NORGES TEKNISK-NATURVITENSKAPELIGE
UNIVERSITET

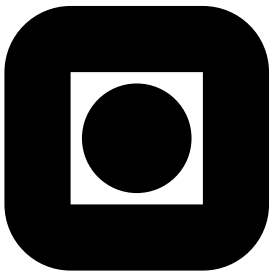
Working paper no. N2-2010
Department of Mathematical Sciences, NTNU

Estimation of extreme values of time series with heavy tails

by

Arvid Naess

PREPRINT
STATISTICS NO. 14/2010



NORWEGIAN UNIVERSITY OF SCIENCE AND
TECHNOLOGY
TRONDHEIM, NORWAY

This preprint has URL
http://www.math.ntnu.no/preprint/statistics/2010/S14-2010.pdf
Arvid Naess has homepage: http://www.math.ntnu.no/∼arvidn
E-mail: arvidn@math.ntnu.no
Address: Department of Mathematical Sciences, Norwegian University of Science and Technology, NO-7491 Trondheim, Norway.

Estimation of extreme values of time series with heavy tails

Arvid Naess*

Abstract

The paper focuses on the development of a new method for extreme value estimation based on sampled financial time series. Of particular concern is the case when the extreme values asymptotically follow the Fréchet distribution. The method is designed to account for statistical dependence between the data points of the time series in a rational way. The proposed procedure avoids the problem of declustering of data to ensure independence, which is a common problem for the peaks-over-threshold method. The goal has been to establish an accurate method for prediction of e.g. the VaR based on recorded historical data.

KEYWORDS: Extreme value estimation; financial time series; approximation by conditioning; average conditional exceedance rate.

*Department of Mathematical Sciences, Norwegian University of Science and Technology, NO-7491 Trondheim, Norway, e-mail: arvidn@math.ntnu.no

1 Introduction

Extreme value statistics, even in applications, have very often been based on asymptotic results. This is done either by assuming that the episodical extremes, for example yearly extreme values, are distributed according to the generalized (asymptotic) extreme value distribution with unknown parameters to be estimated on the basis of the observed data. Or it is assumed that the exceedances above high thresholds follow a generalized (asymptotic) Pareto distribution with parameters to be estimated from the data. The major problem with both of these approaches is that the asymptotic extreme value theory itself cannot be used in practice to decide to what extent it is applicable for the observed data. Hence, the assumption that an asymptotic extreme value distribution is the appropriate distribution for the observed data is based more or less on faith or convenience.

In an effort to ameliorate this situation, we have developed an approach to this problem that is less restrictive and more flexible than the one based on asymptotic theory. In particular, it has the capability to capture sub-asymptotic behaviour of the large data, which seems to be of some importance for accurate prediction. The proposed approach presented in this paper is developed under the assumption that the appropriate asymptotic extreme value distribution is the Fréchet distribution.

2 Cascade of Conditioning Approximations

Consider a stochastic process $Z(t)$, which has been observed over a time interval, $(0, T)$ say. Assume that values $X_1, \ldots, X_N$, which have been derived from the observed process, are allocated to the discrete times $t_1, \ldots, t_N$ in $(0, T)$. This could be simply the observed values of $Z(t)$ at each t_j, $j = 1, \ldots, N$, or it could be average values or peak values over smaller time intervals centered at the t_j's. Our goal in this paper is to accurately determine the distribution function of the extreme value $M_N = \max\{X_j\,; j = 1, \ldots, N\}$. Specifically, we want to estimate $P(\eta) = \text{Prob}(M_N \leq \eta)$ accurately for large values of η. An underlying premise for the development in this paper is that a rational approach to the study of the extreme values of the sampled time series is to consider exceedances of the individual random variables X_j above given thresholds, as in classical extreme value theory. The alternative approach of considering the exceedances by upcrossing of given thresholds by a continuous stochastic process has been developed by the authors [1, 2]. The approach taken in this paper would seem to be the appropriate way to deal with the recorded data time series of, for example, the daily largest wind speeds observed at a given location.

In the following we outline possible approaches for practical implementation of a cascade of approximations based on conditioning, where the first is a Markov-like approximation in the sense that it is a one-step memory approximation. This approximation concept is described in [3, 4].

From the definition of $P(\eta)$ it follows that

$$\begin{aligned} P(\eta) &= \text{Prob}\{X_1 \leq \eta, \ldots, X_N \leq \eta\} \\ &= \text{Prob}\{X_N \leq \eta | X_1 \leq \eta, \ldots, X_{N-1} \leq \eta\} \text{Prob}\{X_1 \leq \eta, \ldots, X_{N-1} \leq \eta\} \\ &= \prod_{j=2}^{N} \text{Prob}\{X_j \leq \eta | X_1 \leq \eta, \ldots, X_{j-1}\} \cdot P(X_1 \leq \eta) \end{aligned} \tag{1}$$

In general, the variables X_j are statistically dependent. Hence, instead of assuming that all the X_j are statistically independent, which leads to the classical approximation

$$P(\eta) \approx \prod_{j=1}^{N} P(X_j \leq \eta), \tag{2}$$

the following Markov-like, or one-step memory, assumption will to a certain extent account for dependence between the X_j,

$$\text{Prob}\{X_j \leq \eta | X_1 \leq \eta, \ldots, X_{j-1} \leq \eta\} \approx \text{Prob}\{X_j \leq \eta | X_{j-1} \leq \eta\}, \tag{3}$$

for $2 \leq j \leq N$. This can be extended to

$$\text{Prob}\{X_j \leq \eta | X_1 \leq \eta, \ldots, X_{j-1} \leq \eta\} \approx \text{Prob}\{X_j \leq \eta | X_{j-2} \leq \eta, X_{j-1} \leq \eta\}, \tag{4}$$

for $3 \leq j \leq N$, and so on.

Eqs. (3) and (4) represent refinements of the Poisson assumption. One would expect that such approximations would be increasingly more able to capture statistical dependence between neighboring data in the time series. As will be seen in the examples in the following section, $P(\eta)$ computed using Eq. (4) is quite close to the value obtained using Eq. (3). This indicates that in practice, Eq. (3) is often able to capture the effect of statistical dependence in e.g. wind speed data with good accuracy. However, there is no noticeable increase of numerical effort by using Eq. (4), or its further refinements by including three or more preceding peaks.

Combining Eq. (1) with Eq. (3), the following relation is obtained

$$P(\eta) \approx \frac{\prod_{j=2}^{N} p_{2j}(\eta)}{\prod_{j=2}^{N-1} p_{1j}(\eta)} \tag{5}$$

where we have introduced the notation $p_{kj}(\eta) = \text{Prob}\{X_{j-k+1} \leq \eta, \ldots, X_j \leq \eta\}$ for $j \geq k$.

It is of interest to compare the values for $P(\eta)$ obtained by using Eq. (5) as compared to Eq. (2). Now, Eq. (2) can be rewritten in the form

$$P(\eta) \approx \prod_{j=1}^{N} \big(1 - \alpha_{1j}(\eta)\big), \tag{6}$$

where

$$\alpha_{1j}(\eta) = \text{Prob}\{X_j > \eta\} = 1 - p_{1j}(\eta). \tag{7}$$

Then

$$P(\eta) \approx P_1(\eta) = \exp\big(-\sum_{j=1}^{N} \alpha_{1j}(\eta)\big) \tag{8}$$

Alternatively, Eq. (5) gives

$$P(\eta) \approx \prod_{j=2}^{N} \big(1 - \alpha_{2j}(\eta)\big)\, p_{11}(\eta), \tag{9}$$

where $\alpha_{kj}(\eta) = 1 - p_{kj}(\eta)/p_{k-1,j-1}(\eta)$, for $j \geq k \geq 2$. That is

$$\alpha_{kj}(\eta) = \text{Prob}\{X_j > \eta \,|\, X_{j-k+1} \leq \eta, \ldots, X_{j-1} \leq \eta\} \tag{10}$$

denotes the exceedance probability conditional on $k-1$ previous non-exceedances. From Eq. (9) it is obtained that,

$$P(\eta) \approx P_2(\eta) = \exp\big(-\sum_{j=2}^{N} \alpha_{2j}(\eta) - \alpha_{11}(\eta)\big), \tag{11}$$

since $p_{11}(\eta) \approx \exp(-\alpha_{11}(\eta))$.

Conditioning on two previous observations X_{j-2}, X_{j-1} preceding X_j gives

$$P(\eta) \approx P_3(\eta) = \exp\big(-\sum_{j=3}^{N} \alpha_{3j}(\eta) - \alpha_{22}(\eta) - \alpha_{11}(\eta)\big), \tag{12}$$

while conditioning on three prior observations leads to the equation

$$P(\eta) \approx P_4(\eta) = \exp\big(-\sum_{j=4}^{N} \alpha_{4j}(\eta) - \alpha_{33}(\eta) - \alpha_{22}(\eta) - \alpha_{11}(\eta)\big), \tag{13}$$

and so on. Therefore, extreme value prediction by the conditioning approach described above reduces to estimation of (combinations) of the $\alpha_{kj}(\eta)$ functions. For most practical applications $N >> 1$, so that ($k \geq 2$)

$$P_k(\eta) \approx \exp\big(-\sum_{j=k}^{N} \alpha_{kj}(\eta)\big). \tag{14}$$

Going back to Eq. (8), and the definition of $\alpha_{1j}(\eta)$, it follows that $\sum_{j=1}^{N} \alpha_{1j}(\eta)$ is equal to the expected number of exceedances of the threshold η during the time interval $(0, T)$. Eq. (8) therefore expresses the approximation that the stream of exceedance events constitute a (non-stationary) Poisson process. This opens for an understanding of Eq. (11) and subsequent approximations by interpreting the expressions $\sum_{j=k}^{N} \alpha_{kj}(\eta) + \alpha_{k-1,k-1}(\eta) + \ldots + \alpha_{11}(\eta) \approx \sum_{j=k}^{N} \alpha_{kj}(\eta)$ as the expected effective number of (independent) exceedances provided by conditioning on $k-1$ previous observations.

3 Empirical Estimation of the Mean Exceedance Rates

It is expedient to introduce the concept of average conditional exceedance rates (ACER) as follows,

$$\varepsilon_k(\eta) = \frac{1}{N-k+1} \sum_{j=k}^{N} \alpha_{kj}(\eta)\,,\ k = 1, 2, \ldots \tag{15}$$

In practice there are typically two scenarios for the underlying process $Z(t)$. Either we may consider it to be a stationary process, or, in fact, even an ergodic process. The other alternative is to view $Z(t)$ as a process that depends on certain parameters whose variation in time may be modelled as an ergodic process in its own right. For each set of values of the parameters, the premise is that $Z(t)$ can be modelled as an ergodic process. This would be the scenario that can be used to model long-term statistics [5].

For both these scenarios, the empirical estimation of the ACER function $\overline{\varepsilon}_k(\eta)$ proceeds in a completely analogous way by counting the total number of favourable incidents, that is, exceedances conditional on the requisite number of preceding non-exceedances, for the total data time series and then finally dividing by $N-k+1 \approx N$. This can be shown for the long-term situation by using a similar analysis as in [5].

A few more details on the numerical estimation of $\varepsilon_k(\eta)$ for $k \geq 2$ are useful. We start by introducing the following random functions,

$$A_{kj}(\eta) = \mathbf{1}\{X_j > \eta, X_{j-1} \leq \eta, \ldots, X_{j-k+1} \leq \eta\}\,,\ j = k, \ldots, N,\ k = 2, 3, \ldots \tag{16}$$

and

$$B_{kj}(\eta) = \mathbf{1}\{X_{j-1} \leq \eta, \ldots, X_{j-k+1} \leq \eta\}\,,\ j = k, \ldots, N,\ k = 2, \ldots\,, \tag{17}$$

where $\mathbf{1}\{\mathcal{A}\}$ denotes the indicator function of some event $\mathcal{A}$. Then

$$\alpha_{kj}(\eta) = \frac{\mathrm{E}[A_{kj}(\eta)]}{\mathrm{E}[B_{jk}(\eta)]}\,,\ j = k, \ldots, N,\ k = 2, \ldots\,, \tag{18}$$

where $\mathrm{E}[\cdot]$ denotes the expectation operator. Assuming an ergodic process, then obviously $\varepsilon_k(\eta) = \alpha_{kk}(\eta) = \ldots = \alpha_{kN}(\eta)$, and it may be assumed that for the time series at hand

$$\varepsilon_k(\eta) = \lim_{N\to\infty} \frac{\sum_{j=k}^{N} A_{kj}(\eta)}{\sum_{j=k}^{N} B_{jk}(\eta)}\,. \tag{19}$$

Clearly, $\lim_{\eta\to\infty} \sum_{j=k}^{N} B_{jk}(\eta) = N-k+1 \approx N$. Hence, $\lim_{\eta\to\infty} \tilde{\varepsilon}_k(\eta)/\varepsilon_k(\eta) = 1$, where

$$\tilde{\varepsilon}_k(\eta) = \lim_{N\to\infty} \frac{\sum_{j=k}^{N} A_{kj}(\eta)}{N-k+1}\,. \tag{20}$$

In the following we shall use $\tilde{\varepsilon}_k(\eta)$ instead of $\varepsilon_k(\eta)$ for $k \geq 2$. The advantage of using the modified ACER function $\tilde{\varepsilon}_k(\eta)$ for $k \geq 2$ is that it is easier to use for non-stationary or long-term statistics than $\varepsilon_k(\eta)$. Since our focus is on

the values of the ACER at the extreme levels, we may use any function that provides correct estimates of the ACER function at the extreme levels.

For both stationary and non-stationary time series, the sample estimate of $\tilde{\varepsilon}_k(\eta)$ would be,

$$\hat{\varepsilon}_k(\eta) = \frac{1}{R} \sum_{r=1}^{R} \hat{\varepsilon}_k^{(r)}(\eta)\,, \tag{21}$$

where R is the number of realizations (samples), and

$$\hat{\varepsilon}_k^{(r)}(\eta) = \frac{\sum_{j=k}^{N} A_{kj}^{(r)}(\eta)}{N-k+1}\,, \tag{22}$$

where the index (r) refers to realization no. r.

It is of interest to note what events are actually counted for the calculation of the various $\hat{\varepsilon}_k(\eta)$, $k \geq 2$. Let us start with $\hat{\varepsilon}_2(\eta)$. It follows from the definition of $\tilde{\varepsilon}_2(\eta)$ that $\tilde{\varepsilon}_2(\eta)\,(N-1)$ can be interpreted as the expected number of exceedances above the level η satisfying the condition that an exceedance is counted only if it is immediately preceded by a non-exceedance. A reinterpretation of this is that $\hat{\varepsilon}_2(\eta)\,(N-1)$ equals the average number of clumps of exceedances above η for the realizations considered, where a clump of exceedances is defined as a maximum number of consecutive exceedances above η. In general, $\hat{\varepsilon}_k(\eta)\,(N-1)$ then equals the average number of clumps of exceedances above η separated by at least $k-1$ non-exceedances.

Now, let us look at the problem of estimating a confidence interval for $\tilde{\varepsilon}_k(\eta)$. The sample standard deviation $\hat{s}_k(\eta)$ can be estimated by the standard formula,

$$\hat{s}_k(\eta)^2 = \frac{1}{R-1} \sum_{r=1}^{R} \Big(\hat{\varepsilon}_k^{(r)}(\eta) - \hat{\varepsilon}_k(\eta) \Big)^2 . \tag{23}$$

Assuming that realizations are independent, for a suitable number R, e.g. $R \geq 20$, Eq. (23) leads to a good approximation of the 95 % confidence interval CI $= \big(\mathrm{CI}^-(\eta), \mathrm{CI}^+(\eta)\big)$ for the value $\tilde{\varepsilon}_k(\eta)$, where

$$\mathrm{CI}^{\pm}(\eta) = \hat{\varepsilon}_k(\eta) \pm 1.96\, \hat{s}_k(\eta)/\sqrt{R}\,. \tag{24}$$

The approach to extreme value prediction presented in this paper derives from an assumption about the sampled time series to be used as a basis for prediction. This assumption derives from an underlying premise concerning the relevant asymptotic extreme value distribution, which is assumed here to be of Fréchet type. For independent data, this assumption can be expressed in terms of the ACER function $\varepsilon_1(\eta)$ as,

$$\varepsilon_1(\eta) \approx \big[1 + \xi\big(a(\eta - b)\big)\big]^{-\frac{1}{\xi}}\ ,\ \ \eta \geq \eta_0, \tag{25}$$

for a suitable asymptotic tail marker η_0, where $a > 0$, b, $\xi > 0$ are constants.

The implication of this assumption on the possible subasymptotic functional forms of $\varepsilon_k(\eta)$ cannot easily be decided. However, using the asymptotic form as a guide, it is assumed that the behaviour of the mean exceedance rate in

the subasymptotic part of the tail will follow a function largely of the form $\left[1+\xi\big(a(\eta-b)^c\big)\right]^{-\frac{1}{\xi}}$ $(\eta \geq \eta_1 \geq b)$ where $a>0$, b, $c>0$ and $\xi>0$ are suitable constants, and η_1 is an appropriately chosen tail level. Hence, it will be assumed that

$$\varepsilon_k(\eta) \approx q_k(\eta)\ \left[1+\xi_k\big(a_k(\eta-b_k)^{c_k}\big)\right]^{-\frac{1}{\xi_k}}\ ,\ \eta \geq \eta_1, \tag{26}$$

where the function $q_k(\eta)$ is weakly varying compared with the function $\left[1+\xi_k\big(a_k(\eta-b_k)^{c_k}\big)\right]^{-\frac{1}{\xi_k}}$ and $a_k>0$, b_k, $c_k>0$ and $\xi_k>0$ are suitable constants, that in general will be dependent on k. Note that the values $c_k=1$ and $q_k(\eta)=1$ corresponds to the asymptotic limit, which is then a special case of the general expression given in Eq. (26).

An alternative form to Eq. (26) would be to assume that

$$\varepsilon_k(\eta) \approx\ \left[1+\xi_k\big(a_k(\eta-b_k)^{c_k}+d_k(\eta)\big)\right]^{-\frac{1}{\xi_k}}\ ,\ \eta \geq \eta_1, \tag{27}$$

where the function $d_k(\eta)$ is weakly varying compared with the function $a_k(\eta-b_k)^{c_k}$. However, for estimation purposes, it turns out that the form given by Eq. (26) is preferrable as it leads to simpler estimation procedures. This aspect will be discussed later in the paper.

For practical identification of the ACER functions given by Eq. (26), it expedient to assume that the unknown function $q_k(\eta)$ varies sufficiently slowly to be replaced by a constant. In general, $q_k(\eta)$ is not constant, but its variation in the tail region is usually sufficiently slow to allow for its replacement by a constant. Hence, it is in effect assumed that $q_k(\eta)$ can be replace by a constant for $\eta \geq \eta_1$, for an appropriate choice of tail marker η_1. For simplicity of notation, in the following we shall suppress the index k on the ACER functions, which will then be written as,

$$\varepsilon(\eta) \approx q\ [1+\tilde{a}\,(\eta-b)^c]^{-\gamma}\ ,\ \eta \geq \eta_1, \tag{28}$$

where $\gamma=1/\xi$, $\tilde{a}=a\xi$.

In the practical analysis of data, first the tail marker η_1 is provisionally identified from visual inspection of the log plot $(\eta, \ln\hat{\varepsilon}_k(\eta))$. The value chosen for η_1 corresponds to the beginning of regular tail behaviour in a sense to be discussed below.

The optimization process to estimate the parameters is done relative to the log plot. The mean square error function to be minimized is written as

$$F(\tilde{a},b,c,q,\gamma)=\sum_{j=1}^{N} w_j\ \big|\log\hat{\varepsilon}(\eta_j)-\log q+\gamma\,[1+\tilde{a}(\eta_j-b)^c]\ \big|^2\,, \tag{29}$$

where $w_j=\big(\log CI^+(\eta_j)-\log CI^-(\eta_j)\big)^{-2}$ denotes a weight factor that puts more emphasis on the more reliable data points. The choice of weight factor is, of course, to some extent arbitrary, and if it is considered more appropriate to put a slightly stronger emphasis on the larger data, this can be simply achieved by replacing the exponent -2 by, for example, -1 in the definition of w_j.

One of the options for estimating the five parameters $\tilde{a}, b, c, q, \gamma$ is to use the Levenberg-Marquardt least squares optimization method, which seems well suited for the task at hand [6], especially if a simplified optimization approach is implemented. This is obtained by observing that if $\tilde{a}$, b and c are fixed in Eq. (29), the optimization problem reduces to a standard weighted linear regression problem. That is, with $\tilde{a}$, b and c fixed, the optimal values of γ and $\log q$ are found using closed form weighted linear regression formulas in terms of w_j, $y_j = \log \hat{\varepsilon}(\eta_j)$ and $x_j = 1 + \tilde{a}(\eta_j - b)^c$. In that light, it can also be concluded that the best linear unbiased estimators (BLUE) are obtained for $w_j = \sigma_{yj}^{-2}$, where $\sigma_{yj}^2 = \mathrm{Var}[y_j]$ (empirical) [7, 8]. Unfortunately, this is not a very practical weight factor for the kind of problem we have here because it seems that the summation in Eq. (29) then typically would have to stop at undesirably small values of η_j.

It is obtained that the optimal values of γ and $\log q$ are given by the relations,

$$\gamma^*(\tilde{a}, b, c) = -\frac{\sum_{j=1}^N w_j (x_j - \overline{x})(y_j - \overline{y})}{\sum_{j=1}^N w_j (x_j - \overline{x})^2}\,, \tag{30}$$

and

$$\log q^*(\tilde{a}, b, c) = \overline{y} + \gamma^*(\tilde{a}, b, c)\overline{x}\,. \tag{31}$$

To calculate the final optimal set of parameters, the Levenberg-Marquardt method may now be used on the function $\tilde{F}(\tilde{a}, b, c) = F(\tilde{a}, b, c, q^*(\tilde{a}, b, c), \gamma^*(\tilde{a}, b, c))$ to find the optimal values $\tilde{a}^*$, b^* and c^*, and then Eqs. (30) and (31) are used to calculate the corresponding γ^* and q^*. The optimal values of the parameters may e.g also be found by a sequential quadratic programming (SQP) method [9].

For estimation of the confidence interval for the predicted return value provided by the optimal curve, the empirical confidence band is reanchored to the optimal curve. The range of fitted curves that stay within the reanchored confidence band will determine an optimized confidence interval of the predicted return value. As a final point, it has been observed that the predicted return value is not very sensitive to the choice of η_1. However, this sensitivity should always be checked to verify the robustness of the obtained predictions.

4 Concluding remarks

This paper proposes a new method for extreme value estimation in time series with heavy tails leading to Fréchet type asymptotic extreme value distributions. It is based on the concept of average conditional exceedance rate (ACER). It is demonstrated that the exact extreme value distribution given by the data can be captured within the inherent statistical uncertainty by using the ACER functions.

To provide predictions of the high quantiles in the extreme value distribution, a representation of the ACER functions by a particular class of parametric functions is proposed. This class of functions is constructed in such a way that they are capable to some extent of capturing the subasymptotic behaviour of the extreme value distribution. By this, more data may become available for analysis, and better and more accurate predictions may ensue.

5 Acknowledgements

Part of this work was supported by the Research Council of Norway (NFR) through the Centre for Ships and Ocean Structures (CeSOS) at the Norwegian University of Science and Technology.

References

[1] A. Naess and O. Gaidai. Monte Carlo methods for estimating the extreme response of dynamical systems. *Journal of Engineering Mechanics, ASCE*, 134(8):628–636, 2008.

[2] A. Naess, O. Gaidai, and S. Haver. Efficient estimation of extreme response of drag dominated offshore structures by Monte Carlo simulation. *Ocean Engineering*, 34(16):2188–2197, 2007.

[3] A. Naess. The joint crossing frequency of stochastic processes and its application to wave theory. *Applied Ocean Research*, 7(1):35–50, 1985.

[4] A. Naess. Approximate first-passage and extremes of narrow-band Gaussian and non-Gaussian random vibrations. *Journal of Sound and Vibration*, 138(3):365–380, 1990.

[5] A. Naess. On the long-term statistics of extremes. *Applied Ocean Research*, 6(4):227–228, 1984.

[6] P. Gill, W. Murray, and M. H. Wright. *Practical Optimization*. Academic Press, London, 1981.

[7] N. R. Draper and H. Smith. *Applied Regression Analysis*. Wiley-Interscience, New York, NY, 1998.

[8] D. C. Montgomery, E. A. Peck, and G. G. Vining. *Introduction to Linear Regression Analysis*. Elsevier Science Publishers B. V., Amsterdam, The Netherlands, 2002.

[9] Numerical Algorithms Group. *NAG Toolbox for Matlab*. NAG Ltd, Oxford, UK, 2010.