

# Estimation de quantiles conditionnels extrêmes : Package *Extremefit*

G. Durrieu \*

I. Grama †

## Résumé

La modélisation et l'estimation des valeurs extrêmes jouent un rôle important dans de nombreux domaines tel que la finance, la biologie, l'écologie, etc. Nous introduisons ici un package, *extremefit*, pour la détermination de quantiles extrêmes conditionnels. Nous présentons brièvement la méthode d'estimation des quantiles conditionnels extrêmes et nous expliquons les concepts statistiques afférents. Nous donnons un exemple d'utilisation du package *extremefit* pour la surveillance de la qualité des eaux littorales et la détection des effets du changement climatique.

**Mots-clefs :** Valeurs extrêmes – Modèle semi-paramétrique – Statistique – Écologie – Package R

## 1 Développement

### 1.1 Modèle et estimation

Le modèle utilisé dans le package est décrit dans [3], [4], [5] et [6]. Nous donnons ici un bref résumé. Soit  $F_t(x) = P(X \leq x | T = t)$  la distribution conditionnelle d'une variable aléatoire  $X$  sachant que la covariable  $T$  prend la valeur  $t \in [0, T_{max}]$  où  $T_{max} > 0$ . La distribution conditionnelle  $F_t$  est supposée avoir son support sur  $[x_0, \infty)$ ,  $x_0 > 0$  avec une densité  $f_t$  strictement positive. Le but principal est d'estimer ponctuellement la fonction de survie  $S_t(x) = 1 - F_t(x)$  et les  $p$ -quantiles extrêmes  $F_t^{-1}(p)$  pour  $t \in [0, T_{max}]$ .

L'approche utilisée est celle du "Peak-Over-Threshold" (voir section 5.3 dans [1]). Nous ajustons pour un certain seuil  $\tau > x_0$  la fonction de répartition d'excès  $F_{t,\tau}(x) = 1 - \frac{1-F_t(x)}{1-F_t(\tau)}$ ,  $x \in [\tau, \infty)$  par la fonction de répartition de Pareto  $G_{\tau,\theta}(x) = 1 - (x/\tau)^{-1/\theta}$ ,  $x \in [\tau, \infty)$ . Nous adoptons donc le modèle semi-paramétrique défini par :

$$F_{t,\tau,\theta}(x) = \begin{cases} F_t(x) & x \in [x_0, \tau], \\ 1 - (1 - F_t(\tau))(1 - G_{\tau,\theta}(x)) & x > \tau. \end{cases} \quad (1)$$

Ce modèle couvre essentiellement des observations dont la distribution est dans le domaine d'attraction de la loi de Fréchet. Selon la partie (i) du théorème 2.1 dans [1], une condition nécessaire et suffisante est  $1 - F_{t,\tau}(\tau x) \rightarrow x^{-\frac{1}{\theta}}$  quand  $\tau \rightarrow \infty$  ce qui implique que  $G_{\tau,\theta}(x)$  peut être utilisée pour estimer  $F_{t,\tau}(x)$  pour  $x \in [\tau, \infty)$ . Ceci justifie notre approche pour l'estimation de la fonction de répartition  $F_t$  sur l'intervalle  $[x_0, \tau]$ . Nous estimons  $F_t$  par la fonction de répartition empirique  $\hat{F}_t$ , alors qu'au-delà du seuil  $\tau$ , nous utilisons la probabilité ajustée  $(1 - \hat{F}_t(\tau))(1 - G_{\tau,\hat{\theta}_t}(x))$  où  $\hat{\theta}_t$  est l'estimateur de Hill pondéré décrit dans [1]. Le choix du seuil  $\tau$  est un problème pour la qualité des estimations dans les modèles des valeurs extrêmes. Nous donnons un choix automatique du seuil en fonction des observations basé sur une suite de tests d'ajustement du modèle paramétrique proposé (voir [3] et [4]).

Dans l'article [5], nous visons à estimer les distributions de survie conditionnellement à une covariable dans le modèle de risques proportionnels de Cox, dans le cas où les probabilités estimées se

---

\*LMBA Université Bretagne Sud, gilles.durrieu@univ-ubs.fr

†LMBA Université Bretagne Sud, ion.grama@univ-ubs.fr

situent en dehors de la plage des données observées en utilisant la modélisation des valeurs extrêmes. Notre analyse est également basée sur la méthode “Peak-Over-Threshold” qui permet d’estimer la queue d’une distribution au-delà d’un seuil. Les applications de cette méthode peuvent être observées dans divers domaines tels que l’assurance, la biologie, les prévisions météorologiques et écologie [4].

## 1.2 Package *extremefit*

Le package *extremefit* [2] implémente la méthode ci-dessus pour estimer ponctuellement les quantiles extrêmes ainsi que les probabilités de survie.

Tout d’abord, les valeurs critiques associées aux tests d’ajustement pour le choix du seuil  $\tau$  sont déterminées à l’aide d’une fonction contenue dans le package, pour un noyau choisi parmi plusieurs noyaux proposés ou définis par l’utilisateur. Ensuite, une fonction permet de choisir automatiquement le seuil  $\tau$  ainsi que d’estimer le paramètre  $\theta_t$  en fonction de la covariable  $T = t$  dans le cas d’estimation des quantiles et probabilités conditionnelles. Les prédictions des quantiles et/ou des probabilités de survie sont disponibles via la fonction *predict* qui agit sur les objets résultant de la fonction précédente.

Certaines fonctions contenues dans le package permettent de calculer des intervalles de confiance ponctuels par bootstrap ou encore de visualiser les statistiques du test d’ajustement. Les méthodes présentées ci-dessus sont disponibles dans une version non-conditionnelle, i.e. pour des observations issues d’une fonction de répartition. Nous donnerons une application sur des données écologiques en précisant les apports de ce package.

## Références

- [1] Jan Beirlant, Yuri Goegebeur, Johan Segers, and Jozef L Teugels. *Statistics of extremes : theory and applications*. John Wiley & Sons, 2006.
- [2] Gilles Durrieu, Ion Grama, Kevin Jaunatre, Quang-Khoai Pham, and Jean-Marie Tricot. *Extremefit : a package for extreme quantiles*. *Journal of Statistical Software*, 87 :1–20, 2018.
- [3] Gilles Durrieu, Ion Grama, Quang-Khoai Pham, and Jean-Marie Tricot. Nonparametric adaptive estimator of extreme conditional tail probabilities quantiles. *Extremes*, 18 :437–478, 2015.
- [4] Gilles Durrieu, Quang-Khoai Pham, Anne-Sophie Foltete, Valérie Maxime, Ion Grama, Véronique Le Tilly, Helene Duval, Jean-Marie Tricot, Chiraz Ben Naceur, and Olivier Sire. Dynamic extreme values modeling and monitoring by means of sea shores water quality biomarkers and valvometry. *Environmental monitoring and assessment*, 188 :1–8, 2016.
- [5] Ion Grama and Kévin Jaunâtre. Estimation of extreme survival probabilities with cox model. *Statistics*, 53(4) :807–838, 2019.
- [6] Ion Grama, Jean-Marie Tricot, and Jean-François Petiot. Estimation of the extreme survival probabilities from censored data. *Buletinul Academiei de Ştiinţe a Moldovei. Matematica*, 74(1) :33–62, 2014.