Inference for the Sample Maximum in the Presence of Serial Correlation and Heavy Tailed Distributions

Tucker McElroy
University of California, San Diego
email: tmcelroy@math.ucsd.edu

Dimitris N. Politis
University of California, San Diego
email: politis@math.ucd.edu

Abstract

We consider data from an infinite order moving average time series model with inputs in the α-stable domain of attraction, for $\alpha \in (0, 2)$. The sample maximum of the data is of interest in settings such as insurance and finance; we produce a normalization for this statistic, which, in conjunction with subsampling methods, will allow for asymptotically correct estimation of its cumulative distribution function. A concrete application to the concept of "safety-first" portfolio selection is given.

1 Introduction

One topic of classical extreme value theory is the exploration of the asymptotic properties of the sample maximum ($\max_{1 \leq t \leq n} X_t$). The cumulative distribution function (cdf) of the sample maximum is a quantity of interest in the financial and insurance communities, since it has applications to risk analysis.

This paper examines the limit behavior of the sample maximum for linear heavy-tailed time series. Heavy-tailed random variables have become increasingly popular as models for financial and insurance data, and the literature has expanded in the past few decades to accommodate the growing interest: see Davis and Resnick (1985, 1986), Davis and Hsing (1995), Embrechts et al. (1997), and Resnick (1997). This movement was induced by the discovery of high kurtosis in financial and insurance data (as well as in teletraffic data and many data coming from natural phenomena) which was not adequately explained by light-tailed distributions such as the normal. Linear time series have continued to be a popular model due to both the ease of analysis as well as the extent of applicability; thus our model is relevant and important.

We consider an observed stretch $\{X_1, X_2, \cdots, X_n\}$ of a strictly stationary time series which can be written as an infinite order moving average of independent identically distributed (*iid*) inputs Z_t (denote a common version by Z), as follows:

$$X_t = \sum_{j \in \mathbb{Z}} \psi_j Z_{t-j},$$

i.e. $\{X_t\}$ is a linear time series. The random variables Z_t are assumed to belong to the domain of attraction of an α-stable random variable, where α is an *unknown* parameter (strictly) between zero and two. The fact that the exact value of α is not known to the modeler is important for the applications that we have in mind.

The above time series has been studied in two papers by Davis and Resnick (1985, 1986), wherein they establish important results about the weak convergence of the sample mean, autocovariances and autocorrelations. One drawback to their theorems is that explicit knowledge of the parameter α is required to make use of these statistics (α appears in the normalizing rate of convergence); worse, the limiting distribution is also dependent on α, and is quite complicated. This seriously hinders the practical construction of confidence intervals: one must first estimate α (perhaps via the well-known Hill's estimator), and then try to calculate the quantiles of the complicated limit law. A solution to this difficulty in the case of the sample mean was presented by McElroy and Politis (2001); self-normalization together with subsampling removed the enigmatic α from the convergence rate of the statistic, and allowed estimation of the limit cdf. Self-normalization was introduced effectively for heavy-tailed data by Logan et al (1973), while subsampling was introduced in Politis and Romano (1994), and is extensively discussed in the book by Politis, Romano, and Wolf (1999).

It turns out that the sample maximum of the time series converges to a nondegenerate distribution at a rate which depends on α. To deal with this difficulty we normalize by the sample standard deviation, and apply subsampling to obtain an approximation to the limit cdf. This is the main result of the paper; the proof is a straight-forward application of the Point process methods – see Resnick (1986, 1987). The following example illustrates the utility of such a result:

Example 1: Insurance Consider a positive claim amount X_t to be made at time $t \in \mathbb{N}$ (cf. Chapter 1 of Embrechts et al (1997), where the claim times come according to a Poisson Process). If an insurance company has an amount of wealth x, they may be interested in knowing the probability that the largest claim amount (i.e. the sample maximum) exceeds x over a time horizon of length n. Assuming that we are at time 1, we wish to estimate

$$\mathbb{P}\left[\max_{1\leq t\leq n} X_t > x\right];$$

seen another way, the insurance company may wish to find x such that the above quantity is smaller than a specified probability p, which amounts to calculating the pth quantile of the maximum's cumulative distribution function.

Example 2: Finance This example comes from the literature of "safety first" portfolio selection. Here we interpret X_t as the value of some asset at time t; we wish to know the probability that the sample maximum stays above some minimum quantity x. In this case we must estimate the same probability as above, and calculate the pth quantile. This approach is called a "worst case scenario" analysis.

Section two of this paper contains background material on heavy-tailed linear time series, and the main result mentioned. The third section discusses the applications to insurance and finance, and how subsampling can be used to estimate the desired quantiles.

2 Theory

2.1 Background

Consider a moving average process of infinite order ($MA(\infty)$ for short) which has *iid* (independent and identically distributed) inputs:

$$X_t := \sum_{j\in\mathbb{Z}} \psi_j Z_{t-j} \quad \forall t \in \mathbb{N}.$$

Throughout this paper we will use the term "linear" to denote this infinite order moving average. We require that the filter coefficients $\{\psi_j\} \in l_p$ for some $p \in (0,1] \cap (0,\alpha)$ – see Chap. 13 of Brockwell and Davis (1991) – in order to ensure that the sum converges almost surely; the random variables $\{Z_t\}$ (for any $t \in \mathbb{Z}$) satisfy the following two properties for some $\alpha \in (0,2)$:

$$\mathbb{P}[|Z_t| > x] = x^{-\alpha} L(x) \tag{1}$$

$$\frac{\mathbb{P}[Z_t > x]}{\mathbb{P}[|Z_t| > x]} \to p, \qquad \frac{\mathbb{P}[Z_t \leqslant -x]}{\mathbb{P}[|Z_t| > x]} \to q \tag{2}$$

as $x \to \infty$. Here p and q are arbitrary fixed numbers between 0 and 1 and add up to 1. $L(x)$ is a "slowly varying" function, i.e. $L(ux)/L(x) \to 1$ as $x \to \infty$ for any fixed u. An example of a slowly varying function is the logarithmic function. We will denote a common version of the Z_t's by Z. Note that it easily follows that the right and left tails of Z behave like

$$\mathbb{P}[Z > x] \sim px^{-\alpha} L(x), \qquad \mathbb{P}[Z \leqslant -x] \sim qx^{-\alpha} L(x)$$

where "$\sim$" denotes that the ratio of the left and right sides tends to unity as $x \to \infty$.

A random variable that satisfies conditions (1) and (2) is said to have "heavy tails." Such random variables are in the domain of attraction of an α-stable law, i.e. if we take an *iid* sequence of such Z_t's, then there exist real-valued sequences $a_n > 0$ and b_n such that

$$a_n^{-1} \left(\sum_{t=1}^{n} Z_t - b_n \right) \stackrel{\mathcal{L}}{\Longrightarrow} S,$$

where S is an α-stable random variable. The symbol $\stackrel{\mathcal{L}}{\Longrightarrow}$ denotes convergence in distribution as $n \to \infty$. We then write $Z \in DOM(\alpha)$, which means that Z is in the α-stable domain of attraction. It is common to see condition (2) written as $Z \sim F$

(which means, in this context, that Z has cumulative distribution function $\mathrm{F}(\cdot)$) and as $x \to \infty$,

$$x^{\alpha}(1 - F(x)) \to p, \qquad x^{\alpha} F(-x) \to q.$$

An example of the above is given by Z in the "normal" domain of attraction of an α-stable law, which means that we can take $a_n = n^{\frac{1}{\alpha}}$. If Z is itself α-stable, then (1) and (2) are certainly satisfied. If in addition Z is symmetric (written Z is sαs), then X_t has the law of a sαs as well, but scaled by $(\sum_j |\psi_j|^{\alpha})^{\frac{1}{\alpha}}$. (This quantity is finite, since $\{\psi_j\} \in l_p$ for $p \in (0, \alpha)$.) Note that no generality is lost if one places a constant $C > 0$ on the right hand side of equations (1) and (2); this constant is called the "dispersion" of Z (written $disp(Z)$). If Z has the $S_{\alpha}(\sigma, \beta, \mu)$ law, i.e. Z is α-stable with scale σ, skewness β, and location μ (see Samorodnitsky and Taqqu (1994)), then we can take $p = \frac{\beta+1}{2}$ and $q = \frac{\beta-1}{2}$, and the dispersion is

$$C = \begin{cases} \sigma/\Gamma(1-\alpha)\cos\frac{\pi\alpha}{2} & \alpha \neq 1 \\ 2\sigma/\pi & \alpha = 1 \end{cases}$$

There are a few facts about the choice of a_n : firstly, the sequence should satisfy

$$n\mathbb{P}[|Z| > a_n x] \to x^{-\alpha} \tag{3}$$

as $n \to \infty$ for every positive x. (In particular, if we take a_n that satisfies this, then we can prove the limit result for the domain of attraction.) It is easy to check that $a_n := \inf\{x : \mathbb{P}[|Z| > x] \leqslant n^{-1}\}$ satisfies this condition. Condition (3) will be very important in what follows. A change of variable argument, using (1), implies that $a_n = n^{\frac{1}{\alpha}} L(n)$, where L is slowly varying, but not neccessarily the same slowly varying function in (1); thus the "normal" domain of attraction has $L(n) \equiv 1$. Given this, a suitable choice for b_n is $\mathbb{E}[Z; |Z| \leqslant a_n]$. This definition is interesting – it suggests a "natural" truncation for Z.

The following notation will be used: Ψ will denote the whole sequence of $\{\psi_j\}$, and Ψ_p will denote its l_p norm, i.e.

$$\Psi_p = (\sum_{j \in \mathbb{Z}} |\psi_j|^p)^{\frac{1}{p}}.$$

Notice that since $\{\psi_j\} \in l_p$, they are also in l_α since $p < \alpha$, so $(\sum_j |\psi_j|^{\alpha})^{\frac{1}{\alpha}} < \infty$. It is true that $\{X_t\}$ forms a strictly stationary sequence, since applying a shift operator to the law for the Z-series does not affect the distribution.

2.2 Point Process Techniques

We now give a brief introduction to point process techniques – see Resnick (1987) for additional details.

Let E be the space $(0,\infty) \times \mathbb{R} \setminus \{0\}$ with the σ-field $\mathcal{E}$ of open sets, and define

$$\varepsilon_x(F) := 1_F(x) = \begin{cases} 1 & x \in F \\ 0 & x \notin F \end{cases}$$

for $x \in E$ and $F \in \mathcal{E}$.

A **point measure** m has the form $\sum_{i \in I} \varepsilon_{x_i}$, and is defined on all relatively compact subsets of E. The class of such measures is denoted by $M_p(E)$, and $\mathcal{M}_p(E)$ is the smallest σ-field such that the evaluation mapping for $F \in \mathcal{E}$

$$m \mapsto m(F),$$

is measurable for each $m \in M_p(E)$.

A **Point Process** on E is a measurable map from a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ into $(M_p(E), \mathcal{M}_p(E))$. Thus a point process is a random point measure.

Let $C_K^+(E)$ be the set of continuous functions from E to $\mathbb{R}^+$ that have compact support. A useful topology for $M_p(E)$ is the "vague" topology, under which $M_p(E)$ is a complete separable metric space. If $\mu_n \in M_p(E)$, then $\mu_n \longrightarrow \mu_0$ "vaguely" iff $\mu_n(f) \longrightarrow \mu_0(f) \quad \forall f \in C_K^+(E)$, in which case we write $\mu_n \xrightarrow{v} \mu_0$.

A Poisson process on $(E, \mathcal{E})$ with mean measure μ is a point process ξ satisfying for all $F \in \mathcal{E}$:

$$\mathbb{P}\left[\xi(F) = k\right] = \begin{cases} exp\{-\mu(A)\} \frac{(\mu(A))^k}{k!} & if\ \mu(A) < \infty \\ 0 \quad if\ \mu(A) = \infty \end{cases}$$

which has the independent scattering property, i.e. if $F_1, \cdots, F_n \in \mathcal{E}$ are all disjoint, then $\xi(F_1), \cdots, \xi(F_n)$ are independent random variables. We say that ξ is a Poisson Random Measure with mean measure μ, which is abbreviated $PRM(\mu)$.

Now suppose we have an *iid* sequence $\{Z_t\}$ which satisfies the tail behavior described by conditions (1), (2), (3) . By defining the Levy measure ν on $\mathbb{R} \setminus \{0\}$ as follows

$$\nu(dx) := \alpha p x^{-\alpha-1} 1_{(0,\infty)}(x)dx + \alpha q(-x)^{-\alpha-1} 1_{(-\infty,0)}(x)dx,$$

we can formulate condition (3) of the previous section as

$$n\mathbb{P}\left[a_n^{-1} Z \in \cdot\right] = nF(a_n \cdot) \xrightarrow{v} \nu,$$

or in other words (for any real number $x > 0$)

$$n\mathbb{P}\left[a_n^{-1} Z \leqslant x\right] = nF(a_n x) \to \nu((-\infty, x]).$$

We define the measure $\mu(dt, dx) := dt \times \nu(dx)$ on $(0,\infty) \times \mathbb{R} \setminus \{0\}$. Then the following convergence from Davis and Resnick (1985) holds true:

$$\sum_{k=1}^{\infty} \varepsilon_{\left(\frac{k}{n}, \frac{Z_k}{a_n}\right)} \overset{\mathcal{L}}{\Longrightarrow} \sum_{k=1}^{\infty} \varepsilon_{(t_k, j_k)}$$

on $M_p\left((0,\infty) \times \mathbb{R} \setminus \{0\}\right)$, where the limit is a $PRM(\mu)$. We use here the convention that a point that falls outside the state space does not contribute to the sum.

Now we cite a result from Davis and Resnick (1985) for linear time series $\{X_t\}$ which satisfy our assumptions (1) , (2) , and (3) :

$$\sum_{k=1}^{\infty} \varepsilon_{\left(\frac{k}{n}, \frac{X_k}{a_n}\right)} \overset{\mathcal{L}}{\Longrightarrow} \sum_{k=1}^{\infty} \sum_{i \in \mathbb{Z}} \varepsilon_{(t_k, \psi_i j_k)}, \tag{4}$$

where $\{(t_k, j_k)\}$ are, as above, the points of a $PRM(\mu)$ on the space $E = (0,\infty) \times (\mathbb{R} \setminus \{0\})$.

2.3 Main Results

Define the sample maximum and sample standard deviation by

$$M_n := \max_{1 \leqslant t \leqslant n} X_t$$

and

$$\hat{\sigma}_n := \sqrt{\sum_{t=1}^{n} X_t^2}$$

respectively. In addition, define the following features of the filter (where $a \vee b$ denotes the maximum of a and b):

$$\psi_+ := \max_{j \in \mathbb{Z}} (\psi_j \vee 0) \quad \psi_- := \max_{j \in \mathbb{Z}} (-\psi_j \vee 0).$$

Our main result below is a limit theorem for the self-normalized maximum.

Theorem 1 *Let $\alpha \in (0,2)$, and let $\{X_t\}$ be a linear time series which satisfies conditions (1), (2), and (3). Also require that either $\psi_+^\alpha p > 0$ or $\psi_-^\alpha q > 0$. Then the following weak convergence holds:*

$$\frac{M_n}{\hat{\sigma}_n} \stackrel{\mathcal{L}}{\Longrightarrow} \frac{Y}{\Psi_2 \sqrt{S_2}},$$

where S_2 is an $\frac{\alpha}{2}$-stable totally right skewed random variable, and Y has the following distribution:

$$\mathbb{P}[Y \leqslant x] = \begin{cases} \exp\{-x^{-\alpha}(\psi_+^\alpha p + \psi_-^\alpha q)\} & x > 0 \\ 0 \quad x \leqslant 0 \end{cases} \tag{5}$$

Proof We begin with the Poisson process convergence result (4) of the previous section. We define the mappings T_1 and T_2^δ as follows:

$$T_1\left(\sum_{k=1}^{\infty} \varepsilon_{(u_k,v_k)}\right) := \bigvee_{u_k \leqslant 1} v_k$$

and

$$T_2^\delta\left(\sum_{k=1}^{\infty} \varepsilon_{(u_k,v_k)}\right) := \sum_{u_k \leqslant 1} v_k^2 1_{[v_k^2 > \delta]}$$

for each positive δ. Notice that $\bigvee$ denotes the maximum of that collection of random variables. These maps are continuous with respect to the limit point process – viz. Resnick (1986, 1987). It follows that the joint mapping (T_1, T_2^δ) is continuous, and by applying this to the convergence (5) above, we obtain the following joint convergence:

$$\left(\bigvee_{k=1}^{n} a_n^{-1} X_k, \sum_{k=1}^{n} a_n^{-2} X_k^2 1_{[X_k^2 > a_n^2 \delta]}\right) \stackrel{\mathcal{L}}{\Longrightarrow} \left(\bigvee_{t_k \leqslant 1}\left(\bigvee_{i \in \mathbb{Z}} \psi_i j_k\right), \sum_{t_k \leqslant 1} \sum_{i \in \mathbb{Z}} \psi_i^2 j_k^2 1_{[|\psi_i j_k|^2 > \delta]}\right)$$

$$= \left(\bigvee_{t_k \leqslant 1} (\psi_+ j_k \vee (-\psi_-) j_k), \sum_{i \in \mathbb{Z}} \psi_i^2 \sum_{t_k \leqslant 1} j_k^2 1_{[\psi_i^2 j_k^2 > \delta]}\right)$$

Next we wish to let δ tend to zero on both sides of this convergence. The validity of this follows the exact same argument as in Theorem 4.2 of Davis and Resnick (1985), where they take the limit as δ tends to zero of the same map T_2^δ. The first component of the weak convergence does not depend on δ, so we obtain the joint convergence

$$\left(a_n^{-1} M_n, a_n^{-2} \sum_{k=1}^{n} X_k^2\right) \stackrel{\mathcal{L}}{\Longrightarrow} \left(\bigvee_{t_k \leqslant 1} (\psi_+ j_k \vee (-\psi_-) j_k), \Psi_2^2 \sum_{t_k \leqslant 1} j_k^2\right).$$

Let the first random variable on the right hand be denoted by Y; it follows that Y has the stated cdf – see Theorem 3.1 of Davis and Resnick (1985). As for $\sum_{t_k \leqslant 1} j_k^2$, we know that the j_k are the points of a $PRM(\nu)$ on the space $\mathbb{R}\backslash\{0\}$ (via projecting the $PRM(\mu)$ onto the second coordinate). As discussed in Resnick (1987), the mapping

$$\sum_k \varepsilon_{j_k} \mapsto \sum_k \varepsilon_{j_k^2}$$

induces a corresponding transformation on the mean measures:

$$\nu(dx) \mapsto \frac{\alpha}{2} x^{-\frac{\alpha}{2}} 1_{(0,\infty)}(x)\, dx =: \nu_2(dx).$$

Thus we see that $\sum_{t_k \leqslant 1} j_k^2$ is a $PRM(\nu_2)$, and by the Ito representation is an $\frac{\alpha}{2}$ stable totally right skewed random variable, which we will denote by S_2. Finally, it is easy to apply the continuous function $f(x, y) := \frac{x}{\sqrt{y}}$ to obtain the weak convergence result stated in the theorem. †

3 Application: Insurance

Now we interpret the random variables X_t as a claim amount, and we are interested in the maximum of the claim amounts over a certain time horizon. Some minor modifications will produce the finance example, but we shall focus on **Example 1**. Since the claim amount is positive, the series $\{X_t\}$ must be non-negative for every t (almost surely); to model this, we must impose that all the filter coefficients are non-negative, and that the inputs $\{Z_t\}$ are totally right skewed. This latter situation can be characterized by decreeing that $p = 1$ and $q = 0$ in lines (1) and (2), and thus the limit Y in Theorem 1 has cdf

$$\mathbb{P}[Y \leqslant x] = \begin{cases} \exp\{-x^{-\alpha}(\psi_+^\alpha)\} & x > 0 \\ 0 \quad x \leqslant 0 \end{cases}$$

We are therefore interested in the cumulative distribution function (cdf) of M_n, which we will denote by $F_M(x) = \mathbb{P}[M_n \leqslant x]$; knowing F_M, we can calculate all other interesting parameters/quantities of M_n. F_M also has the following interpretation: for a given threshhold x, the number $F_M(x)$ represents the probability that no claim amount exceeds that threshhold.

The theory from the previous section allows us to estimate a quantity related to F_M. For each time t between 1 and n, define

$$Y_t := \frac{X_t}{\hat{\sigma}_n},$$

a "normalized" version of the claim amount. Due to the positivity of the sample variance, we have

$$\frac{M_n}{\hat{\sigma}_n} = \max_{1 \leq t \leq n} Y_t =: \tilde{M}_n$$

from which it follows (by Theorem 1) that

$$F_{\tilde{M}}(x) := \mathbb{P}\left[\tilde{M}_n \leqslant x\right] = \mathbb{P}\left[\frac{M_n}{\hat{\sigma}_n} \leqslant x\right] \to \mathbb{P}\left[W \leqslant x\right],$$

where $W := \frac{Y}{\Psi_2 \sqrt{S_2}}$ is the limit random variable discussed in Theorem 1 above. Therefore, so long as we know the cdf of W and n is fairly large, we have a way to estimate $F_{\tilde{M}}$:

$$F_{\tilde{M}}(x) \approxeq \mathbb{P}\left[W \leqslant x\right]. \tag{6}$$

In order to make use of the approximation (6), we must know the cdf of W. However, W is a complicated random variable, which depends through ψ_+ , ψ_-, and Ψ_2 on the unknown filter coefficients; therefore it is impractical to assume knowledge of the exact distribution of W. Subsampling methods have proved very effective at estimating such limit distributions in similar contexts; see for example Politis, Romano, and Wolf (1999) or McElroy and Politis (2000). We now give a subsampling-based approach to this problem.

Let T_n denote $\frac{M_n}{\sqrt{\hat{\sigma}}}$, and define the "subsampling distribution estimator" of T_n to be the following empirical distribution function (edf):

$$K_b(x) := \frac{1}{n-b+1} \sum_{i=1}^{n-b+1} 1_{\{T_{b,i} \leqslant x\}}$$

where $T_{b,i}$ is essentially the statistic T_b evaluated on the subseries $\{X_i, \cdots, X_{b+i-1}\}$; in other words,

$$T_{b,i} := \frac{M_{b,i}}{\hat{\sigma}_{b,i}}.$$

The precise definitions of $M_{b,i}$ and $\hat{\sigma}_{b,i}$ are as follows:

$$M_{b,i} := \bigvee_{t=i}^{b+i-1} X_t, \qquad \hat{\sigma}_{b,i} := \sqrt{\frac{1}{b} \sum_{t=i}^{b+i-1} X_t^2}.$$

The object of subsampling is to use the subsampling distribution estimator as an approximation of the limit distribution in Theorem 1; it also approximates the sampling distribution. For more details and background on these methods, see the book by Politis, Romano, and Wolf (1999).

Strong mixing is a sufficient condition on the dependence structure which insures the validity of subsampling. The strong mixing assumption requires that $\alpha_X(k) \to 0$ as $k \to \infty$; here $\alpha_X(k) := sup_{A,B}|\mathbb{P}[A \cap B] - \mathbb{P}[A]\mathbb{P}[B]|$, where A and B are events in the σ-fields generated by $\{X_t, t \leq l\}$ and $\{X_t, t \geq l+k\}$, respectively, for any $l \geq 0$. General conditions for a linear series to be strong mixing are given by Withers (1981); they require that the ψ_j tend to zero fast enough (with j), and that the Z_t's have an absolutely continuous distribution. The strong mixing condition is easily seen to be satisfied if the series is an $MA(m)$ model for some $m \in \mathbb{N}$, i.e., when only a finite number of the filter coefficients ψ_j are nonzero. In addition, if the series has an $AR(1)$ representation, i.e. if

$$X_t = \psi X_{t-1} + Z_t$$

for some ψ bounded by one in absolute value, then the time series is strong mixing; see Pham and Tran (1985).

Now define the sampling distribution of T_n to be $W_n(x)$:

$$W_n(x) := \mathbb{P}\left[T_n \leqslant x\right]$$

and let $W(x)$ be the cdf of the random variable W from Theorem 1 :

$$W(x) := \mathbb{P}\left[W \leqslant x\right].$$

Then the following corollary results from a standard application of the subsampling theory of Politis, Romano, and Wolf (1999) to our Theorem 1 :

Corollary 1 *Under the $MA(\infty)$ model with the strong mixing hypothesis, the subsampling distribution estimator K_b is consistent as an estimator of the true sampling distribution of T_n, denoted by $W_n(x)$. In other words, if $b \to \infty$ as $n \to \infty$ but with $b/n \to 0$, we have*

$$\sup_x |K_b(x) - W_n(x)| \xrightarrow{P} 0 \tag{7}$$

and in addition

$$K_b^{-1}(t) \xrightarrow{P} W^{-1}(t) \tag{8}$$

for any $t \in (0,1)$*, where* $G^{-1}(t)$ *denotes the* t*th quantile of a given cdf* G*. Finally, we can also write*

$$W_n\left(c_{n,b}(1-t)\right)\mathbb{P}\left[T_n \leqslant c_{n,b}(1-t)\right] \to 1-t \tag{9}$$

as $n \to \infty$*, where* $c_{n,b}(1-t) = inf\{x : K_b(x) \geqslant 1-t$ *for any* $t \in (0,1)$*.*

Proof We need to check the hypotheses of Theorem 11.3.1 of Politis, Romano, and Wolf (1999). In that theorem, the rate τ_n is just $\frac{a_n}{a_n} = 1$, and $\frac{a_b}{a_n} \to 0$ as $n \to \infty$ (recall that $a_n = n^{\frac{1}{\alpha}} L(n)$). The limit random variable in our Theorem 1 has no point masses, and its cdf $W(\cdot)$ is certainly continuous everywhere, so Assumption 11.3.1 of Politis, Romano, and Wolf (1999) is satisfied. Thus line (7) holds, so that (8) and (9) follow from this. †

This result allows us to approximate the sampling cdf $W_n(x)$ by $K_b(x)$, from which we can draw quantiles:

$$F_{\tilde{M}}(x) \approxeq K_b(x).$$

Note that the subsampling distribution is a bona fide statistic, i.e. it can be computed from the data alone.

The good thing about this procedure is that the rate a_n need not be known nor estimated in order to form the above approximation. If the $\{Z_t\}$ sequence were in the normal domain of attraction, then $a_n = n^{\frac{1}{\alpha}}$. The problem with using this rate is that the exact value of α is unknown. One can either estimate α empirically, or use self-normalization. The Hill estimator may be used to first estimate α, but its implementation requires the choice of a bandwidth parameter, and is not robust under dependence (cf. Resnick 1997). However, in the case that the domain of attraction is not normal, there will also be a slowly varying function to estimate, which is a problem of a higher order of difficulty. Providentially, self-normalization avoids the need to estimate a_n ; in essence, self-normalization implicitly performs the rate estimation.

An obvious drawback is that we estimate the cdf of $\tilde{M}_n$ rather than the actual cdf of M_n. Some way of circumventing the appearance of the $\hat{\sigma}_n$ term is yet to be discovered. Since the sample variance diverges at a fast rate, with high probability $M_n \geq \tilde{M}_n$. From this we have heuristically

$$\mathbb{P}[M_n > x] \geq \mathbb{P}\left[\tilde{M}_n > x\right],$$

which says that the probability of exceeding amount x over n periods is bounded below by the probability that the "normalized" claims over n periods exceeds x; this latter quantity is estimable via subsampling.

References

[1] Beran, J. (1994) *Statistics for Long-Memory Processes* . Chapman and Hall.

[2] Beran, J., Sherman R., Taqqu, M., and Willinger W. (1995) Long-range dependence in Variable-bit rate video traffic. *IEEE Trans. Comm.* **43** 1566–1579.

[3] Billingsley, P. (1995) *Probability and Measure.* John Wiley and Sons.

[4] Brockwell, P. and Davis, R. (1991) *Time Series: Theory and Methods.* Springer.

[5] Davis, R. and Hsing, T. (1995) Point Process and Partial Sum Convergence for Weakly Dependent Random Variables with Infinite Variance. *Annals of Probability* **23**, No. 2, 879–917.

[6] Davis, R., and Mikosch, T. (1998) The sample autocorrelations of heavy-tailed processes with applications to ARCH. *Annals of Statistics* **26**, No. 5, 2049–2080.

[7] Davis, R., and Resnick, S. (1985) Limit theory for moving averages of random variables with regularly varying tail probabilities. *Annals of Probability* **13**, No. 1, 179–195.

[8] Davis, R., and Resnick, S. (1986) Limit theory for the sample covariance and correlation functions of moving averages. *Annals of Statistics* **14**, No. 2, 533–558.

[9] Durrett, R. (1996) *Probability: Theory and Examples.* Duxbury Press.

[10] Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997) *Modeling Extrema Events for Insurance and Finance.* Springer-Verlag, Berlin.

[11] 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.

[12] McElroy, T., and Politis, D. (2001) Robust Inference for the Mean in the Presence of Serial Correlation and Heavy Tailed Distributions. To appear in *Econometrica*.

[13] Pham, Tuan D. and Tran, Lanh T. (1985) Some Mixing Properties of Time Series Models. *Stochastic Processes and Their Applications* **19**, 297–303.

[14] Politis, D. and Romano, J. (1994) Large Sample Confidence Regions Based on Subsamples Under Minimal Assumptions. *Annals of Statistics* **22**, No. 4, 2031–2050.

[15] Politis, D., Romano, J., and Wolf, M. (1999) *Subsampling.* Springer, New York.

[16] Resnick, S. (1986) Point processes, regular variation, and weak convergence. *Advances in Applied Probability* **18**, 66–138.

[17] Resnick, S. (1987) *Extreme Values, Regular Variation, and Point Processes.* Springer–Verlag.

[18] Resnick, S. (1997) Special Invited Paper: Heavy Tail Modeling and Teletraffic Data. *Annals of Statistics* **25**, No. 5, 1805–1849.

[19] Resnick, S., and Stairca, C. (1998) Tail index estimation for dependent data. *Annals of Applied Probability* **8**, No. 4, 1156–1183.

[20] Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proceedings of the National Academy of Sciences* **42**, 43–47.

[21] Rosenblatt, M. (1984). Asymptotic normality, strong mixing and spectral density estimates, *Annals of Probability* **12**, 1167–1180.

[22] Rosenblatt, M. (1985). *Stationary Sequences and Random Fields.* Birkhäuser, Boston.

[23] Samorodnitsky, G. and Taqqu, M. (1994) *Stable Non-Gaussian Random Processes. Chapman and Hall.*

[24] Withers, C. S. (1981) Conditions for Linear Processes to be Strong Mixing. *Probability Theory and Related Fields* **57**, 477–480