On estimating extreme tail probabilities of the integral of a stochastic process *

Ana Ferreira
Instituto Superior de Agronomia, UTL and CEAUL

Laurens de Haan
University of Tilburg, Erasmus University Rotterdam and CEAUL

Chen Zhou
De Nederlandsche Bank and Erasmus University Rotterdam

Spatial Extremes, Theory and Applications, Lisbon, 6-8 April, 2009

*Research partially supported by FCT, Project PTDC/MAT/64924/2006
One-dimensional situation

$X$ is a r.v. with d.f. $F_X$.

**Failure probability:**

$$p_n = P(X > q_n),$$

with $q_n \to \sup\{x : F_X(x) < 1\}$ or $p_n \to 0$ ($n \to \infty$).

**EV condition:** $X_1, \ldots, X_n$ i.i.d. If there exists $a_n > 0$ and $b_n \in \mathbb{R}$ such that $\max_{1 \leq i \leq n} (X_i - b_n)/a_n$ converges, in distribution, to some r.v., then this r.v. is $GEV_\gamma$ ($\gamma \in \mathbb{R}$) distributed.

Equivalently,

$$\lim_{t \to \infty} t P\left(\frac{X - b(t)}{a(t)} > x\right) = (1+\gamma x)^{-1/\gamma}, \quad 1+\gamma x > 0, \quad \gamma \in \mathbb{R}.$$ 

**Estimation:** on the basis of an i.i.d. sample $X_1, \ldots, X_n$,

$$\hat{p}_n = \frac{k}{n} \left(1 + \hat{\gamma} \frac{q_n - \hat{b}(\frac{n}{k})}{\hat{a}(\frac{n}{k})}\right)^{-1/\hat{\gamma}},$$

with $\hat{\gamma}$, $\hat{a}(\frac{n}{k})$ and $\hat{b}(\frac{n}{k})$ appropriate estimators.

Theoretically $k$ must be an intermediate sequence: $k = k(n) \to \infty$, $n/k \to 0$, as $n \to \infty$. 

2
**Infinite-dimensional situation**

$X := \{X(s)\}_{s \in S}$ is an a.s. continuous stoch. proc., in some compact subset $S$ of $\mathbb{R}^2$, with non-degenerate marginals.

**Failure probability:**

$$p_n := P\left(\int_S X(s) ds > q_n\right)$$

with $q_n \to \sup\{x : F_\int_S X(s) ds (x) < 1\}$ or $p_n \to 0$ ($n \to \infty$).

**Some EVT that we need:**

Let $C(S)$ be the space of continuous functions on $S$, equipped with the supremum norm, $|f|_\infty = \sup_{s \in S} |f(s)|$. The stoch. proc. $X$ is assumed to be on $C(S)$.

$X_1, X_2, \ldots$, i.i.d. copies of $X$. Suppose there are continuous functions $a_s(n) > 0$ and $b_s(n) \in R$, such that, for some limiting stoch. proc. $Y := \{Y(s)\}_{s \in S}$,

$$\left\{\max_{1 \leq i \leq n} \frac{X_i(s) - b_s(n)}{a_s(n)}\right\}_{s \in S} \overset{d}{\to} \{Y(s)\}_{s \in S},$$

in $C(S)$ ($n \to \infty$).
Consequences of (2):

A. Convergence of the marginals. W.l.o.g., $a_s(n)$ and $b_s(n)$ are chosen such that

$$P(Y(s) \leq x) = \exp \left( - (1 + \gamma(s)x)^{-1/\gamma(s)} \right),$$

$1 + \gamma(s)x > 0$, $\gamma(s) \in \mathbb{R}$, for all $s \in S$.

B. The index function, $\gamma(s)$, is continuous and real.

C. There exists the exponent measure, $\nu$, on the space

$$C^+(S) := \{ f \in C(S) : f \geq 0 \},$$

such that for each Borel subset $E$ of $C^+(S)$ with $\inf\{|f|_\infty : f \in E\} > 0$ and $\nu(\partial E) = 0$,

$$\lim_{t \to \infty} tP \left( \left\{ \left( 1 + \gamma(s) \frac{X(s) - b_s(t)}{a_s(t)} \right)^{1/\gamma(s)} \right\}_{s \in S} \in E \right) = \nu(E) \quad (3)$$

is finite.

Property of the exponent measure:

$$\nu(cE) = c^{-1}\nu(E), \quad \text{for all } c > 0.$$
D. There exists the spectral measure, $\rho$, finite on

$$\bar{C}_1^+(S) := \{ g \in C^+(S) : |g|_\infty = 1 \}$$

such that

$$\nu(E) = \int \int_{rg \in E} \frac{dr}{r^2} d\rho(g) \quad (4)$$

and satisfying the side conditions

$$\int_{\bar{C}_1^+(S)} g(s) d\rho(g) = 1, \quad \text{for all } s \in S. \quad (5)$$

**Note:** We can always consider the transformation:

$$f \in C^+(S) \rightarrow (|f|_\infty, f/|f|_\infty), \quad \text{i.e. } C^+(S) = (0, \infty) \times \bar{C}_1^+(S).$$

For $B_{r,A} := (r, \infty) \times A$, with $r > 0$ and $A$ a Borel set of $\bar{C}_1^+(S)$, we have $B_{r,A} = rB_{1,A}$ and

$$\nu(B_{r,A}) = r^{-1}\nu(B_{1,A}) = r^{-1}\rho(A).$$
**Theorem 1.** Under (2) with \( a_s(t) \) such that
\[
\sup_{s \in S} \left| \frac{a_s(t)}{a(t)} - A(s) \right| \to 0, \quad \text{as} \quad t \to \infty, \tag{6}
\]
for some functions \( a(t) > 0 \) and \( A(s) \geq 0 \), and with
\[
\rho\{ g \in \bar{C}_{1}^{+}(S) : \inf_{s \in S} g(s) = 0 \} = 0, \tag{7}
\]
we have,
\[
\lim_{t \to \infty} tP \left( \frac{\int_{S} X(s) ds - \int_{S} b_{s}(t) ds}{a(t)} > x \right) = \theta_{\gamma}(1 + \gamma x)^{-1/\gamma}, \tag{8}
\]
where \( x > 0 \), and
\[
\theta_{\gamma} := \int_{\bar{C}_{1}^{+}(S)} \left( \int_{S} A(s) g^{\gamma}(s) ds \right)^{1/\gamma} d\rho(g); \tag{9}
\]
for \( \gamma = 0 \) the right-hand side of (8) should be read as \( \theta_{0}e^{-x} \) and the right-hand side of (9) as
\[
\int_{\bar{C}_{1}^{+}(S)} \exp \left( \int_{S} A(s) \log g(s) ds \right) d\rho(g).
\]
Note: The right-hand side of (9) is continuous in \( \gamma \).
In Coles and Tawn (1996), $\theta_\gamma$ was named the *areal coefficient* and interpreted as the effect of spacial dependence.

**Proposition 1.** *Under the conditions of Theorem 1,*

1. $0 < \theta_\gamma \leq 1$, $\gamma \leq 1$,

2. $1 \leq \theta_\gamma \leq \rho \left( \bar{C}_1^+(S) \right)$, $\gamma \geq 1$.

**Remark:** Define for $p \in \mathbb{R}$ and $g \in \{ f \in C(S) : f > 0, |f|_\infty = 1 \}$,

$$L_p(g) := \begin{cases} 
\left( \int_S g^p(s) A(s) ds \right)^{1/p}, & p \neq 0 \\
\exp \left( \int_S (\log g(s)) A(s) ds \right), & p = 0 
\end{cases} \quad (10)$$

where $A > 0$ satisfies $\int_S A(s) ds = 1$.

**Proposition 2.** 1. $L_p(g)$ is continuous and non-decreasing in $g$ for all $p$.

2. $L_p(g)$ is continuous and non-decreasing in $p$ for all $g$.  

7
Estimation of $p_n$: on the basis of an i.i.d. sample $X_1, \ldots, X_n$,

$$\hat{p}_n = \frac{k}{n} \tilde{\theta} \left( 1 + \tilde{\gamma} q_n - \int_S \hat{b}_s(\frac{n}{k}) ds \right)^{-1/\tilde{\gamma}}$$

(11)

and

$$\hat{\theta} = \int_{\tilde{C}_1(S)} \left( \int_S \hat{A}(s) \hat{\gamma}(s) ds \right)^{1/\tilde{\gamma}} d\hat{\rho}(g).$$

(12)

We use,

$$\tilde{\gamma} := \int_S \tilde{\gamma}(s) ds / |S| = \left( \int_S \tilde{\gamma}_+(s) ds + \int_S \tilde{\gamma}_-(s) ds \right) / |S|$$


$$\tilde{\gamma}_+(s) = M_n^{(1)}(s), \quad \tilde{\gamma}_-(s) = 1 - \frac{1}{2} \left\{ 1 - \left( \frac{M_n^{(1)}(s)}{M_n^{(2)}(s)} \right)^2 \right\}^{-1}$$

and ($j = 1, 2$)

$$M_n^{(j)}(s) = k^{-1} \sum_{i=0}^{k-1} (\log X_{n-i,n}(s) - \log X_{n-k,n}(s))^j.$$
Also,

\[ \hat{b}_s \left( \frac{n}{k} \right) = X_{n-k,n}(s) \]

\[ \hat{a}_s \left( \frac{n}{k} \right) = X_{n-k,n}(s) \hat{\gamma}_+(s) (1 - \hat{\gamma}_-(s)) \]

We take

\[ \hat{a} \left( \frac{n}{k} \right) = \int_S \hat{a}_s \left( \frac{n}{k} \right) ds \]

\[ \hat{A}(s) = \frac{\hat{a}_s \left( \frac{n}{k} \right)}{\int_S \hat{a}_s \left( \frac{n}{k} \right) ds} \]
It is known that,
\[ \sqrt{k} \left( \frac{\hat{a}(\frac{n}{k})}{a(\frac{n}{k})} - 1 \right) = O_p(1) \]
and
\[ \sqrt{k} \frac{\hat{b}(\frac{n}{k}) - b(\frac{n}{k})}{a(\frac{n}{k})} = O_p(1), \]
uniformly in \( s \in S \), as \( n \to \infty \).

Then, since \( a(\frac{n}{k}) = \int_S a_s(\frac{n}{k}) \, ds \),
\[ \sqrt{k} \left( \frac{\hat{a}(\frac{n}{k})}{a(\frac{n}{k})} - 1 \right) = \int_S \sqrt{k} \left( \frac{\hat{a}_s(\frac{n}{k})}{a_s(\frac{n}{k})} - 1 \right) \frac{a_s(\frac{n}{k})}{a(\frac{n}{k})} \, ds = O_p(1) \]
and
\[ \sqrt{k} \left[ \int_S \hat{b}_s(\frac{n}{k}) \, ds - \int_S b_s(\frac{n}{k}) \, ds \right] \frac{a(\frac{n}{k})}{a(\frac{n}{k})} = \int_S \sqrt{k} \frac{\hat{b}_s(\frac{n}{k}) - b_s(\frac{n}{k})}{a_s(\frac{n}{k})} \frac{a_s(\frac{n}{k})}{a(\frac{n}{k})} \, ds = O_p(1). \]

Moreover, uniformly in \( s \),
\[ \sqrt{k} (\hat{A}(s) - A(s)) = O_P(1). \]
Estimation of the spectral measure $\rho$:

Condition (2) implies, with $\xi(s) := 1 / (1 - F_s(X(s)))$,

$$\lim_{t \to \infty} tP(t^{-1} \xi \in E) = \nu(E),$$

for every Borel set $E$ of $C^+(S)$ such that $\inf \{|f|_\infty : f \in E\} > 0$ and $\nu(\partial E) = 0$.

Hence, with $B_{1,A} := (1, \infty) \times A$, with $A$ a Borel set of $\bar{C}_1^+(S)$, if $\rho(\partial A) = 0$,

$$\lim_{t \to \infty} tP(t^{-1} \xi \in B_{1,A}) = \rho(A).$$

(13)

Therefore,

$$\hat{\rho}(A) := \frac{n}{k} \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{\left\{ \frac{1}{n} |\hat{\xi}_i|_\infty > 1 \text{ and } \{\hat{\xi}_i(s)/|\hat{\xi}_i|_\infty\}_{s \in S} \in A \right\}} =$$

$$\frac{1}{k} \sum_{i=1}^{n} \mathbf{1}_{\left\{ \sup_{s \in S} R(X_i(s)) > n+1-k \text{ and } \left\{ \frac{s+1-\sup_{s \in S} R(X_i(s))}{s+1-R(X_i(s))} \right\}_{s \in S} \in A \right\}},$$

with

$$\hat{\xi}_i(s) := \frac{n}{(n+1 - R(X_i(s)))}, \quad s \in S,$$

$R(X_i(s))$ the rank of $X_i(s)$ among $(X_1(s), \ldots, X_n(s))$.

**Theorem 2.** Under (2), $\hat{\rho} \rightarrow^P \rho$, in the space of finite measures on $\bar{C}_1^+(S)$, with $k = k(n) \to \infty$, $k/n \to 0$, as $n \to \infty$. 

11
Consistency of $\hat{p}_n$:

**Theorem 3.**

$$\frac{\hat{p}_n}{p_n} \xrightarrow{P} 1, \quad n \to \infty;$$

**If:**

- **the basic first order condition (2) with** $\gamma(s) \equiv \gamma > -1/2$.

- **second order condition:**

  *There exists a function $\alpha(t)$, positive or negative with $\alpha(\cdot)$ regularly varying of index $\tilde{\rho} \leq 0$, or $\tilde{\rho} = 0$ if $\gamma < 0$, and $\lim_{t \to \infty} \alpha(t) = 0$ such that

\[
\lim_{t \to \infty} \frac{U(tx) - \int_0^x b_s(t) \, ds}{a(t)} \rightarrow \frac{(\theta, x)^{\gamma - 1}}{\gamma},
\]

exists for $x > 0$, with $U$ the inverse function of $1/P \left( \int_S X(s) \, ds > x \right)$; moreover,
\[
\sqrt{k} \left( \hat{\gamma} - \gamma, \frac{\hat{a}(\frac{n}{k})}{a(\frac{n}{k})} - 1, \frac{\int_S \hat{b}_s(\frac{n}{k}) ds - \int_S b_s(\frac{n}{k}) ds}{a(\frac{n}{k})} \right) \\
= (O_p(1), O_p(1), O_p(1)), \quad (14)
\]

with \( k = k(n) \rightarrow \infty, k/n \rightarrow 0, \)

- \( d_n = k/(np_n) \rightarrow \infty, \)
- \( w_\gamma(d_n)/\sqrt{k} \rightarrow 0 \) with \( w_\gamma(t) := t^{-\gamma} \int_1^t s^{\gamma-1} \log s ds, \)
- \( \sqrt{k} \alpha(\frac{n}{k}) \rightarrow \lambda, \) finite,

and \( \hat{\rho} \rightarrow^P \rho \) in the space of finite measures on \( \bar{C}_{1+}^+(S). \)
Application

Evaluate extreme rainfall in a low-lying flat area in the northwest of the Netherlands (North Holland).

We have daily rainfall data at 32 monitoring stations, over the 30-year period 1971-2000.

What is the amount of rain on one day that is exceeded once in 100 year (i.e. the 100-year quantile of the total rainfall in this area)?

Sample size = $91 \times 30$ years = 2730 observations
References


