

# EV-GAN: Simulation of extreme events with ReLU neural networks

Michaël Allouche   Stéphane Girard   Emmanuel Gobet

EcoSta2022

2022, June 4th

<https://hal.archives-ouvertes.fr/hal-03250663>

*To be published in Journal of Machine Learning Research (JMLR)*



# Motivations

▷ Risk management: **Low probability / High impact**



Lehman Brothers  
2005 - 2008

credit: edelit



credit: latimes

# Motivations

▷ Risk management: **Low probability / High impact**



Lehman Brothers  
2005 - 2008

credit: edelit



credit: latimes

▷ Existing methods (IS, MCMC, IPS) work well but **high complexity cost**

# Motivations

▷ Risk management: **Low probability / High impact**



credit: edelit



credit: latimes

▷ Existing methods (IS, MCMC, IPS) work well but **high complexity cost**

Focus on new paradigm of **data-based** generative models trained on

- **Dataset** (historical financial returns)
- **Black-box** model (meteorological model)

+ **Fast simulation**

- **Few extreme data**

# Generative model

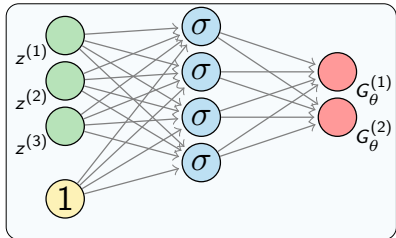
If  $X$  denotes the r.v. taking values in some space  $\mathcal{X} \subseteq \mathbb{R}^d$  from which we have observations  $(X_1, \dots, X_n, \dots)$ , the problem is to find a function  $G : \mathbb{R}^{d'} \rightarrow \mathcal{X}$  and a **latent probability distribution**  $\pi$  on  $\mathbb{R}^{d'}$  such that

$$X \stackrel{d}{=} G(Z) \text{ and } Z \sim \pi$$

**Existence:** see the Kuratowski's Theorem

**Questions:** which  $\pi$  ? which  $d'$  ? how to approximate  $G$  ?

**Learning  $G$ :** Generative Adversarial Networks (GANs) [Goodfellow et al., 2014]



Generator with  $d' = 3$  and  $d = 2$

$$G_\theta^{(m)}(z^{(1)}, \dots, z^{(d')}) = \sum_{j=1}^J a_j^{(m)} \sigma \left( \sum_{i=1}^{d'} w_j^{(i)} z^{(i)} + b_j \right), \quad (1)$$

for  $m = \{1, \dots, d\}$  and

$\sigma(x) = \max(x, 0)$ : **ReLU function**

## Extreme-value theory

Let  $X$  be a r.v. with a continuous and strictly increasing c.d.f.  $F$ , and its associated quantile function  $q(u) := \inf \{x : F(x) \geq u\}$

( $H_1$ ): The distribution is supposed **heavy-tailed**, i.e. the survival function  $\bar{F} := 1 - F$  is **regularly varying** with index  $-1/\gamma$ ,  $\bar{F} \in \mathcal{RV}_{-1/\gamma}$ :

$$\bar{F}(x) = x^{-\frac{1}{\gamma}} \ell(x), \quad x > 0,$$

where

- $\gamma > 0$  is called the **tail-index**; thus  $F$  is said to belong to the Fréchet Maximum domain of attraction,  $F \in \text{MDA}(\text{Fréchet})$
- $\ell$  is a slowly-varying function,  $\ell \in \mathcal{RV}_0$ :

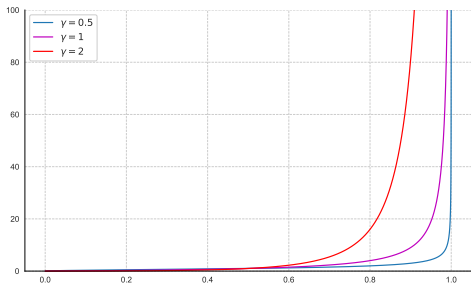
$$\frac{\ell(\lambda x)}{\ell(x)} \rightarrow 1, \quad \text{as } x \rightarrow \infty \text{ for all } \lambda > 0 \quad (2)$$

$\Rightarrow$  As a consequence [de Haan and Ferreira, 2006, Proposition B.1.9.9]

$$t \mapsto q(1 - 1/t) \in \mathcal{RV}_\gamma, \quad t > 1$$

# Problem Statement

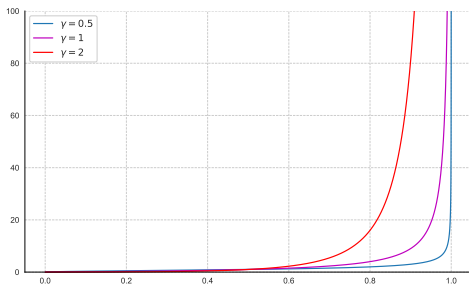
**Goal:** Approximate a quantile function  $q(1 - 1/\cdot) \in \mathcal{RV}_\gamma, \gamma > 0$ , for extreme level  $u \in [0, 1)$  as  $u \rightarrow 1$  with a generative one hidden layer neural network  $G$



Quantile function of a Burr distribution  $u \mapsto q(u)$  with parameters  $\gamma = \{0.5, 1, 2\}$  and  $\rho = -1$

# Problem Statement

**Goal:** Approximate a quantile function  $q(1 - 1/\cdot) \in \mathcal{RV}_\gamma, \gamma > 0$ , for extreme level  $u \in [0, 1)$  as  $u \rightarrow 1$  with a generative one hidden layer neural network  $G$



Quantile function of a Burr distribution  $u \mapsto q(u)$  with parameters  $\gamma = \{0.5, 1, 2\}$  and  $\rho = -1$

## Problem:

▷ The Universal Approximation Theorem [Cybenko, 1989] holds for continuous functions in a **compact set**

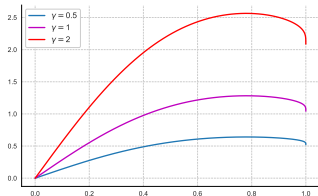
▷ If  $Z$  is either bounded or a Gaussian vector: **by no means**  $G_\theta(Z) \stackrel{d}{=} X$



# Solution

1. Find a Tail-index function (**TIF**)  $f^{\text{TIF}}$  **continuous** and **bounded** on  $[0, 1]$  for all **heavy-tailed** distributions s.t.  $f^{\text{TIF}}(u) \rightarrow \gamma$  as  $u \rightarrow 1$

$$f^{\text{TIF}}(u) = -\frac{\log(q(u))}{\log\left(\frac{1-u^2}{2}\right)}$$

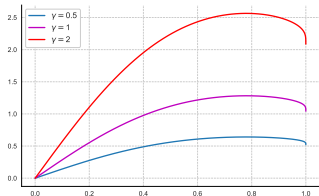


TIF associated with Burr distribution ( $\rho = -1$ )

# Solution

1. Find a Tail-index function (**TIF**)  $f^{\text{TIF}}$  **continuous** and **bounded** on  $[0, 1]$  for all **heavy-tailed** distributions s.t.  $f^{\text{TIF}}(u) \rightarrow \gamma$  as  $u \rightarrow 1$

$$f^{\text{TIF}}(u) = -\frac{\log(q(u))}{\log\left(\frac{1-u^2}{2}\right)}$$

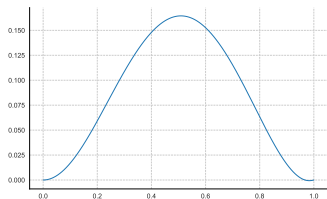


TIF associated with Burr distribution ( $\rho = -1$ )

2. For better approximation, find a Correction TIF (**CTIF**)

$$f^{\text{CTIF}}(u) = f^{\text{TIF}}(u) - \sum_{k=1}^6 \kappa_k e_k(u)$$

which enjoys higher regularity around  $u = 1$



CTIF with  $\gamma = 0.5$  and  $\rho = -3$

## Tail-index function (TIF)

▷ Consider  $d' = d = 1$ . Since  $q(1 - 1/\cdot) \in \mathcal{RV}_\gamma$ , then

$$q(u) = (1 - u)^{-\gamma} L\left(\frac{1}{1 - u}\right), \quad u \in (0, 1), L \in \mathcal{RV}_0.$$

▷ Taking the log yields

$$\log q(u) = -\gamma \log(1 - u)(1 + o(1)), \quad \text{as } u \rightarrow 1,$$

since  $\log L(v)/\log v \rightarrow 0$  as  $v \rightarrow \infty$  [Bingham et al., 1987, Proposition 1.3.6].

▷ Noting that  $\log((1 - u^2)/2) = \log(1 - u)(1 + o(1))$  as  $u \rightarrow 1$ , define

$$f^{\text{TIF}}(u) := -\frac{\log(q(u))}{\log\left(\frac{1-u^2}{2}\right)}$$

## Second-order condition

**Karamata representation** ([de Haan and Ferreira, 2006, Definition B 1.6]).

The function  $L \in \mathcal{RV}_0$  if and only if  $L$  can be written as

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

where  $c(x) \rightarrow c_\infty$  and  $\varepsilon(x) \rightarrow 0$  as  $x \rightarrow \infty$ .

## Second-order condition

**Karamata representation** ([de Haan and Ferreira, 2006, Definition B 1.6]).

The function  $L \in \mathcal{RV}_0$  if and only if  $L$  can be written as

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

where  $c(x) \rightarrow c_\infty$  and  $\varepsilon(x) \rightarrow 0$  as  $x \rightarrow \infty$ .

**(H<sub>2</sub>)**:  $c(x) = c_\infty$  for all  $x \geq 1$  and  $\varepsilon(x) = x^\rho l(x)$  with  $l \in \mathcal{RV}_0$  and  $\rho < 0$  is referred to the **second-order parameter**

## Second-order condition

**Karamata representation** ([de Haan and Ferreira, 2006, Definition B 1.6]).

The function  $L \in \mathcal{RV}_0$  if and only if  $L$  can be written as

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

where  $c(x) \rightarrow c_\infty$  and  $\varepsilon(x) \rightarrow 0$  as  $x \rightarrow \infty$ .

**(H<sub>2</sub>)**:  $c(x) = c_\infty$  for all  $x \geq 1$  and  $\varepsilon(x) = x^\rho l(x)$  with  $l \in \mathcal{RV}_0$  and  $\rho < 0$  is referred to the **second-order parameter**

As a consequence:

- $F$  satisfies the **second-order condition**
- $L$  is **normalized** and it ensures its **differentiability**

# TIF - First derivative

## Proposition

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

$$\partial_u f^{\text{TIF}}(0) = c',$$

$$\partial_u f^{\text{TIF}}(u) = \sum_{j=0}^3 c_j \varphi_j(u) - \frac{\varepsilon \left( \frac{1}{(1-u)} \right)}{(1-u) \log(1-u)} (1 + o(1)) + \mathcal{O} \left( \frac{(1-u)}{\log(1-u)} \right),$$

as  $u \rightarrow 1$ , where  $c', c_1, \dots, c_3$  are constants and

$$\varphi_0(u) = \frac{1}{(1-u) (\log(1-u))^2} \text{ and } \varphi_j(u) = \frac{1}{(\log(1-u))^j}, j = 1, 2, 3$$

# Correction TIF (CTIF)

$$f^{\text{CTIF}}(u) := f^{\text{TIF}}(u) - g(u) \sum_{j=0}^3 c_j \Phi_j(u) - \gamma g(u) - \partial_u f^{\text{TIF}}(0) h(u),$$

with, for all  $u \in (0, 1)$ ,

$$g(u) = -4u^5 + 5u^4,$$

$$h(u) = u^3 - 2u^2 + u,$$

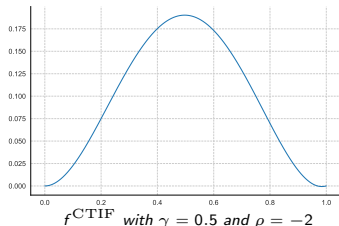
$$\Phi_0(u) = \varphi_1(u),$$

$$\Phi_1(u) = -\text{li}(1-u),$$

$$\Phi_2(u) = \Phi_1(u) + (1-u)\varphi_1(u),$$

$$\Phi_3(u) = \left( \Phi_1(u) + (1-u)(\varphi_1(u) + \varphi_2(u)) \right) / 2,$$

with the logarithmic integral function  $\text{li}(x) := \int_0^x \frac{1}{\log(t)} dt$ .





## CTIF - Second derivative

---

(H<sub>3</sub>):  $l$  is normalized.

## CTIF - Second derivative

---

(H<sub>3</sub>):  $l$  is normalized.

As a consequence:

- $l$  is **differentiable** on  $(0,1)$
- $L$  and  $q$  are **twice differentiable** on  $(0,1)$

## CTIF - Second derivative

(**H**<sub>3</sub>):  $\ell$  is normalized.

As a consequence:

- $\ell$  is **differentiable** on  $(0, 1)$
- $L$  and  $q$  are **twice differentiable** on  $(0, 1)$

### Proposition (part 1)

(i) If (**H**<sub>1</sub>) holds, then

$$\lim_{u \rightarrow 0} f^{\text{CTIF}}(u) = \lim_{u \rightarrow 1} f^{\text{CTIF}}(u) = 0.$$

(ii) If, moreover, (**H**<sub>2</sub>) holds with  $\rho < -1$ , then  $f^{\text{CTIF}} \in \mathcal{C}^1([0, 1])$  and

$$\lim_{u \rightarrow 0} \partial_u f^{\text{CTIF}}(u) = \lim_{u \rightarrow 1} \partial_u f^{\text{CTIF}}(u) = 0.$$

## CTIF - second derivative

### Proposition (part 2)

(iii) If, moreover,  $(\mathbf{H}_3)$  holds, then  $f^{\text{CTIF}} \in \mathcal{C}^2([0, 1])$  and

$$\partial_{uu}^2 f^{\text{CTIF}}(u) = \bar{c}_1 - \frac{(1 + \rho)\varepsilon \left( \frac{1}{(1-u)} \right)}{(1-u)^2 \log(1-u)} (1 + o(1)) + \mathcal{O} \left( \frac{1}{\log(1-u)} \right),$$

as  $u \rightarrow 1$  and

$$\lim_{u \rightarrow 0} \partial_{uu}^2 f^{\text{CTIF}}(u) = \bar{c}_2$$

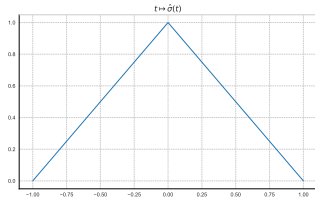
(iv) If, moreover,  $\rho < -2$ , then  $f^{\text{CTIF}} \in \mathcal{C}^2([0, 1])$  and

$$\lim_{u \rightarrow 1} \partial_{uu}^2 f^{\text{CTIF}}(u) = \bar{c}_1$$

# Approximation error

**Triangular function.** Built with 3 translated ReLU  $\sigma(x) := \max(0, x)$

$$\hat{\sigma}(t) := \sigma(t + 1) - 2\sigma(t) + \sigma(t - 1)$$



**Piecewise linear approximation.** For all  $M \in \mathbb{N}_0$ , let  $\delta = 1/M$ ,  $t_j = j/M$  for  $j = 0, \dots, M$ . If  $f \in \mathcal{C}^2([0, 1])$ , then

$$\sup_{t \in [0, 1]} \left| f(t) - \sum_{j=0}^M f(t_j) \hat{\sigma} \left( \frac{t - t_j}{\delta} \right) \right| \leq cM^{-2}$$

$\Rightarrow$  Similarly with  $J = 3(M + 1)$  ReLU  $\sigma$

# Main result

## Theorem

Assume  $(\mathbf{H}_1)$ ,  $(\mathbf{H}_2)$  and  $(\mathbf{H}_3)$  hold. For all  $m = \{1, \dots, d\}$ , if  $\rho^{(m)} < -2$  then there exists a space  $\mathcal{G}_J$  of one hidden layer generators

$$G_{\psi}^{(m)}(z) = \sum_{j=1}^J a_j^{(m)} \sigma \left( \sum_{i=1}^{d'} w_j^{(i)} z^{(i)} + b_j \right) + \sum_{k=1}^6 \kappa_k^{(m)} e_k \left( z^{(m)} \right)$$

with  $J \geq 6$  neurons such that

$$\inf_{G \in \mathcal{G}_J} \sup_{m=1, \dots, d} \sup_{z \in [0,1]^{d'}} \left| f^{\text{TIF},(m)}(z^{(m)}) - G_{\psi}^{(m)}(z) \right| = \mathcal{O}(J^{-2})$$

# Experiments

## Evaluation.

- Marginals:

- ▷ *criteria:* Mean squared logarithmic error (MSLE) at prob. level ( $\xi$ )
- ▷ *graph:*  $-\log(1 - u) \mapsto \log q(u)$

- Dependence:

- ▷ *criteria:* Absolute Kendall error (AKE) using the Kendall's function

$$K(t) = \mathbb{P}(F(x) \leq t),$$

which can be estimated [Genest and Rivest, 1993] by

$$\hat{K}_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}\{Y_i \leq t\} \text{ with}$$

$$Y_i = \frac{1}{n-1} \sum_{j \neq i}^n \mathbb{1}\{X_j^{(1)} < X_i^{(1)}, \dots, X_j^{(d)} < X_i^{(d)}\},$$

for all data  $i \in \{1, \dots, n\}$

- ▷ *graph:*  $t \mapsto t - \hat{K}_n(t)$

## Experiments - Real data

### Settings.

- Log-returns on selected heavy-tailed financial stock market indexes
- Processing of missing data: compute and synchronize log-returns

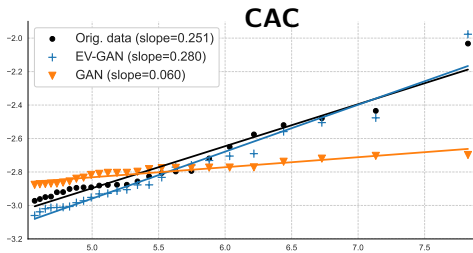
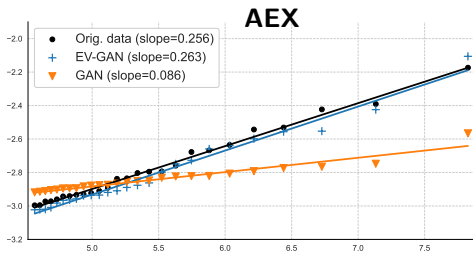
ticker	NKX		Europe		Asia		World	
dimension $d$	1		2		3		6	
sample size $n$	3173		2504		1378		548	
MSLE(0.90)	0.473	<b>0.133</b>	3.860	<b>0.132</b>	2.353	<b>0.677</b>	3.306	<b>0.874</b>
MSLE(0.95)	0.742	<b>0.103</b>	4.925	<b>0.178</b>	1.481	<b>0.579</b>	4.467	<b>1.219</b>
MSLE(0.99)	1.381	<b>0.200</b>	2.792	<b>0.320</b>	1.023	<b>0.538</b>	5.000	<b>1.960</b>
AKE	—	—	16.807	<b>4.697</b>	9.760	<b>4.872</b>	24.781	<b>3.533</b>

Performance comparison between the best GAN (left) and EV-GAN (right)

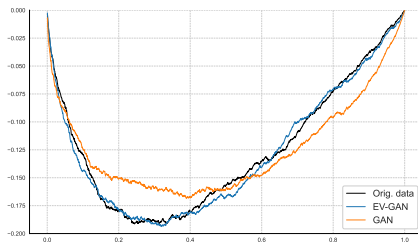


# Experiments - European stock indexes

Marginals fitting at  $\xi = 0.99$



Dependence fitting




# Conclusion

---

- EV-GAN: new parametrization of a **multivariate generative model** dedicated to **extreme events** with an **approximation rate** related to the **second-order parameter**
- Further work: investigate mathematically how **dependence** structure is preserved, leveraging **multivariate extreme-value theory**

# References I

---

-  Bingham, N. H., Goldie, C. M., and Teugels, J. L. (1987).  
*Regular variation*, volume 27 of *Encyclopedia of Mathematics and its Applications*.  
Cambridge University Press, Cambridge.
-  Cybenko, G. (1989).  
Approximation by superpositions of a sigmoidal function.  
*Mathematics of Control, Signals, and Systems*, 2(4):303–314.
-  de Haan, L. and Ferreira, A. (2006).  
*Extreme value theory*.  
Springer Series in Operations Research and Financial Engineering.  
Springer, New York.
-  Genest, C. and Rivest, L.-P. (1993).  
Statistical inference procedures for bivariate Archimedean copulas.  
*Journal of the American Statistical Association*, 88(423):1034–1043.

## References II

---



Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014).

Generative adversarial nets.

In *Advances in neural information processing systems*, pages 2672–2680.