Bootstrap and other resampling methodologies in Statistics of Extremes *

D. Prata Gomes

Universidade Nova de Lisboa, FCT and CMA

M. Manuela Neves

Universidade Técnica de Lisboa, ISA and CEAUL

Abstract

In Statistics of Extremes the estimation of parameters of extreme or even rare events is usually done under a semi-parametric framework. The estimators are based on the largest k ordered statistics in the sample or on the excesses over a high level u and although showing good asymptotic properties, most of them present a strong dependence on k or u with high bias when the k increases or the level u decreases. The use of resampling methodologies has revealed to be promising in the reduction of the bias and in the choice of k or u. Different approaches for resampling need to be considered depending on whether we are in an independent or in a dependent setup. A great amount of investigation has been performed for the independent situation. The

*Research partially supported by National Funds through **FCT**—Fundação para a Ciência e a Tecnologia, projects PEst-OE/MAT/UI0006/2011, PEst-OE/AGR/UI0239/2011, PTDC/FEDER and CMA.

main objective of this paper is to use bootstrap and jackknife methods in the context of dependence to obtain more stable estimators of a parameter that appears characterizing the degree of local dependence on extremes, the so-called *extremal index*. A simulation study illustrates the application of those methods.

Keywords. Bias reduction, bootstrap, jackknife, semi-parametric estimation, statistics of extremes.

1 Introduction and Motivation

In many environmental situations we are concerned with the occurrence of events more extremes than any that have already been observed.

Let us suppose now we want do study the behaviour of $M_n = \max\big(X_1, X_2, \cdots, X_n\big)$ (or $m_n = \min\big(X_1, X_2, \cdots, X_n\big)$) where $X_1, X_2, \cdots, X_n$ is a sequence of independent random variables having a common distribution function (d.f.) F.

Whereas the d.f. of M_n may be written down exactly, $P\big\{M_n \le x\big\} = F^n(x)$, it strongly depends on the form of F, so the interest relies on the asymptotic distribution of M_n.

The central result of classical extreme value theory, due in varying degrees of generality to Fréchet, Fisher and Tippet and Gnedenko, establishes that the sequence of maximum values, M_n, linearly normalized, converges weakly towards a nondegenerate random variable.

Theorem 1.1. *(Extremal types theorem) Let* $M_n = \max\big(X_1, X_2, \cdots, X_n\big)$, *where* X_i *are i.i.d. If*

$$P\Big\{\big(M_n - b_n\big)/a_n \le x\Big\} \xrightarrow[n\to\infty]{d} G(x),$$

holds for a non-degenerate d.f. G *and* $a_n > 0$ *and* b_n *normalizing constants, then* G *belongs to one of the Gumbel, Fréchet or Weibull families that can be* *combined into a single family of models having d.f. of the form*

$$G(x) = \begin{cases} \exp\left[-\left\{1+\gamma\left(\frac{x-\mu}{\sigma}\right)\right\}^{-\frac{1}{\gamma}}\right]; & 1+\gamma\left(\frac{x-\mu}{\sigma}\right) > 0,\ \gamma \neq 0, \\ \\ \exp\left[-\exp\left(-\frac{x-\mu}{\sigma}\right)\right]; & \gamma = 0, \end{cases}$$

with $\mu \in R$ and $\sigma > 0$. This is the *extreme value*, $EV(\mu, \sigma, \gamma)$, family of distributions. This theorem considers an i.i.d. random sample, but it holds true if the original scheme being no longer i.d. still remains independent. However, in many practical situations, this assumption is not valid. For example, for the amount of rain in a given location on consecutive days, it is obvious that the chance of rain after a rainy day is much higher than the chance of rain after a dry day. But now the complexity of situations that can be considered is enormous, so some kind of dependence must be assumed.

Unlike in independent case, where only a limited family can arise as limit distributions, in the dependent case any distribution can arise as the limit. Some important dependent sequences have been studied and the limit distributions of their order statistics under some dependence structures are then known. Stationary sequences are examples of those sequences and are realistic for many real problems.

Dependence in stationary sequences can assume several forms, so some conditions are needed to be imposed. The first condition, known as *the* $D(u_n)$ *dependence condition* makes precise the notion that extreme events being sufficiently distant are nearly independent. Let us denote $F_{i_1,i_2,\cdots,i_p}(u_1, u_2, \cdots, u_p) := P\{X_{i_1} \leq u_1, X_{i_2} \leq u_1, \cdots, X_{i_p} \leq u_p\}$, the joint d.f. of $(X_{i_1}, X_{i_2}, \cdots, X_{i_p})$ for any arbitrary positive integers $(i_1, i_2, \cdots, i_p)$.

Definition 1.1. *($D(u_n)$ dependence condition, Leadbetter(1974)). Let $\{u_n\}$ be a real sequence. The condition $D(u_n)$ is said to be hold if for any set of integers $i_i < i_2 < \cdots < i_p$ and $j_i < j_2 < \cdots < j_q$ such that $j_1 - i_p > \ell$, we* *have*

$$\Big|F_{i_1,i_2,\cdots,i_p,j_1,j_2,\cdots,j_q}(u_n,u_n,\cdots,u_n) \\ - F_{i_1,i_2,\cdots,i_p}(u_n,u_n,\cdots,u_n)F_{j_1j_2,\cdots,j_q}(u_n,u_n,\cdots,u_n)\Big| \le \alpha_{n,\ell}, \quad (1)$$

where $\alpha_{n,\ell}$ *is nondecreasing and* $\lim_{n\to\infty} \alpha_{n,\ell_n} = 0,$ *for some sequence* ℓ_n *such that* $\ell_n/n \to 0$ *as* $n \to \infty$.

For a specific sequence of thresholds u_n that increase with n, the $D(u_n)$ condition ensures that, for sets of variables that are far enough apart, the difference of d.f., as in (1), while not zero, is sufficiently close to zero to have no effect on the limit laws for extremes. This is summarized by the following result.

Theorem 1.2. *(Limit distributions of maxima under* $D(u_n)$ *condition, Leadbetter (1974)) Let* $\{X_n\}$ *be a stationary sequence and define* $M_n = \max\big(X_1, X_2, \cdots, X_n\big)$*. Then if* $\{a_n > 0\}$ *and* $\{b_n\}$ *are sequences of constants such that*

$$P\big\{M_n \le u_n\big\} \underset{n\to\infty}{\longrightarrow} H(x)$$

where H *is a non-degenerate distribution function, and the* $D(u_n)$ *condition is satisfied with* $u_n = a_n x + b_n$ *for every real* x*,* H *is an EV distribution.*

This result implies that, provided a sequence has limited long-range dependence at extreme levels, maxima of stationary sequence follow the same distributional limit laws as those of independent sequence. However, the parameters of the limit distribution are affected by the dependence in the sequence. A summary of the result is given in following Theorem.

Theorem 1.3. *(Coles, 2001) Let* $\{X_n\}$ *be a stationary sequence and* $\{\widehat{X}_n\}$ *be a sequence of independent variables with the same marginal distribution. Define* $M_n = \max\big(X_1, X_2, \cdots, X_n\big)$ *and* $\widehat{M}_n = \max\big(\widehat{X}_1, \widehat{X}_2, \cdots, \widehat{X}_n\big)$*. Under suitable regularity conditions,*

$$P\Big\{\big(\widehat{M}_n - b_n\big)/a_n \le x\Big\} \underset{n\to\infty}{\longrightarrow} G(x),$$

for normalizing sequences $\{a_n > 0\}$ *and* $\{b_n\}$, *where* G *is a non-degenerate distribution function, if only if*

$$P\Big\{\big(M_n - b_n\big)/a_n \leq x\Big\} \underset{n\to\infty}{\longrightarrow} H(x),$$

where

$$H(x) = G^{\theta}(x) \tag{2}$$

for a constant θ *such that* $0 < \theta \leq 1$.

Thus if G is an EV d.f. with parameters λ, δ, γ then so is H with parameters $(\lambda_\theta, \delta_\theta, \gamma_\theta)$, where

$$\lambda_\theta = \lambda - \delta\frac{1-\theta^{\gamma}}{\gamma}(\leq \lambda), \qquad \delta_\theta = \delta\theta^{\gamma}, \qquad \gamma_\theta = \gamma.$$

The quantity θ, that appears in (2), is termed *extremal index* and is a parameter that needs to be adequately estimated by itself and because its influence in other parameters. In Section 2, the definition and some characterizations of the extremal index will be provided as well as some of the classical estimators and their properties. Despite of good asymptotic properties, for finite samples the estimate depends crucially of the upper level u_n. This section will end with an example showing the difficulties on estimating θ.

As much as we know, in a dependent setup and specially for the extremal index estimation, we have not found references regarding the application of resampling methodologies for dealing with bias and the level u_n. In Section 3 we shall propose a methodology based on bootstrap and jackknife procedures for the situation where observations are dependent and classical resampling procedures cannot be applied.

In Section 4, some of the results of an extensive simulation study will be presented. For the simulation study several examples of stationary sequences, for which the extremal index is known, were considered.

2 The extremal index

2.1 Definition and different characterizations

The extremal index, θ, measures the relationship between the dependence structure of the data and the behaviour of the exceedances over a high threshold u_n. This threshold u_n is such that, with τ fixed, the underlying d.f. F verifies

$$F(u_n) = 1 - \tau/n + o(1/n), \qquad n \to \infty. \tag{3}$$

The notion of the extremal index was briefly mentioned in the introduction. Let us see the definition:

Definition 2.1. *(Definition of θ, Leadbetter* et. al *(1983)) Let $\{X_n\}$ be a strictly stationary sequence with marginal distribution function F and $M_n = \max\big(X_1, X_2, \cdots, X_n\big)$. We say that the process has extremal index, $\theta \in [0,1]$, if for every $\tau > 0$, there exists a sequence of thresholds $\{u_n(\tau)\}_{n\geq 1}$ such that $nP\big\{X_n > u_n(\tau)\big\} \to \tau$ and $P\big\{M_n \leq u_n(\tau)\big\} \to \exp(-\theta\tau)$ as $n \to \infty$.*

This definition does not involve any dependence restriction on the sequence $\{X_n\}$. If, however, $\{X_n\}$ is a stationary sequence with $D(u_n(\tau))$ holding for each $\tau > 0$ ($u_n(\tau)$ satisfying (3)) it may be shown that, if $P\big\{M_n \leq u_n(\tau)\big\}$ converges for some $\tau > 0$, then $P\big\{M_n \leq u_n(\tau)\big\} \underset{n\to\infty}{\longrightarrow} \exp(-\theta\tau)$, for all $\tau > 0$ and $\{X_n\}$ has extremal index, $\theta \in [0,1]$, (Leadbetter and Rootzén, 1988).

Several interpretations of the extremal index have appeared, leading to several suggestions for its estimation. One way of interpreting the extremal index of a stationary sequence is in terms of the tendency of the process to cluster at extreme levels. Leadbetter (1983) interpreted θ as the inverse of the limiting mean cluster size, where a cluster is defined as the set of exceedances of the threshold u_n that occur in an arbitrary block of length r_n, with $r_n = o(n)$, given that at least one exceedance occurs in the block.

With this definition of a cluster, the cluster size distribution π_n at level u_n is defined as

$$\pi_n(j) = P\Big\{ \sum_{i=1}^{r_n} I(X_i > u_n) = j | M_{r_n} > u_n \Big\}, \quad j \in \mathbb{N},$$

and if it exists, the limiting cluster size distribution is written as $\pi(j) = \lim_{n\to\infty} \pi_n(j)$.

Given a sequence $\{u_n\}$ and with $r_n = o(n)$, the extremal index can be defined in terms of exceedances in a block of length r_n through $\theta = \lim_{n\to\infty} \theta_n$, where θ_n, can have several interpretations. One of them is:

$$\begin{aligned} \big(\theta_n^B\big)^{-1} &= \frac{r_n\big(1 - F(u_n)\big)}{P\big\{M_{r_n} > u_n\big\}} \\ &= E\Big(\sum_{i=1}^{r_n} I(X_i > u_n) | M_{r_n} > u_n \Big) \\ &= \sum_{j=1}^{r_n} j\pi_n(j). \end{aligned} \tag{4}$$

The clusters of exceedances may be identified asymptotically as runs of consecutive exceedances and cluster sizes as run lengths. Under regularity conditions the conditional expected run length is approximately equal to $1/\theta$ (Nandagopalan, 1990). Thus the extremal index can also be defined as the limit of runs of exceedances over u_n, θ_n^R, defined by

$$\theta_n^R = P\big\{ \max(X_2, \cdots, X_{r_n}) \le u_n | X_1 > u_n \big\},$$

i.e, as the probability that a high threshold exceedance is the last in a cluster of exceedances (O' Brien, 1987).

Now, identifying clusters by the occurrence of downcrossings or upper-crossings we can write, respectively,

$$\theta_n^{DC} = P\big\{X_2 \le u_n | X_1 > u_n\big\} \quad \text{or} \quad \theta_n^{UC} = P\big\{X_2 > u_n | X_1 \le u_n\big\}.$$

Under mild conditions one can show that $\lim_{n\to\infty} \theta_n^B = \lim_{n\to\infty} \theta_n^R = \theta$. Clustering increases the mean distance between clusters at high thresholds by factor $1/\theta$, relative to an independent sequence with the same marginal distribution.

There is statistical consequences of clustering. Let us see that if we estimate the tail of marginal distribution F by fitting, for example, to block maxima, then $P\{M_n \le x\} \approx F(x)^{n\theta} \approx H(x)$, where H is EV with parameters $(\lambda_\theta, \delta_\theta, \gamma_\theta)$. The marginal quantiles are approximately $F^{-1}(p) \approx H^{-1}(p^{n\theta}) \ge H^{-1}(p^n)$ so they may be much larger than would be the case with $\theta = 1$. Then ignoring θ can lead to under-estimating a return level computed from F.

These results suggest that identifying independent clusters and estimate θ is fundamentally important for statistical applications of stationary sequences.

2.2 Extremal index estimation

Classical estimators of θ have been developed based on the characterizations for θ of Leadbetter (1983) and O'Brien (1987). A general form for those estimators is given by

$$\widehat{\Theta}_n \sim \frac{C_n(u_n)}{N_n(u_n)}, \quad \text{as} \quad n \to \infty,$$

where $N_n(u_n)$ is the number of exceedances of a high threshold u_n and $C_n(u_n)$ is the number of independent clusters above u_n. Thus, estimating θ is equivalent to identify independent clusters.

As a first attempt to identify clusters, $C_n(u_n)$ can be considered as the number of down-crossing of u_n (or up-crossing of u_n), what gives the naive estimators,

$$\widehat{\Theta}_n^{DC}(u_n) := \frac{\sum_{i=1}^{n} I\left(X_i > u_n, X_{i+1} \le u_n\right)}{\sum_{i=1}^{n} I\left(X_i > u_n\right)}, \tag{5}$$

$$\widehat{\Theta}_n^{UC}(u_n) := \frac{\sum_{i=1}^{n} I\left(X_i \le u_n < X_{i+1}\right)}{\sum_{i=1}^{n} I\left(X_i > u_n\right)}, \tag{6}$$

called Down-Crossing estimator (or Up-Crossing estimator), (Nandagopalan, 1990), (Gomes,1990, 1992, 1993).

Nandagopalan showed that u_n must verify condition (5) in order to have consistency of estimators (5) and (6). Given the sample $\mathbf{X}_n := (X_1, \ldots, X_n)$ and the associated ascending order statistics, $X_{1:n} \leq \cdots \leq X_{n:n}$, we shall consider the level u_n as a deterministic level $u \in [X_{n-k:n}, X_{n-k+1:n}[$. The Up-Crossing estimator can now be written as a function of k, the number of top order statistics above the chosen threshold,

$$\widehat{\Theta}_n^{UC}(k) := \frac{1}{k} \sum_{i=1}^{n-1} I\,(X_i \leq X_{n-k:n} < X_{i+1}).$$

For many dependent structures, the bias of $\widehat{\Theta}^{UC}(k)$ (or of $\widehat{\Theta}^{DC}(k)$) has two dominant components of orders k/n and $1/k$, i.e.

$$Bias\big(\widehat{\Theta}_n^{UC}(k)\big) = \varphi_1(\theta)\left(\frac{k}{n}\right) + \varphi_2(\theta)\left(\frac{1}{k}\right) + o\left(\frac{k}{n}\right) + o\left(\frac{1}{k}\right)$$

The Generalized Jackknife methodology has the property of estimating the bias and the variance of estimators and so of allowing to build estimators with bias and mean squared error smaller than those of an initial set of estimators.

The Generalized Jackknife methodology states that if the bias has two main terms we would like to reduce, we need to have access to three estimators, with the same type of bias.

Definition 2.2. *(Gray and Schucany, 1972) Given three biased estimators of θ $T_n^{(1)}$, $T_n^{(2)}$ and $T_n^{(3)}$ such that*

$$E[T_n^{(i)} - \theta] = b_1(\theta)\varphi_1^{(i)}(n) + b_2(\theta)\varphi_2^{(i)}(n) \qquad i = 1, 2, 3,$$

the generalized jackknife statistic (of order 2) is given by

$$T_n^{GJ} := \frac{\begin{Vmatrix} T_n^{(1)} & T_n^{(2)} & T_n^{(3)} \\ \varphi_1^{(1)}(n) & \varphi_1^{(2)}(n) & \varphi_1^{(3)}(n) \\ \varphi_2^{(1)}(n) & \varphi_2^{(2)}(n) & \varphi_2^{(3)}(n) \end{Vmatrix}}{\begin{Vmatrix} 1 & 2 & 3 \\ \varphi_1^{(1)}(n) & \varphi_1^{(2)}(n) & \varphi_1^{(3)}(n) \\ \varphi_2^{(1)}(n) & \varphi_2^{(2)}(n) & \varphi_2^{(3)}(n) \end{Vmatrix}}.$$

Using this methodology Gomes *et al.* (2008) studied a reduced-bias Generalized Jackknife estimator of order 2, $\widehat{\Theta}_n^{GJ}$, based on the estimator $\widehat{\Theta}^{UC}$ computed at the three levels, k, $\lfloor k/2 \rfloor + 1$ and $\lfloor k/4 \rfloor + 1$, ($\lfloor x \rfloor$ denotes, as usual, the integer part of x), given by

$$\widehat{\Theta}_n^{GJ} := 5\widehat{\Theta}^{UC}(\lfloor k/2 \rfloor + 1) - 2\big(\widehat{\Theta}^{UC}(\lfloor k/4 \rfloor + 1) + \widehat{\Theta}^{UC}(k)\big). \tag{7}$$

Other estimators were defined depending on the way that clusters are defined. Two common methods used to define clusters are the blocks and the runs declustering schemes.

A complete description, as well as the study of the asymptotic properties, such as consistency and normality of the blocks and runs estimators can be seen in Hsing (1991, 1993), Smith and Weissman (1994) and Weissman and Novak (1998). These estimators are consistent under certain conditions, and asymptotically normal if $\theta < 1$. Where comparisons have been made, the runs method is generally found to be preferable because is less biased than the blocks estimator.

In addition to the runs and blocks estimators of θ, more estimators have recently been proposed: maximum likelihood estimator (Ancona-Navarrete and Tawn, 2000), two-threshold estimator (Laurini and Tawn, 2003), intervals estimator (Ferro and Segers, 2003) and K-gaps estimator (Süveges and Davison, 2010), among others.

Although several estimators of the extremal index θ have been proposed, most of them require the choice of a threshold u_n or a block length, because they show a high variance for high levels and a high bias when the level decreases. Regarding the compromise between these two measures given by the mean squared error, MSE, a resampling scheme and an adaptive procedure have revealed to perform quite well for estimating the optimal number of ordered statistics to be used in the estimation of parameters of rare events. However as much as we know no work has been done regarding θ estimation.

This paper tries to use resampling techniques for stabilizing the estimates when u_n increases. The adaptive choice of k or u_n is out of the scope of the paper.

For illustrating what we are claiming, let us consider the simplest moving-maximum process - the two-dependent model (hereafter referred to as the Model MB1), defined by

Model MB1 $- X_i = \max\{Z_i, Z_{i+1}\}, \quad i \in \mathbb{N}$, where $\{Z_i\}$ are i.i.d. random variables with standard exponential distribution, $F(z) = \exp(-z)$, $z > 0$. $\{X_n\}$ is stationary with d.f. defined by $P\{X_i \leq x\} = \big(P\{Z_i \leq x\}\big)^2 = F^2(x)$.

Choose $u_n(\tau)$ such that $nP(X_1 > u_n) \underset{n\to\infty}{\longrightarrow} \tau$, then $nP(Z_1 > u_n) \underset{n\to\infty}{\longrightarrow} \tau/2$ and $P\{M_n \leq u_n(\tau)\} = P\{\max\big(Z_1, Z_2, \cdots, Z_n\big) \leq u_n(\tau)\}P\{Z_{n+1} \leq u_n(\tau)\} \underset{n\to\infty}{\longrightarrow} \exp(-\tau/2)$, so $\{X_n\}$ has extremal index $\theta = 1/2$.

Figure 1 shows a realization of model MB1 and a realization of i.i.d. random variables Y_i with the same d.f as the marginal of that model.

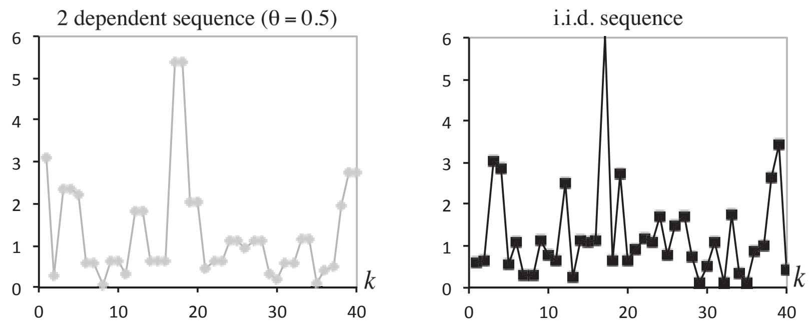

Figure 1: A realization of model MB1 $\{X_n\}$ (left) with d.f. $F(x) = (1-\exp(-x))^2$ and a sequence of i.i.d. random variables $\{Y_n\}$ (rigth) with the same d.f. $F(y) = (1-\exp(-y))^2$.

In Figure 2 illustrates the behaviour of $\widehat{\Theta}^{UC}$ and $\widehat{\Theta}^{JG}$, for 1000 replicas of a sample of size $n = 1000$ from model MB1. All quantities are plotted against k.

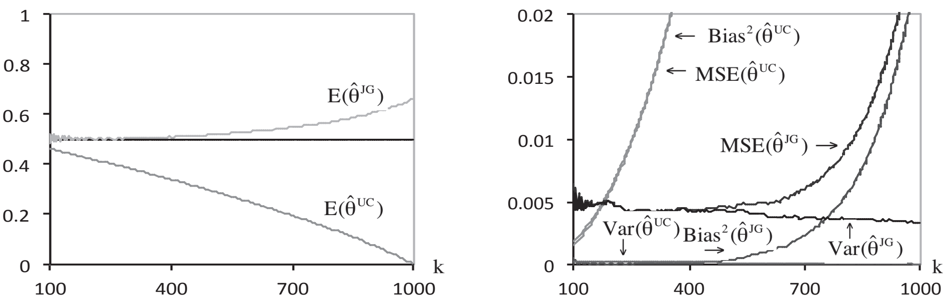

Figure 2: Simulated mean values (left) and MSE, Var and $Bias^2$ (right) of the estimators $\widehat{\Theta}^{UC}$ and $\widehat{\Theta}^{JG}$, for a sample of size $n = 1000$ from model MB1.

3 Resampling-based methods in the estimation of θ

Computer-intensive methods, that emerged when computers became more powerful, have been developed in the last decades. The most well known are perhaps the jackknife (Quenouille, 1949, 1956; Tukey, 1958) and the bootstrap (Efron, 1979; Efron and Tibshirani, 1993) methodologies. Recently, these two methodologies have been used with success in Extreme Value Theory overcoming the difficulties that appear in the semi-parametric estimation of parameters of extreme events.

In their classical form, as first proposed by Efron (1979), bootstrap methods are designed for being used in samples collected under an independent set-up. In context of dependent data, the situation is more complicated since population is not characterized entirely by the one-dimensional marginal distribution F alone, but requires the knowledge of the joint distribution of the whole sequence $X_1, \cdots, X_n$. Singh (1981) presented an example on the inadequacy of the classical bootstrap under dependence. Several attempts have been made to extend the bootstrap method to the dependent case. A breakthrough was achieved when resampling of single observations was replaced by block resampling, an idea that was put forward by Hall (1985), Carlstein (1986), Künsch (1989), Liu and Singh (1992) and others in various forms and in different inference problems. Several ways of blocking appeared. Here we briefly describe the moving block bootstrap method.

3.1 Nonparametric Resampling Schemes for Dependent Data: The Moving Block Bootstrap

Different block bootstrap methods that have been proposed attempt to reproduce different aspects of the dependence structure of the observed data in the resampled data. Künsch (1989), Liu and Singh (1992) independently formulated a substantially new resampling scheme, called the moving block bootstrap (MBB), applicable to dependent data. In contrast to resampling a single observation at a time, the MBB resamples blocks of (consecutive) observations at a time. This method can be described briefly as follows:

Given the sample $\underline{x}_n = \big(x_1, \cdots, x_n\big)$ and with $b \equiv b_n \in [1, n]$ denoting the block length to be resampled from $\underline{x}_n$ and $n_b = n - b + 1$, the MBB method resamples blocks randomly, with replacement, from the overlapping blocks $\big\{\mathcal{B}(i, b) = \big(X_i, \cdots, X_{i+b-1}\big),\ i = 1, \cdots, n_b\big\}$.

From the $I_1, \cdots, I_{\lfloor n/b \rfloor}$ conditionally i.i.d. random variables with discrete uniform distribution on $\big\{1, \cdots, n_b\big\}$, a random sample is drawn with replacement. Arranging the elements in all $\lfloor n/b \rfloor$ blocks in a sequence, we get the bootstrap sample $\underline{x}_n^* = \big(x_1^*, \cdots x_{\lfloor n/b \rfloor \times b}^*\big)$.

The accuracy of block bootstrap estimators, critically depending on the block length, must be supplied by the user.

3.2 Empirical choice of the block length

The orders of magnitude of the optimal block sizes are known in some inference problems (Bühlmann and Künsch, 1999, Hall *et. al.*, 1995, Künsch, 1989, Lahiri, 1999 and Lahiri *et. al.*, 2007). It turns out that optimal block length depends very much on the context. According to those authors three different settings of practical importance can be identified: estimation of the bias or variance, estimation of a one-sided distribution function and estimation of a two-sided distribution function.

The optimal block length in the above situations are of different size, being $b \sim Cn^{1/k}, \quad k = 3, 4$ or 5, respectively where n is the sample size.

This result, of practical and theoretical interest, will be used here as the basis for choosing the "optimal" block length. Two main approaches can be pointed out: a cross validation method proposed by Hall *et. al.* (1995) and a plug-in method based on a recent work of Lahiri *et. al.* (2007).

3.2.1 The nonparametric plug-in method in the estimation of θ

As it was said before, extremal index estimators show usually a high bias that is, in most cases, the main component of the MSE. There is then a need for bias reduction. Based on a recent work of Lahiri *et. al.* (2007), a nonparametric plug-in (denoted here NPPI) method for selecting the "optimal" block length in order to reduce the bias, will be considered. Unlike traditional plug-in rules, this method employs nonparametric resampling procedures to estimate the relevant constants in the leading term of the optimal block length and, hence, does not require the knowledge and/or derivation of explicit analytical expressions for the constants.

Given the sample $\underline{X}_n = (X_1, X_2, \ldots, X_n)$ from an unknown model F, let us consider $\widehat{\Theta}_n$, any estimator of θ and $\widehat{\Theta}^*_n(b)$. the corresponding bootstrap estimator based on blocks of size b.

Let us denote by $\phi_n \equiv Bias\big(\widehat{\Theta}_n\big) = E\big(\widehat{\Theta}_n\big) - \theta$, the bias of $\widehat{\Theta}_n$ and $\widehat{\phi}^*_n(b) \equiv$

$\widehat{Bias}(b) = E_*\big(\widehat{\Theta}^*_n(b)\big) - \widehat{\Theta}_n$ the corresponding block bootstrap estimator, based on blocks of length b, where E_* denotes the conditional expected value given the data.

It is known (Hall *et. al.*, 1995) that the variance of block bootstrap estimator is an increasing function of the block length b while its bias is a decreasing function of b. As a result, for each block bootstrap estimator, there is a critical value, b_n^0, that minimizes MSE. The value of b that minimizes the leading term in the expansion of the MSE is denominated *MSE-optimal block length.* Under suitable regularity conditions, the variance and the bias of a block bootstrap estimator admit expansions of the form, see Lahiri (1999),

$$\begin{aligned} n^{2a}Var\big(\widehat{\phi}^*_n(b)\big) &= C_1 n^{-1} b^r + o\big(n^{-1}b^r\big) \quad \text{as} \quad n\to\infty, & (8)\\ n^{a}Bias\big(\widehat{\phi}^*_n(b)\big) &= C_2 b^{-1} + o\big(b^{-1}\big) \quad \text{as} \quad n\to\infty, & (9)\end{aligned}$$

over a suitable set of possible block lengths $b \in \{2, \cdots, n\}$, where C_1, C_2 are population parameters, $r \geq 1$ is an integer, and $a \in [0, \infty[$ is a known constant. For $\phi_n \equiv Bias$, Hall *et. al.* (1995) consider that (8) and (9) hold with $r = 1$ and $a = 1$. From (8) and (9) an expansion for MSE$(\widehat{\phi}^*_n(b))$ is obtained and leads to the asymptotic *MSE-optimal block length*, $b^0 \equiv b^0_n$:

$$b_n^0 = \left(\frac{2C_2^2}{C_1}\right)^{1/(r+2)} n^{1/(r+2)}(1+o(1)). \tag{10}$$

In (10) C_1 and C_2 need to be estimated. That estimation, under the NPPI method, is done considering the leading part of (8) and (9):

$$C_1 \sim nb^{-r}n^{2a}Var\big(\widehat{\phi}^*_n(b)\big) \quad \text{and} \quad C_2 \sim bn^a Bias\big(\widehat{\phi}^*_n(b)\big).$$

This suggests the use of consistent estimators of $Var\big(\widehat{\phi}^*_n(b)\big)$ and $Bias\big(\widehat{\phi}^*_n(b)\big)$ and define estimators of the parameters C_1 and C_2 as

$$\widehat{C}_1 = nb^{-r}n^{2a}\widehat{Var}_n \quad \text{and} \quad \widehat{C}_2 = bn^a\widehat{Bias}_n. \tag{11}$$

The NPPI estimator $\widehat{b}_n^0$ of the optimal block length b_n^0 is then obtained from (10) and (11) as

$$\widehat{b}_n^0 = \left(\frac{2\widehat{C}_2^2}{\widehat{C}_1} \right)^{1/(r+2)} n^{1/(r+2)}(1 + o(1)). \tag{12}$$

In the next subsections we describe the plug-in method of Lahiri *et. al.* (2007) who used the Jackknife-After-Bootstrap (JAB) method of Efron (1992) and Lahiri (2002) for estimating $Var\big(\widehat{\phi}_n^*(b)\big)$ and constructed an estimator of $Bias\big(\widehat{\phi}_n^*(b)\big)$ by combining two block bootstrap estimators of ϕ_n, (Lahiri et. al., 2007).

3.2.2 The JAB estimator of the Variance

Efron (1992) showed that for an i.i.d. setup the jackknife estimate of standard error of bootstrap values can be computed from the original bootstrap replications, with no further resampling requirement. The procedure is called the JAB method of Efron (1992) and particularly attractive for deriving variance estimators of bootstrap quantities in the i.i.d. case.

A modified version of the method for block bootstrap estimators in the case of dependent data was proposed by Lahiri (2002). The JAB method for dependent data applies a version of the block jackknife method (see Lahiri, 2002) to a block bootstrap estimator and can be described as follows:

Let $\widehat{\phi}_n^*(b)$ be the MBB estimator of the ϕ_n based on (overlapping) blocks of length b from $\underline{X}_n = \big(X_1, \ldots, X_n\big)$, where $\phi_n \equiv Bias\big(\widehat{\Theta}_n\big)$ and $\big\{\mathcal{B}(i, b) = (X_i, \cdots, X_{i+b-1}),\ i = 1, \cdots, n_b\big\}$ (with $n_b = n - b + 1$) denote the collection of all overlapping blocks contained in $\underline{X}_n$.

Let $m \equiv m_n$ be a sequence of integers such that

$$m^{-1} + n^{-1}m = o(1) \ \text{ as } \ n \to \infty,$$

and let $M \equiv n_b - m + 1$. Here, m denotes the number of bootstrap blocks to be deleted. Since there are n_b observed blocks of length b, the first step of JAB method is to define a jackknife version, $\widehat{\phi}_n^{(i)*}(b)$, for $\widehat{\phi}_n^*(b)$, for each $i \in \{1, \cdots, M\}$ by deleting the m blocks $\big\{\mathcal{B}(i,b), \cdots, \mathcal{B}(i+m-1,b)\big\}$.

For $i = 1, \cdots, M$, let $I_i = \{1, \cdots, n_b\} \backslash \{i, \cdots, i+m-1\}$ denote the index set of all blocks of length b obtained by deleting the m blocks. Then, the ith block deleted jackknife point value $\widehat{\phi}_n^{(i)*}(b)$ is obtained by resampling $\lfloor n/b \rfloor$ blocks randomly, with replacement from the reduced collection $\big\{\mathcal{B}(j,b) \,:\, j \in I_i\big\}$.

The JAB estimator of the variance of $\widehat{\phi}_n^*(b)$ is defined as

$$\widehat{VAR}_{JAB}(\widehat{\phi}_n^*(b)) = \frac{m}{(n_b - m)M} \sum_{i=1}^{M} \Big(\widetilde{\phi}_n^{(i)*}(b) - \widehat{\phi}_n^*(b)\Big)^2 \tag{13}$$

where $\widetilde{\phi}_n^{(i)*}(b) = m^{-1}\Big(n_b \widehat{\phi}_n^*(b) - (n_b - m)\widehat{\phi}_n^{(i)*}(b)\Big)$ denotes the ith JAB pseudo-value corresponding to $\widehat{\phi}_n^*(b)$.

3.2.3 The Bias estimator

As a motivation to the definition of the bias estimator, $\widehat{Bias}_n$, we consider the relation (9) that may be rewritten as

$$E\big(\widehat{\phi}_n^*(b)\big) = \phi_n + \frac{C_2}{n^a b} + o\left(n^{-a} b^{-1}\right) \quad \text{as} \quad n \to \infty. \tag{14}$$

If (14) holds for the sequences $\{b_1\} \equiv \{b_{1n}\}_{n \geq 1}$ and $\{2b_1\} \equiv \{2b_{1n}\}_{n \geq 1}$ and considering the corresponding expansion of $E\big(\widehat{\phi}_n^*(b_1)\big)$ and $E\big(\widehat{\phi}_n^*(2b_1)\big)$, we have, as $n \to \infty$

$$\begin{aligned} 2E\big[\widehat{\phi}_n^*(b_1) - \widehat{\phi}_n^*(2b_1)\big] &= 2\bigg[\Big(\phi_n + \frac{C_2}{n^a b_1} + o\left(n^{-a} b_1^{-1}\right)\Big) - \\ &- \Big(\phi_n \frac{C_2}{2n^a b_1} + o\left(n^{-a} b_1^{-1}\right)\Big)\bigg] = \\ &= \frac{C_2}{n^a b_1} = Bias\big(\widehat{\phi}_n^*(b_1)\big) + o\left(n^{-a} b_1^{-1}\right). \end{aligned}$$

This suggests the consistent estimator of $Bias\big(\widehat{\phi}_n^*(b_1)\big)$ as

$$\widehat{Bias}_n \equiv \widehat{Bias}_n(b_1) = 2\big(\widehat{\phi}_n^*(b_1) - \widehat{\phi}_n^*(2b_1)\big). \tag{15}$$

3.2.4 The Optimal block length estimator

The nonparametric plug-in method, described in Section 3.2.1 suggests (12) as an estimator of the optimal block length, where $\widehat{C}_1 = nb^{-r}n^{2a}\widehat{Var}_n$ and $\widehat{C}_2 = bn^a\widehat{Bias}_n$ are estimators of parameters C_1 and C_2, respectively and $\widehat{Var}_n \equiv \widehat{Var}_n(b_1)$ and $\widehat{Bias}_n \equiv \widehat{Bias}_n(b_1)$ are some consistent estimators of the variance and bias of the block bootstrap estimator $\widehat{\phi}_n^*(b)$ based on some suitable initial block length b_1.

Lahiri *et. al.* (2007) suggest using the JAB variance estimator $\widehat{VAR}_{JAB}(\widehat{\phi}_n^*(b_1))$ in (13) for estimating C_1 and using the bias estimator $\widehat{Bias}_n$ of (15) to estimate C_2 and then obtain the plug-in estimator of the optimal block length b_n^0.

Although the nonparametric plug-in method produces a consistent estimator of the optimal block length, finite sample performance of the estimator depends on the choice of the smoothing parameter b_1 and on the JAB blocking parameter m. Lahiri *et. al.* (2007) show that the optimal choice of b_1 is of the form, $b_1 = C_3 n^{1/(r+4)}$, where $r \geq 1$ is an integer and C_3 is a population parameter. As for the other smoothing parameter, an heuristic argument in Lahiri (2002) suggests that a reasonable choice of the JAB parameter m is given by $m = C_4 n^{1/3} b_1^{2/3}$, for some constant C_4. Numerical results of Lahiri *et. al.* (2007) show that the choice $C_3 = 1$ for the initial block length b_1 gives the best result for different functionals of interest, while the value of C_4 for calculating m is $C_4 = 1$ for the bias and variance functionals.

4 Monte-Carlo simulations

A Monte-Carlo simulation of the mean value (E), the mean squared error (MSE), the variance (VAR) and the squared bias (BIAS^2) of several estimators was performed. Here we only present results for the *Up-crossing estimator*, $\widehat{\Theta}^{UC}$, and the *Generalized-Jackknife estimator*, $\widehat{\Theta}^{JG}$, for two of the models (MA1 and MB1, see description below) and a few values of θ. The simulations were also done for several sample sizes and 1000 replicas, but we only present plots for $n = 1000$.

The following stationary sequences were considered in our study, however only results from two models will be showed in this paper - model MA1 and the two-dependent model, model MB1.

A. Max-autoregressive processes

We consider a first-order max-autoregressive processes (ARMAX processes), ARMAX(1), which are a special case of the general ARMAX-MAX(p,q) processes introduced by Davis and Resnick (1989). ARMAX processes became quite popular in extreme value theory, specially because their nice treatment in what concerns extremal behavior. In literature ARMAX processes appear formulated in different ways.

Model MA1 – $X_0 = Z_0, \quad X_i = \max\big\{\beta X_{i-1}, (1-\beta)Z_i\big\}, \quad i = 1, \cdots, n,$ where $0 \leq \beta < 1$ and $\{Z_i\}$ is an i.i.d. sequence of random variables, which we assume having d.f. $F(z) = \exp(-z^{-1})$, $z > 0$. So the marginal distribution of the process $\{X_i\}$ is unit Fréchet and for $0 < x < \infty$ and $u_n = nx$, $P\big\{M_n \leq u_n\big\} \underset{n\to\infty}{\longrightarrow} \exp\big(-(1-\beta)/x\big)$, for example, see Beirlant et. al. (2004). The extremal index is then $\theta = 1 - \beta$.

Model MA2 – $X_0 = Z_0, \quad X_i = \max\big\{\alpha X_{i-1}, Z_i\big\}, \quad i = 1, \cdots, n,$ where $0 \leq \alpha < 1$ and $\{Z_i\}$ is an i.i.d. sequence of random variables with d.f. $F(z) = \exp\big(-(1-\alpha)z^{-1}\big)$ for $z > 0$. The marginal distribution of the process $\{X_i\}$ is unit Fréchet. The extremal index of this process is easily shown to be $\theta = 1 - \alpha$, for example, see Leadbetter (1983).

Model MA3 – $X_0 = Z_0, \quad X_i = \beta \max\big\{X_{i-1}, Z_i\big\}, \quad i = 1, \cdots, n,$ where $0 < \beta < 1$ and $\{Z_i\}$ is an i.i.d. sequence of random variables with d.f. $F(z) = \exp\big(-z^{-\alpha}\big)$ and X_0 a random variable with dd.f. $H_0(x) =$

$\exp\big(-x^{-\alpha}(\beta^{-\alpha}-1)\big)$. The extremal index of this process is $\theta = 1-\beta^{\alpha}$, see, for example, see Alpuim (1989) and Canto e Castro (1992).

B. Moving-maximum processes

These models, developed by Newell (1964) and Deheuvels (1983), are defined by

$$X_i = \max\big\{Z_i, \psi_1 Z_{i-1}, \cdots, \psi_q Z_{i-q}\big\}, \quad i \in \mathbb{Z},$$

where $\psi_j \geq 0,\ (1 \leq j \leq q)$ and $\{Z_i\}$ is an i.i.d. sequence of random variables with d.f. F. Details on the behaviour of these extremal models for various cases can be found in Weissman and Cohen (1995).

Model MB1 – see Section 2.2.

Model MB2 – $X_0 = Z_0, \quad X_i = (a+1)^{-1} \max\big\{aZ_{i-1}, Z_i\big\}, \quad i = 1, \cdots, n,$ where $a \geq 0$ and $\{Z_i\}$ is an i.i.d. sequence of random variables with d.f. $F(z) = \exp\big(-z^{-1}\big)$, $z > 0$. The extremal index of this process is $\theta = \max\{1, a\}/(a+1)$ and lies in interval $[1/2, 1]$, see Davison (2011).

Model MB3 – $X_0 = Z_0, \quad X_i = \max\big\{aZ_{i-1}, Z_i\big\}, \quad i = 1, \cdots, n,$ where $0 \leq a \leq 1$ and $\{Z_i\}$ is an i.i.d. sequence of random variables with d.f. $F(z) = \exp\big(-1/(a+1)z\big)$, $z > 0$. The marginal distribution of $\{X_i\}$ is standard Fréchet. It is easy to check that the sequence is stationary and the extremal index is $\theta = 1/(a+1)$, see Coles (2001).

4.1 The simulation study

Here we will use simulated samples, generated from our models and for some values of parameter θ. If we know the true value of θ, we can illustrate the behavior of our estimators and we can control the performance of the resampling techniques explained in Section 3. All the aforementioned models were included in an extensive simulation study for checking the properties of the estimators as well as for controlling the application of resampling techniques. Figure 3 shows the mean values of $\widehat{\Theta}^{UC}$ and $\widehat{\Theta}^{GJ}$, for model MA1, with three values of θ and a sample of size $n = 1000$. For each simulated case, MSE, Var and Bias^2 are plotted in Figure 4.

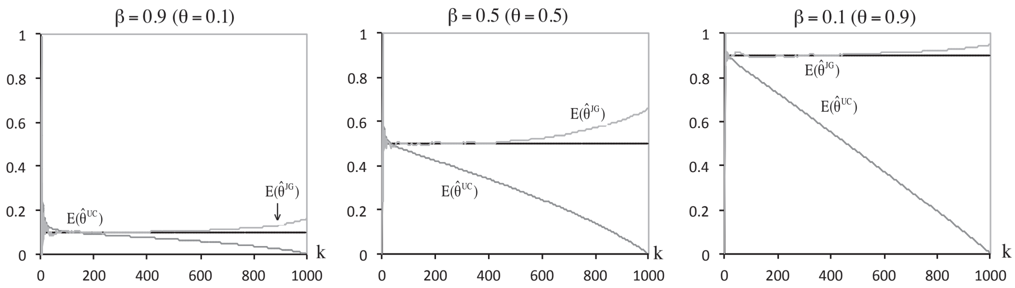

Figure 3: Simulated mean values of the estimators $\widehat{\Theta}^{UC}$ and $\widehat{\Theta}^{JG}$, for a sample of size $n = 1000$ from model MA1.

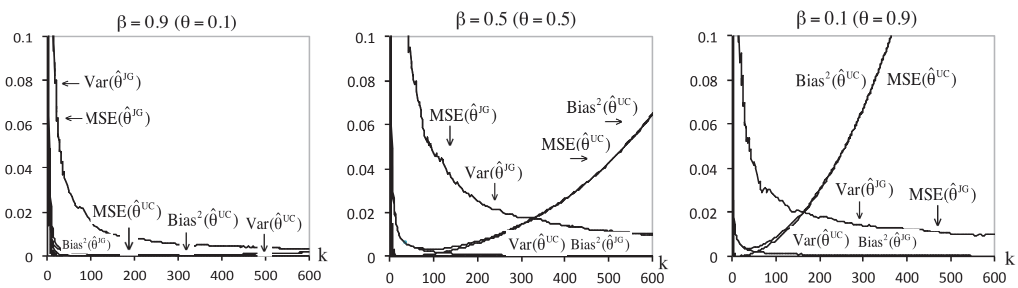

Figure 4: Simulated MSE, Var and $Bias^2$ of the estimators $\widehat{\Theta}^{UC}$ and $\widehat{\Theta}^{JG}$, for a sample of size $n = 1000$ from model MA1.

For one of the samples generated, a simple path of the estimators $\widehat{\Theta}^{UC}$ and $\widehat{\Theta}^{JG}$ is calculated, see Figure 5 (top). Those samples are then used to obtain the *optimal block length* for the moving block bootstrap. For the sample coming from model MA1, we saw that if $\theta = 0.9$ (near independence) the optimal block size was 2 ou 3 and for $\theta = 0.1$ we obtained $\widehat{b}_n^0 = 10$.

Block bootstrap estimates are plotted in Figure 5. A more stable path was obtained for the $\widehat{\Theta}^{JG}$ estimator, although our results claim for a more complete simulation study for the optimal block size, possibly jointly with alternative adaptive estimation methods.

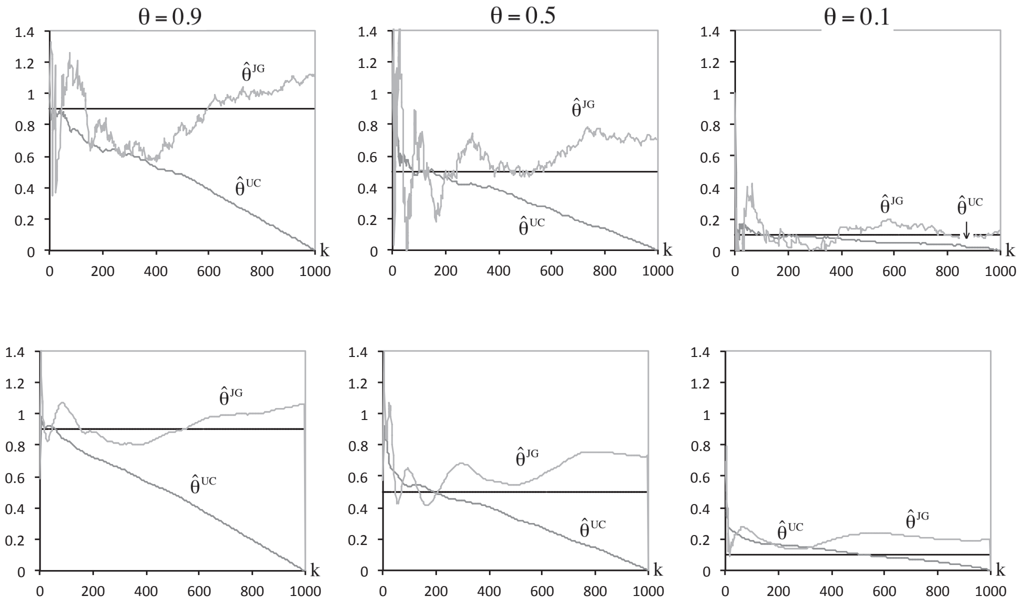

Figure 5: One sample path of the estimators $\widehat{\Theta}^{UC}$ and $\widehat{\Theta}^{JG}$, for a sample of size $n = 1000$ from model MA1 (top); block bootstrap estimates using $\widehat{b}_n^0 = 2$ for $\theta = 0.9$ and $\widehat{b}_n^0 = 10$ for $\theta = 0.1$ and $\theta = 0.5$ (bottom).

5 Concluding remarks

In this paper, a general method for estimating the optimal block size for bootstrap estimation in situation of dependence was presented. It was applied to two estimators of the extremal index. Generalized Jackknife estimator presented promising results, showing a more stable path. However more estimators should be compared and procedures for an adaptive choice of the high level need to be considered.

References

[1] Ancona-Navarrete, M. and Tawn, J. (2001). Comparison of methods for estimating the extremal index. *Extremes* 3:5–38.

[2] Alpuim, M. (1989). An extremal markovian sequence. *J. Appl. Prob.* 26:219–222.

[3] Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J., Waal, D. and Ferro, C. (2004). *Statistics of Extremes: Theory and Applications*. John Wiley & Sons.

[4] Bühlmann, P. and Künsch, H. (1999). Block length selection in the bootstrap for time series. *Comput. Statist. and Data Ana.* 31:295–310.

[5] Canto e Castro, L. (1992). Sobre a teoria assinttica de extremos. *Tese de Doutoramento.* DEIO. Faculdade de Cincias da Universidade de Lisboa.

[6] Carlstein, E. (1986). The use of subseries methods for estimating the variance of a general statistics from a stationary times series. *The Annals of Statist.* 14:1171–1179.

[7] Coles, S. (2001). *An introduction to statistical modeling of extreme values.* Springer-verlag.

[8] Davis, R. and Resnick, S. (1989). Basic properties and prediction of Max-ARMA processes. *Adv. Appl. Prob.* 21:781–803.

[9] Deheuvels, P. (1983). Point processes and multivariate extreme values. *J. multivariate analysis* 13:257–272.

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

[11] Efron, B. (1992). Jackknife-after-bootstrap standard errors and influence functions (with discussions). *Journal of Royal Statistical Series B* 54:83–111.

[12] Efron, B. and Tibshirani, R. (1993). *An Introduction to the Bootstrap.* New York: Chapman and Hall.

[13] Ferro, C. and Segers, J. (2003). Inference for clusters of extreme values. *Journal of Royal Statistical Series B* 65:545-556.

[14] Gomes, M.I. (1990). Statistical inference in an extremal markovian model. *COMPSTAT* 257–262.

[15] Gomes, M.I. (1992). Modelos extremais em esquemas de dependência. *I Congresso Ibero-Americano de Esdadistica e Investigacion Operativa* 209–220.

[16] Gomes, M.I. (1993). On the estimation of parameters of rare events in environmental time series. Em *Statistics for environment* (Vic. Barnett and K. Feridun Turkman) 225–241. John Wiley and Sons Ltd.

[17] Gomes, M.I., Hall, A. and Miranda, C. (2008). Subsampling techniques and the Jackknife methodology in the estimation of the extremal index . *J. Comput. Statist. and Data Analysis* 52:2022–2041.

[18] Gray, H.L., and Schucany, W.R. (1972). *The Generalized Jackknife Statistic.* Marcel Dekker.

[19] Hall, P. (1985). Resampling a coverage pattern. *Stoc. Proc. and Their App.* 20:231–246.

[20] Hall, P., Horowitz, J. L. and Jing, B.-Y. (1995). On blocking rules for the bootstrap with dependent data. *Biometrika* 50:561–574.

[21] Hsing, T. (1991). Estimating the parameters of rare events. *Stoc. Proc. and Their App.* 37:117–139.

[22] Hsing, T. (1993). Extremal index estimation for a weakly dependent stationary sequence. *Ann. Statist.* 21:2043–2071.

[23] Künsch, H. (1989). The jackknife and the bootstrap for general stationary observations. *Ann. Math.* 17:1217–1241.

[24] Lahiri, S. (1999). Theoretical comparisons of block bootstrap methods. *Annals of Statistics*27:386–404.

[25] Lahiri, S. (2002). On the jackknife after bootstrap method for dependent data and its consistency properties. *Econometric Theory*,18:79–98.

[26] Lahiri, S., Furukawa, K. and Lee, Y-D. (2007). Nonparametric plug-in method for selecting the optimal block lengths. *Statistical methodology* 4:292–321.

[27] Laurini, F. and Tawn, J. (2007). New estimators for the extremal index and other cluster characteristics. *Extremes* 28:289–303.

[28] Leadbetter, M. R. (1974). On extreme values in stationary sequences. *Z. Wahrsch. Verw. Gebiete* 65: 291–306.

[29] Leadbetter, M. R. (1983). Extremes and local dependence in a stationary sequence. *Z. Wahrsch. Verw. Gebiete* 65: 291–306.

[30] Leadbetter, M., Lindgren, G. e Rootzén, H. (1983). *Extremes and related properties of random sequences and series.* Springer-Verlag, New York.

[31] Leadbetter, M. R. and Nandagopalan, L. (1989). On exceedance point process for stationary sequences under mild oscillation restrictions. *In Extreme Value*

Theory: Proceedings, Oberwolfach 1987, ed J. Hüsler and R. D. Reiss, Lecture Notes in Statistics 52: 69-80. Springer-Verlag, Berlim.

[32] Leadbetter, M. R. and Rootzén, H. (1988). Extremal theory for stochastic processes. *The Annals of Prob.* 16:431–478.

[33] Liu, R. and Singh, K. (1992). Moving blocks jackknife and bootstrap capture weak dependence. In *Exploring the Limits of Bootstrap* (R. Lepage and L. Billard, Eds.), Wiley, New York, 225-248.

[34] Nandagopalan, S. (1990). *Multivariate Extremes and Estimation of the Extremal Index*. PhD Thesis, University of North Carolina, Chapel Hill.

[35] Newell, G. (1964). Asymptotic extremes for m-dependent random variables. *Ann. Math. Statist.* 35:1322–1325.

[36] O'Brien, G. (1987). Extreme values for stationary and markov sequences. *Ann. Probab.*15:281–289.

[37] Quenouille, M.H. (1949). Approximate tests of correlation in time series. *J. Royal Statist. Soc. B* 11:68–84.

[38] Quenouille, M.H. (1956). Notes on bias in estimation. *Biometrika* **43**, 353–360.

[39] Singh, K. (1981). On the asymptotic accuracy of the Efron's bootstrap. *Ann. of Stat.* 9: 345–362.

[40] Smith, R. and Weissman, I. (1994). Estimating the extremal index. *J. R. Statist. Soc. B* 56:515–528.

[41] Süveges, M. and Davison, A. (2010). Model misspecification in peaks over threshold analysis. *The Ann. of App. Stat.* 4:203–221.

[42] Tukey, J.W. (1958). Bias and confidence in not quite large samples (Abstract). *Ann. Math. Statist.* 29:614.

[43] Weissman, I. and Cohen, U. (1995). The extremal index and clustering of high values for derived stationary sequences. *J. Appl. Probab.* 32:972–981.

[44] Weissman, I. and Novak, S. (1995). On blocks and runs estimators of the extremal index. *J. Statist. Plann. Inf.* 66:281–288.