On the approximation of extreme quantiles with neural networks

Michaël Allouche [1] , Stéphane Girard [2] & Emmanuel Gobet [3]

[1] *Centre de Mathématiques Appliquées (CMAP), Ecole Polytechnique and CNRS, Université Paris-Saclay, Route de Saclay, 91128 Palaiseau Cedex, France; michael.allouche@polytechnique.edu.*
[2] *Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, 38000 Grenoble, France; stephane.girard@inria.fr.*
[3] *Centre de Mathématiques Appliquées (CMAP), Ecole Polytechnique and CNRS, Université Paris-Saclay, Route de Saclay, 91128 Palaiseau Cedex, France; emmanuel.gobet@polytechnique.edu.*

Résumé. Dans cette étude nous proposons une nouvelle paramétrisation du générateur d'un réseau antagoniste génératif (GAN) adaptée aux données issues d'une distribution à queue lourde. Nous apportons une analyse de l'erreur d'approximation en norme uniforme d'un quantile extrême par le GAN ainsi construit. Des simulations numériques sont réalisées sur des données réelles et simulées.
Mots-clés. Théorie des valeurs extrêmes, réseau de neurones, modèle génératif

Abstract. In this study, we propose a new parametrization for the generator of a Generative adversarial network (GAN) adapted to data from heavy-tailed distributions. We provide an analysis of the uniform error between an extreme quantile and its GAN approximation. Numerical experiments are conducted both on real and simulated data.
Keywords. Extreme value theory, neural networks, generative models

1 Introduction

In this paper we are interested in approximating the quantile function q_X defined by $q_X(u) := F_X^{\leftarrow}(u) = \inf\{x : F_X(x) \geq u\}$, for all level of quantiles $u \in [0,1)$, where F_X is an unknown cumulative distribution function on $\mathcal{X} \subseteq \mathbb{R}$. Clearly, extreme quantiles are observed as $u \to 1$ and will be our region of interest. The objective is, starting from an i.i.d. sample set $\{X_i \in \mathcal{X}\}_{i=1}^n$, to build a generator $G_\theta : \mathcal{Z} \to \mathcal{X}$ from a parametric family of functions $\mathcal{G} = \{G_\theta\}_{\theta\in\Theta}$ and mapping a random variable $Z : \mathcal{Z} \to \mathbb{R}^d$ with known density p_Z to the data support $\mathcal{X}$. In this study, we shall consider $Z \sim \mathcal{U}(0, \mathbf{I}_d)$ and denote by p_θ the density associated with $G_\theta(Z)$. Then, for each p_θ, $\theta \in \Theta$, is a potential candidate to approximate $p_X = F_X'$. In a neural network architecture, this setting is related to Generative Adversarial Networks (GAN) [7].

Let X be a random variable associated with F_X supposed to be continuous and strictly increasing. We focus on the case of heavy-tailed distributions, *i.e.* when F_X is attracted to the maximum domain of Pareto-type distributions with tail-index $\gamma > 0$. From [2], the survival function $\bar{F}_X := 1 - F_X$ of such a heavy-tailed distribution can be expressed as

($\mathbf{H_1}$): $\bar{F}_X(x) = x^{-1/\gamma}\ell_X(x)$, where ℓ_X is a slowly-varying function at infinity *i.e.* such that $\ell_X(\lambda x)/\ell_X(x) \to 1$ as $x \to \infty$ for all $\lambda > 0$.

In such a case, $\bar{F}_X$ is said to be regularly-varying with index $-1/\gamma$ at infinity, which is denoted for short by $\bar{F}_X \in RV_{-1/\gamma}$. The tail-index γ tunes the tail heaviness of the distribution function F_X. Assumption $(\mathbf{H}_1)$ is recurrent in risk assessment, since actuarial and financial data are most of the time heavy-tailed, see for instance the recent studies [1, 3] or the monographs [6, 9]. As a consequence of the above assumptions, the tail quantile function $x \mapsto q_X(1 - 1/x)$ is regularly-varying with index γ at infinity, see [5, Proposition B.1.9.9], or, equivalently,

$$q_X(u) = (1-u)^{-\gamma} L\left(\frac{1}{1-u}\right), \tag{1}$$

for all $u \in (0,1)$ with L a slowly-varying function at infinity. Clearly $q_X(u) \to \infty$ as $u \to 1$ but is pretty smooth elsewhere. This type of function does not seem to be consistent with a neural network approximation framework since the latter mainly consists in making a linear combination of bounded functions, which is very unlikely to approximate diverging functions. This argument is confirmed by the Universal Approximation Theorem [4] stating that a one hidden layer neural network can approximate any continuous function on a **compact set**.

2 Contribution

In order to build a tail-index function which may be well approximated by a neural network, the quantile function (1) is rewritten in a logarithmic scale and normalized to avoid exploding issues at the boundaries. Without of loss of generality, one can assume that $\eta := \mathbb{P}(X \geq 1) \neq 0$ and, since, we focus on the upper tail behavior of X, introduce the random variable $Y = X$ given $X \geq 1$. It follows that the quantile function of Y is given by $q_Y(u) = q_X(1-(1-u)\eta)$, for all $u \in (0,1)$. Thus, we define the Tail-Index function (TIF) as

$$f^{\mathrm{TIF}}(u) := -\frac{\log q_X\big(1-(1-u)\eta\big)}{\log(1-u^2) - \log 2},$$

for all $u \in (0,1)$. The main contribution of this work is to combine the TIF analysis based on Extreme Value Theory with GANs in order to address the general issue of neural network approximation for extremes. Let $\varphi^{\mathrm{TIF}}:\mathbb{R}^2\to\mathbb{R}$ be the non-linear TIF transformation with

$$\varphi^{\mathrm{TIF}}(x,u):=\left(\frac{1-u^2}{2}\right)^{-x}.$$

In addition, let $\mathbf{e}\in\mathbb{R}^6$ be a vector of functions mapping from $[0,1]$ to $\mathbb{R}$ and $\alpha\in\mathbb{R}^6$ be a vector of parameters. Thus, we define the **Tail-GAN** with a generator $G_\psi^{\mathrm{TIF}}:\mathbb{R}^d\to\mathbb{R}$ where the j-th output component is defined as

$$G_\psi^{\mathrm{TIF(j)}}(Z)=\varphi^{\mathrm{TIF}}\left(G_\psi^{(j)}(Z),Z^{(j)}\right),$$

and $\psi=(\theta,\alpha)$ with

$$G_\psi^{(j)}(Z)=G_\theta^{(j)}(Z)+\left\langle \boldsymbol{e}\left(Z^{(j)}\right),\,\alpha\right\rangle.$$

The selection of functions in $\boldsymbol{e}$ and optimal parameter α is based on the following approximation results dealing with the regularity properties of f^{TIF} and the construction of its regularized extension.

3 Approximation results

Our first result describes the behaviour of f^{TIF} in the neighborhood of 0 and 1.

Proposition 1 *Under* $(\mathbf{H}_1)$*,* f^{TIF} *is a continuous and bounded function on* $[0,1]$*. Besides,* $f^{\mathrm{TIF}}(0)=0$ *and* $f^{\mathrm{TIF}}(u)\to\gamma$ *as* $u\to 1$*.*

Focusing on the behavior of the first derivative of the TIF, extra assumptions on F_X, or equivalently on L, are necessary such that f^{TIF} is differentiable. Consider the Karamata representation of the slowly-varying function L [5, Definition B1.6]:

$$L(x)=c(x)\exp\left(\int_1^x\frac{\varepsilon(t)}{t}\,\mathrm{d}t\right),$$

where $c(x)\to c_\infty$ as $x\to\infty$ and ε is a measurable function such that $\varepsilon(x)\to 0$ as $x\to\infty$. Our second main assumption then writes:

$(\mathbf{H_2})$: $c(x)=c_\infty>0$ for all $x\geq 1$ and $\varepsilon(x)=x^\rho\ell(x)$ with $\ell\in RV_0$ and $\rho<0$.

The assumption that c is a constant function is equivalent to assuming that L is normalized [8] and ensures that L is differentiable. The condition $\varepsilon\in RV_\rho$ with $\rho<0$ entails that $L(x)\to L_\infty\in(0,\infty)$ as $x\to\infty$. Besides, $(\mathbf{H}_2)$ entails that F_X satisfies the so-called second-order condition which is the cornerstone of all proofs of asymptotic normality in extreme-value statistics. We shall also consider the assumption:

($\mathbf{H_3}$): ℓ is normalized.

The latter condition ensures that ℓ is differentiable on $(0,1)$ and thus that L and q_X are twice differentiable on $(0,1)$. Our second result provides the behaviour of the first order derivative of f^{TIF} in the neighborhood of 0 and 1.

Proposition 2 *Assume* $(\mathbf{H}_1)$ *and* $(\mathbf{H}_2)$ *hold. Then,* f^{TIF} *is continuously differentiable on* $(0,1)$ *and*

$$\begin{aligned}\partial_u f^{\mathrm{TIF}}(0) &= \frac{\gamma + \varepsilon\left(1/\eta\right)}{\log(2)}, \\ \partial_u f^{\mathrm{TIF}}(u) &\to \infty \text{ as } u \to 1. \end{aligned} \tag{2}$$

It is possible to build regularized version of f^{TIF} by removing the diverging components in the neighborhood of $u = 1$. To this end, consider

$$f^{\mathrm{R}}(u) := f^{\mathrm{TIF}}(u) - \langle \boldsymbol{e}(u),\, \alpha \rangle\,, \tag{3}$$

where $\boldsymbol{e} : \mathbb{R} \to \mathbb{R}^6$ is not described here for the sake of conciseness. Regularity properties of f^{R} are established in the next Proposition.

Proposition 3

(i) If $(\mathbf{H_1})$ *holds, then*

$$\lim_{u\to 0} f^{\mathrm{R}}(u) = \lim_{u\to 1} f^{\mathrm{R}}(u) = 0. \tag{4}$$

(ii) If, moreover, $(\mathbf{H_2})$ *holds with* $\rho < -1$*, then* f^{R} *is continuously differentiable on* $[0,1]$ *and*

$$\lim_{u\to 0} \partial_u f^{\mathrm{R}}(u) = \lim_{u\to 1} \partial_u f^{\mathrm{R}}(u) = 0. \tag{5}$$

(iii) If, moreover, $(\mathbf{H_3})$ *holds with* $\rho < -2$*, then* f^{R} *is twice continuously differentiable on* $[0,1]$.

Given the above regularity properties, it is possible to establish the convergence rate of the uniform error for a one hidden layer neural network depending on the parameter ρ.

Theorem 4 *Let* σ *be a ReLU function. There exists a neural network with* J *neurons and real coefficients* $\{\gamma_j, \lambda_j, b_j\}_{j=1,\ldots,J}$ *such that:*

1. *For* $-2 < \rho < -1$,

$$\sup_{t\in[0,1]} \left| f^{\mathrm{R}}(t) - \sum_{j=1}^{J} \gamma_j \sigma\left(\lambda_j t + b_j\right) \right| = \mathcal{O}\left(J^{\rho}\right),$$

2. for $\rho \leq -2$,

$$\sup_{t\in[0,1]} \left| f^{\mathrm{R}}(t) - \sum_{j=1}^{J} \gamma_j \sigma\left(\lambda_j t + b_j\right) \right| = \mathcal{O}\left(J^{-2}\right).$$

Numerical experiments will be presented in the communication in order to compare the realizations of traditional GANs with our Tail-GAN in two situations: simulated data from heavy-tailed distributions and real financial data from public source.

References

[1] J. Alm. Signs of dependence and heavy tails in non-life insurance data. *Scandinavian Actuarial Journal*, 2016(10):859–875, 2016.

[2] N. H. Bingham, C. M. Goldie, and J. L. Teugels. *Regular variation*, volume 27 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1987.

[3] V. Chavez-Demoulin, P. Embrechts, and S. Sardy. Extreme-quantile tracking for financial time series. *Journal of Econometrics*, 181(1):44–52, 2014.

[4] G. Cybenko. Approximation by superpositions of a sigmoidal function. *Math. Control Signals Systems*, 2(4):303–314, 1989.

[5] L. de Haan and A. Ferreira. *Extreme value theory*. Springer Series in Operations Research and Financial Engineering. Springer, New York, 2006. An introduction.

[6] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling Extremal Events for Insurance and Finance*. Springer-Verlag, Berlin, 1997.

[7] I. Goodfellow, J. Pouget-Abadie, M. Mirza, Bing X., D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[8] E. Kohlbecker. Weak asymptotic properties of partitions. *Transactions of The American Mathematical Society*, 88(2):346–365, 1958.

[9] S. Resnick. *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. Springer, 2007.