a sufficiently large constant A implies that $\beta_n = O(n^{-1})$ and yields the rate in (14).

Now, similarly to the proof of Theorem 3.1 in Politis & Romano (1994), it can be shown that $Var(L_b(x)) = O(b/n)$ due to the geometric mixing rate (11). Now we have using (14)

$$q^{-1}\sum_{i=1}^{q} 1\{\sqrt{b}\frac{T_{b,i}-\theta}{s_{b,i}} \le x\} = E(1\{\sqrt{b}\frac{T_{b,1}-\theta}{s_{b,1}} \le x\}) + O_P(\sqrt{b/n})$$

$$= \Phi(x) + \frac{k_1 p(x)}{b^{1/2}\sigma_1^{3/2}} + O_P(\frac{\log b}{b^{1-2/s}}) + O_P(\sqrt{\frac{b}{n}}).$$

The above coupled with the fact that $\sqrt{b}(\theta - T_n)/s_{b,i} = O_P(\sqrt{\frac{b}{n}})$ yields that

$$L_b(x) = \Phi(x) + \frac{k_1 p(x)}{b^{1/2}\sigma_1^{3/2}} + O_P(\sqrt{b/n}) + O_P(\log n / b^{1-2/s}). \tag{15}$$

It follows that taking $b = \Theta(\sqrt{n})$; we minimize the Mean Squared Error (MSE) of $L_b(x)$, thus having $L_b(x) = \Phi(x) + O_P(n^{-1/4})$. The remainder of the proof is straighforward by using (15) and minimizing the remainders respectively in the interpolation and the extrapolation.

Remark 1: The choice of b is nearly optimal up to a log factor. However the rate of the interpolation is still worse than the best rate of the block-bootstrap obtained by Götze & Künsch (1996) which can be made close to $O_P(n^{-\frac{3s-4}{4s}})$ that gives $O_P(n^{-3/4})$ when s is infinite. Nonetheless, note that for the interpolation to be correct we need only $s \ge 5$ whereas at least $s \ge 24$ is needed for the block bootstrap to be second order correct; see Götze & Künsch (1996).

Remark 2 The fact that the second order correctness is not attain may be explained by the need of some finite population correction factor as in the i.i.d. case. Recall that the finite population correction factor in the i.i.d. case was in particular due to the fact that

$$Var(b^{1/2}(\bar{X}_{b,i} - \bar{X}_n)) = (1 - \frac{b}{n})\sigma^2;$$

where $\bar{X}_{b,i}$ is the sample mean of subsample Y_i. In the strong mixing case it is quite interesting to see that a similar relation holds thus indicating that the same finite population correction factor (surprisingly of same form as in the i.i.d. case) may be appropriate. A straightforward calculation shows that we have

$$Var(b^{1/2}(\bar{X}_{b,i} - \bar{X}_n)) = (1 - \frac{b}{n})\sigma_1^2 + O(\frac{b}{n^2} + \frac{1}{b} + \frac{b^2}{n^3}). \tag{16}$$

So it is obvious that taking into account the finite population correction factor is generally advisable as it will reduce the error of the subsampling distribution in the mixing case as well. However it is not clear that the correction will improve the order of the extrapolation.

Remark 3 The preceding discussion has made apparent that perhaps second order accuracy may be too much to hope for from extrapolated distributions in the case of strong mixing data, because of the "bad" effect of the standardization.

Thus it may be interesting in that case to use unstandardized distributions. Using the conditions (2.3), (2.5) and(2.6) of Götze & Hipp (1983) as well as our conditions (11) and (12) –with $s > 3$ it is easy to see that regular statistics (functions of moments) admits an Edgeworth expansion uniformly in x. Using the results in Bertail (1997) (conditions $A2[l]$, $\alpha 3[l]$, *and* $\alpha 4[l]$), the extrapolation of l undersampling distribution is clearly first order correct and improves over one undersample distribution. Moreover in regular cases (function of means) if one choose $b = n^{1/l}/\log n$ then as l grows we come close to the first order rate $O(n^{-1/2})$ (without achieving it). See Bertail & Politis (1996) for further details. Obviously, extrapolated subsampling will definitely not be second order correct in the unstudentized case, therefore it will be inferior to the studentized block-bootstrap *when the block-bootstrap applies as well*; see Künsch (1989), Liu & Singh (1992), Götze & Künsch (1996). However the fact that extrapolated subsampling distributions can be a robust and more accurate asymptotic approximation under very weak assumptions is rather remarkable.

4. Some simulation results for ARMA processes.

A straighforward application of our result concerns confidence intervals for parameters of ARMA processes using pseudo-maximum likelihood or robust methods (see for instance Künsch(1984)). In this section, we give some simulations results on ARMA processes estimated by pseudo-maximum likelihood assuming that the underlying likelihood is normal. The model considered here is an ARMA(2,1) model

$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} = \epsilon_t - \theta_1 \epsilon_{t-1}$$

We are interested in confidence intervals for ϕ_1. In the following tables we compare the asymptotic distribution with the interpolated distribution and our extrapolation technique, taking into account the "finite population correction". The exact quantiles of the distribution of the maximum likelihood are computed by Monte-Carlo replications by generating 100000 processes. The mean, the median of the bounds and estimated coverage probability of the interpolated and extrapolated distributions are calculated over 10000 iterations of the procedure.

In Table 1 and 2, the true parameters are $\phi_1 = 0.5; \phi_2 = 0.3$ and $\theta_1 = 0.6$ and the residuals are $N(0;1)$. Table 1 and 2 give respectively the results for an observed stretch of size 50 and 100.

Distribution	α %	2.5	5.0	95.0	97.5
True	$K_n^{-1}(\alpha)$	-7.127	-4.603	1.564	2.511
Asymptotic	$\Phi^{-1}(\alpha)$	-1.960	-1.645	1.645	1.960
	(β%)	12.5	15.4	94.9	96.3
Interpolation $b_n = 7$	Moy	-5.174	-2.332	1.907	4.929
	Med	-3.636	-2.270	1.672	2.436
	(β%)	10.8	14.6	94.6	96.8
Extrapolation $b_1 = 7;\ b_2 = 13$	Moy	-7.393	-3.776	3.244	9.425
	Med	-4.132	-2.632	1.672	2.813
	(β%)	11.4	13.7	94.7	97.1

Table1 : Confidence intervals for ϕ_1; :normal residuals, $n = 50$.

Distribution	® %	2:5	5:0	95:0	97:5
True	$K_n^{¡1}$(®)	¡ 3:853	¡ 2:312	1:754	2:306
Asymptotic	©$^{¡1}$(®) (¯%)	¡ 1:960 7:8	¡ 1:645 10:2	1:645 95:0	1:960 96:9
Interpolation $b_n = 10$	Moy Med (¯%)	¡ 2:753 ¡ 2:848 5:2	¡ 1:994 ¡ 2:008 8:4	1:669 1:618 95:5	2:211 2:135 97:9
Extrapolation $b_1 = 10; b_2 = 21$	Moy Med (¯%)	¡ 3:424 ¡ 3:300 5:9	¡ 2:279 ¡ 2:422 8:3	1:704 1:549 93:5	2:646 2:106 97:1

Table 2 : Con…dence intervals for $Á_2$; :normal residuals, n = 100:

The …rst striking feature of these simulation results is how far the asymptotic quantiles are from the true quantile. Both the extrapolation and the interpolation succeed in catching the asymmetry of the true distribution. Curiously the extrapolation gives better results in terms of average estimation of the quantile than the interpolation. However in term of coverage probability, the interpolation gives a better result as predicted by the theoretical results. Both in terms of quantile estimation and coverage probability, the extrapolation and the interpolation outperform the asymptotic distribution.

In Table 3 and 4 , the model is actually AR(1), thus corresponding to $Á_1 = 0:9$; $Á_2 = 0$ and $¹_1 = 0$: Moreover the true residuals have a log normal distribution recentered at 0. These simulations are used to test the e¤ect of the asymmetry of the distribution of the residuals on the pseudo-likelihood estimator and the robustness of the extrapolations.

Distribution	® %	2:5	5:0	95:0	97:5
True	$K_n^{¡1}$(®)	¡ 1:802	¡ 1:521	2:978	4:404
Asymptotic	©$^{¡1}$(®) (¯%)	¡ 1:960 1:7	¡ 1:645 4:3	1:645 89:5	1:960 92:0
Interpolation $b = 7$	Moy Med (¯%)	¡ 1:851 ¡ 1:751 2:7	¡ 1:499 ¡ 1:406 6:5	2:238 2:259 94:2	3:332 3:301 96:6
Extrapolation $b_1 = 7; b_2 = 13$	Moy Med (¯%)	¡ 1:836 ¡ 1:635 3:2	¡ 1:503 ¡ 1:270 6:7	2:448 2:464 92:9	3:561 3:372 94:7

Table 3 : Con…dence intervals for $Á_1$; log-normal residuals,n = 50.

Distribution	® %	2:5	5:0	95:0	97:5
True	$K_n^{¡1}$(®)	¡ 1:820	¡ 1:538	2:177	3:046
Asymptotic	©$^{¡1}$(®) (¯%)	¡ 1:960 1:6	¡ 1:645 4:4	1:645 92:4	1:960 94:0
Interpolation $b = 10$	Moy Med (¯%)	¡ 1:843 ¡ 1:792 2:6	¡ 1:506 1:452 6:2	2:095 2:114 94:6	2:881 2:951 97:4
Extrapolation $b_1 = 10; b_2 = 21$	Moy Med (¯%)	¡ 1:821 ¡ 1:683 2:8	¡ 1:510 ¡ 1:379 6:4	2:336 2:254 96:0	3:205 3:015 97:7

Table 4 : Con…dence intervals for $Á_1$; log-normal residuals, n = 100.

In comparison to the previous simulations, the true distribution exhibits an opposite asymmetric behavior. Once again the interpolation and the extrapolation catches this asymmetry. For $n = 100$, the coverage probability are very close to the nominal level but the interpolation which is known to be second order correct in that case clearly exhibits a better behavior in terms of coverage probability.

Appendix: Accurate variance estimation in the strong mixing case

Many estimators of $\sigma_\infty^2 = \lim_{n\to\infty} Var(\sqrt{n}\bar{X}_n) = \sum_{s=-\infty}^{\infty} |R(s)| < \infty$ (under our mixing condition (11)) have been proposed in the literature; probably the most popular one (under many different names and variations, see Politis & Romano (1995)) is

$$\hat{\sigma}^2_{M;n} = \frac{M}{Q}\sum_{i=1}^{Q}(\bar{X}_{i;M;L} - \bar{X}_n)^2; \tag{17}$$

where $\bar{X}_{i;M;L} = M^{-1}\sum_{t=L(i-1)+1}^{L(i-1)+M} X_t$ is the mean of the block $\{X_{L(i-1)+1}, \ldots, X_{L(i-1)+M}\}$ of the data, the numbers L, M are integers depending on the sample size n, and $Q = [\frac{n-M}{L}] + 1$, with $[\cdot]$ being the integer part; M is the block's size here, L is the amount of 'lag' between the starting points of block i and block $i+1$, and Q is the total number of such blocks that can be extracted from the data. If $L = M$, there is no overlap between block i and block $i+1$. The full-overlap case corresponding to $L = 1$ is recommended (see, e.g. Künsch (1989)); thus we set $L = 1$ in what follows.

Under regularity conditions, $\hat{\sigma}^2_{M;n}$ is a consistent and asymptotically normal estimator. The regularity conditions are moment and mixing conditions, and conditions on the design parameters; typically $M \to \infty$, but with $M/n \to 0$. Consistency is immediate considering the first two moments of $\hat{\sigma}^2_{M;n}$ that can be asymptotically calculated to be

$$Bias(\hat{\sigma}^2_{M;n}) = E\hat{\sigma}^2_{M;n} - \sigma^2_\infty = O(1/M) + O(M/n); \tag{18}$$

$$Var(\hat{\sigma}^2_{M;n}) = 2c\frac{M}{n}\sigma^4_\infty + o(M/n):$$

Realizing that the poor rate of convergence of $\hat{\sigma}^2_{M;n}$ is due to its bias, Politis & Romano (1995) in the more general case of estimation of the spectral density $g(w)$ proposed a bias-corrected version which in our setting is given by

$$\bar{\sigma}^2_{m;M;n} \equiv (h+1)\hat{\sigma}^2_{M;n} - h\hat{\sigma}^2_{m;n}; \tag{19}$$

here h is some chosen positive constant, and m is chosen as $m = hM/(1+h)$. The choice $h = 1$, leading to $m = M/2$ is proposed as a simple solution, and an empirical data-driven method for choosing M is presented in Politis & Romano (1995). Equation (19) can be interpreted as an extrapolation of the two subsampling variance estimators $\hat{\sigma}^2_{M;n}$ and $\hat{\sigma}^2_{m;n}$ that has an improved asymptotic performance.

The difference between the set-up of variance estimation considered here and the set-up of spectral density estimation considered in Politis and Romano (1995) is that in the usual spectral density estimation practice, the true mean μ is assumed known, and used (in place of $\bar{X}_n$) in constructing Bartlett's estimator; the implication is that improper centering leads to some -usually negligible- "edge effects". For example, the added $O(M/n)$ bias term in equation (18) above is due to this improper centering in the construction of $\hat{\sigma}^2_{M;n}$, i.e., centering the data at $\bar{X}_n$ instead of μ.

Similarly to what shown by Politis & Romano (1995), for the bias-corrected Bartlett estimator, it may be shown that not only the bias of $\bar{\sigma}^2_{m;M;n}$ becomes

$o(1/M)$, but a more spectacular bias correction is achieved: namely, under the exponential strong mixing assumption (11) we have that in the case of the mean

$$\text{(20)} \qquad Bias(\hat{\sigma}^2_{m,M,n}) = O(1/n)$$

if $M = A \log n$, for some sufficiently large constant A. In other words, we have (for $h = 1$, say) that

$$\text{(21)} \qquad \hat{\sigma}^2_{0.5A \log n, A \log n, n} = \sigma^2_\infty + O_P(\sqrt{\log n/n});$$

thus $\hat{\sigma}^2_{0.5A \log n, A \log n, n}$ may be used whenever an accurate variance estimator is needed. For example, it may be used for studentization in the context of subsampling distributions discussed here, or block-bootstrap distributions in Götze & Künsch (1996).

Notice that if $\alpha_X(k) = 0$ for all $|k|$ bigger than some $K > 0$, i.e., if the data are K-dependent, then it can be shown additionally that $Bias(\hat{\sigma}^2_{m,M,n}) = O(1/n)$, even when m and M are constants, satisfying $M \geq m \geq K$. Consequently, taking $K = m = M/2$ we have that $\hat{\sigma}^2_{K,2K,n} = \sigma^2_\infty + O_P(1/\sqrt{n})$; in other words, $\hat{\sigma}^2_{K,2K,n}$ achieves the parametric $\sqrt{n}$ rate in this case!

As a final practical comment, note that $\hat{\sigma}^2_{m,M,n}$ is not almost surely nonnegative; this is not a problem with $\hat{\sigma}^2_{m,M,n}$ in particular, but rather of all higher-order accurate variance (or spectral) estimators (see Politis & Romano (1995)). Although this problem disappears asymptotically, in finite samples it might pose a real problem, especially if we want to studentize using $\hat{\sigma}^2_{m,M,n}$. If we happen to compute a $\hat{\sigma}^2_{m,M,n} < 0$ and the obvious solution to take zero as our estimate is not acceptable (e.g. in the studentization set-up) there are two practical ways out:
(a) try out different choices for M (or for A, if we take $M = A \log n$), and use the corresponding $\hat{\sigma}^2_{m,M,n}$ if it turns out to be positive, or
(b) use a fraction of $\bar{\sigma}^2_{M,n}$ (e.g. $\lambda\bar{\sigma}^2_{M,n}$ for some $\lambda \in (0,1]$) as our variance estimator, i.e., 'shrink' the estimator $\bar{\sigma}^2_{M,n}$ towards zero.

Regarding (b) note that $\hat{\sigma}^2_{m,M,n}$ can be interpreted as $\bar{\sigma}^2_{M,n} - \widehat{Bias}(\bar{\sigma}^2_{M,n})$; a negative $\hat{\sigma}^2_{m,M,n}$ indicates that our estimate of the bias of $\bar{\sigma}^2_{M,n}$ is positive and large (actually: too large). Nevertheless, we might take the hint that $\bar{\sigma}^2_{M,n}$ has a positive bias and attempt to reduce it by taking $\lambda\bar{\sigma}^2_{M,n}$ as our estimator. The additional difficult question of choosing λ actually prompts us to favor method (a) of solving the problem of a negative estimate. Simulations results (see Bertail & Politis (1996)) further support suggestion a), that a negative $\hat{\sigma}^2_{m,M,n}$ can occur as a result of a poor choice of M.

References

[1] Babu, G. & Singh, K.(1985). Edgeworth expansions for sampling without replacements from …nite populations. *J. Multivariate Analysis* 17 261-278.

[2] Barbe, Ph. & Bertail, P. (1995). *The Weighted Bootstrap*. Lecture Notes in Statistics No. 98, Springer Verlag.

[3] Bhattacharya, R.N. and Ghosh J. (1978). On the validity of Edgeworth expansion. *Ann. Statist.* 6 434-451.

[4] Bertail, P. (1997), Second order properties of an extrapolated bootstrap without replacement: the i.i.d. and the strong mixing cases, *Bernoulli*, 3, 149-179.

[5] Bertail, P. , Politis, D.N., & Romano, J.P. (1999), On subsampling estimators with unknown rate of convergence, *JASA*, 94, 569-579.

[6] Bertail, P. , Politis, D.N. (1996), Extrapolation of subsampling distributions, Technical report 9604, INRA-Corela.

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

[8] Bickel, P.J. & Yahav, J.A. (1988). Richardson extrapolation and the bootstrap. *J. Amer. Statist. Assoc.* 83 387-393.

[9] Bloznelis, M. & Götze F. (2000). An Edgeworth Expansion for Finite Population U-statistics, *Bernouilli*, 6, 729-760.

[10] Booth, J.G. & Hall P. (1993). An improvement of the jackknife distribution function estimator. *Ann. Statist.* 21 1476-1485

[11] 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

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

[13] Götze, F. & Hipp, C. (1983), Asymptotic expansions for sums of weakly dependent random vectors, *Zeit. Wahrschein. verw. Geb.*, vol 64, pp. 211-239.

[14] Götze, F. & Künsch, H.R. (1996), Second order correctness of the blockwise bootstrap for stationary observations, *Ann. Statist.*, 24, 1914-1933.

[15] Hall, P. (1985), Resampling a coverage pattern, *Stochastic Processes and Their Applications,* 20, 231-246.

[16] Hall, P. (1992), *The Bootstrap and Edgeworth expansion*, Sprinver Verlag, N.Y.

[17] Hall, P., Jing, B. (1996), On sample reuse methods for dependent data. *Journal of the Royal Statistical Society* Series B, 58, 727-737.

[18] Künsch, H.R. (1984). In…nitesimal robustness for autoregressive processes. *Ann. Statist.* 12 843-863.

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

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

[21] Politis, D.N. & Romano, J.P. (1994). 'Large Sample Con…dence Regions Based on Subsamples under Minimal Assumptions', *Ann. Statist.*, vol. 22, No. 4, 2031-2050.

[22] Politis, D.N. & Romano, J.P. (1995), Bias-corrected nonparametric spectral estimation, *J. Time Ser. Anal.*, vol. 16, no. 1, pp. 67-104.

[23] Shao, J. & Tu, D. (1995), *The jackknife and the bootstrap*, Springer Verlag, New York.

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

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

[26] Sherman, M. & 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.

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

Acknowledments : This research paper was partially supported by an AIP contract of the laboratory Corela at INRA We would like to thank the referees and Pr Genest for their comments and suggestions which greatly improved the presentation of this paper.

Adresses :

Pr. P. Bertail, INRA-Corela and University Paris X, MODALX

Université Paris X, UFR Segmi , Bat. G, 200, ave de la république, 92001 Nanterre, France

email : Patrice.Bertail@u-paris10.fr

Pr. D. Politis, Department of Mathematics, University of California et San Diego, La Jolla. CA 92093, USA

email : politis@math.ucsd.edu