MATHEMATIQUES DE L'ASSURANCE NON-VIE

TOME II: TARIFICATION ET PROVISIONNEMENT

CHAPITRE 14. THEORIE DES EXTREMES ET COUVERTURE DES CATASTROPHES

MICHEL DENUIT

Institut des Sciences Actuarielles

denuit@stat.ucl.ac.be

Comportement (asymptotique) des sommes

- La tarification est basée sur la loi des grands nombres.
- La loi des grands nombres permet d'écrire, pour une suite de variable i.i.d. d'espérance finie, que

$$\frac{1}{n}\big(X_1 + ... + X_n\big) \overset{p.s.}{\to} \mathbb{E}[X], \text{ quand } n \to \infty.$$

- En notant $S_n = X_1 + ... + X_n$,

$$\frac{1}{n}\big(S_n - n\mathbb{E}[X]\big) \overset{p.s.}{\to} 0, \text{ quand } n \to \infty.$$

Comportement (asymptotique) des extrêmes

- On considèrera un échantillon $X_1, ..., X_n$ i.i.d., de fdr F, et on définit la statistique d'ordre associée $X_{1:n}, ..., X_{n:n}$ où

$$\min\{X_1, ..., X_n\} = X_{1:n} \leq X_{2:n} \leq ...$$

$$\leq X_{n-1:n} \leq X_{n:n} = \max\{X_1, ..., X_n\}.$$

- En notant $x_F = \sup\{x \in \mathbb{R} | F(x) < 1\} \leq \infty$, comme $\Pr[X_{n:n} \leq x] = F^n(x)$, on voit que

$$X_{n:n} \overset{p.s.}{\to} x_F, \text{ quand } n \to \infty.$$

Comportement (asymptotique) des extrêmes

- Est-il possible de trouver des constantes de normalisation a_n et $b_n > 0$ telles que

$$\Pr\left[\frac{X_{n:n} - a_n}{b_n} \leq x\right] = F^n(a_n + b_n x) \to H(x).$$

- La fdr H est dite max-stable si pour tout $n \geq 2$, il existe des constantes $a_n \in \mathbb{R}$ et $b_n > 0$ telles que

$$H^n(a_n + b_n x) = H(x), \text{ pour tout } x \in \mathbb{R}.$$

- La classe des lois max-stables coïncide avec la classes des limites possibles, non-dégénérées, pour les maximas normalisés de variables i.i.d.

Comportement (asymptotique) des extrêmes

La fdr limite H est du même type qu'une des lois suivantes (données par leur fonction de répartition):

1. Loi de Fréchet,

$$\Phi_\xi(x) = \exp\left(-x^{-\xi}\right) \mathbb{I}[x > 0], \;\; \xi > 0$$

2. Loi de Weibull,

$$\Psi_\xi(x) = \exp\left(-x^{-\xi}\right) \text{ si } x \leq 0\text{, et } 1 \text{ sinon}, \;\; \xi > 0$$

3. Loi de Gumbel,

$$\Lambda(x) = \exp\left(-\exp(-x)\right).$$

Comportement (asymptotique) des extrêmes

Ces trois lois sont en fait les trois cas particuliers de la loi GEV (pour Generalized Extreme Value) dont la fonction de répartition peut se mettre sous la forme

$$H(x) = \begin{cases} \exp\left(-\left(1-\xi\frac{x-\mu}{\sigma}\right)^{1/\xi}\right) & \text{si } \xi \neq 0 \\ \exp\left(-\exp\left(-\frac{x-\mu}{\sigma}\right)\right) & \text{si } \xi = 0 \end{cases}$$

si $\mu + \xi x/\sigma > 0$ (représentation de Jenkinson-von Mises).

La notion de variation régulière

- Une fonction positive $f :]0, +\infty[\to \mathbb{R}$ est dite à variation régulière (en $+\infty$) d'indice α si f vérifie

$$\lim_{x\to\infty} \frac{f(tx)}{f(x)} = t^{\alpha} \text{ pour tout } t > 0.$$

- Si $\alpha = 0$, on parlera alors de fonction à variation lente (notée $\mathcal{L}$).
- Cette propriété est stable par convolution: si $\Pr[X > x] = x^{-\alpha}\mathcal{L}_X(x)$ et $\Pr[Y > y] = y^{-\alpha}\mathcal{L}_Y(y)$ où X et Y sont indépendantes alors

$$\Pr[X + Y > s] \sim s^{-\alpha}\big(\mathcal{L}_X(s) + \mathcal{L}_Y(s)\big), \text{ quand } s \to \infty.$$

Lien entre maximum et somme: les lois sous-exponentielles

- Soient $X_1, X_2, \ldots$ une suite de variables i.i.d., de fdr F.
- La loi correspondant à F est dite subexponentielle si
$$\lim_{x\to\infty} \frac{\Pr\left[X_1 + ... + X_n > x\right]}{\Pr\left[\max\left\{X_1, ..., X_n\right\} > x\right]} = 1,\ n \geq 1.$$
- Autrement dit, le comportement dans les queues d'une somme est alors essentiellement déterminé par la loi du maximum.

Comportement (asymptotique) des excès

- Soit G la fonction de répartition définie par

$$G(x) = \begin{cases} 1 - (1+\xi x)^{-1/\xi} & \text{si } \xi \neq 0 \\ 1 - \exp(-x) & \text{si } \xi = 0, \end{cases}$$

  pour $x \geq 0$ si $\xi \geq 0$, et $0 \leq x \leq -1/\xi$ si $\xi < 0$.

- Cette fonction de répartition est appelée loi de Pareto généralisée (GPD) de paramètre ξ.
- Une généralisation à trois paramètres μ, σ et ξ est obtenue en remplaçant x par $(x-\mu)/\sigma$.
- Notons $G_{\xi,\sigma}(x) = 1 - \left(1+\xi\frac{x}{\sigma}\right)^{-1/\xi}$.

Comportement (asymptotique) des excès

Schématiquement, on retiendra la distinction suivante en fonction du paramètre de queue ξ:

- $\xi = 0$, loi à queue fine,
- $\xi \in]0, 1/2[$, loi à queue épaisse, espérance finie, variance finie,
- $\xi \in [1/2, 1[$, loi à queue épaisse, espérance finie, variance infinie,
- $\xi \in [1, \infty[$, loi à queue épaisse, espérance infinie.

Comportement (asymptotique) des excès

- La loi GPD correspond à la loi limite des excès (théorème de Pickands-Balkema-de Haan).
- Plus précisément, $F \in MDA\left(H_\xi\right)$ si, et seulement si,

$$\lim_{u \to x_F} \sup_{0<x<x_F} \left\{ \left| \Pr[X - u \leq x | X > u] - G_{\xi,\sigma(u)}\left(x\right) \right| \right\} = 0,$$

  où $\sigma\left(\cdot\right)$ est une fonction positive.
- Le comportement limite de la loi des excès permet d'approximer la queue de distribution.

Analogie entre les deux approches

- Bien que différentes, les lois limites GEV et GPD font intervenir toutes deux un paramètre dit paramètre de queue.
- Soit $\xi \in \mathbb{R}$, les deux résultats suivants sont équivalents:

  RESULTAT 1
  $F \in MDA\left(H_\xi\right)$, i.e. il existe des suites (a_n) et (b_n) telles que

$$\lim_{n\to\infty} \Pr\left[X_{n:n} \leq a_n x + b_n\right] = H_\xi\left(x\right), \; x \in \mathbb{R}.$$

Analogie entre les deux approches

RESULTAT 2
Il existe une fonction positive $a(\cdot)$ telle que pour $1+\xi x>0$,

$$\begin{aligned}
\lim_{u\to\infty}\frac{\overline{F}(u+xa(u))}{\overline{F}(u)} &= \lim_{u\to\infty}\Pr\left[\frac{X-u}{a(u)}>x\middle|X>u\right] \\
&= \begin{cases}(1+\xi x)^{-1/\xi} & \text{if } \xi\neq 0, \\ \exp(-x) & \text{if } \xi=0,\end{cases} \\
&= G(x).
\end{aligned}$$

Cas de la loi exponentielle

- Considérons la loi exponentielle de paramètre 1, $F(x) = 1 - e^{-x}$.
- En posant $a_n = 1$ et $b_n = \log n$, alors

$$F^n(a_n x + b_n) = \left(1 - \frac{e^{-x}}{n}\right)^n \rightarrow \exp\left(-e^{-x}\right),$$

  c'est-à-dire que la loi limite du maximum normalisé est la loi de Gumbel (de paramètre GEV $\xi = 0$).
- Aussi, les lois dans le domaine d'attraction de la loi de Gumbel sont parfois appelées de type exponentielle.

Cas de la loi exponentielle

- Pour l'approche par seuil, notons que,

$$\Pr\left[X > u + y | X > u\right] = \frac{1 - F\left(u + y\right)}{1 - F\left(u\right)} = \frac{e^{-(u+y)}}{e^{-u}} = e^{-y},$$

  pour tout $y > 0$.
- Aussi, la loi limite est la loi GPD de paramètre $\xi = 0$, avec $\sigma_u = 1$.
- Notons que dans le cas, la loi GPD n'est pas simplement la loi limite, mais il s'agit de la loi exacte pour tout u.
- Notons que pour ces deux approches, on retrouve le cas limite de la loi de paramètre ξ nul.

Cas de la loi de Pareto

- Considérons la loi de Pareto, $1 - F(x) = cx^{-\alpha}$, où $c > 0$, $\alpha > 0$.
- En notant $b_n = 0$ et $a_n = (nc)^{1/\alpha}$ alors pour $x > 0$,

$$F^n(a_n x) = \left(1 - \frac{x^{-\alpha}}{n}\right)^n \to \exp\left(-x^{-\alpha}\right),$$

  où la limite correspond à la loi de Fréchet.
- Aussi, les lois dans le domaine d'attraction de la loi de Fréchet sont parfois appelées de type Pareto.

Cas de la loi de Pareto

- Pour l'approche par seuil, posons $\sigma_u = ub$ où $b > 0$, de telle sorte que

$$\begin{aligned} F_u(\sigma_u z) &= \frac{F(u+ubz)-F(u)}{1-F(u)} \\ &\sim \frac{cu^{-\alpha}-c(u+ubz)^{-\alpha}}{cu^{-\alpha}} = 1-(1+bz)^{-\alpha}. \end{aligned}$$

- En posant $\xi = \frac{1}{\alpha} > 0$ et $b = \xi$ la limite est alors la loi GPD de paramètre $\xi > 0$.

Lois à queue épaisse, étude de MDA(Φ_ξ)

Parmi les lois appartenant au max-domaine d'attraction de la loi de Fréchet ($\xi > 0$), on notera

- la loi de Cauchy $f(x) = (\pi(1+x^2))^{-1}$, $x \in \mathbb{R}$,
- les lois α-stable,
- la loi de Pareto pour $\alpha > 1$ $f(x) = \alpha x^{-(\alpha+1)}$, $x > 1$,
- la loi log-gamma $f(x) = \alpha^\beta \Gamma(\beta)^{-1} (\log x)^{\beta-1} x^{-\alpha-1}$, $x > 1$,
- la loi de Student $f(x) = \Gamma((n+1)/2) \left(1 + x^2/n\right)^{-(n+1)/2} / \sqrt{n\pi}\Gamma(n/2)$, $x \in \mathbb{R}$.

Lois à queue épaisse, étude de MDA(Φ_ξ)

- Rappelons que $\overline{\Phi}_\xi(x) = \exp(-x^{1/\xi})) \sim x^{-1/\xi}$ quand $x \to \infty$ (décroissance en fonction puissance).
- Pour $\xi > 0$, $F \in MDA(H_\xi)$ si et seulement si $\overline{F}(x) = x^{-1/\xi}\mathcal{L}(x)$, où $\mathcal{L}(\cdot)$ est une fonction à variation lente.
- Si F est absolument continue, de densité f, telle que

$$\lim_{x\to\infty} \frac{xf(x)}{\overline{F}(x)} = \lim_{x\to\infty} xr(x) = \frac{1}{\xi} > 0,$$

alors $F \in MDA(H_\xi)$.

Lois à queue fine, étude de MDA(Λ)

- Un simple développement de Taylor à l'ordre 1 permet de montrer que $1 - \Lambda(x) \sim \exp(-x)$ quand $x \to \infty$.
- Parmi les lois appartenant au max-domaine d'attraction de la loi de Gumbel ($\xi = 0$), on notera
  - la loi exponentielle $f(x) = \exp(-x)$ $x > 0$,
  - la loi gamma $f(x) = \Gamma(\alpha)^{-1} x^{\alpha-1} \exp(-x)$, $x > 0$,
  - la loi normale $f(x) = \exp(-x^2/2)/\sqrt{2\pi}$, $x \in \mathbb{R}$,
  - la loi log-normale $f(x) = \exp(-[\log x]^2/2)/\sqrt{2\pi} x$, $x > 0$.

Estimation de l'indice de queue

$$\xi_{n,m}^{Pickands} = \frac{1}{\log 2} \log \frac{X_{n-m:n} - X_{n-2m:n}}{X_{n-2m:n} - X_{n-4m:n}}$$

$$\xi_{n,m}^{Hill} = \frac{1}{m} \sum_{i=1}^{m} \log X_{n-i:n} - \log X_{n-m+1:n}$$

$$\xi_{n,m}^{DEdH} = \xi_{n,m}^{H(1)} + 1 - \frac{1}{2} \left(1 - \frac{\left(\xi_{n,m}^{H(1)}\right)^2}{\xi_{n,m}^{H(2)}} \right)^{-1},$$

où $\xi_{n,m}^{H(r)} = \frac{1}{m} \sum_{i=1}^{m-1} \left(\log X_{n-i:n} - \log X_{n-m:n}\right)^r$, $r = 1, 2, \ldots$

Estimation de l'indice de queue

- Il est possible de montrer que ces estimateurs convergent quand $n \to \infty$, $m \to \infty$ et $m/n \to 0$, et

$$\sqrt{m}\left(\xi_{n,m}^{Pickands} - \xi\right) \to_{\text{Loi}} \mathcal{N}or\left(0, \frac{\xi^2\left(2^{\xi+1}+1\right)}{\left(2\left(2^{\xi}-1\right)\log 2\right)^2}\right)$$

$$\sqrt{m}\left(\xi_{n,m}^{Hill} - \xi\right) \to_{\text{Loi}} \mathcal{N}or\left(0, \xi^2\right)$$

$$\sqrt{m}\left(\xi_{n,m}^{DEdH} - \xi\right) \to_{\text{Loi}} \mathcal{N}or\left(0, 1+\xi^2\right) \text{ pour } \xi \geq 0$$

- L'estimateur de Hill présentant une variance asymptotique plus faible que les trois autres (*dans le cas où* $\xi \geq 0$) il est généralement le plus utilisé.

Utilisation de l'approximation GPD pour estimer une VaR

- Notons $N_u = \sum_{i=1}^{n} \mathbb{I}(X_i > u)$, le nombre de dépassement de u dans un échantillon $X_1, ..., X_n$.
- Pour $x > u$,

$$\overline{F}(x) = \overline{F}(u) \Pr[X > x | X > u],$$

où $\Pr[X > x | X > u] = \overline{F}_u(x - u)$ avec

$$F_u(t) = \Pr[X - u \leq t | X > u] \approx G_{\xi,\beta}(t),$$

pour des valeurs ξ et β appropriées.

Utilisation de l'approximation GPD pour estimer une VaR

- Aussi, un estimateur naturel de $\overline{F}(x)$ repose sur l'utilisation d'un estimateur empirique de $\overline{F}(u)$, et de l'approximation GPD de $\overline{F}_u(x)$, i.e.

$$\widehat{F}(x) = 1 - \frac{N_u}{n}\left(1 + \widehat{\xi}\frac{x-u}{\widehat{\beta}}\right)^{-1/\widehat{\xi}}$$

pour tout $x > u$, et u suffisamment grand.

- Un estimateur naturel de VaR$[X; p]$ est

$$\widehat{x}_p = u + \frac{\widehat{\beta}}{\widehat{\xi}}\left(\left(\frac{n}{N_u}(1-p)\right)^{-\widehat{\xi}} - 1\right).$$

Alternative basée sur l'estimateur de Hill

- Un estimateur naturel de la fonction de répartition est

$$\widehat{F}(x) = 1 - \frac{k}{n}\left(\frac{x}{X_{n-k:n}}\right)^{-1/\xi_{n,k}^{Hill}}, \text{ pour } x \geq X_{n-k:n}.$$

- En considérant l'inverse de cette fonction, on en déduit l'estimation naturelle de la VaR$[x.P]$:

$$\widehat{x}_p^{Hill} = X_{n-k:n}\left(\frac{n}{k}(1-p)\right)^{-\xi_{n,k}^{Hill}},$$

pour $p > 1 - k/n$.

Comparaison

- Notons que cet estimateur peut également s'écrire

$$\widehat{x}_{p,k}^{Hill} = X_{n-k:n} + X_{n-k:n} \left(\left(\frac{n}{k}(1-p) \right)^{-\xi_{n,k}^{Hill}} - 1 \right),$$

qui peut ainsi être comparé à l'estimateur obtenu par maximum de vraisemblance sur le modèle GPD,

$$\widehat{x}_{p,k} = X_{n-k:n} + \frac{\widehat{\beta}_k}{\widehat{\xi}_k} \left(\left(\frac{n}{k}(1-p) \right)^{-\widehat{\xi}_k} - 1 \right).$$

- Rappelons que l'estimateur de Hill n'est pertinent que si $\xi > 0$ (et donc pas si $\xi = 0$).