Modèles multivariés pour l'indépendance asymptotique des extrêmes

Néjib Dalhoumi [1] & Jean-Noel Bacro [2] & Gwladys Toulemonde [3]

[1] *I3M, Université de Montpellier, mohamed-nejib.dalhoumi@univ-montp2.fr*
[2] *I3M, Université de Montpellier, jean-noel.bacro@univ-montp2.fr*
[3] *I3M, Université de Montpellier, gwladys.toulemonde@univ-montp2.fr*

Résumé.
L'absence d'un cadre paramétrique exhaustif pour les extrêmes multivariés constitue un obstacle majeur dans l'étude de la dépendance des extrêmes, en particulier en présence d'indépendance asymptotique. En effet les modèles max-stables sont assez restrictifs en ce qui concerne l'indépendance asymptotique des réalisations les plus extrêmes puisqu'ils se limitent à la seule notion usuelle d'indépendance stricte. Ces derniers n'offrent pas la flexibilité souhaitée pour modéliser des données asymptotiquement indépendantes et faire l'inférence statistique. Ledford et Tawn (1996, 1997) ont proposé un modèle pour le comportement de queue d'une distribution bivariée permettant une décroissance plus lente dans le cas de variables asymptotiquement indépendantes. Ce modèle a été à la base de différents développements pour la caractérisation de l'indépendance asymptotique, comme par exemple la variation régulière cachée (Resnick (2002), Resnick et Maulik (2002, 2004), Ramos et Ledford (2009)). En 2011, Ramos et Ledford ont utilisé le processus ponctuel introduit par Ledford et Tawn (1997) pour proposer un modèle permettant d'unifier les cadres de dépendance et d'indépendance asymptotique. En nous appuyant sur cette approche, nous montrons comment des modèles paramétriques multivariés flexibles, valides dans des situations de dépendance ou d'indépendance asymptotique, peuvent être construits.
Mots-clés. Extrêmes multivariés, (in)dépendance asymptotique, variation régulière cachée.

Abstract. The absence of a comprehensive framework for parametric multivariate extreme is a major obstacle in the study of extreme dependence, especially in the presence of asymptotic independence. Indeed, the max-stable models are quite restrictive regarding the asymptotic independence of the most extreme achievements as it concerns only the usual notion of strict independence. This does not provide the desired flexibility to model asymptotically independent data and make statistical inference. Ledford and Tawn (1996, 1997) proposed a model for the tail behavior of a bivariate distribution allowing a slower decrease of asymptotically independent variables. This model was the basis for various developments for the characterization of the asymptotic independence, such as hidden regular variation (Resnick (2002), Resnick and Maulik (2002, 2004), Ramos and

Ledford (2009)). In 2011, Ramos and Ledford have used the point process introduced by Ledford and Tawn (1997) to introduce a model to unify the dependence and asymptotic independence. Using this approach, we show how flexible multivariate parametric models in situations of dependence or asymptotic independence can be built.
Keywords. Multivariate extremes, asymptotic (in)dependence, hidden regular variation.

1 Insuffisance du cadre max-stable

Dans cette section on se place dans un cadre bivarié, suffisant pour montrer que les approches classiques des extrêmes max-stables sont inopérantes dans le cas de l'indépendance asymptotique (IA). Soit F la distribution jointe d'un vecteur aléatoire (Y_1, Y_2) et soit F_j la distribution marginale de Y_j pour $j = 1, 2$. Sans perte de généralité, il est classique de supposer que les lois marginales F_1 et F_2 sont Fréchet unitaires, i.e. $F_1(t) = F_2(t) = \exp(-1/t), t > 0$. En effet on peut toujours s'y ramener moyennant une transformation des marges, en posant $Z_j = -\frac{1}{\log(F_j(Y_j))}$ $j = 1, 2$. Soient $(Z_{1,1}, Z_{2,1}), .., (Z_{1,n}, Z_{2,n})$ n copies indépendantes de (Z_1, Z_2) et $M_{1,n} = \max_{i=1,..,n} Z_{1,i}$, $M_{2,n} = \max_{i=1,..,n} Z_{2,i}$ pour $n \geq 1$. Si :

$$P\left(\frac{M_{1,n}}{n} \leq x, \frac{M_{2,n}}{n} \leq y\right) = F^n(nx, ny) \overset{n\to\infty}{\longrightarrow} G(x, y), \tag{1}$$

avec G non dégénérée, alors G est une loi max-stable, i.e. $G^t(tx, ty) = G(x, y) \quad \forall t > 0$. Il existe une mesure finie μ sur le cône convexe $[0, \infty)^2 \setminus \{\mathbf{0}\}$ telle que :

$$G(x, y) = \exp[-V(x, y)] \quad \text{où} \quad V(x, y) = \mu\left\{(u, v) : u > x \quad ou \quad v > y\right\}.$$

En décomposant μ en composantes radiale et angulaire on obtient :

$$V(x, y) = \int_0^1 \left(\frac{\omega}{x}, \frac{1-\omega}{y}\right) \,\mathrm{d}H(\omega),$$

avec $H(\cdot)$ mesure positive finie sur $[0, 1]$ telle que $\int_0^1 \omega \,\mathrm{d}H(\omega) = \int_0^1 (1-\omega) \,\mathrm{d}H(\omega) = 1$. $V(\cdot, \cdot)$, appelée fonction de dépendance extrêmale, est homogène d'ordre -1 , i.e. $V(tx, ty) = t^{-1}V(x, y), \forall t > 0$. Lorsque H charge seulement les valeurs 0 et 1, on a le cas particulier suivant :

$$\mu\left\{(u, v) : u > x \quad ou \quad v > y\right\} = \frac{1}{x} + \frac{1}{y} = \mu\left\{(u, v) : u > x\right\} + \mu\left\{(u, v) : v > y\right\},$$

et G s'écrit comme le produit de ses marges. On dit alors que le vecteur (Z_1, Z_2) est asymptotiquement indépendant (AI).

Sibuya (1960) a défini un coefficient χ caractérisant cette notion. Pour un vecteur bivarié (Y_1, Y_2) de distributions marginales F_1 et F_2,

$$\chi = \lim_{v \to 1} P\left(F_2(Y_2) > v | F_1(Y_1) > v\right).$$

Les variables Y_1 et Y_2 sont respectivement AI ou asymptotiquement dépendantes (AD) si $\chi = 0$ ou $\chi > 0$. En d'autres termes, l'IA exprime le fait que la probabilité que deux composantes soient simultanément grandes est négligeable par rapport à la probabilité qu'une seule le soit. Un exemple typique est celui de la loi du maximum d'un vecteur bivarié gaussien de coefficient de corrélation ρ : pour toute valeur $|\rho| < 1$, on montre qu'asymptotiquement les marges sont indépendantes.
En présence d'IA, si l'on s'intéresse à la probabilité que les maxima dépassent simultanément un seuil élevé, la convergence (1) est d'intérêt limité. En effet, en passant au logarithme on montre aisément que :

$$nP\left(\frac{Z_1}{n} \geq x, \frac{Z_2}{n} \geq y\right) = \ln G(x,y) - \ln G(x,\infty) - \ln G(\infty, y),$$

et, dans le cas d'une indépendance asymptotique, cette quantité est nulle.

2 Modèles de Ledford et Tawn (1996-1997)

Les modèles de Ledford et Tawn (1996,1997) constituent la base de la majorité des résultats présentés dans la suite. D'une façon intuitive, Ledford et Tawn (1996) ont compris que le problème de l'incapacité de détecter les résidus de dépendance dans le cas de l'IA, c'est à dire la dépendance qui s'exprime à taille d'échantillon finie, est lié à la vitesse de convergence dans la relation (1). Une nouvelle modélisation de la queue de distribution est donc nécessaire pour alléger un peu cette vitesse, sauf que cette nouvelle modélisation sort du cadre max-stable. En effet la nouvelle normalisation est bien appropriée pour la stabilisation de la fonction de survie mais reste incapable de stabiliser les maxima dans le cas de l'IA. Ledford et Tawn (1996) ont présenté un modèle s'appuyant sur l'hypothèse que la fonction de survie d'un vecteur aléatoire (Z_1, Z_2) avec marges Fréchet unitaires est à variation régulière au voisinage de l'infini :

$$P\left(Z_1 > r, Z_2 > r\right) \sim \mathcal{L}(r) r^{\frac{-1}{\eta}} \text{ quand } r \longrightarrow \infty, \tag{2}$$

où $\mathcal{L}$ est une fonction à variation lente, c'est-à-dire $\dfrac{\mathcal{L}(tx)}{\mathcal{L}(x)} \longrightarrow 1$ quand $x \longrightarrow \infty$ pour tout $t > 0$ fixé.
Le paramètre η est une constante qui détermine la vitesse de décroissance de la fonction de survie jointe pour de grandes valeurs de r. Il fournit une mesure de la dépendance entre les queues des marginales et on l'appelle le coefficient de dépendance de queue. Le

modèle de Ledford et Tawn (1997) généralise celui de 1996 en donnant un développement d'ordre supérieur de la fonction de survie jointe d'un vecteur aléatoire de marges Fréchet unitaires. Pour des valeurs élevées de z_1 etz_2, la fonction de survie jointe de (Z_1, Z_2) est modélisée par :

$$\begin{aligned}\overline{F}(z_1, z_2) &= P(Z_1 > z_1, Z_2 > z_2) \\ &= \mathcal{L}_1(z_1, z_2) z_1^{-c_1} z_2^{-c_2} + \mathcal{L}_2(z_1, z_2) z_1^{-(c_1+d_1)} z_2^{-(c_2+d_2)} + ...\end{aligned}$$

où $c_1 + c_2 = \frac{1}{\eta}$, $d_j \geq 0$ et $\mathcal{L}_j(z_1, z_2) \neq 0$ $j = 1, 2$ est une fonction bivariée à variation lente (BVL). On a les résultats suivants : (i) si $\eta = 1$ et $\lim_{t \to \infty} \mathcal{L}(t) \neq 0$, les variables marginales sont AD ; (ii) si $0 < \eta < \dfrac{1}{2}$, les variables marginales sont AI, avec une association négative ; (iii) si $\eta = \dfrac{1}{2}$, et $\mathcal{L} \underset{\infty}{\sim} 1$, les variables marginales sont strictement indépendantes ; (iv) si $\dfrac{1}{2} < \eta < 1$, les variables marginales sont AI, avec une association positive.

La théorie des extrêmes multivariés max-stables est caractérisée par la convergence de mesure : pour tout borélien compact $B \subset [0, \infty]^d \backslash \{\mathbf{0}\}$, si la limite :

$$\lim_{n \to \infty} P\left(\frac{Z}{n} \in B\right) = \mu(B),$$

existe, est non dégénérée et vérifie $\mu(\partial B) = 0$, alors le vecteur Z est dans le domaine d'attraction d'une distribution des extrêmes multivariée. La mesure μ est nécessairement homogène d'ordre -1 et sa masse est concentrée dans le sous-cône $(0, \infty]^d \subset [0, \infty]^d \backslash \{0\}$ dans le cas de la dépendance asymptotique. Lorsque Z présente une IA, μ n'a pas de masse dans $(0, \infty]^d$ et cela pose un problème dans les applications statistiques.
En utilisant leur nouvelle modélisation, Ledford et Tawn (1997) ont défini un nouveau processus ponctuel :

$$\{(Z_{1,i}/b_n, Z_{2,i}/b_n : i = 1, ..., n)\},$$

où b_n vérifie $n\bar{F}(b_n, b_n) \sim 1$. Ce processus converge faiblement vers un processus poissonien dont l'intensité est définie par une nouvelle mesure μ_0 sur le nouveau cône convexe $(0, \infty]^2$, ce qui permet de considérer une nouvelle structure de convergence dans $(0, \infty]^2$ (cf Resnick, 1987, chapitre 5), ouvrant la voie à plusieurs approches de modélisation.

3 Approches Ramos et Ledford (2009, 2011)

Durant la dernière décennie, plusieurs approches ont essayé de combler les lacunes de la théorie des valeurs extrêmes dans le cas de l'IA : variation régulière cachée (Resnick (2002), Resnick et Maulik (2004)), modélisation de la fonction de survie (Ramos et Ledford (2009, 2011)), dépendance résiduelle (de Haan et Zhou (2012)). Dans la suite on se focalise sur les résultats de Ramos et Ledford (2009, 2011).

3.1 Modèle paramétrique de Ramos et Ledford (2009)

Ramos et Ledford (2009) introduisent une nouvelle modélisation et fournissent un développement théorique de la limite sous-jacente au modèle de Ledford et Tawn (1997) permettant d'assurer la validité de la distribution de survie. Partant d'une version simplifiée de la fonction de survie jointe définie dans le modèle de Ledford et Tawn (1997), Ramos et Ledford (2009) introduisent un modèle paramétrique. Ce nouveau modèle permet de gérer les cas de dépendance et d'indépendance asymptotiques. Soit (Z_1, Z_2) un vecteur aléatoire de marges Fréchet unitaires et de fonction de survie jointe :

$$\bar{F}(x,y) \quad = \quad P\left(Z_1 > x, Z_2 > y\right) = \frac{\mathcal{L}(x,y)}{(xy)^{\frac{1}{2\eta}}}, \qquad \eta \in (0,1],$$

où $\mathcal{L}$ est une fonction BVL. En introduisant un vecteur bivarié (S,T) défini par $(S,T) = \lim_{u\to\infty} \{(Z_1/u, Z_2/u) \mid (Z_1 > u, Z_2 > u)\}$ Ramos et Ledford factorise la mesure μ_{WR} (Resnick, 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 $(0,1)$. La condition de normalisation sur la mesure H_η s'écrit alors sous la forme suivante :

$$\eta^{-1} = \int_0^{1/2} \omega^{1/\eta} dH_\eta(\omega) + \int_{1/2}^1 (1-\omega)^{1/\eta}\, dH_\eta(\omega). \tag{3}$$

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

3.2 Loi du maximum de Ramos et Ledford (2011)

Ramos et Ledford (2011) ont considéré le processus ponctuel introduit par Ledford et Tawn (1997) pour retrouver leurs résultats de 2009 sur la mesure H_η puis, grâce à cette nouvelle mesure, ont défini une loi de maxima sur le cône $(0,\infty)^2$. La démarche et le résultat sont brièvement rappelé ci-dessous.
On désigne par $R_\varepsilon = \{(x,y)\,/x > \varepsilon, y > \varepsilon\}$ et $R_\varepsilon\,(x,y) = R_\varepsilon \diagdown \{[\varepsilon, x] \times [\varepsilon, y]\}$ et par $M_{1,n,\varepsilon}$ et $M_{2,n,\varepsilon}$ les composantes de maxima des points $(Z_{1,1}, Z_{2,1})\,, ..., (Z_{1,n}, Z_{2,n})$ dans la région du plan R_ε. Ainsi pour tout $x > \varepsilon$ et $y > \varepsilon$, en utilisant la relation $n\bar{F}\,(n^\eta, n^\eta) = 1$, on obtient :

$$\lim_{\varepsilon\to 0}\lim_{n\to\infty} P\left(\frac{M_{1,n,\varepsilon n^\eta}}{n^\eta} \le x \quad , \frac{M_{2,n,\varepsilon n^\eta}}{n^\eta} \le y\right) = G_\eta(x,y) = \exp\left(-V_\eta(x,y)\right),$$

où

$$V_\eta(x,y) = \int_0^1 \max\left(\frac{\omega}{x}, \frac{1-\omega}{y}\right)^{1/\eta} dH_\eta(\omega).$$

4 Généralisation

Une étude sur simulations couvrant un large spectre de type de dépendances extrêmales montre que le modèle de Ramos et Ledford (2009, 2011) est d'intérêt pour les applications multivariées ($d \geq 2$). Un modèle paramétrique couvrant les cas de dépendance et indépendance asymptotique a été proposé mais la condition de normalisation (3) est un frein pour proposer d'autres modèles d'intérêt. Cette condition est difficile à vérifier pour les distributions définies sur le simplexe :

$$S_d = \left\{ (\omega_1, .., \omega_d) \in \mathbb{R}^d; \omega_1 + \omega_2 +, .., \omega_d = 1 \right\}.$$

En nous appuyant sur l'approche de Ramos et Ledford (2011), nous proposons une approche permettant de modifier toute mesure positive définie sur le simplexe S_d pour assurer la condition de normalisation (3). De nouveaux modèles paramétriques, permettant de gérer les situations présentant de la dépendance et/ou de l'indépendance asymptotiques peuvent alors être construits. Les résultats obtenus sont, par exemple, d'intérêt dans un cadre spatial où les dépendances extrêmales entre paires de sites peuvent évoluer avec les distances considérées.

Bibliographie

[1] de Haan, L. (1985), Extremes in higher dimensions: the model and some statistics, *Sess. Int. Statist. Inst. Paper 26.3. The Hague: International Statistical Institute.*

[2] de Haan, L. et Zhou, C. (2012), Extremes residual dependence for random vectors and processes, *Adv.Appl.Prob, 43:217-242.*

[3] Ledford, A. W. et Tawn, J. A. (1996), Statistics for near independence in multivariate analysis, *Biometrika, 83:169-187.*

[4] Ledford, A. W. et Tawn, J. A. (1997), Modelling depedence within joint tail regions, *J.ROY.Statist.Soc.Ser.B, 59(2):475-499.*

[5] Ramos, A. et Ledford, A. (2009), A new class of models for bivariate joint tails, *J.R.Stat.Ser.B Stat.Methodol, 71(1):219-241.*

[6] Ramos, A. et Ledford, A. (2011), An Alternative Point Process Framework for Modeling Multivariate Extreme Values, *Communications in Statistics - Theory and Methods, 40:12, 2205-2224.*

[7] Resnick, S.I. (1987), *Extreme values, regular variation, and point processes, volume 4 Probability*, A Series of the Applied Probability Trust. Springer-Verlag. New York.

[8] Resnick, S.I. (2002), Hidden regular variation, second order regular variation and asymptotic independence, *Extremes, 5, 303-336.*

[9] Resnick, S. et Maulik, K. (2004), Characterisation and examples of hidden regular variation, *Extremes, 7, 31-67.*

[10] Sibuya, M. (1960), Bivariate extreme statistics, *I.Ann.Inst.Statist.Math.Tokyo, 11:195-210.*