Chapitre 2. Analyses des extrêmes dans des cadres dynamique et multivarié

1. Lois limites pour le maximum et les dépassements de seuils dans un cadre dynamique

Les extrêmes d'une série temporelle peuvent être très différents de ceux d'une suite de variables aléatoires IID.

La dépendance temporelle peut non seulement affecter l'amplitude des extrêmes mais aussi leur dynamique: les évènements de grande ampleur sont souvent regroupés en groupe (clusters).

Soit $X_1, X_2, \dots$ une série strictement stationnaire, i.e. pour tous entiers $h \geq 0$ et $n \geq 1$, la distribution du vecteur $(X_{h+1}, \dots, X_{h+n})$ ne dépend pas de h. On note F la distribution stationnaire d'un X_i.

Soit $\tilde{X}_1, \tilde{X}_2, \dots$ la série associée de variables aléatoires IID de distribution F.

Pour le maximum $M_n = \max_{i=1,\ldots,n} X_i$, nous recherchons une distribution limite de $(M_n - b_n)/a_n$ pour des choix judicieux des constantes $a_n > 0$ et b_n.

Dans le chapitre 1, nous avons montré que, pour des variables aléatoires IID, les seules limites possibles non-dégénérées étaient les distributions GEV.

Ceci reste vrai pour les séries temporelles strictement stationnaires pour autant que la dépendance des extrêmes ne soit pas trop forte.

Cependant la distribution limite diffère de $\tilde{M}_n = \max_{i=1,\ldots,n} \tilde{X}_i$.

La différence est due à l'indice extrêmal, θ, qui mesure la l'intensité avec laquelle les extrêmes se regroupent dans le temps.

Définition: Soit

$$\begin{aligned}\beta_{n,l}(\tau) &= \sup\left|\mathbb{P}\left(X_i \le u_n(\tau), i \in A \cup B\right)\right. \\ &\quad \left. - \mathbb{P}\left(X_i \le u_n(\tau), i \in A\right)\mathbb{P}\left(X_i \le u_n(\tau), i \in B\right)\right|,\end{aligned}$$

où les ensembles A et B sont tels que: $A \subset \{1, ..., k\}$, $B \subset \{k+l, ..., n\}$, et $1 \le k \le n - l$.

Condition $D(u_n(\tau))$: elle est satisfaîte s'il existe une suite $l_n = o(n)$ telle que $l_n \to \infty$ et $\beta_{n,l_n}(\tau) \to 0$ quand $n \to \infty$.

La condition $D(u_n(\tau))$ dit que deux évènements du type $\{\max_{i \in I_1} X_i \le u_n(\tau)\}$ et $\{\max_{i \in I_2} X_i \le u_n(\tau)\}$ deviennent asymptotiquement indépendants lorsque n augmente et que les ensembles d'indices $I_i \subset \{1, ..., n\}$ sont séparés par une distance $l_n = o(n)$.

Ainsi la condition $D(u_n(\tau))$ limite la dépendance à longue portée des extrêmes.

Proposition: Supposons que $n\bar{F}(u_n(\tau)) \to \tau$ et que la condition $D(u_n(\tau))$ est satisfaîte. Si $P(M_n < u_n(\tau))$ converge, alors il existe θ ($0 \leq \theta \leq 1$) tel que

$$P(M_n \leq u_n(\tau)) \to \exp(-\theta\tau).$$

θ ($0 < \theta \leq 1$) est appelé l'indice extrêmal du processus (X_n).

Remarquez que pour la série associée: $P\left(\tilde{M}_n < u_n(\tau)\right) \to \exp(-\tau)$.

Sauf si l'indice extrêmal est égal à un, les distributions limites sont différentes pour la série stationnaire et la série associée.

Si $\theta > 0$, $u_n(\tau) = a_n x + b_n$ où $\tau = (-\ln G(x))$ et G a une distribution GEV de paramètres (μ, σ, ξ), alors

$$P\left(\frac{M_n - b_n}{a_n} \leq x\right) \to G^*(x)$$

où G^* a une distribution GEV de paramètre (μ^*, σ^*, ξ^*) caractérisé par

$$\xi = \xi^*,\ \sigma = \sigma^* \theta^{\xi},\ \mu = \mu^* - \sigma^* \frac{1 - \theta^{\xi}}{\xi}.$$

Remarquez que le paramètre de forme ξ n'est pas modifié.

Condition $D'(u_n(\tau))$: elle est satisfaîte si

$$\lim_{k\to\infty} \limsup_{n\to\infty} n \sum_{j=2}^{[n/k]} P\left(X_1 > u_n(\tau), X_j > u_n(\tau)\right) = 0.$$

Contrairement à la condition $D(u_n(\tau))$ qui contrôle la dépendance à longue portée, la condition $D'(u_n(\tau))$ limite la dépendance à courte portée du processus au delà du seuil. En particulier, cette condition permet d'assurer que la probabilité d'observer plus d'un dépassement de seuil dans un bloc de $[n/k]$ observations est négligeable.

Théorème: Supposons que les conditions $D(u_n(\tau))$ et $D'(u_n(\tau))$ sont satisfaîtes,

alors

$$P(M_n \leq u_n(\tau)) \to \exp(-\tau)$$

et $\theta = 1$.

Le processus ponctuel de dépassement de seuil

Soit $\mathcal{F}_{p,q} = \mathcal{F}_{p,q}(\tau)$ la σ-algèbre générée par les évènements $\{X_i > u_n(\tau)\}$, $p \leq i \leq q$, et

$$\alpha_{n,l}(\tau) \equiv \sup \left| \mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B) : A \in \mathcal{F}_{1,t}, B \in \mathcal{F}_{t+l,\infty}, t \geq 1 \right|.$$

Condition $\Delta(u_n(\tau))$: Elle est satisfaîte si il existe une suite $l_n = o(n)$ telle que $l_n \to \infty$ et $\alpha_{n,l_n}(\tau) \to 0$ quand $n \to \infty$.

Rappelons que le processus ponctuel de dépassements de seuil $N_n^{(\tau)}(\cdot)$ est défini par

$$N_n^{(\tau)}(B) = \sum_{i=1}^{n} \mathbb{I}_{\{i/n \in B, X_i > u_n(\tau)\}},$$

pour tout ensemble $B \subset E := (0, 1]$.

Théorème: Supposons que la condition $\Delta\left(u_n\left(\tau\right)\right)$ soit satisfaîte pour une suite $(u_n(\tau))$ telle que $\lim_{n\to\infty} n\bar{F}(u_n(\tau)) = \tau$. Si le processus ponctuel limite $N_n^{(\tau)}$ existe, c'est nécessairement un processus de Poisson composé homogène d'intensité $\tau\theta\left|\cdot\right|$ et de distribution des clusters π.

Soit $\pi_n\left(m; q_n, u_n\left(\tau\right)\right) = P\left(N_n^{(\tau)}\left(\left(0; q_n/n\right]\right) = m \middle| N_n^{(\tau)}\left(\left(0; q_n/n\right]\right) > 0\right)$.

Proposition: Supposons que l'indice extrêmal θ existe, alors une condition nécessaire et suffisante de convergence de $N_n^{(\tau)}$ est donnée par

$$\lim_{n\to\infty} \pi_n\left(m; q_n, u_n\left(\tau\right)\right) = \pi\left(m\right),$$

où (q_n) est une suite telle qu'il existe une suite (l_n) satisfaisant $l_n = o\left(q_n\right)$, $q_n = o\left(n\right)$ et $nq_n^{-1}\alpha_{n,l_n}\left(\tau\right) \to 0$.

Remarque: Si la condition $\Delta(u_n(\tau))$ est satisfaîte pour tout $\tau > 0$, alors θ et π ne dépendent pas de τ. De plus θ^{-1} est égal à l'inverse de la moyenne de π, $\sum_{k=1}^{\infty} k\pi\left(k\right)$.

2. Lois limites pour les maxima des composantes d'un vecteur

Soit $\left(\mathbf{X}_i = \left(X_{i,1}, ..., X_{i,m}\right)\right)$ une suite de vecteurs de variables aléatoires IID de distribution (multivariée) F. Les maxima par composante sont définis par $M_{j,n} = \max(X_{1,j}, ..., X_{n,j})$.

Les distributions des extrêmes multivariées naissent des lois jointes limites des maxima par composante normalisés:

$$\lim_{n\to\infty} P\left(M_{j,n} \leq c_{j,n}x_j + d_{j,n}, j = 1, ..., m\right) = G(x_1, ..., x_m).$$

Remarquez que les distributions marginales de G, G_j, doivent être des distributions des extrêmes univariées.

Contrairement au cas univarié, il n'existe pas de familles paramétriques naturelles pour caractériser les distributions des extrêmes multivariées G.

La distribution G peut être caractérisée par plusieurs formes semi-parametriques équivalentes.

Supposons que les distributions marginales de G sont des Fréchet unitaires $\Phi_1(x) = \exp\left(-x^{-1}\right)$, $x > 0$, la *représentation de Pickands de* G est donnée par

$$G(x_1, ..., x_m) = \exp\left(-\int_{S_m} \max_{1 \leq j \leq m} \left(w_j x_j^{-1}\right) \mu\left(dw\right)\right),$$

où μ est une mesure finie positive sur

$$S_m = \left\{ y_j \geq 0, j = 1, ..., m : \sum_{j=1}^{m} y_j = 1 \right\}$$

telle que

$$\int_{S_m} w_j \mu\left(dw\right) = 1, \qquad j = 1, ..., m.$$

▷ Quelques exemples dans le cas $m = 2$

- Si μ est une mesure ponctuelle telle que $\mu(0) = \mu(1) = 1$, alors

$$G(x_1, x_2) = \exp\left(-x_1^{-1}\right) \exp\left(-x_2^{-1}\right)$$

et il y a indépendance entre les deux composantes.

- Si μ est une mesure ponctuelle telle que $\mu(0,5) = 2$, alors

$$G(x_1, x_2) = \exp\left(-\max\left(x_1^{-1}, x_2^{-1}\right)\right)$$

et il y a parfaite dépendance entre les deux composantes.

3. Caractérisation de la dépendance extrêmale dans le cas bivarié

La dépendance dans les extrêmes peut apparaître dans différents contextes:

- proximité dans l'espace: ex. les températures maximales annuelles d'Oxford et de Worthing sont liées à cause de la proximité géographique;

- proximité dans le temps: ex. les rentabilités extrêmes d'un indice boursier sont très dépendantes d'une journée sur l'autre car elles proviennent des mêmes périodes de crise;

- des covariables communes: ex. les comportements stochastiques de la hauteur de la mer et de la hauteur des vagues du même site sont influencés par les mêmes conditions météorologiques.

▷ <u>Distributions marginales et structure de dépendence</u>

Il est difficile de faire des comparaisons de la dépendance entre deux variables aléatoires lorsqu'elles n'ont pas les mêmes distributions marginales.

$\Longrightarrow$Transformer les marginales permet de se concentrer sur la structure de dépendance.

∘ Standardisation des distributions marginales et copules

<u>Définition:</u> Une copule est une fonction de répartition d'une distribution bivariées ayant des distributions de loi uniforme sur $[0, 1]$ pour ces distributions marginales.

Une distribution bivariée $F_{X,Y}(x,y)$ de fonctions de distribution marginales $F_X(x)$ et $F_Y(y)$ peut être exprimée de la manière suivante:

- soient $U = F_X(X) = F_{X,Y}(X,\infty)$ et $V = F_Y(Y) = F_{X,Y}(\infty,Y)$, alors U et V ont des distributions uniformes sur l'intervalle $[0,1]$;

- sous des conditions de continuité, il existe une unique fonction $C : [0,1] \times [0,1] \rightarrow [0,1]$ satisfaisant:

$$F_{X,Y}(x,y) = C(u,v) = C\left(F_X(x), F_Y(y)\right).$$

La copule C contient ainsi toute l'information sur la distribution jointe $F_{X,Y}(x,y)$, exceptée l'information concernant la structure des distributions marginales.

Les copules sont des exemples de distributions bivariées avec des distributions marginales standardisées.

Naturellement, on peut choisir d'autres distributions pour les marginales que celle uniforme.

Il est possible de standardiser des distributions marginales continues F_X et F_Y en des distributions marginales continues G_X et G_Y en utilisant la transformation:

$$X^* = G_X^{-1}\left(F_X(X)\right) \text{ et } Y^* = G_Y^{-1}\left(F_Y(Y)\right).$$

Dans la théorie des valeurs extrêmes multivariées, il est ainsi courant de transformer les marginales pour avoir des distribution de Fréchet unitaire

$$F(x) = \exp(-1/x).$$

▷ Comportement limite de dépendance

Soit (X, Y) un v ecteur de v ariables aléatoires de distribution jointe F, av ec des marginales de distribution Fréchet unitaire. Nous disons que X et Y sont asympto-tiquement indépendants si

$$\Pr(Y > t | X > t) \to 0 \text{ quand } t \to \infty,$$

et que X et Y sont asymptotiquement dépendants si

$$\Pr(Y > t | X > t) \to c > 0 \text{ quand } t \to \infty.$$

○ Coefficient de dépendance de queue

Il est possible de caractériser plus finement la dépendance dans les extrêmes en introduisant le modèle suivant:

$$\Pr(X > t, Y > t) \sim \mathcal{L}(t) \, [\Pr(X > t)]^{1/\eta} \text{ pour des grands } t,$$

où $\mathcal{L}(t)$ est une fonction à variations lentes quand $t \to \infty$ et η est le coefficient de dépendance de queue appartenant à l'intervalle $(0, 1]$.

Le coefficient décrit les types de dépendance possibles entre X et Y, et $\mathcal{L}(t)$ mesure en deuxième ordre l'intensité de cette dépendance:

- indépendance asymptotique: $0 < \eta < 1$,
- indépendance: $\eta = 1/2$ et $\mathcal{L}(t) = 1$,
- parfaite dépendance: $\eta = 1$ et $\mathcal{L}(t) = 1$,
- dépendance asymptotique: $\eta = 1$ et $\mathcal{L}(t) \to c > 0$ quand $t \to \infty$.

- Indépendance asymptotique: $0 < \eta < 1$

Il existe trois sous-types pour cette classe:

- $1/2 < \eta < 1$ et $\mathcal{L}(t) \to c > 0$ ou $\eta = 1$ et $\mathcal{L}(t) \to 0$: cela correspond à des observations pour lesquelles les évènements du type $X > t$ et $Y > t$ pour des grandes valeurs de t se produisent plus souvent que sous l'hypothèse d'indépendance exacte;

- $\eta = 1/2$: les extrêmes de X et Y sont presque indépendants et si X et Y sont parfaitement indépendants alors $\mathcal{L}(t) = 1$;

- $0 < \eta < 1/2$: cela correspond à des observations pour lesquelles les évènements du type $X > t$ et $Y > t$ pour des grandes valeurs de t se produisent moins souvent que sous l'hypothèse d'independance exacte;

Le degré de dépendance est déterminé par la combinaison de η et $\mathcal{L}(t)$. Cependant, le paramètre de premier ordre est η.

○ Exemple: la distribution logistique bivariée

Cette distribution avec des marges de type Fréchet unitaire a la fonction de distribution suivante:

$$\mathsf{Pr}(X \leq x, Y \leq y) = \exp\left(-(x^{-1/\alpha} + y^{-1/\alpha})^{\alpha}\right)$$

pour $0 < \alpha < 1$.

Pour identifier η, il faut faire un développement limité de la fonction de survie jointe:

$$\begin{aligned}
\mathsf{Pr}\left(X > t, Y > t\right) &= 1 - 2\exp(-1/t) + \mathsf{Pr}(X < t, Y < t) \\
&\sim (2 - 2^{\alpha})\, t^{-1} + (2^{2\alpha-1} - 1)t^{-2} \text{ quand } t \to \infty
\end{aligned}$$

et ainsi on trouve que $\eta = 1$ et $\mathcal{L}(t) = (2 - 2^{\alpha})$.

○ Les mesures de dépendance χ et $\bar{\chi}$

Pour décrire plus finement la dépendance extrêmale, on introduit les mesures χ et $\bar{\chi}$ qui sont reliées à η et $\mathcal{L}(t)$.

Après la transformation de (X, Y) en (U, V) pour avoir des distributions marginales uniformes, nous obtenons

$$\Pr(V > u | U > u) = 2 - \frac{1 - \Pr(U < u, V < u)}{1 - \Pr(U < u)} \sim 2 - \frac{\log C(u,u)}{\log u}$$

quand $u \to 1$.

Ainsi on peut définir la mesure de dépendance $\chi(u)$ de la manière suivante

$$\chi(u) = 2 - \frac{\log C(u,u)}{\log u} \text{ pour } 0 \leq u \leq 1.$$

La fonction $\chi(u)$ peut être interprétée comme une mesure de dépedance de type quantile:

- le signe de $\chi(u)$ détermine si les variables sont positivement ou négativement associées pour le quantile de niveau u

- $\chi(u)$ est bornée de la manière suivante:

$$2 - \log(2u-1)/\log(u) \leq \chi(u) \leq 1.$$

La borne basse est interprétée comme $-\infty$ si $u \leq 1/2$, et 0 si $u = 1$.

Le paramètre qui est utilisé pour mesurer la dépendance est donné par

$$\chi = \lim_{u \to 1} \chi(u).$$

Schématiquement, χ est la probabilité que l'une des variables est extrême sachant que l'autre l'est.

- Dépendance asymptotique

Dans le cas où $\chi = 0$, les variables sont asympotitquement indépendantes.

Des niveaux assez différents de “dépendance” peuvent tout de même être donnés pour des sous cas.

Pour cela, nous avons besoin d'une mesure de dépendance complémentaire pour mesurer la dépendance dans cette classe des distributions qui sont asymptotiquement indépendants.

La copule de survie est donnée par

$$\bar{C}(u,v) = 1 - u - v + C(u,v).$$

Par analogie avec la définition de $\chi(u)$, une comparaison des fonctions de distributions marginales et jointes de (U, V) conduit à introduire:

$$\bar{\chi}(u) = \frac{2 \log \Pr(U > u)}{\log \Pr(U > u, V > u)} - 1 = \frac{2 \log(1-u)}{\log \bar{C}(u,u)} - 1 \text{ pour } 0 \leq u \leq 1,$$

où $-1 < \bar{\chi}(u) < 1$ pour tout $0 \leq u \leq 1$.

Pour caractériser complétement la dépendance, nous définissons également la mesure

$$\bar{\chi} = \lim_{u \to 1} \bar{\chi}(u)$$

pour laquelle $-1 < \bar{\chi} < 1$.

Les mesures χ et $\bar{\chi}$ sont reliées de la manière suivante, via η et $\mathcal{L}(t)$:

$$\begin{aligned} \bar{\chi} &= 2\eta - 1 \\ \chi &= \begin{cases} c & \text{si } \bar{\chi} = 1 \text{ et } \mathcal{L}(t) \to c > 0 \text{ quand } t \to \infty, \\ 0 & \text{si } \bar{\chi} = 1 \text{ et } \mathcal{L}(t) \to 0 \text{ quand } t \to \infty, \\ 0 & \text{si } \bar{\chi} < 1. \end{cases} \end{aligned}$$

En résumé:

- $\chi \in [0,1]$; l'ensemble $(0,1]$ correspond à la dépendance asymptotique;

- $\bar{\chi} \in [-1,1]$; l'ensemble $[-1;1)$ correspond à l'indépendance asymptotique.

Ainsi la paire $(\chi, \bar{\chi})$ est nécessaire pour caractériser complètement la dépendance extrêmale:

- $(\chi > 0; \bar{\chi} = 1)$ signifie dépendance asymptotique et χ mesure l'intensité de la dépendance dans cette classe;

- $(\chi = 0; \bar{\chi} < 1)$ signifie indépendance asymptotique et $\bar{\chi}$.

En pratique, il faut d'abord regarder (estimer) la valeur de $\bar{\chi}$

- $\bar{\chi} < 1 \Rightarrow$ indépendance asymptotique;

- $\bar{\chi} = 1 \Rightarrow$ dépendance asymptotique: il faut alors regarder (estimer) χ.