

# Subsampling, Symmetrization, and Robust Interpolation

Dimitris N. Politis

Department of Mathematics  
University of California, San Diego  
La Jolla, CA 92093-0112, USA  
politis@euclid.ucsd.edu

Joseph P. Romano

Department of Statistics  
Stanford University  
Stanford, CA 94305, USA  
romano@stat.stanford.edu

Michael Wolf

Departamento de Estadística y Econometría  
Universidad Carlos III, Madrid, Spain  
mwolf@est-econ.uc3m.es

## Abstract

The recently developed subsampling methodology has been shown to be valid for the construction of large-sample confidence regions for a general unknown parameter  $\theta$  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 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 a (partial) symmetrization 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.

**Key words:** Extrapolation, jackknife, large sample inference, resampling.

# 1 Introduction

Let  $\underline{X}_n = (X_1, \dots, X_n)$  be an observed stretch of a (strictly) stationary, strong mixing sequence of random variables  $\{X_t, t \in Z\}$  taking values in some general space  $S$ ; 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  $\sigma$ -algebras generated by  $\{X_t, t < 0\}$  and  $\{X_t, t \geq k\}$  respectively; the case where  $X_1, \dots, X_n$  are independent, identically distributed (i.i.d.) is an important special case where  $\alpha_X(k) = 0$  for all  $k > 0$ .

In Politis and Romano (1994), a general subsampling methodology was put forth for the construction of large-sample confidence regions for a general unknown parameter  $\theta = \theta(P)$  under very minimal conditions; see also Wu (1990) where subsampling distribution estimators were first considered in the particular case where  $\theta$  is the sample mean of real-valued, i.i.d. data.

The subsampling methodology hinges on approximating the sampling distribution of a statistic  $T_n = T_n(\underline{X}_n)$  that is consistent for  $\theta$  at some known rate  $\tau_n$ . Note that, in general, the rate of convergence  $\tau_n$  depends on  $P$  as well, although this dependence will not be explicitly denoted; the case with unknown convergence rate is studied in Bertail et al. (1999).

In the present paper, we make the simplifying assumption that  $T_n$  and  $\theta$  are real-valued. To obtain asymptotically pivotal (or at least, scale-free) statistics, a standardization or ‘studentization’ is often required. Thus, we also introduce a statistic  $\hat{\sigma}_n = \hat{\sigma}_n(\underline{X}_n)$  with the purpose of estimating the ‘scale’ of  $T_n$ .

In the case of i.i.d. data, subsampling may be seen as a delete- $d$  (with  $d = n - b$ ) jackknife (cf. Shao and Wu (1989), or Shao and Tu (1995)), but also as resampling (bootstrap) *without* replacement with a resampling size  $b$  smaller than the original sample size  $n$  (cf. Politis and Romano (1993), or Bickel et al. (1997)). In the case of stationary data (time series or random fields), subsampling is closely related to the blocking methods of Carlstein (1986), Künsch (1989), Liu and Singh (1992), and Sherman and Carlstein (1994, 1996).

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; of course, if one is unsure regarding the independence assumption, it is safer (and more robust) to operate under the general strong mixing assumption.

- **General case (strong mixing data).** Define  $Y_i$  to be the subsequence  $(X_i, X_{i+1}, \dots, X_{i+b-1})$ , for  $i = 1, \dots, q$ , and  $q = n - b + 1$ ; note that  $Y_i$  consists of  $b$  consecutive

observations from the  $X_1, \dots, X_n$  sequence, and the order of the observations is preserved.

- **Special case (i.i.d. data).** Let  $Y_1, \dots, Y_q$  be equal to the  $q = \frac{n!}{b!(n-b)!}$  subsets of size  $b$  chosen from  $\{X_1, \dots, X_n\}$ , and then ordered in any fashion; here the subsets  $Y_i$  consist of unordered observations.

In either case, let  $T_{b,i}$  and  $\hat{\sigma}_{b,i}$  be the values of statistics  $T_b$  and  $\hat{\sigma}_b$  as calculated from just subsample  $Y_i$ . The subsampling distribution of the root  $\tau_n \hat{\sigma}_n^{-1}(T_n - \theta)$ , based on a subsample of size  $b$ , is defined by

$$K_b(x) \equiv q^{-1} \sum_{i=1}^q 1\{\tau_b \hat{\sigma}_{b,i}^{-1}(T_{b,i} - T_n) \leq x\}. \quad (1)$$

Under the assumption that, as  $n \rightarrow \infty$ ,

$$\hat{\sigma}_n \xrightarrow{P} \sigma > 0, \quad (2)$$

where  $\sigma$  is some constant, and assuming that there is a well-defined asymptotic distribution for the centered, ‘studentized’ statistic  $\tau_n \hat{\sigma}_n^{-1}(T_n - \theta)$ , i.e., assuming that there is a distribution  $K(x, P)$ , continuous in  $x$ , such that

$$K_n(x, P) \equiv \Pr_P\{\tau_n \hat{\sigma}_n^{-1}(T_n - \theta) \leq x\} \rightarrow K(x, P) \quad (3)$$

as  $n \rightarrow \infty$ , for any real number  $x$ , the subsampling methodology was shown to ‘work’ asymptotically, provided also that the integer “subsample size”  $b$  satisfies

$$b \rightarrow \infty \quad (4)$$

and

$$\max\left(\frac{b}{n}, \frac{\tau_b}{\tau_n}\right) \rightarrow 0 \quad (5)$$

as  $n \rightarrow \infty$ . In other words, subsampling “works” in the sense that

$$\sup_x |K_b(x) - K(x, P)| = o_P(1) \quad (6)$$

and subsequently also that

$$\sup_x |K_b(x) - K_n(x, P)| = o_P(1) \quad (7)$$

as  $n$  tends to infinity; cf. Politis and Romano (1994). Equation (6) can then be used to construct confidence intervals for  $\theta$  of asymptotically correct coverage.

Nevertheless, the rate of convergence of  $K_b(x)$  to  $K(x, P)$  is of some concern to us in order to see how large the sample size should be such that the approximation of  $K(x, P)$  or of  $K_n(x, P)$  by  $K_b(x)$  is reasonably accurate. The benchmark for comparison is provided by the Berry-Esseen theorem in the sample mean case stating that, under some regularity conditions (e.g., finite third moment), there is an approximation (namely the normal) that is in error by  $O_P(1/\sqrt{n})$  from  $K_n(x, P)$ . The subsampling distribution in this case 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; see Wu (1990). This phenomenon may be explained by the fact that the Berry-Esseen bound for the subsampling distribution based on subsamples of size  $b$  gives an error of size  $O(1/\sqrt{b})$ . Nevertheless, an interpolation idea, first introduced by Booth and Hall (1993) for the particular case of the sample mean of i.i.d data, was shown to be able to improve the accuracy of subsampling.

To describe the interpolation idea for general statistics, note that Bertail (1997) proved that, if  $K_n(x, P)$  admits an Edgeworth expansion of the type

$$K_n(x, P) = K(x, P) + f_1(n)^{-1}p(x, P) + O(f_2(n)^{-1}) \quad (8)$$

for some increasing functions  $f_1$  and  $f_2$  satisfying  $f_1(n) = o(f_2(n))$ , and where the  $O(f_2(n)^{-1})$  term in (8) is uniform in  $x$ , then —under some extra conditions as well as possible restrictions on the subsampling size  $b$ — the subsampling distribution also admits the *same* Edgeworth expansion but in powers of  $b$  instead of  $n$ ; that is,

$$K_b(x) = K(x, P) + f_1(b)^{-1}p(x, P) + O_P(f_2(b)^{-1}), \quad (9)$$

where the  $O_P(f_2(b)^{-1})$  term in (9) is uniform in  $x$ . This result has a straightforward consequence when there exists a standardization  $\hat{\sigma}_n$  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:

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

As pointed out in Bertail (1997), the ‘interpolation’  $K_b^{int}(x)$  is closely related to the generalized jackknife (see Gray, Schucany and Watkins(1972)) involving the two non-second order correct estimators  $K_b(x)$  and  $K(x)$ ;  $K_b^{int}(x)$  is generally a more accurate estimator of the limit  $K(x)$  as compared to the subsampling distribution  $K_b(x)$ . Nevertheless, the generality of the subsampling methodology lies in the fact that  $K(x)$  does

not have to be known in order for subsampling to work. Therefore, it is of interest to seek an interpolation method that does *not* explicitly involve  $K(x)$  but still yields improvements over the initial distribution estimator. The next section shows how this goal can indeed be achieved under the sole assumption that the limit distribution  $K(x, P)$  is *symmetric*, i.e.,  $K(x, P) = 1 - K(-x, P)$ , for all  $x$ . Finally, section 3 discusses an interesting application: the sample mean of i.i.d. data with possibly heavy tails.

## 2 Subsampling and partial symmetrization

To construct the ‘robust interpolation’ subsampling distribution, we start by letting  $K_b^{flip}(x) = 1 - K_b(-x)$ , for all  $x$ ; also let

$$K_b^{symm}(x) = (K_b(x) + K_b^{flip}(x))/2 = (K_b(x) + 1 - K_b(-x))/2.$$

Finally construct the partially symmetrized subsampling distribution

$$K_b^{rob}(x) = \left(1 - \frac{f_1(b)}{f_1(n)}\right) K_b^{symm}(x) + \frac{f_1(b)}{f_1(n)} K_b(x); \quad (10)$$

the following theorem justifies the title ‘robust interpolation’ for the partially symmetrized subsampling distribution  $K_b^{rob}(x)$ .

**Theorem 2.1** *Assume conditions (2), (3), (4), and (5), where the limit  $K(x, P)$  is continuous in  $x$ , and symmetric, i.e.,  $K(x, P) = 1 - K(-x, P)$ , for all  $x$ ; then*

$$(a) \quad \sup_x |K_b^{rob}(x) - K_n(x, P)| = o_P(1)$$

as  $n \rightarrow \infty$ .

*If in addition it so happens that the two Edgeworth expansions (8) and (9) hold true for some increasing functions  $f_1$  and  $f_2$ —satisfying  $f_1(n) = o(f_2(n))$ , and a symmetric function  $p(x, P)$ , i.e.,  $p(x, P) = p(-x, P)$ , for all  $x$ , then*

$$(b) \quad \sup_x |K_b^{rob}(x) - K_n(x, P)| = O_P(f_2(b)^{-1})$$

as  $n \rightarrow \infty$ .

**Remark 1(a)** Part (a) of Theorem 2.1 shows that  $K_b^{rob}(x)$  is consistent under very general conditions, namely: existence of a continuous and symmetric limit for the sampling distribution  $K_n(x, P)$ . Since existence of an asymptotic distribution is a *sine qua non* assumption in large-sample theory, the only real restriction is the symmetry (and continuity) of the limit. Incidentally, the continuity restriction can be somewhat relaxed; see,

e.g., Politis and Romano (1994), or Politis, Romano, and Wolf (1999) for more details.

**Remark 1(b)** Part (b) of Theorem 2.1 shows that, under some extra conditions, the consistency of  $K_b^{rob}(x)$  will occur at a fast rate. Of course, part (b) of Theorem 2.1 becomes of real interest if the functions  $f_1$  and  $f_2$  are such that allow us to pick the subsample size  $b$  to satisfy

$$f_2(b)^{-1} = o(f_1(n)^{-1}), \quad (11)$$

while at the same time satisfying conditions (4) and (5). In that case, it is apparent that  $K_b^{rob}(x)$  becomes a “higher-order accurate” estimator of the sampling distribution  $K_n(x, P)$ , in the sense that  $K_b^{rob}(x)$  will provide a more accurate approximation to  $K_n(x, P)$  as compared to the asymptotic distribution  $K(x, P)$ . The combination of parts (a) and (b), i.e., general validity in conjunction with higher-order accuracy (when higher-order accuracy is possible), shows that  $K_b^{rob}(x)$  achieves higher-order accuracy in a robust way.

**Remark 1(c)** In Section 3, we will discuss the particular example of the sample mean of i.i.d. data, where  $f_1(n) = 1/\sqrt{n}$ , and  $f_2(n) = 1/n$ . In that case, to satisfy equation (11), a choice of  $b$  such that  $\sqrt{n} = o(b)$  would be required; a typical choice would be to take  $b$  proportional to  $n^\gamma$ , for some fixed  $\gamma \in (1/2, 1)$ .

**Proof of Theorem 2.1.** First note that, for any real  $x$ , the results of Politis and Romano (1994) imply that  $K_b(x)$  is consistent for  $K(x, P)$ . Since  $K(x, P)$  is symmetric in  $x$ , it follows that  $K_b^{flip}(x)$  is also consistent for  $K(x, P)$ . Consequently,  $K_b^{symm}(x)$ , as well as  $K_b^{rob}(x)$ , are both consistent estimators of  $K(x, P)$ , each being a convex combination of consistent estimators. Part (a) then follows from the continuity of  $K(x, P)$ , and Polya’s theorem.

Regarding part (b), note that equation (9) implies that:

$$K_b(-x) = K(-x, P) + f_1(b)^{-1}p(x, P) + O_P(f_2(b)^{-1}),$$

where the symmetry of  $p(x, P)$  was used. Hence,

$$\begin{aligned} K_b^{flip}(x) &= 1 - K_b(-x) = \\ &= 1 - K(-x, P) - f_1(b)^{-1}p(x, P) + O_P(f_2(b)^{-1}), \end{aligned}$$

and therefore:

$$K_b^{symm}(x) = (K_b(x) + K_b^{flip}(x))/2 = K(x, P) + O_P(f_2(b)^{-1}).$$

Finally, note that, under the assumed Edgeworth expansions (8) and (9), it follows that

$$K_b^{rob}(x) = K(x, P) + f_1(n)^{-1}p(x, P) + O_P(f_2(b)^{-1}), \quad (12)$$

and part (b) follows.  $\square$

It is quite interesting that condition (2) is not really necessary: subsampling has been shown to be asymptotically valid even in the case of ‘self-normalized’ statistic where typically the rate  $\tau_n$  is of simple form but the statistic  $\hat{\sigma}_n$  used for studentization has a nondegenerate asymptotic distribution; see, e.g., Romano and Wolf (1998a,b), or Politis, Romano and Wolf (1999).

In order to relax the assumption (2) we impose the following: let  $a_n$  and  $d_n$  be positive sequences satisfying

$$\tau_n = a_n/d_n. \quad (13)$$

Suppose that, as  $n \rightarrow \infty$ ,

$$a_n(T_n - \theta) \xrightarrow{\mathcal{L}} V, \quad (14)$$

and

$$d_n\hat{\sigma}_n \xrightarrow{\mathcal{L}} W, \quad (15)$$

where  $V$  and  $W$  are some random variables with the distribution of  $W$  not having positive mass at 0.

For the general robust interpolation theorem given below we also need to replace condition (5) with the stronger (16) that reads:

$$\max\left(\frac{b}{n}, \frac{\tau_b}{\tau_n}, \frac{a_b}{a_n}\right) \rightarrow 0 \quad (16)$$

as  $n \rightarrow \infty$ .

**Theorem 2.2** *Assume conditions (3), (4), (13), (14), (15), and (16), where the limit  $K(x, P)$  is continuous in  $x$ , and symmetric, i.e.,  $K(x, P) = 1 - K(-x, P)$ , for all  $x$ ; then*

$$(a) \quad \sup_x |K_b^{rob}(x) - K_n(x, P)| = o_P(1)$$

as  $n \rightarrow \infty$ .

*If in addition it so happens that the two Edgeworth expansions (8) and (9) hold true for some increasing functions  $f_1$  and  $f_2$ —satisfying  $f_1(n) = o(f_2(n))$ , and a symmetric function  $p(x, P)$ , i.e.,  $p(x, P) = p(-x, P)$ , for all  $x$ , then*

$$(b) \quad \sup_x |K_b^{rob}(x) - K_n(x, P)| = O_P(f_2(b)^{-1})$$

as  $n \rightarrow \infty$ .

**Proof of Theorem 2.2.** The proof is very similar to the proof of Theorem 2.1 taking into account Theorem 11.3.1 (i.i.d. case) and Theorem 12.2.2 (strong mixing case) of Politis, Romano and Wolf (1999) that guarantee the consistency of  $K_b(x)$  under the assumed conditions.  $\square$

**Remark 2** In the i.i.d. case, it has been observed that subsampling is nothing other than sampling *without* replacement from the finite population  $\{X_1, \dots, X_n\}$ . Therefore, it is not surprising that the finite population correction might be helpful; see, e.g., Shao and Wu (1989), Wu (1990), Booth and Hall (1993), Shao and Tu (1995), or Bickel et al. (1997). In essence, the finite population correction amounts to using  $\tau_r$  instead of  $\tau_b$  in constructing the subsampling distribution  $K_b(x)$ , where  $r^{-1} = b^{-1} - n^{-1}$ . By condition (5) or (16), it follows that  $\tau_r/\tau_b \rightarrow 1$ , and consequently the finite population correction has no effect on the (first-order) asymptotic validity; it does, however, come into play when higher-order properties are concerned, and its use is highly recommended. A similar situation occurs in the case of strong mixing data where the use of the finite population correction (surprisingly, of the *same* form as in the i.i.d. case) is also recommended; see Bertail and Politis (1996), and Politis, Romano and Wolf (1999). Instead of repeating Theorems 2.1 and 2.2 to include the use of the finite population correction factor, we illustrate its use in the following section where an interesting application is discussed.

### 3 Application: The sample mean of i.i.d. data with possibly heavy tails

In this section we impose the following assumption.

**Assumption A.** Assume  $X_1, X_2, \dots$  is an i.i.d. sequence of real-valued random variables such that the centered sequence  $X_1 - \theta, X_2 - \theta, \dots$  lies in the normal domain of attraction of a *symmetric* stable distribution  $J_\alpha(\cdot, P)$  with some index of stability  $1 < \alpha \leq 2$ ; here,  $\theta = EX_t$ , and the symmetry assumption for the limit law means  $J_\alpha(x, P) = 1 - J_\alpha(-x, P)$ , for all  $x$ .

The case  $\alpha = 2$  corresponds to the case of a Gaussian limit law, while if  $\alpha < 2$ , then we have a ‘heavy-tailed’ distribution. Assumption A is tantamount to assuming that

$$\Pr_P\{n^{1-1/\alpha} \sum_{i=1}^n (X_i - \theta) \leq x\} \rightarrow J_\alpha(x, P) \quad (17)$$



uniformly in  $x$  as  $n \rightarrow \infty$ ; cf. Feller (1966).

Consider the problem of estimating the mean  $\theta$  by the sample mean  $T_n = \bar{X}_n = n^{-1} \sum_{i=1}^n X_i$ . We also define the usual sample variance by

$$\hat{\sigma}_n^2 = (n-1)^{-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2.$$

Note that the value of  $\alpha$  is generally *unknown* in practice. To obtain confidence intervals for  $\theta$ , it is possible to use the subsampling distribution of the unstudentized sample mean  $\bar{X}$  in conjunction with an estimated value of the unknown parameter  $\alpha$ ; see Bertail et al. (1999) for details on this approach which is essentially based on equation (17).

Alternatively, we may look at the sampling distribution of the studentized sample mean, i.e., look at  $K_n(x, P) \equiv \Pr_P\{n^{1/2}\hat{\sigma}_n^{-1}(T_n - \theta) \leq x\}$ . It is actually remarkable that, although for  $\alpha < 2$  the estimator  $\hat{\sigma}_n$  diverges, the studentized distribution  $K_n(x, P)$  possesses a well-defined limit  $K(x, P)$  under our general Assumption A. In other words, equation (3) holds true here for some well-defined continuous distribution  $K(x, P)$ , and for  $\tau_n = \sqrt{n}$  *not depending on  $\alpha$* . The reason for this interesting phenomenon is that  $\hat{\sigma}_n$  grows like  $n^{1/\alpha-1/2}$  with high probability so that studentization has a ‘self-normalization’ effect in this case; see Logan et al. (1973) for more details.

It is equally remarkable that the subsampling distribution  $K_b(x)$ , defined exactly as in equation (1), is consistent for  $K(x, P)$  in this general case; see Romano and Wolf (1998a), or Politis, Romano, and Wolf (1999).

Note, that in the regular case  $\alpha = 2$ , under some additional conditions (the Cramér condition and existence of higher moments — see, e.g., Bhattacharya and Ghosh, 1978) we have the Edgeworth expansion:

$$\begin{aligned} K_n(x, P) &= P \left\{ n^{1/2} \hat{\sigma}_n^{-1} (\bar{X}_n - \theta(P)) \leq x \right\} \\ &= \Phi(x) + n^{-1/2} p_1(x, P) \phi(x) + O(n^{-1}) \end{aligned} \tag{18}$$

with

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

where  $k_3 \equiv \frac{E(X_1 - EX_1)^3}{(E(X_1 - EX_1)^2)^{3/2}}$  is the skewness of  $X_1$ ; in the above,  $\Phi(\cdot)$  and  $\phi(\cdot)$  represent the standard normal distribution and density function respectively.

We will now employ the symmetrization ideas of our previous sections to ‘have our cake and eat it too’: we will construct a distribution estimator that is not only consistent for  $K(x, P)$  in the fat-tailed case, but higher-order accurate as well in case higher-order accurate estimation is possible, i.e., if regularity conditions happen to be satisfied; all this will be achieved *without* requiring knowledge of the parameter  $\alpha$ .

Start by constructing the finite-population corrected subsampling distribution  $\tilde{K}_b(x)$  by

$$\tilde{K}_b(x) \equiv q^{-1} \sum_{i=1}^q 1\{\tau_r \hat{\sigma}_{b,i}^{-1}(T_{b,i} - T_n) \leq x\}$$

where  $r^{-1} = b^{-1} - n^{-1}$  and  $\tau_n = n^{1/2}$ . Note that  $\tilde{K}_b(x) = K_b(x\tau_b/\tau_r)$ ; therefore,  $\tilde{K}_b(x)$  is consistent for  $K(x, P)$  whenever  $K_b(x)$  is itself consistent since the factor  $\tau_b/\tau_r \rightarrow 1$  under conditions (4) and (5). Nevertheless, the finite-population correction is important for higher-order consistency properties.

To elaborate, recall that in the regular case  $\alpha = 2$ , (and under the Cramér condition together with assuming that  $E|X_i|^{8+\eta} < \infty$ , for some  $\eta > 0$ ), Babu and Singh (1985) give the following Edgeworth expansion for sampling without replacement from a finite population, with  $b/n \xrightarrow[n \rightarrow \infty]{} 0$ . For any  $\epsilon > 0$ ,

$$\begin{aligned} \tilde{K}_b(x) &= \Phi(x) + b^{-1/2} p_1(x, P) \phi(x) + b^{-1} p_2(x, P) \phi(x) \\ &\quad + O(b^{1/2} n^{-1}) + O_P(b^{-1/2} n^{-1/2+\epsilon}) + o(b^{-1}), \end{aligned} \quad (19)$$

where

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

and

$$k_4 = \frac{E(X_1 - EX_1)^4}{(E(X_1 - EX_1)^2)^2}.$$

To construct the robust interpolation, we start by letting  $\tilde{K}_b^{flip}(x) = 1 - \tilde{K}_b(-x)$ , and let

$$\tilde{K}_b^{symm}(x) = (\tilde{K}_b(x) + \tilde{K}_b^{flip}(x))/2 = (\tilde{K}_b(x) + 1 - \tilde{K}_b(-x))/2.$$

Finally construct the partially symmetrized subsampling distribution

$$\tilde{K}_b^{rob}(x) = \sqrt{b/n} \tilde{K}_b(x) + (1 - \sqrt{b/n}) \tilde{K}_b^{symm}(x). \quad (20)$$

The following theorem justifies calling  $\tilde{K}_b^{rob}(x)$  a ‘‘robust interpolation’’.

**Theorem 3.1** *Under Assumption A, and conditions (4) and (5) we have*

$$(a) \quad \sup_x |\tilde{K}_b^{rob}(x) - K_n(x, P)| = o_P(1)$$

as  $n \rightarrow \infty$ .

*If in addition it so happens that  $\alpha = 2$ , and the two Edgeworth expansions (18) and (19) hold true (for some small constant  $\epsilon > 0$ ), then*

$$(b) \quad \sup_x |\tilde{K}_b^{rob}(x) - K_n(x, P)| = o_P(1/\sqrt{n})$$

*if  $b \geq \text{const.} n^{1/2+\epsilon}$  as  $n \rightarrow \infty$ .*

**Proof of Theorem 3.1.** First note that Logan et al. (1973) showed that, under our Assumption A, equation (3) holds true with  $\tau_n = \sqrt{n}$ , and for some well-defined continuous and *symmetric* distribution  $K(x, P)$  satisfying  $K(x, P) = 1 - K(-x, P)$  for all  $x$ .

Now, for any real  $x$ , Proposition 11.4.3 of Politis, Romano and Wolf (1999) implies that  $K_b(x)$  is consistent for  $K(x, P)$ . Since  $\tilde{K}_b(x) = K_b(x\tau_b/\tau_r)$ , the consistency of  $\tilde{K}_b(x)$  follows from noting that  $\tau_b/\tau_r \rightarrow 1$  under conditions (4) and (5). Since  $K(x, P)$  is symmetric in  $x$ , it follows that  $\tilde{K}_b^{flip}(x)$  is also consistent for  $K(x, P)$ . Consequently,  $\tilde{K}_b^{symm}(x)$ , as well as  $\tilde{K}_b^{rob}(x)$ , are both consistent estimators of  $K(x, P)$ , each being a convex combination of consistent estimators. Part (a) then follows from Polya's theorem.

Regarding part (b), note that equation (19) together with the condition  $b \geq \text{const}.n^{1/2+\epsilon}$  implies that:

$$\tilde{K}_b(x) = \Phi(x) + b^{-1/2}p_1(x, P)\phi(x) + b^{-1}p_2(x, P)\phi(x) + O_P(b^{1/2}n^{-1}),$$

and thus:

$$\tilde{K}_b(-x) = \Phi(-x) + b^{-1/2}p_1(x, P)\phi(x) - b^{-1}p_2(x, P)\phi(x) + O_P(b^{1/2}n^{-1}),$$

where it was used that  $p_1(x, P) = p_1(-x, P)$ , and  $p_2(x, P) = -p_2(-x, P)$  for all  $x$ .

Hence,  $\tilde{K}_b^{flip}(x) = 1 - \tilde{K}_b(-x) =$

$$= \Phi(x) - b^{-1/2}p_1(x, P)\phi(x) + b^{-1}p_2(x, P)\phi(x) + O_P(b^{1/2}n^{-1}),$$

and therefore:  $\tilde{K}_b^{symm}(x) = (\tilde{K}_b(x) + \tilde{K}_b^{flip}(x))/2 =$

$$= \Phi(x) + b^{-1}p_2(x, P)\phi(x) + O_P(b^{1/2}n^{-1}).$$

Now note that  $p_2(x, P)\phi(x)$  is bounded; hence, under the assumed Edgeworth expansions it follows that

$$\tilde{K}_b^{rob}(x) = \Phi(x) + n^{-1/2}p_1(x, P)\phi(x) + O_P(b^{-1}) + O_P(\sqrt{b}/n) \quad (21)$$

and part (b) follows.  $\square$

**Remark 3(a)** Observe that equation (21) implies that the most accurate approximation by  $\tilde{K}_b^{rob}(x)$  is achieved by taking  $b = \text{const}.n^{2/3}$ , yielding an approximation error of order  $O_P(n^{-2/3})$ . As is to be expected, this is slightly worse than the error of size  $O_P(n^{-5/6})$  obtained by Booth and Hall (1993) in their (non-robust) interpolation; nevertheless,  $\tilde{K}_b^{rob}(x)$  retains its asymptotic validity even in the case of a non-Gaussian limit, i.e., the case  $\alpha < 2$ .

**Remark 3(b)** To elaborate on the usefulness of the subsampling-based robust interpolation note that a practitioner can stick to working with the usual studentized sample mean for inference regarding the location parameter  $\theta$ . As long as this inference (confidence intervals and/or hypothesis tests) is based on the robust interpolation estimator of the sampling distribution, the practitioner is assured of higher-order accurate inferences under the usual regularity conditions (finite variance, asymptotic Gaussian limit, etc.); at the same time the first-order asymptotic validity is maintained even if the aforementioned regularity conditions break down, under the sole provision of existence of a symmetric  $\alpha$ -stable limit law. Notably, all this is achieved *without* knowledge (or explicit estimation) of the unknown parameter  $\alpha$ .

**Acknowledgement.** Research partially supported by NSF grants DMS97-03964 and DMS97-04487. Many thanks are due to Patrice Bertail (INRA-France) for many helpful discussions on the subject of general Edgeworth expansions.

## References

- [1] Babu, G. and Singh, K.(1985). Edgeworth expansions for sampling without replacement from finite populations. *Journal of Multivariate Analysis*, **17**, 261-278.
- [2] 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.
- [3] Bertail, P. and Politis, D.N. (1996). Extrapolation of subsampling distributions in the i.i.d. and strong mixing cases, Technical Report 9604, INRA-CORELA, Ivry, France.
- [4] Bertail, P., Politis, D.N., and Romano, J.P. (1999), On subsampling estimators with unknown rate of convergence, to appear in *Journal of the American Statistical Association*, June 1999.
- [5] Bickel, P., Götze, F. and van Zwet, W.R. (1997), Resampling fewer than n observations: Gains, losses and remedies for losses, *Statistica Sinica*, **7**, 1-31.
- [6] Booth, J.G. and Hall P. (1993). An improvement of the jackknife distribution function estimator. *Annals of Statistics*, **21**, 1476-1485
- [7] Carlstein, E.(1986), The use of subseries values for estimating the variance of a general statistic from a stationary sequence, *Annals of Statistics*, **14**, 1171-1179

- [8] Feller, W. (1966). *An introduction to probability theory and its applications, Vol. II*, John Wiley, New York.
- [9] Gray, H., Schucany, W. and Watkins, T.(1972). *The Generalized Jackknife Statistics*, Marcel Dekker, New York.
- [10] Logan, B.F., Mallows, C.L., Rice, S.O., and Shepp, L.A. (1973). Limit distributions of self-normalized sums. *Annals of Probability*, **1**, 788–809.
- [11] Politis, D.N. and Romano, J.P.(1993). ‘Estimating the Distribution of a Studentized Statistic by Subsampling’, *Bulletin of the International Statistical Institute*, 49th Session, Firenze, August 25 -September 2, 1993, Book 2, pp. 315-316.
- [12] Politis, D.N. and Romano, J.P. (1994). ‘Large Sample Confidence Regions Based on Subsamples under Minimal Assumptions’, *Annals of Statistics*, **22**, 2031-2050.
- [13] Politis, D.N., Romano, J.P. and Wolf, M. (1999). *Subsampling*, Springer Verlag, New York. (to appear: August 1999)
- [14] Romano, J.P. and Wolf, M. (1998a), Inference for the mean in the heavy-tailed case, Technical Report, Dept. of Statistics, Stanford University.
- [15] Romano, J.P. and Wolf, M. (1998b), Subsampling confidence intervals for the autoregressive root, Technical Report, Dept. of Statistics, Stanford University.
- [16] Shao, J. and Tu, D. (1995), *The jackknife and the bootstrap*, Springer Verlag, New York.
- [17] Shao, J. and Wu, C.F.J. (1989). A general theory for jackknife variance estimation. *Annals of Statistics*, **17**, 1176–1197.
- [18] Wu, C.F.J. (1990). On the asymptotic properties of the jackknife histogram. *Annals of Statistics*, **18**, 1438-1452.