

extremefit : un package pour estimer les probabilités et quantiles conditionnels extrêmes

G. Durrieu^a, I. Grama^a, K. Jaunâtre^a, Q. K. Pham^b and J.-M. Tricot^a

^aLaboratoire de Mathématiques de Bretagne Atlantique (LMBA)
Université de Bretagne Sud et UMR CNRS 6205
Vannes, France

^bDepartment of Mathematics, Forestry University of Hanoi
Hanoi, Vietnam

gilles.durrieu@univ-ubs.fr, ion.grama@univ-ubs.fr, kevin.jaunatre@univ-ubs.fr,
quangkhoaihd@gmail.com, jean-marie.tricot@univ-ubs.fr

Mots clefs : Valeurs extrêmes, Modèle semi-paramétrique, Peak-over-threshold.

Résumé

La modélisation et l'estimation des valeurs extrêmes jouent un rôle important dans de nombreux domaines tel que la finance, la biologie, etc. Des packages **R** tels que **evir**, **evd**, **fExtremes**, traitent déjà de certains problèmes liés aux extrêmes. Nous introduisons ici un nouveau package, **extremefit**, pour la détermination des quantiles et probabilités *conditionnels* extrêmes.

Nous présentons brièvement la méthode d'estimation des quantiles conditionnels extrêmes utilisée et nous expliquons les concepts statistiques afférents. Nous donnons un exemple d'utilisation du package **extremefit** et nous indiquons son intérêt par rapport aux autres packages.

Modèle et estimation

Le modèle utilisé dans le package est décrit dans [2], [3] et [4] : 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 τ , nous utilisons la probabilité

ajustée $(1 - \widehat{F}_t(\tau))(1 - G_{\tau, \widehat{\theta}_t}(x))$ où $\widehat{\theta}_t$ est l'estimateur de Hill pondéré décrit dans [2]. Le choix du seuil τ est un problème recurrent 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]).

Package **extremefit**

Le package **extremefit** implémente la méthode ci-dessus pour estimer ponctuellement les quantiles extrêmes ainsi que les probabilités de survie en utilisant les objets **S3** du langage **R**.

Tout d'abord, la valeur critique correspondant aux tests d'ajustement pour le choix du seuil τ est déterminée 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 τ ainsi que d'estimer le paramètre θ_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 donnons quelques exemples d'application sur des données réelles et indiquons les apports nouveaux de ce package.

Références

- [1] Beirlant, J., Goegebeur, Y., Teugels, J. and Segers, J. (2004). *Statistics of Extremes : Theory and Applications*, Wiley.
- [2] Durrieu, G., Grama, I., Pham, Q. and Tricot, J.- M. (2015). Nonparametric adaptive estimator of extreme conditional tail probabilities quantiles. *Extremes*, Vol. 18, 437-478.
- [3] Grama, I. and Spokoiny, V. (2008). Statistics of extremes by oracle estimation. *Annals of Statistics*, Vol. 36 (4), 1619-1648.
- [4] Grama, I., Tricot, J.-M. and Petiot, J.-F. (2014). Estimation of the extreme survival probabilities from censored data, *Buletinul Academiei de Stiinte a Republicii Moldova*, Vol. 74 (1), 33-62.