Dépendance extrémale: un modèle général pour la fonction de survie multivariée

Néjib Dalhoumi [1] & Jean-Noel Bacro [1] & Gwladys Toulemonde [1]

[1] *IMAG, Université de Montpellier,*
mohamed-nejib.dalhoumi@umontpellier.fr,jean-noel.bacro@umontpellier.fr
gwladys.toulemonde@umontpellier.fr

Résumé.

Dans le cas de l'indépendance asymptotique, la théorie des valeurs extrêmes classique peut conduire à des résultats d'intérêt limité pour l'estimation de probabilités attachées à des réalisations extrêmes. Plutôt que de travailler sur des lois de maxima, Ledford et Tawn (1996, 1997) ont introduit un modèle pour le comportement de queue d'une distribution bivariée permettant de gérer la décroissance des probabilités de survie. Nous introduisons une nouvelle modélisation de queue bivariée avec de marges Pareto. Ce modèle est inspiré de celui de Wadsworth et Tawn (2013). Une nouvelle variation régulière multivariée non-standard de coefficient une fonction à deux variables $(\beta, \gamma) \in (0, \infty)^2$ est introduite, permettant de généraliser les deux approches de modélisation proposées par Ramos et Ledford (2009) et Wadsworth et Tawn (2013).

En nous appuyant sur cette nouvelle modélisation, nous proposons une nouvelle approche non paramétrique pour l'extrapolation multivariée selon des trajectoires orientées par la valeur de $\alpha = \frac{\beta}{\beta+\gamma}$. Nous considérons aussi une nouvelle classe de modèles paramétriques construits grâce à une mesure non-négative satisfaisant une contrainte qui généralise celle de Ramos et Ledford (2009). Ces nouveaux modèles sont flexibles et conviennent tant pour les situations de dépendance que d'indépendance asymptotique.
Mots-clés. Extrêmes multivariés, (in)dépendance asymptotique, extrapolation.

Abstract. In the case of asymptotic independence, the classical theory of extreme values is usually inefficient for estimating probabilities attached to any extreme event. Ledford and Tawn (1996, 1997) proposed a model for the tail behavior of a bivariate distribution to a slower decrease in the case of asymptotically independent variables. We introduce a new modeling tail of a bivariate distribution with Pareto margins. This model is inspired from the Wadsworth and Tawn (2013) one. A new non-standard multivariate regular variation with index a function of two variables $(\beta, \gamma) \in (0, \infty)^2$ is thus introduced to generalize the two modeling approaches proposed by Ramos and Ledford (2009) and Wadsworth and Tawn (2013).

Building on this new approach we propose a new class of non-parametric models allowing multivariate extrapolation along paths determined by the values of $\alpha = \frac{\beta}{\beta+\gamma}$. Similarly we consider parametric models built with a non-negative measure satisfying a

constraint which generalizes that of Ramos and Ledford (2009). These new models are flexible and valid in situations of dependence or asymptotic independence.
Keywords. Asymptotic (in)dependence, multivariate extremes, extrapolation.

1 Modèles pour la fonction de survie bivariée

1.1 Modèle paramétrique de Ramos et Ledford (2009)

Ramos et Ledford (2009) ont proposé un modèle pour la queue de la fonction de survie jointe. Ce modèle s'appuie sur une version simplifiée du modèle de Ledford et Tawn (1997) et permet de gérer les cas de dépendance et d'indépendance asymptotiques.
Soit (X_F, Y_F) un vecteur aléatoire de marges Fréchet unitaires et de fonction de survie jointe :

$$\bar{F}(x,y) \;=\; P\left(X_F > x, Y_F > y\right) = \frac{\mathcal{L}(x,y)}{(xy)^{\frac{1}{2\eta}}}, \qquad \eta \in (0,1], \tag{1}$$

où $\mathcal{L}$ est une fonction à variation lente bivariée i.e. il existe une fonction g telle que pour tout $x, y > 0$ et $c > 0$

$$g(x,y) = \lim_{r\to\infty} \left\{ \frac{\mathcal{L}(rx,ry)}{\mathcal{L}(r,r)} \right\}, \quad g(cx,cy) = g(x,y).$$

La fonction g est constante sur tout rayon $y = ax$, $a > 0$, par conséquent une fonction g_* dite fonction de ray-dépendance pourra être définie par

$$g\left(\frac{x}{x+y}, \frac{y}{x+y}\right) = g(\omega, 1-\omega) \equiv g_*(\omega), \quad \omega \in [0,1].$$

La fonction à variation lente $\mathcal{L}$ est dite ray-dépendante, respectivement ray indépendante, si $g_*(\omega)$ varie, respectivement est constante, en fonction de ω.
En introduisant un vecteur bivarié (S,T) défini par $(S,T) = \lim_{u\to\infty} \{(X_F/u, Y_F/u) \mid (X_F > u, Y_F > u)\}$ Ramos et Ledford factorisent la mesure μ_{WR} (Resnick et Maulik, 2004) associée au vecteur (R,W) où $R = S + T$ et $W = S/R$ de la façon suivante : $(dr, dw) = r^{-(1+1/\eta)} dr dH_\eta(\omega)$ où H_η est une mesure définie sur $(0,1)$.

En identifiant l'expression de $\bar{F}_{S,T}(s,t)$ déduite directement de (1) à celle obtenue par la transformation en coordonnées pseudo-polaires, on obtient pour $\omega \in (0,1)$:

$$\eta^{-1} g_*(\omega) = \left(\frac{1-\omega}{\omega}\right)^{1/2\eta} \int_0^\omega z^{1/\eta} dH_\eta(z) + \left(\frac{\omega}{1-\omega}\right)^{1/2\eta} \int_\omega^1 (1-z)^{1/\eta} dH_\eta(z).$$

Comme $\bar{F}_{S,T}(1,1) = g_*(\frac{1}{2}) = 1$, H_η doit satisfaire la condition de normalisation suivante:

$$\eta^{-1} = \int_0^{1/2} z^{1/\eta} dH_\eta(z) + \int_{1/2}^1 (1-z)^{1/\eta}\, dH_\eta(z).$$

Cette condition de normalisation généralise celle de la théorie classique des extrêmes bivariée $\int_0^1 \omega \, \mathrm{d}H(\omega) = \int_0^1 (1-\omega) \, \mathrm{d}H(\omega) = 1$.

La mesure H_η est analogue à la mesure classique de la théorie des extrêmes multivariée H. Les deux mesures H_η et H sont liées : si (X, Y) est max-stable, $\eta = 1$ et $dH(\omega) = \left(2 - \int_0^1 \max(\omega, 1-\omega)\right) dH_1(\omega) = \theta dH_1(\omega)$, où θ désigne le coefficient extrémal de (X, Y).

Pour un seuil élevé u et pour $(x, y) \in (u, \infty)^2$, Ramos et Ledford (2009) ont déduit le modèle suivant pour la fonction de survie:

$$\bar{F}_{X_F,Y_F}(x, y) = pu^{\frac{1}{\eta}} g_*(x/(x+y))(xy)^{-\frac{1}{2\eta}} \tag{2}$$

où $p = P(X_F > u, Y_F > u)$. Chaque choix d'une fonction g_* dans (2) fournit un modèle paramétrique de la fonction de survie. En se basant sur une modification du modèle logistique, Ramos et Ledford (2009) ont défini le modèle η-asymétrique logistique.

La fonction de survie de ce modèle est définie par:

$$\bar{F}_{X_F,Y_F}(x, y) = \frac{pu^{1/\eta}}{N_\rho} \left[(\rho x)^{-1/\eta} + \left(\frac{y}{\rho}\right)^{-1/\eta} - \left\{ (\rho x)^{-1/\delta} + \left(\frac{y}{\rho}\right)^{-1/\delta} \right\}^{\delta/\eta} \right] \tag{3}$$

avec $\rho > 0$, $\delta > 0$ tels que $\delta \not> 1$ et $N_\rho = \rho^{-1/\eta} + \rho^{1/\eta} - \left(\rho^{-1/\delta} + \rho^{1/\delta}\right)^{\delta/\eta}$.

1.2 Modèle de Wadsworth et Tawn (2013)

Wadsworth et Tawn (2013) ont introduit le modèle de queue suivant: pour $(\beta, \gamma) \in \mathbb{R}_+^2 \backslash \{(0, 0)\}$ et (X_P, Y_P) un vecteur bivarié avec des marginales Pareto standard

$$P(X_P > n^\beta, Y_P > n^\gamma) = n^{-\kappa(\beta,\gamma)} L(n; \beta, \gamma) \tag{4}$$

où L est une fonction à variation lente univariée de variable n. La fonction strictement positive $\kappa(\beta, \gamma)$ détermine la vitesse de décroissance de la fonction de survie et permet de contrôler la dépendance entre les queues des marginales. Les fonctions κ et L doivent vérifier certaines propriétés (voir Wadsworth et Tawn (2013) pour plus de détails). En particulier :

- la fonction κ est homogène d'ordre 1, par conséquent en posant $\alpha = \beta/(\beta + \gamma) \in [0, 1]$, $\kappa(\beta, \gamma) = (\beta + \gamma)\kappa(\alpha, 1 - \alpha)$, il suffit donc de considérer le modèle avec $\kappa(\alpha, 1 - \alpha)$. En posant $\alpha = 1/2$, $\kappa(1, 1) = 2\kappa(\frac{1}{2}, \frac{1}{2}) = \frac{1}{\eta}$, on retrouve le coefficient η des modèles de Ledford et Tawn (1996) et Ramos et Ledford (2009).

- la fonction L satisfait

$$L(n^h;\beta,\gamma) = L(n;h\beta,h\gamma) \quad h>0. \tag{5}$$

Si l'on suppose que κ est différentiable et que la fonction L vérifie

$$\lim_{n\to\infty} \frac{L(n;\alpha+\log x/\log n, 1-\alpha+\log y/\log n)}{L(n;\alpha,1-\alpha)} = 1,$$

on peut déduire le modèle suivant pour $x, y \geq 0$ (Wadsworth et Tawn (2013)):

$$P(X_P > n^\alpha x, Y_P > n^{1-\alpha} y) = n^{-\kappa(\alpha,1-\alpha)} x^{-\kappa_1(\alpha)} y^{-\kappa_2(\alpha)} L\left(n;\alpha+\frac{\log x}{\log n}, 1-\alpha+\frac{\log y}{\log n}\right) \tag{6}$$

où $\{\kappa_1(\alpha),\kappa_2(\alpha)\} = \left\{\frac{\partial\kappa}{\partial\gamma},\frac{\partial\kappa}{\partial\beta}\right\}|_{(\alpha,1-\alpha)}$.

En particulier en utilisant (5) on a

$$\begin{aligned} P(X_P > nx, Y_P > ny) &= P(X_P > (n^2)^{1/2}x, Y_P > (n^2)^{1/2}y) \\ &= (n^2)^{-\kappa(1/2,1/2)} x^{-\kappa_1(1/2)} y^{-\kappa_2(1/2)} L\left(n^2;\tfrac{1}{2}+\tfrac{\log x}{2\log n},\tfrac{1}{2}+\tfrac{\log y}{2\log n}\right) \\ &= n^{-1/\eta} x^{-\kappa_1(1/2)} y^{-\kappa_2(1/2)} L\left(n;1+\tfrac{\log x}{\log n},1+\tfrac{\log y}{\log n}\right). \end{aligned}$$

En posant $c_1 = \kappa_1(1/2)$, $c_2 = \kappa_2(1/2)$ et $\mathcal{L}(nx,ny) \sim L(n;1+\log x/\log n, 1+\log y/\log n)$, $n \to \infty$ on retrouve le modèle de Ledford et Tawn (1997). Dans ce cas, la fonction $\mathcal{L}(nx,ny)$ prend une forme particulière et le modèle (2) reconstruit à partir de cette fonction est alors limité à un cadre bien particulier de fonction à variation lente à savoir le cadre ray-indépendant. En d'autres termes, les seuls modèles paramétriques disponibles pour évaluer $P(X_P > nx, Y_P > ny)$ correspondent à des modèles pour lesquels g_* est constante. En utilisant une modification de l'approche (4), nous levons l'hypothèse de ray-indépendance. Ceci permet de proposer une modélisation plus générale de la probabilité concerné, permettant la construction de modèles paramétriques pour la fonction de survie.

2 Une nouvelle classe de modèles paramétriques bivariés

Soit (X_P, Y_P) un vecteur bivarié avec des distributions marginales Pareto standard. On suppose que pour $(\beta,\gamma) \in \mathbb{R}_+^2\backslash\{\mathbf{0}\}$ et $(x,y) \in [1,\infty)^2$:

$$P(X_P > n^\beta x, Y_P > n^\gamma y) = \mathcal{L}(n^\beta x, n^\gamma y) n^{-\kappa(\beta,\gamma)} x^{\frac{-\kappa(\beta,\gamma)}{2\beta}} y^{\frac{-\kappa(\beta,\gamma)}{2\gamma}} \tag{7}$$

où κ est la fonction définie dans le modèle de Wadsworth-Tawn (2013) et $\mathcal{L}$ est une fonction à variation lente bivariée vérifiant

$$\lim_{\min(a,b)\to\infty} \frac{\mathcal{L}(ax, by)}{\mathcal{L}(a, b)} = g(x, y)$$

voir Bingham *et al.* (1987). En particulier pour tout $(\beta, \gamma) \in \mathbb{R}_+^2 \backslash \{\mathbf{0}\}$ on a

$$\lim_{\min(n^\beta, n^\gamma)\to\infty} \frac{\mathcal{L}(n^\beta x, n^\gamma y)}{\mathcal{L}(n^\beta, n^\gamma)} = \lim_{n\to\infty} \frac{\mathcal{L}(n^\beta x, n^\gamma y)}{\mathcal{L}(n^\beta, n^\gamma)} = g_{(\beta,\gamma)}(x, y),$$

avec $g_{(\beta,\gamma)}$ homogène d'ordre zéro non standard *i.e* pour tout $c > 0$ et $(x, y) \in (0, \infty)^2$,

$$g_{(\beta,\gamma)}(c^\beta x, c^\gamma y) = g_{(\beta,\gamma)}(x, y).$$

Pour tout $0 < x < \infty$, on a alors $g_{(\beta,\gamma)}(x^\beta, x^\gamma) = g_{(\beta,\gamma)}(1, 1) = 1$. En particulier pour $\beta = \gamma = 1$, on retrouve la définition de l'homogénéité d'ordre zéro classique.
La fonction $\mathcal{L}(n^\beta, n^\gamma)$ peut être considérée comme une fonction à variation lente univariée de variable n, c'est à dire $\mathcal{L}(n^\beta, n^\gamma) \equiv L(n, \beta, \gamma)$. Par conséquent en posant $x = y = 1$ dans (7) on retrouve le modèle (2) de Wadsworth et Tawn (2013). De plus pour $(\beta, \gamma) = (1, 1)$ comme $\kappa(1, 1) = \frac{1}{\eta}$, en considérant l'approximation $\exp(-1/x) \approx 1 - 1/x$ on reconnait le modèle (1) de Ramos et Ledford (2009).

Pour tout couple $(\beta, \gamma) \in (0, \infty)^2$, notre modèle (7) est une nouvelle représentation des queues de distributions qui constituera le point de départ pour la construction de nouveaux modèles semi-paramétriques de la fonction de survie. Une généralisation des résultats de Ramos et Ledford (2009) est alors possible.
Dans chacun des cas non paramétrique ou semi-paramétrique, le passage en marges exponentielles nous permettra d'ajuster des modèles d'extrapolation suivant n'importe quel angle variant dans $(0, \frac{\pi}{2})$.

Bibliographie

[1] Bingham, N H., Goldie, C. M., Teugels, J. L., 1987. *Regular variation. vol. 27 of Encyclopedia of Mathematics and its Applications.* Cambrridge university Press, Cambridge.
[2] Ledford, A. W. , Tawn, J. A. 1996, Statistics for near independence in multivariate analysis, *Biometrika, 83:169-187.*
[3] Ledford, A. W. , Tawn, J. A. 1997, Modelling depedence within joint tail regions, *J.R. Stat. Soc. Ser.B, 59(2):475-499.*
[4] Ramos, A. , Ledford, A. 2009, A new class of models for bivariate joint tails, *J.R. Stat. Soc. Ser.B, 71(1):219-241.*
[5] Resnick, S. , Maulik, K. 2004, Characterisation and examples of hidden regular variation, *Extremes, 7, 31-67.*
[6] Wadsworth, J. , Tawn, J. A, 2013, A new representation for multivariate tail probabilities, *Bernoulli 19, 2689-2714.*