

# Un nouvel estimateur des quantiles extrêmes basé sur le modèle “log Weibull-tail” généralisé

Clément Albert, Anne Dutfoy, Laurent Gardes, Stephane Girard

► **To cite this version:**

Clément Albert, Anne Dutfoy, Laurent Gardes, Stephane Girard. Un nouvel estimateur des quantiles extrêmes basé sur le modèle “log Weibull-tail” généralisé. 50èmes Journées de Statistique organisées par la Société Française de Statistique, May 2018, Saclay, France. pp.1-6. <hal-01807672>

**HAL Id: hal-01807672**

**<https://hal.inria.fr/hal-01807672>**

Submitted on 5 Jun 2018

**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.

# UN NOUVEL ESTIMATEUR DES QUANTILES EXTRÊMES BASÉ SUR LE MODÈLE “LOG WEIBULL-TAIL” GÉNÉRALISÉ

Clément ALBERT <sup>1</sup>, Anne DUTFOY <sup>2</sup>, Laurent GARDES <sup>3</sup> & Stéphane GIRARD <sup>1</sup>

<sup>1</sup> *Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, 38000 Grenoble, France*

<sup>2</sup> *EDF R&D dept. Périclès, 91120 Palaiseau, France*

<sup>3</sup> *Université de Strasbourg & CNRS, IRMA, UMR 7501, 7 rue René Descartes, 67084 Strasbourg Cedex, France*

*clement.albert@inria.fr, anne.dutfoy@edf.fr, gardes@unistra.fr, stephane.girard@inria.fr*

**Résumé.** L'estimation de quantiles extrêmes demeure un problème statistique majeur. Dans cette communication, le problème est abordé dans le cadre du modèle “log Weibull-tail” généralisé, où le logarithme de l'inverse du taux de hasard cumulé est supposé à variation régulière étendue. Après une discussion sur les conséquences de cette hypothèse, nous proposons un nouvel estimateur des quantiles extrêmes basé sur ce modèle. La normalité asymptotique dudit estimateur est alors établie et son comportement en pratique est évalué sur données simulées.

**Mots-clés.** Estimation de quantiles extrêmes, théorie des valeurs extrêmes, variation régulière, propriétés asymptotiques

**Abstract.** Extreme quantile estimation remains a major statistical challenge. In this communication, the problem is addressed in the framework of the so-called “log-Generalized Weibull tail limit”, where the logarithm of the inverse cumulative hazard rate function is supposed to be of extended regular variation. Based on this model, a new estimator of extreme quantiles is proposed. Its asymptotic normality is established and its behavior in practice is illustrated on simulated data.

**Keywords.** Extreme quantiles estimation, extreme-value theory, regular variation, asymptotic properties

## 1 Modèle

Soit  $X$  une variable aléatoire de fonction de répartition  $F(\cdot) = \mathbb{P}(X \leq \cdot)$  et de fonction de survie  $S(\cdot) := 1 - F(\cdot)$ . Supposons  $S(1) = 1$  et

$$S(x) =: \exp[-V^{\leftarrow}(\ln x)], \quad x > 1, \quad (1)$$

où  $V$  est une fonction continue et strictement croissante, avec  $V^{\leftarrow}(\cdot) := \inf\{y; V(y) \geq \cdot\}$  l'inverse généralisé de  $V$ .

**Modèle.** La fonction  $V$  est supposée être à variation régulière étendue d'indice  $\theta \in \mathbb{R}$ . Plus précisément, il existe une fonction positive  $a$  (appelée fonction auxiliaire) telle que, quel que soit  $t > 0$

$$\lim_{x \rightarrow \infty} \frac{V(tx) - V(x)}{a(x)} = \int_1^t u^{\theta-1} du =: L_\theta(t). \quad (2)$$

La classe des fonctions à variation régulière étendue est notée  $\mathcal{ERV}(\theta)$ . Selon [7, Corollaire 1.1.10], une condition suffisante pour (2) est  $V$  est dérivable avec pour dérivée  $V'$  satisfaisant

$$\lim_{x \rightarrow \infty} \frac{V'(tx)}{V'(x)} = t^{\theta-1}. \quad (3)$$

Une telle fonction  $V'$  est dite à variation régulière d'indice  $\theta - 1$  et cette propriété est notée  $V' \in \mathcal{RV}(\theta - 1)$ . De plus, sous (3), un choix possible dans (2) est  $a(t) = tV'(t)$ . Le modèle (2) ou (3) est désigné par “log Weibull-tail” généralisé. Ce modèle a été introduit et développé dans [8, 9, 10].

**Propriétés et exemples.** En termes de domaine d'attraction (MDA), le résultat suivant est tiré de [1, Proposition 4] :

**Lemme 1** *Supposons  $F$  dérivable.*

- (i) *Si (3) est vérifiée avec  $\theta < 1$  alors  $F \in \text{MDA}(\text{Gumbel})$ .*
- (ii) *Si  $F \in \text{MDA}(\text{Fréchet})$  alors (3) est vérifiée avec  $\theta = 1$ .*
- (iii) *Si (3) est vérifiée avec  $\theta > 1$  alors  $F$  n'appartient à aucun domaine d'attraction.*

Le modèle “log Weibull-tail” généralisé avec  $\theta \leq 1$  est donc particulièrement intéressant dans le sens où il est associé avec la plupart des lois appartenant à  $\text{MDA}(\text{Gumbel})$  ou  $\text{MDA}(\text{Fréchet})$ . Le cas  $\theta > 1$ , qui ne correspond à aucun domaine d'attraction, est quelque fois cité sous le nom de “super-heavy tails”, voir par exemple [2]. Les exemples suivants sont tirés de [1, Proposition 3] :

**Exemple 1** *Soit  $x^* := \sup\{x \geq 1, F(x) < 1\}$  le point terminal de  $F$  et soit  $F'$  la densité associée,  $F'$  étant supposée monotone à l'infini.*

- (i) *Si  $V^{\leftarrow}(\ln \cdot) \in \mathcal{RV}(1/\beta)$ ,  $\beta > 0$ , alors (3) est vérifiée avec  $\theta = 0$ . Dans ce cas,  $F$  est appelée une “Weibull tail-distribution”, voir par exemple [5, 6]. Parmi ce type de lois, on peut citer les lois Normale, Gamma, Exponentielle ou encore Weibull stricte.*
- (ii)  *$V^{\leftarrow} \in \mathcal{RV}(1/\beta)$ ,  $0 < \beta < 1$  si et seulement si (3) est vérifiée avec  $\theta = \beta > 0$ . Ici,  $F$  est appelée une “log-Weibull tail-distribution”, voir [2, 4, 6], la plus populaire étant la loi lognormale.*
- (iii)  *$1 \leq x^* < \infty$  et  $V^{\leftarrow}(\ln x^* + \ln(1 - 1/\cdot)) \in \mathcal{RV}_{-1/\beta}$ ,  $\beta < 0$  si et seulement si (3) est vérifiée avec  $\theta = \beta < 0$ . Ce cas correspond aux lois qui présentent un comportement en queue de type Weibull au voisinage d'un point terminal fini.*

## 2 Inférence

Soient  $X_1, \dots, X_n$   $n$  copies indépendantes d'une variable aléatoire  $X$  de fonction de survie  $S$  donnée par (1). Les statistiques d'ordre associées sont notées  $X_{1,n} \leq \dots \leq X_{n,n}$ . Notre but est l'estimation des quantiles extrêmes de  $F$ , *i.e.*  $Q(u) := S^{\leftarrow}(u) = \exp[V(\ln(1/u))]$  quand  $u \rightarrow 0$ . Deux situations pour le niveau  $u$  sont considérées.

**Cas intermédiaire.** Si  $u = \alpha_n$  avec  $\alpha_n$  un niveau intermédiaire satisfaisant  $\alpha_n \rightarrow 0$  et  $n\alpha_n \rightarrow \infty$  lorsque  $n \rightarrow \infty$ , un estimateur naturel est obtenu en remplaçant  $Q$  par son estimateur empirique  $\hat{Q}_n$ . Plus précisément,  $Q(\alpha_n)$  est estimé par

$$\hat{Q}_n(\alpha_n) = X_{n-\lfloor n\alpha_n \rfloor, n}. \quad (4)$$

**Cas extrême.** Si  $u = \beta_n$  avec  $\beta_n$  un niveau extrême tel que  $n\beta_n \rightarrow c \geq 0$  lorsque  $n \rightarrow \infty$ , une simple statistique d'ordre ne suffit plus. L'extrapolation au-delà de l'échantillon est nécessaire. Partant alors d'un niveau intermédiaire  $\alpha_n := k_n/n$  avec  $k_n \rightarrow \infty$  et  $k_n/n \rightarrow 0$ , nous proposons d'estimer  $Q(\beta_n)$  par

$$\check{Q}_n(\beta_n) := \hat{Q}_n\left(\frac{k_n}{n}\right) \exp\left[\hat{a}_n[\ln(n/k_n)]L_{\hat{\theta}_n}\left(\frac{\ln \beta_n}{\ln(k_n/n)}\right)\right], \quad (5)$$

où  $\hat{\theta}_n$  et  $\hat{a}_n[\ln(n/k_n)]$  sont des estimateurs de  $\theta$  et  $a[\ln(n/k_n)]$ . La construction de (5) est basée sur (2) qui signifie que, pour  $\alpha$  proche de 0 et pour tout  $t > 0$ ,

$$\ln Q(t\alpha) \approx \ln Q(\alpha) + a[\ln(1/\alpha)]L_\theta\left(1 + \frac{\ln(t)}{\ln(\alpha)}\right).$$

L'estimateur (5) est alors obtenu en prenant  $\alpha = k_n/n$  et  $t = n\beta_n/k_n$  et en remplaçant les quantités inconnues  $Q(k_n/n)$ ,  $a[\ln(n/k_n)]$  et  $\theta$  par leur estimateur associé. Puisque  $k_n/n$  est un niveau intermédiaire,  $Q(k_n/n)$  est estimé par  $\hat{Q}_n(k_n/n) = X_{n-k_n, n}$ .

**Estimation des paramètres.** Nous proposons maintenant de nouveaux estimateurs de  $\theta$  et  $a[\ln(n/k_n)]$ . Dans ce cadre, pour  $j \in \{1, 2\}$ , nous définissons la statistique suivante :

$$M_n^{(j)} := \frac{1}{k_n} \sum_{i=0}^{k_n-1} [\ln_2(X_{n-i, n}) - \ln_2(X_{n-k_n, n})]^j,$$

où nous avons introduit  $\ln_2 := \ln \ln$ , ainsi que les fonctions

$$\mu_b(t, \zeta) := \int_0^1 \left[ L_\zeta \left( 1 + \frac{\ln(1/s)}{t} \right) \right]^b ds \text{ et } \Psi_t(\zeta) := \frac{\mu_1^2(t, \zeta)}{\mu_2(t, \zeta)},$$

définies pour  $t > 0$ ,  $b \in \mathbb{N} \setminus \{0\}$  et  $\zeta < 1$ . De plus, il est possible de montrer que  $\Psi_t$  est une fonction décroissante, au moins pour  $t$  assez grand, et par conséquent sa réciproque est bien définie pour des  $t$  suffisamment grands. Les statistiques suivantes sont introduites :

$$\hat{\theta}_{n,+}^{(M)} := \frac{M_n^{(1)}}{\mu_1[\ln(n/k_n), 0]}, \quad (6)$$

$$\hat{\theta}_{n,-}^{(M)} := \Psi_{\ln(n/k_n)}^{\leftarrow} \left( \frac{[M_n^{(1)}]^2}{M_n^{(2)}} \right), \quad (7)$$

$$\hat{\theta}_n^{(M)} := \hat{\theta}_{n,+}^{(M)} + \hat{\theta}_{n,-}^{(M)}, \quad (8)$$

$$\hat{a}_n^{(M)}[\ln(n/k_n)] := \frac{\ln X_{n-k_n, n}}{\mu_1[\ln(n/k_n), \hat{\theta}_{n,-}^{(M)}]} M_n^{(1)}. \quad (9)$$

Nous concluons ce paragraphe en donnant les idées principales qui ont servi à construire les estimateurs (8) et (9) de  $\theta$  et  $a[\ln(n/k_n)]$ . L'estimateur (8) est construit dans le même esprit que l'estimateur des moments proposé par [3]. Sa construction est basée sur les deux résultats suivants. Soient  $\theta_+ := \theta \vee 0$  et  $\theta_- := \theta \wedge 0$ . Soit  $V \in \mathcal{ERV}(\theta)$  une fonction croissante. Alors

$$\lim_{x \rightarrow \infty} \frac{V(x)}{a(x)} \ln \frac{V(tx)}{V(x)} = L_{\theta_-}(t), \quad (10)$$

uniformément localement dans  $(0, \infty)$ , voir [7, Lemme B.3.16]. De plus (voir par exemple [7, Eq. 3.5.5]),

$$\lim_{x \rightarrow \infty} \frac{a(x)}{V(x)} = \theta_+.$$

En remplaçant alors  $x := \ln(1/\alpha)$  et  $t := 1 + \ln(s)/\ln(\alpha)$  dans (10), il vient :

$$\ln_2 Q(s\alpha) - \ln_2 Q(\alpha) \approx \theta_+ L_0 \left( 1 + \frac{\ln s}{\ln \alpha} \right),$$

lorsque  $\alpha \rightarrow 0$  et ce quel que soit  $s \in (0, 1)$ . En intégrant alors par rapport à  $s$  sur  $(0, 1)$ , on obtient :

$$\int_0^1 [\ln_2 Q(s\alpha) - \ln_2 Q(\alpha)] ds \Big/ \int_0^1 L_0 \left( 1 + \frac{\ln s}{\ln \alpha} \right) ds \approx \theta_+.$$

En considérant alors  $\alpha = k_n/n$ , où  $k_n$  est une suite intermédiaire telle que  $k_n \rightarrow \infty$  et  $k_n/n \rightarrow 0$ , et en remplaçant  $Q$  par son estimateur empirique  $\hat{Q}_n$ , nous obtenons l'estimateur (6) de  $\theta_+$ . De façon similaire, (10) mène à l'approximation suivante :

$$\left\{ \int_0^1 [\ln_2 Q(s\alpha) - \ln_2 Q(\alpha)] ds \right\}^2 \Big/ \int_0^1 [\ln_2 Q(s\alpha) - \ln_2 Q(\alpha)]^2 ds \approx \Psi_{\ln(1/\alpha)}(\theta_-),$$

lorsque  $\alpha \rightarrow 0$ . Le remplacement encore une fois de  $\alpha$  par  $k_n/n$  et de  $Q$  par son estimateur empirique nous permet d'exhiber l'estimateur (7) de  $\theta_-$ . Finalement, l'estimateur (9) est obtenu en remarquant que, d'après (10)

$$\frac{\ln Q(\alpha)}{a[\ln(1/\alpha)]} \int_0^1 \ln \frac{\ln Q(s\alpha)}{\ln Q(\alpha)} ds \approx \mu_1[\ln(1/\alpha), \theta_-],$$

pour  $\alpha$  proche de 0. La substitution de  $\alpha$  par  $k_n/n$ ,  $Q$  par  $\hat{Q}_n$  et  $\theta_-$  par  $\hat{\theta}_{n,-}^{(M)}$  donne (9).

### 3 Propriétés des estimateurs

D'un point de vue théorique, des résultats de normalité asymptotique seront présentés pour chacun des estimateurs (4)-(9). D'un point de vue pratique, on comparera sur des données simulées l'estimateur (5) avec l'estimateur proposé dans [9].

## Bibliographie

- [1] Albert, C., Dutfoy, A. and Girard, S. (2018). Asymptotic behavior of the extrapolation error associated with the estimation of extreme quantiles, <https://hal.inria.fr/hal-01692544>.
- [2] Alves, I., de Haan, L. and Neves, C. (2009). A test procedure for detecting super-heavy tails, *Journal of Statistical Planning and Inference*, **139**(2), 213–227.
- [3] Dekkers, A., Einmhal, J. and de Haan, L. (1989). A moment estimator for the index of an extreme-value distribution, *The Annals of Statistics*, **17**(4), 1833–1855.
- [4] El Methni, J., Gardes, L., Girard, S. and Guillou, A. (2012). Estimation of extreme quantiles from heavy and light tailed distributions, *Journal of Statistical Planning and Inference*, **142**(10), 2735–2747.
- [5] Gardes, L. and Girard, S. (2008). Estimation of the Weibull tail-coefficient with linear combination of upper order statistics, *Journal of Statistical Planning and Inference*, **138**(5), 1416–1427.
- [6] Gardes, L., Girard, S. and Guillou, A. (2011). Weibull tail-distributions revisited : a new look at some tail estimators, *Journal of Statistical Planning and Inference*, **141**(1), 429–444.
- [7] de Haan, L., and Ferreira, A. (2006). *Extreme Value Theory : An introduction*, Springer Series in Operations Research and Financial Engineering, Springer.
- [8] de Valk, C. (2016). Approximation of high quantiles from intermediate quantiles, *Extremes*, **19**(4), 661–686.
- [9] de Valk, C. (2016). Approximation and estimation of very small probabilities of multivariate extreme events, *Extremes*, **19**(4), 686–717.

[10] de Valk, C., and Cai, J.-J. (2017). A high quantile estimator based on the log-generalized Weibull tail limit, *Econometrics and Statistics*, <https://doi.org/10.1016/j.ecosta.2017.03.001>