

# On the extrapolation limits of extreme-value theory for risk management

Clément Albert, Anne Dutfoy, Stéphane Girard

► **To cite this version:**

Clément Albert, Anne Dutfoy, Stéphane Girard. On the extrapolation limits of extreme-value theory for risk management. MMR 2017 - 10th International Conference on Mathematical Methods in Reliability, Jul 2017, Grenoble, France. 10th International Conference on Mathematical Methods in Reliability, pp.5. <hal-01571099>

**HAL Id: hal-01571099**

**<https://hal.archives-ouvertes.fr/hal-01571099>**

Submitted on 1 Aug 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# ON THE EXTRAPOLATION LIMITS OF EXTREME-VALUE THEORY FOR RISK MANAGEMENT.

Clément ALBERT<sup>1</sup>, Anne DUTFOY<sup>2</sup> & Stéphane GIRARD<sup>1</sup>

<sup>1</sup> *Inria Grenoble Rhône-Alpes & Laboratoire Jean Kuntzmann, 655,  
avenue de l'Europe, Montbonnot, 38334 Saint-Ismier cedex, France.*

<sup>2</sup> *EDF R&D dpartement Management des Risques Industriels, 7 boulevard  
Gaspard Monge - 91120 Palaiseau, France.  
clement.albert@inria.fr, anne.dutfoy@edf.fr, stephane.girard@inria.fr*

**Abstract.** In the risk management context, the extreme-value methodology consists in estimating extreme quantiles - one hundred years return period or more - from an extreme-value distribution adjusted on data. In this communication, we quantify the extrapolation limits associated with extreme quantile estimations. To this end, we focus on the framework of the block maxima method and we study the behaviour of the relative approximation error of a quantile estimator dedicated to the Gumbel attraction domain. We give necessary and sufficient conditions for the error to converge towards zero and we provide a first order approximation of the latter. We show that extrapolations can be greatly limited depending on the data distribution.

**Keywords.** Extreme-value theory, extreme quantile estimation, asymptotic properties, environmental risks.

## 1. Introduction

The EDF R&D department makes use of the extreme-value theory to perform many statistical studies of extreme events based on meteorological series (temperature, flow, wind speed, ...). These studies are used to size structures against meteorological events such as flood, storm or drought. They consist, given an extreme-value distribution fitted on data, in estimating extreme quantiles of one hundred or more return period.

But what credibility should be attributed to these extrapolations ? Even if several works offer empirical rules for extrapolation limits (see, e.g., [1]), we introduce new mathematical rules based on the convergence analysis towards extreme-value distributions.

In this work, we focus on data from a distribution in the Gumbel maximum domain of attraction ( $F \in DA(Gumbel)$ ).

First, we provide necessary and sufficient conditions for the relative consistency of the estimator of an extreme quantile in the Gumbel maximum domain of attraction. Our second result establishes a first order equivalent of the asymptotic approximation error associated with the previous estimator.

## 2. Extreme quantiles estimation using the block maxima method.

In this study, we consider quantiles from an unknown distribution  $F \in DA(Gumbel)$ . Given a  $n$ -sample, an extreme quantile is a  $(1-p)$ th quantile  $x_p$  of  $F$ , with  $p \rightarrow 0$  as  $n \rightarrow +\infty$ .

The block maxima approach in extreme-value theory, consists in dividing the observation period into nonoverlapping periods of equal size and restricts attention to the maximum observation in each period (see, e.g., [3]). The pseudo observations thus created follow (under domain of attraction conditions) an extreme-value distribution  $G$ . Parametric statistical methods for the extreme-value distributions are then applied to those observations. This method relies on the extreme-value theorem.

**Extreme Value Theorem (see, e.g., [2]):** Let  $X_1, X_2, \dots, X_m$  be independent random variables from a distribution  $F \in DA(Gumbel)$  and  $X_{m,m} = \max(X_1, X_2, \dots, X_m)$ . There exist sequences  $\{a_m > 0\}$  and  $\{b_m\}$  such as :

$$\mathbb{P}\left(\frac{X_{m,m} - b_m}{a_m} \leq x\right) \rightarrow G(x), \quad m \rightarrow \infty,$$

with

$$G(x) = \exp\{-\exp(-x)\}, \quad x \in \mathbb{R}. \quad (1)$$

Interpreting the limit in this Theorem as an approximation for large values of  $m$  suggests the use of the Gumbel distribution for modeling the behaviour of maxima from long sequences.

An extreme quantile approximation  $x_{p_m}$  of  $F$  is then obtained by using the fact that  $\mathbb{P}(X_{m,m} < x) = F^m(x) \approx G\left(\frac{x - b_m}{a_m}\right)$  together with a first order approximation of the log function by inverting equation (1) :

$$\tilde{x}_{p_m} = b_m - a_m \log(mp_m).$$

In the following, we focus on the relative approximation error  $\epsilon_{app_m}$  of  $x_{p_m}$  by  $\tilde{x}_{p_m}$  defined as

$$\epsilon_{app_m} = \frac{x_{p_m} - \tilde{x}_{p_m}}{x_{p_m}}.$$

### 3. Asymptotic results

Let us define the cumulative hazard rate function  $H(x) = -\log(1 - F(x))$  and suppose  $H$  is a strictly increasing and twice differentiable function. Let us also consider  $p_m$  such that  $\frac{\log(1/p_m)}{\log m} = \tau_m$ , with  $\tau_m \xrightarrow{>} \tau \geq 1$ , or equivalently  $p_m = m^{-\tau_m}$ .

Let us define  $K_i(x) = \frac{x^i (H^{-1})^{(i)}(x)}{H^{-1}(x)}$ ,  $i \in \{1, 2\}$  and the associated limits  $l_i = \lim_{x \rightarrow +\infty} K_i(x)$ .

Our first result provides a necessary and sufficient condition such that the relative approximation error tends to zero as  $m$  tends to infinity. This condition links the tail distribution behaviour (via  $K_2$ ) with the rate of convergence towards zero of  $p_m$  (via the  $\tau_m - 1$  difference).

**Theorem 1:** Suppose  $K_1$  is regularly varying with index  $\theta_1$  ( $K_1 \in RV_{\theta_1}$ ),  $\theta_1 \leq 1$ ,  $K_2 \in RV_{\theta_2}$ ,  $\theta_2 \leq 2$  and  $\tau \geq 1$ . Then :

$$\epsilon_{app_m} \xrightarrow{m \rightarrow +\infty} 0 \iff (\tau_m - 1)^2 K_2(\log m) \xrightarrow{m \rightarrow +\infty} 0.$$

The regularly varying hypothesis  $K_1 \in RV_{\theta_1}$  has been first introduced by [4]. In this paper, one can also find a discussion on the distribution types satisfying this hypothesis.

Let us introduce

$$\Psi(t) := \int_0^1 x e^{-tx} dx \text{ et } g(\tau, \theta) := \frac{\tau^\theta - \theta(\tau - 1) - 1}{\theta(\theta - 1)(\tau - 1)^2}, \quad \theta \notin \{0, 1\}, \tau > 1, t \geq 0.$$

Our second result exhibits a first order equivalent associated with the relative approximation error.

**Theorem 2:** Under the assumptions of Theorem 1, we have, as  $m \rightarrow +\infty$  :

(1) if  $l_1 \in \{0, 1\}$  then  $K_2(\log m) \rightarrow 0$ ,  $\theta_2 \leq 0$  and

$$\epsilon_{app_m} \sim (\tau_m - 1)^2 K_2(\log m) \tau^{-l_1} g(\tau, \theta_2 + l_1) \rightarrow 0, \forall \tau \geq 1.$$

(2) if  $l_1 \in [0, +\infty[$ ,  $l_1 \notin \{0, 1\}$ , then  $K_2(\log m) \rightarrow l_1(l_1 - 1)$ ,  $\theta_2 = 0$  and

$$\epsilon_{app_m} \sim (\tau_m - 1)^2 l_1(l_1 - 1) \tau^{-l_1} g(\tau, l_1) \rightarrow 0 \quad \text{if and only if } \tau = 1.$$

(3) if  $l_1 = +\infty$  then  $K_2(\log m) \rightarrow +\infty$  and

$$\begin{aligned} \epsilon_{app_m} &\sim (\tau_m - 1)^2 K_2(\log m) \Psi \left( (\tau_m - 1) \sqrt{K_2(\log m)} \right) \\ &\rightarrow 0 \quad \text{if and only if } (\tau_m - 1)^2 K_2(\log m) \rightarrow 0. \end{aligned}$$

This second Theorem allows to distinguish between three families of distributions. First, distributions verifying  $l_1 \in \{0, 1\}$ , for which there is convergence towards zero of the error without any condition being required on  $\tau$ . There is no extrapolation limits in this case. Such distributions include for example the Exponential and Gamma distribution.

Second, one has distributions verifying  $l_1 \in [0, +\infty[$ ,  $l_1 \notin \{0, 1\}$ . For these distributions, convergence to zero requires  $\tau = 1$ . In such a case, extrapolation is limited to extreme quantiles of order such that  $\log(p_m) \sim \log(1/m)$ , i.e. extreme quantiles not very far from the sample maximum. Such distributions include Weibull and Gaussian distribution.

Finally the most restrictive case is the case where  $l_1 = +\infty$ . In the latter case,  $\tau$  must be equal to 1, but an additional condition is required on the convergence rate of  $\tau$  towards 1. Indeed,  $(\tau_m - 1)^2 K_2(\log m) \rightarrow 0$  implies  $\tau = 1$ . If this condition holds, then  $\epsilon_{app_m} \sim \frac{1}{2} (\tau_m - 1)^2 K_2(\log m)$ . The extrapolation is even more limited than in the previous case, this phenomena occurs for instance with the Lognormal distribution.

#### 4. Numerical illustrations

Previous Theorems allow us to compute the maximum return period  $T_{max}$  up to which one can extrapolate. The maximum return period can be interpreted as follows :

$$T_{max} \approx \frac{1}{mp_{min}},$$

where  $p_{min}$  is the minimum order of an extreme quantile which can be estimated with a relative error smaller than 10%, i.e.

$$p_{min} = \underset{p}{\operatorname{argmin}} \left\{ \left| \frac{x_p - \tilde{x}_p}{x_p} \right| < 0.1 \right\}.$$

The maximum return period  $T_{max}$  in years is displayed in Table 1 for some distributions in the Gumbel maximum domain of attraction.

	$m = 60$	$m = 90$	$m = 120$
<b>Exponential</b>	no limit	no limit	no limit
<b>Gamma(shape = 0.5)</b>	$> 1e + 15$	$> 1e + 15$	$> 1e + 15$
<b>Gaussian</b>	75	126	183
<b>Weibull(shape = 1.5)</b>	1041	2071	3375
<b>Lognormal</b>	7	7	8

This shows that extrapolation is dependent on the data distribution. Extrapolation can be either greatly limited in the Lognormal case for example, or not at all in the Exponential case.

Other numerical illustrations will be provided during the talk.

## References

- [1] MeteoFrance, Climatology department (2016). Extreme rainfall return period. <http://pluiesextremes.meteo.fr/france-metropole/Les-durees-de-retour.html>.
- [2] De Haan, L., & Ferreira, A. (2007). *Extreme value theory: an introduction*. Springer Science & Business Media.
- [3] Ferreira, A., & de Haan, L. (2015). On the block maxima method in extreme value theory: PWM estimators. *The Annals of statistics*, 43(1), 276-298.
- [4] de Valk, C. (2016). Approximation of high quantiles from intermediate quantiles. *Extremes*, 19(4), 661-686.