NORGES TEKNISK-NATURVITENSKAPELIGE UNIVERSITET

Working paper no. N1-2011
Department of Mathematical Sciences, NTNU

A Note on the Bivariate ACER Method

by

Arvid Naess

PREPRINT
STATISTICS NO. 1/2011

NORWEGIAN UNIVERSITY OF SCIENCE AND TECHNOLOGY
TRONDHEIM, NORWAY

This preprint has URL http://www.math.ntnu.no/preprint/statistics/2011/S1-2011.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.

A Note on the Bivariate ACER Method

Arvid Naess*

Abstract

The paper focuses on the extension of the ACER method for prediction of extreme value statistics to the case of bivariate time series. Using the ACER method it is possible to provide an estimate of the exact extreme value distribution of a univariate time series. This is obtained by introducing a cascade of conditioning approximations to the exact extreme value distribution. When this cascade has converged, an estimate of the exact distribution has been obtained. In this paper it will be shown how the univariate ACER method can be extended in a natural way to also cover the case of bivariate data. In fact, the ACER method can in principle be extended to multivariate time series of any dimension. However, the requirements to the requisite statistical analyses would severely hamper a practical implementation for higher dimensional cases.

KEYWORDS: Extreme value estimation; bivariate 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

The extension of extreme value statistics from the univariate to the bivariate case meets with several challenges. First of all, there is no direct generalization of the univariate extreme value types theorem to the bivariate case. The general result on possible bivariate asymptotic extreme value distributions is in a sense too general to be of much practical value. More specifically, there are no precise estimation tools that allows us to decide on the joint distribution of the bivariate extremes from a given set of bivariate data. Of course, the marginal data sets can be used to derive estimates of the marginal extreme value distributions, but the joint distribution is still a long way off.

A popular method of trying to cope with the problem of bivariate extremes is to adopt a copula to represent the joint distribution. For this purpose a range of different copulas have been proposed. The main problem with this approach is that it is rather ad hoc. That is, there is no theoretical justification for choosing one particular copula over the other in a specific case, but rather a more or less subjective decision about which copula that seems to better fit the data.

Efforts have also been invested in trying to extend the peaks-over-threshold (POT) method to the bivariate case. This has not yet resulted in a method of the same prediction capabilities as the univariate POT.

In this paper we shall show that the recently developed average conditional exceedance rate (ACER) method has a direct extension to the bivariate case without the need to introduce any approximations or simplifications. Thus, in principle, the bivariate ACER has the possibility to provide exact estimates of the bivariate extreme value distribution that is given by the data.

2 Cascade of Conditioning Approximations

Consider a bivariate stochastic process $Z(t) = (X(t), Y(t))$ with dependent component processes, which has been observed over a time interval, $(0, T)$ say. Assume that the sampled values $(X_1, Y_1), \ldots, (X_N, Y_N)$ are allocated to the (usually equidistant) discrete times $t_1, \ldots, t_N$ in $(0, T)$. Our goal in this paper is to accurately determine the joint distribution function of the extreme value vector $(\hat{X}_N, \hat{Y}_N)$, where $\hat{X}_N = \max\{X_j\,; j = 1, \ldots, N\}$, and with a similar definition of $\hat{Y}_N$. Specifically, we want to estimate $P(\xi, \eta) = \text{Prob}(\hat{X}_N \leq \xi, \hat{Y}_N \leq \eta)$ accurately for large values of ξ and η.

In the following we outline the implementation of a cascade of approximations based on conditioning, where the first is a one-step memory approximation, which may be considered a Markov-like approximation. This approximation concept is described in [1, 2]. However, it is emphasized that it is not a Markov chain approximation.

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

$$\begin{aligned}
P(\xi,\eta) &= \mathrm{Prob}\big(X_1 \le \xi, Y_1 \le \eta \ldots, X_N \le \xi, Y_N \le \eta\big) \\
&= \mathrm{Prob}\big(X_N \le \xi, Y_N \le \eta \,|\, X_1 \le \xi, Y_1 \le \eta \ldots, X_{N-1} \le \xi, Y_{N-1} \le \eta\big) \\
&\quad \cdot \mathrm{Prob}\{X_1 \le \xi, Y_1 \le \eta \ldots, X_{N-1} \le \xi, Y_{N-1} \le \eta\} \\
&= \prod_{j=2}^{N} \mathrm{Prob}\big(X_j \le \xi, Y_j \le \eta \,|\, X_1 \le \xi, Y_1 \le \eta, \ldots, X_{j-1} \le \xi, Y_{j-1} \le \eta\big) \\
&\quad \cdot \mathrm{Prob}\big(X_1 \le \xi, Y_1 \le \eta\big) \qquad (1)
\end{aligned}$$

We shall start the development of the cascade of conditioning approximations by first looking at the following basic case.

2.1 Independent sample points

In this special case we obtain

$$\begin{aligned}
P(\xi,\eta) &= \prod_{j=1}^{N} \mathrm{Prob}(X_j \le \xi, Y_j \le \eta) \\
&= \prod_{j=1}^{N} \{1 - \mathrm{Prob}(X_j > \xi) - \mathrm{Prob}(Y_j > \eta) + \mathrm{Prob}(X_j > \xi, Y_j > \eta)\} \qquad (2)
\end{aligned}$$

We now introduce the notation $\alpha_{1j}(\xi) = \mathrm{Prob}(X_j > \xi)$, $\beta_{1j}(\eta) = \mathrm{Prob}(Y_j > \eta)$, and $\gamma_{1j}(\xi,\eta) = \mathrm{Prob}(X_j > \xi, Y_j > \eta)$ for $j = 1,\ldots,N$. Eq. (2) can then be rewritten as

$$\begin{aligned}
P(\xi,\eta) &= \prod_{j=1}^{N} \{1 - \alpha_{1j}(\xi) - \beta_{1j}(\eta) + \gamma_{1j}(\xi,\eta)\} \\
&\approx \exp\{-\sum_{j=1}^{N} \big(\alpha_{1j}(\xi) + \beta_{1j}(\eta) - \gamma_{1j}(\xi,\eta)\big)\}\,;\;\; \xi,\eta \to \infty\,. \qquad (3)
\end{aligned}$$

2.2 Conditioning on one previous sample point

The first genuine conditioning approximation is obtained by neglecting all previous conditioning events except the immediate predecessor in Eq. (1). That is, the following one-step memory approximation is adopted,

$$\begin{aligned}
P(\xi,\eta) &= \prod_{j=2}^{N} \mathrm{Prob}\big(X_j \le \xi, Y_j \le \eta \,|\, X_{j-1} \le \xi, Y_{j-1} \le \eta\big) \\
&\quad \cdot \mathrm{Prob}\big(X_1 \le \xi, Y_1 \le \eta\big) \qquad (4)
\end{aligned}$$

This may be rewritten as,

$$P(\xi,\eta) = \prod_{j=2}^{N} \Big\{1 - \mathrm{Prob}\big(X_j > \xi \,|\, X_{j-1} \le \xi, Y_{j-1} \le \eta\big) \\ - \mathrm{Prob}\big(Y_j > \eta \,|\, X_{j-1} \le \xi, Y_{j-1} \le \eta\big) \\ + \mathrm{Prob}\big(X_j > \xi, Y_j > \eta \,|\, X_{j-1} \le \xi, Y_{j-1} \le \eta\big)\Big\} \\ \cdot \big\{1 - \mathrm{Prob}(X_1 > \xi) - \mathrm{Prob}(Y_1 > \eta) + \mathrm{Prob}(X_1 > \xi, Y_1 > \eta)\big\} \tag{5}$$

By introducing the notation $\alpha_{2j}(\xi;\eta) = \mathrm{Prob}\big(X_j > \xi \,|\, X_{j-1} \le \xi, Y_{j-1} \le \eta\big)$, $\beta_{2j}(\eta;\xi) = \mathrm{Prob}\big(Y_j > \eta \,|\, X_{j-1} \le \xi, Y_{j-1} \le \eta\big)$, and $\gamma_{2j}(\xi,\eta) = \mathrm{Prob}\big(X_j > \xi, Y_j > \eta \,|\, X_{j-1} \le \xi, Y_{j-1} \le \eta\big)$, we obtain as in Eq. (3) that,

$$P(\xi,\eta) \approx \exp\{-\sum_{j=2}^{N} \big(\alpha_{2j}(\xi;\eta) + \beta_{2j}(\eta;\xi) - \gamma_{2j}(\xi,\eta)\big)\} \\ \exp\{-\big(\alpha_{11}(\xi;\eta) + \beta_{11}(\eta;\xi) - \gamma_{11}(\xi,\eta)\big)\}\,;\ \ \xi,\eta \to \infty\,. \tag{6}$$

2.3 Conditioning on several previous sample points

It has been observed in the univariate case that conditioning on one previous data point is sometimes enough to capture the effect of dependence in the time series to a large extent [3]. However, there are also cases where this is not sufficient. This can only be ascertained by having available a method that displays the complete picture concerning the importance of dependence on the extreme value distribution. Our proposed solution to this is obtained by introducing a cascade of conditioning approximations beyond the one-step approximation above.

We start by defining the following set of events,

$$\mathcal{C}_{kj}(\xi,\eta) = \big\{X_{j-1} \le \xi, Y_{j-1} \le \eta, \ldots, X_{j-k+1} \le \xi, Y_{j-k+1} \le \eta\big\} \tag{7}$$

Going back to Eq. (1), and conditioning on not more than $k-1$ previous data points, where $k = 2, \ldots, N$ and $j \ge k$, it is obtained that

$$P(\xi,\eta) = \prod_{j=k}^{N} \mathrm{Prob}\big(X_j \le \xi, Y_j \le \eta \,|\, \mathcal{C}_{kj}(\xi,\eta)\big) \cdot \mathrm{Prob}\big(\mathcal{C}_{kk}(\xi,\eta)\big)\,, \tag{8}$$

where

$$\mathrm{Prob}\big(\mathcal{C}_{kk}(\xi,\eta)\big) = \mathrm{Prob}\big(X_{k-1} \le \xi, Y_{k-1} \le \eta \,|\, \mathcal{C}_{k-1,k-1}(\xi,\eta)\big) \\ \cdot \mathrm{Prob}\big(\mathcal{C}_{k-1,k-1}(\xi,\eta)\big)\,. \tag{9}$$

By introducing the notation $\alpha_{kj}(\xi;\eta) = \mathrm{Prob}\big(X_j > \xi \,|\, \mathcal{C}_{kj}(\xi,\eta)\big)$, $\beta_{kj}(\eta;\xi) = \mathrm{Prob}\big(Y_j > \eta \,|\, \mathcal{C}_{kj}(\xi,\eta)\big)$, and $\gamma_{kj}(\xi,\eta) = \mathrm{Prob}\big(X_j > \xi, Y_j > \eta \,|\, \mathcal{C}_{kj}(\xi,\eta)\big)$, it can now be shown that,

$$\prod_{j=k}^{N} \text{Prob}\big(X_j \le \xi, Y_j \le \eta \,|\, \mathcal{C}_{kj}(\xi,\eta)\big)$$
$$\approx \exp\{-\sum_{j=k}^{N}\big(\alpha_{kj}(\xi;\eta) + \beta_{kj}(\eta;\xi) - \gamma_{kj}(\xi,\eta)\big)\}\,;\;\; \xi,\eta \to \infty\,. \tag{10}$$

Similarly, it is found that,

$$\begin{aligned}\text{Prob}\big(\mathcal{C}_{kk}(\xi,\eta)\big) &\approx \exp\{-\big(\alpha_{k-1,k-1}(\xi;\eta) + \beta_{k-1,k-1}(\eta;\xi) - \gamma_{k-1,k-1}(\xi,\eta)\big)\} \\ &\quad\cdot \text{Prob}\big(\mathcal{C}_{k-1,k-1}(\xi,\eta)\big) \\ &\approx \exp\{-\sum_{j=1}^{k-1}\big(\alpha_{jj}(\xi) + \beta_{jj}(\eta) - \gamma_{jj}(\xi,\eta)\big)\}\,;\;\; \xi,\eta \to \infty\,. \end{aligned} \tag{11}$$

Hence, we finally end up with the result,

$$\begin{aligned}P(\xi,\eta) &\approx \exp\{-\sum_{j=k}^{N}\big(\alpha_{kj}(\xi;\eta) + \beta_{kj}(\eta;\xi) - \gamma_{kj}(\xi,\eta)\big)\} \\ &\quad\cdot \exp\{-\sum_{j=1}^{k-1}\big(\alpha_{jj}(\xi;\eta) + \beta_{jj}(\eta;\xi) - \gamma_{jj}(\xi,\eta)\big)\}\,;\;\; \xi,\eta \to \infty\,. \end{aligned} \tag{12}$$

For most applications, $N >> k$, so that the following approximation can be adopted,

$$P(\xi,\eta) \approx \exp\{-\sum_{j=k}^{N}\big(\alpha_{kj}(\xi;\eta) + \beta_{kj}(\eta;\xi) - \gamma_{kj}(\xi,\eta)\big)\}\,;\;\; \xi,\eta \to \infty\,. \tag{13}$$

3 Empirical Estimation of the Average Exceedance Rates

To get a more compact representation, it is expedient to introduce the concept of k'th order average conditional exceedance rate (ACER) functions as follows,

$$\alpha_k(\xi;\eta) = \frac{1}{N-k+1}\sum_{j=k}^{N} \alpha_{kj}(\xi;\eta)\,,\; k = 1,2,\ldots\,, \tag{14}$$

$$\beta_k(\eta;\xi) = \frac{1}{N-k+1}\sum_{j=k}^{N} \beta_{kj}(\eta;\xi)\,,\; k = 1,2,\ldots\,, \tag{15}$$

and

$$\gamma_k(\xi,\eta) = \frac{1}{N-k+1}\sum_{j=k}^{N} \gamma_{kj}(\xi,\eta)\,,\; k = 1,2,\ldots\,. \tag{16}$$

Hence, when $N >> k$, we may write

$$P(\xi, \eta) \approx \exp\{ -(N-k+1)\big(\alpha_k(\xi;\eta) + \beta_k(\eta;\xi) - \gamma_k(\xi,\eta)\big)\}\,; \;\; \xi, \eta \to \infty\,. \tag{17}$$

From this equation follows the result for e.g. $\mathrm{Prob}(\hat{Y}_N \le \xi \,|\, \hat{X}_N \le \xi)$ by writing,

$$\mathrm{Prob}(\hat{Y}_N \le \xi \,|\, \hat{X}_N \le \xi) = \frac{P(\xi,\xi)}{\mathrm{Prob}(\hat{X}_N \le \xi)}$$
$$\approx \exp\{ -(N-k+1)\big(\alpha_k(\xi;\xi) + \beta_k(\xi;\xi) - \varepsilon_k(\xi) - \gamma_k(\xi,\xi)\big)\}\,; \;\; \xi \to \infty\,, \tag{18}$$

where $\varepsilon_k(\xi)$ is the k'th order univariate ACER function for the time series X_j [3].

A few more details on the numerical estimation of the ACER functions are useful. We start by introducing a set of random functions. For $k = 2, \dots, N$, and $k \le j \le N$, let

$$A_{kj}(\xi;\eta) = \mathbf{1}\{X_j > \xi, X_{j-1} \le \xi, Y_{j-1} \le \eta, \dots, X_{j-k+1} \le \xi, Y_{j-k+1} \le \eta\}\,, \tag{19}$$

$$B_{kj}(\eta;\xi) = \mathbf{1}\{Y_j > \eta, X_{j-1} \le \xi, Y_{j-1} \le \eta, \dots, X_{j-k+1} \le \xi, Y_{j-k+1} \le \eta\}\,, \tag{20}$$

$$G_{kj}(\xi,\eta) = \mathbf{1}\{X_j > \xi, Y_j > \eta, X_{j-1} \le \xi, Y_{j-1} \le \eta, \dots, X_{j-k+1} \le \xi, Y_{j-k+1} \le \eta\}\,, \tag{21}$$

where $\mathbf{1}\{\mathcal{A}\}$ denotes the indicator function of some event $\mathcal{A}$. Also, let $C_{kj}(\xi,\eta) = \mathbf{1}\{\mathcal{C}_{kj}(\xi,\eta)\}$.

From these definitions it follows that,

$$\alpha_{kj}(\xi;\eta) = \frac{\mathrm{E}[A_{kj}(\xi;\eta)]}{\mathrm{E}[C_{kj}(\xi,\eta)]}\,, \tag{22}$$

$$\beta_{kj}(\eta;\xi) = \frac{\mathrm{E}[B_{kj}(\eta;\xi)]}{\mathrm{E}[C_{kj}(\xi,\eta)]}\,, \tag{23}$$

$$\gamma_{kj}(\xi,\eta) = \frac{\mathrm{E}[G_{kj}(\xi,\eta)]}{\mathrm{E}[C_{kj}(\xi,\eta)]}\,, \tag{24}$$

where $\mathrm{E}[\cdot]$ denotes the expectation operator.

Assuming an ergodic process, then obviously $\alpha_k(\xi;\eta) = \alpha_{kk}(\xi;\eta) = \ldots = \alpha_{kN}(\xi;\eta)$, and it may be assumed that for the bivariate time series at hand

$$\alpha_k(\xi;\eta) = \lim_{N\to\infty} \frac{\sum_{j=k}^{N} a_{kj}(\xi;\eta)}{\sum_{j=k}^{N} c_{jk}(\xi,\eta)}\,. \tag{25}$$

where $a_{kj}(\xi;\eta)$ and $c_{kj}(\xi,\eta)$ are the realized values of $A_{kj}(\xi;\eta)$ and $C_{kj}(\xi,\eta)$, respectively, for the observed time series.

For multiply recorded stationary time series, the sample estimate of $\alpha_k(\xi;\eta)$ would be,

$$\hat{\alpha}_k(\xi;\eta) = \frac{1}{R}\sum_{r=1}^{R} \hat{\alpha}_k^{(r)}(\xi;\eta)\,, \tag{26}$$

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

$$\hat{\alpha}_k^{(r)}(\xi;\eta) = \frac{\sum_{j=k}^N a_{kj}^{(r)}(\xi;\eta)}{\sum_{j=k}^N c_{kj}^{(r)}(\xi,\eta)}\,, \tag{27}$$

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

Clearly, $\lim_{\xi,\eta\to\infty} \mathrm{E}[C_{kj}(\xi,\eta)] = 1$. Hence, $\lim_{\xi,\eta\to\infty} \tilde{\alpha}_k(\xi;\eta)/\ \alpha_k(\xi;\eta) = 1$, where

$$\tilde{\alpha}_k(\xi;\eta) = \lim_{N\to\infty} \frac{\sum_{j=k}^N \mathrm{E}[A_{kj}(\xi;\eta)]}{N-k+1}\,. \tag{28}$$

The advantage of using the modified ACER function $\tilde{\alpha}_k(\xi;\eta)$ for $k \geq 2$ is that it is easier to use for non-stationary or long-term statistics than $\alpha_k(\xi;\eta)$. Since our focus is on the values of the ACER functions at the extreme levels, we may use any function that provides correct predictions of the appropriate ACER function at these extreme levels. Exactly the same kind of arguments apply to the other ACER functions as well, with the obvious changes.

To see why Eq. (28) may be applicable for nonstationary time series, it is recognized that when $\xi,\eta \to \infty$,

$$\begin{aligned} P(\xi,\eta) &\approx \exp\{-(N-k+1)\big(\alpha_k(\xi;\eta) + \beta_k(\eta;\xi) - \gamma_k(\xi,\eta)\big)\} \\ &\approx \exp\{-(N-k+1)\big(\tilde{\alpha}_k(\xi;\eta) + \tilde{\beta}_k(\eta;\xi) - \tilde{\gamma}_k(\xi,\eta)\big)\} \\ &= \exp\Big\{-\sum_{j=k}^N \big(\mathrm{E}[A_{kj}(\xi;\eta)] + \mathrm{E}[B_{kj}(\eta;\xi)] - \mathrm{E}[G_{kj}(\xi,\eta)]\big)\Big\}\,. \end{aligned} \tag{29}$$

If the nonstationary time series can be segmented into K blocks such that $\mathrm{E}[A_{kj}(\xi;\eta)]$, $\mathrm{E}[B_{kj}(\eta;\xi)]$ and $\mathrm{E}[G_{kj}(\xi,\eta)]$ remain approximately constant within each block and such that $\sum_{j\in C_i} \mathrm{E}[A_{kj}(\xi;\eta)] \approx \sum_{j\in C_i} a_{kj}(\xi;\eta)$, $\sum_{j\in C_i} \mathrm{E}[B_{kj}(\eta;\xi)] \approx \sum_{j\in C_i} b_{kj}(\eta;\xi)$ and $\sum_{j\in C_i} \mathrm{E}[G_{kj}(\xi,\eta)] \approx \sum_{j\in C_i} g_{kj}(\xi,\eta)$ for a sufficient range of ξ,η-values, where C_i denotes the set of indices for block no. i, $i = 1,\ldots,K$, then $\sum_{j=k}^N \mathrm{E}[A_{kj}(\xi;\eta)] \approx \sum_{j=k}^N a_{kj}(\xi;\eta)$. Hence, for a nonstationary bivariate time series it is obtained that $(\xi,\eta \to \infty)$,

$$P(\xi,\eta) \approx \exp\{-(N-k+1)\big(\hat{\alpha}_k(\xi;\eta) + \hat{\beta}_k(\eta;\xi) - \hat{\gamma}_k(\xi,\eta)\big)\}\,, \tag{30}$$

where

$$\hat{\alpha}_k(\xi;\eta) = \frac{1}{N-k+1}\sum_{j=k}^N a_{kj}(\xi;\eta)\,, \tag{31}$$

$$\hat{\beta}_k(\eta;\xi) = \frac{1}{N-k+1}\sum_{j=k}^N b_{kj}(\eta;\xi)\,, \tag{32}$$

$$\hat{\gamma}_k(\xi,\eta) = \frac{1}{N-k+1}\sum_{j=k}^N g_{kj}(\xi,\eta)\,. \tag{33}$$

Now, let us look at the problem of estimating confidence intervals for the ACER functions. The sample standard deviation $\hat{s}_{\alpha,k}(\xi;\eta)$ can be estimated by the standard formula,

$$\hat{s}_{\alpha,k}(\xi;\eta)^2 = \frac{1}{R-1}\sum_{r=1}^{R}\Big(\hat{\alpha}_k^{(r)}(\xi;\eta) - \hat{\alpha}_k(\xi;\eta)\Big)^2 . \tag{34}$$

Assuming that realizations are independent, for a suitable number R, e.g. $R \geq 20$, Eq. (27) leads to a good approximation of the 95 % confidence interval CI $= \big(\mathrm{CI}^-(\eta), \mathrm{CI}^+(\eta)\big)$ for the value $\alpha_k(\xi;\eta)$, where

$$\mathrm{CI}^{\pm}(\eta) = \hat{\alpha}_k(\xi;\eta) \pm 1.96\, \hat{s}_{\alpha,k}(\xi;\eta)/\sqrt{R}\,. \tag{35}$$

An entirely similar procedure is adopted for the other ACER functions.

4 Concluding remarks

In this paper the bivariate ACER method has been described. It is based on an extension of the univariate concept of average conditional exceedance rate (ACER) to the case of bivariate data sets. It has been demonstrated that this leads to a cascade of conditioning approximations to the exact bivariate extreme value distribution. When this cascade has converged, 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 has to be adopted. This problem has not been pursued in this paper. It will be the focus of future research.

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

[2] 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.

[3] A. Naess and O. Gaidai. Estimation of extreme values from sampled time series. *Structural Safety*, 31:325–334, 2009.