

Expected shortfall of claims amounts: some practical aspects[†]

Frédéric Planchet*

Pierre-E. Thérond^α

ISFA – Laboratoire SAF^β
Université de Lyon - Université Claude Bernard Lyon 1
WINTER & Associés^γ

Abstract

In this paper, we focus on the practical aspects of the methodologies which enable the calculation of the expected shortfall of claims amounts. If many theoretical results exist on the subject, the operational setting for applications is seldom discussed. One shows that the choice of the method to adopt requires some particular investigations. Moreover the results obtained are overall not very robust when the only extreme value theory is used (without making an assumption on the structure of the tail distribution).

KEYWORDS: Reserving, extreme values.

Espérance du coût des sinistres au-delà d'un seuil : quelques aspects pratiques

Résumé

Le présent article s'intéresse à la robustesse des méthodes pratiques de calcul de l'espérance du coût des sinistres au-delà d'un seuil. Si de nombreuses réflexions théoriques existent sur le sujet, la mise en œuvre opérationnelle est rarement discutée. On montre que le choix de la méthode à retenir nécessite des investigations particulières et que les résultats obtenus sont globalement peu robustes lorsque l'on s'en tient à l'utilisation de la théorie des valeurs extrêmes.

MOTS-CLEFS : Provisionnement, valeurs extrêmes.

[†] This paper will be presented in English.

* Corresponding author. Contact : fplanchet@winter-associés.fr

^α Contact : ptherond@winter-associés.fr

^β Institut de Science Financière et d'Assurances (ISFA) - 50 avenue Tony Garnier - 69366 Lyon Cedex 07 – France.

^γ WINTER & Associés – 43-47 avenue de la Grande Armée - 75116 Paris et 18 avenue Félix Faure - 69007 Lyon – France.

1. Introduction

Dans de nombreuses situations en assurance il s'avère indispensable de disposer d'une information robuste sur le comportement des « grandes valeurs » d'une série statistique.

On peut bien entendu citer le projet Solvabilité 2 (cf. le projet de directive cadre publié été 2007) qui a modifié de fait profondément les règles de fixation du niveau des fonds propres en assurance en introduisant comme critère explicite le contrôle du risque global supporté par la société, risque qui devra être quantifié au travers de la probabilité de ruine à horizon un an. Le niveau retenu de 99,5 %, implique la nécessité d'estimer convenablement un quantile d'ordre élevé de la distribution d'intérêt (en pratique ici celle du surplus, ou marge actif-passif). La problématique posée par un tel critère est étudiée dans Thérond, Planchet (2007). En dehors de l'Union européenne, le Swiss Solvency Test recourt également à des valeurs extrêmes en utilisant l'expected shortfall (espérance au delà d'un seuil) à 99 % dans le calcul de l'exigence de solvabilité des sociétés d'assurance.

De manière plus traditionnelle, cette question se pose en assurance non-vie au moment du calcul des provisions (cf. Partrat et Besson (2005)) ; il est alors indispensable de pouvoir déterminer efficacement des estimations de l'espérance de la charge sinistre au-delà d'un seuil, le seuil s'interprétant comme la limite entre les « grands sinistres » et les sinistres « standards ». D'autant plus que, dans un certain nombre de cas, les contrats sont réassurés par des traités de type excédent de sinistres (XS).

Dans le présent article, après un bref rappel méthodologique, nous examinons le comportement de différents estimateurs à partir de données simulées dont on connaît la vraie loi. Nous en concluons que les approches fondées exclusivement sur les résultats de la théorie des valeurs extrêmes sont peu robustes, du fait notamment du faible volume de données disponibles pour l'estimation des paramètres, mais aussi parce que les « grandes valeurs » de la distribution d'intérêt ne sont pas nécessairement des valeurs extrêmes au sens de la théorie.

Des approches paramétriques basées sur un modèle mixte dans lequel une attention particulière est portée à la modélisation des « grandes valeurs » de la distribution (cf. Planchet et Thérond (2007) sur ce sujet).

2. Distribution des excès au-delà d'un seuil : rappels théoriques

Étant donné un échantillon X_1, \dots, X_n de variables aléatoires (supposées représenter ici des coûts de sinistres), on s'intéresse, pour un seuil u fixé, à l'espérance au-delà du seuil, soit de manière formelle à :

$$e_X(u) = E(X - u | X > u).$$

Le seuil u est supposé « grand ».

Deux approches sont *a priori* envisageables :

- calculer explicitement, ou de manière approchée pour u suffisamment grand, la valeur de $e_X(u)$ dans le cadre d'une loi paramétrique connue ;
- recourir à une approche non paramétrique faisant l'économie de la spécification de la loi de X .

La première approche est classique et les expressions de $e_X(u)$ pour les lois classiques peuvent être trouvées dans de nombreuses références (cf. par exemple Partrat et Besson (2005) p. 460). Une loi mixte est présentée dans Planchet et Thérond (2007) en abordant les problèmes d'estimation et de tests. Elle ne sera pas développée ici, on se concentrera dans la suite de ce travail sur l'approche non paramétrique.

On considère dans la suite l'échantillon ordonné $X_{(1)} \leq \dots \leq X_{(n)}$; afin d'alléger les notations on notera parfois $X_{k,n} = X_{(n-k+1)}$, de sorte que le sous-échantillon des k valeurs les plus élevées s'écrive : $X_{k,n} \leq \dots \leq X_{1,n}$. On note $S(x) = 1 - F(x)$ la fonction de survie associée à F , la fonction de répartition de X . Cette fonction permet de décrire simplement la loi des excès, au sens où :

$$S_u(x) = \Pr(X - u > x | X > u) = \frac{S(u+x)}{S(u)}$$

Cette approche a été introduite par De Haan et Rootzen (1993). Pour une classe assez large de lois de probabilité¹, le théorème de Pickands (1975) permet d'affirmer qu'il existe une constante de normalisation positive $a(u)$ telle que :

$$\lim_{u \rightarrow \omega(F)} S_u(xa(u)) = S_{\xi, \beta}(x)$$

avec :

$$S_{\xi, \beta}(x) = \begin{cases} \left(1 + \frac{\xi x}{\beta}\right)^{-\frac{1}{\xi}} & \text{si } \xi \neq 0 \\ \exp\left(-\frac{x}{\beta}\right) & \text{si } \xi = 0 \end{cases}$$

et $\beta > 0$. Cette distribution limite s'appelle « distribution de Pareto généralisée » (GPD). ξ est l'indice de queue de la distribution des extrêmes associée (au sens de la loi GED). On rappelle que $\xi = 0$ correspond au domaine d'attraction de Gumbel, qui contient en particulier les distributions exponentielle, normale, log-normale, gamma, Weibull, Benktender, etc. Dans les situations pratiques en assurance dommage, on se trouve plutôt dans le domaine de Fréchet, pour lequel $\xi > 0$, avec les distributions Pareto, log-gamma, Burr, stables, etc.

L'égalité ci-dessus peut être reformulée de la manière suivante : lorsque u est grand (c'est à dire proche du point terminal $\omega_F = \sup\{y \in \mathbb{R}, F(y) < 1\}$), alors on peut trouver $\beta(u) = \beta \times a(u)$ tel que :

$$S_u(x) \approx S_{\xi, \beta(u)}(x).$$

De plus, la fonction $\beta(u)$ est unique à une constante multiplicative près. Dans le cas où F est dans le domaine d'attraction de Gumbel, on peut montrer que l'on peut prendre $\beta(u) = \frac{1}{S(u)} \int_u^{\omega(F)} S(t) dt$. Ce résultat peut s'exprimer en disant que la loi des excès au-delà d'un seuil converge vers une loi GPD lorsque la taille de l'échantillon, et donc le seuil, tendent vers l'infini (au moins dans les cas où le support de la distribution n'est pas borné, *i.e.* $\omega(F) = +\infty$).

2.1. Propriétés de la loi GPD

Les résultats suivants (cf. Embrechts et al. (1997)) sont énoncés pour une variable aléatoire X distribuée selon une loi GPD de paramètre (ξ, β) .

Propriété 1. Si $\xi < 1$, on a

$$\begin{aligned} \mathbb{E}\left(1 + \frac{\xi}{\beta} X\right)^{-r} &= \frac{1}{1 + \xi}, \text{ pour } r > -1/\xi, \\ \mathbb{E}\left[\ln\left(1 + \frac{\xi}{\beta} X\right)\right]^k &= \xi^k k!, \text{ pour } k \in \mathbb{N}, \\ \mathbb{E}\left(X(1 - G_{\xi, \beta}(X))\right)^r &= \frac{\beta}{(r+1-\xi)(r+1)}, \text{ pour } (r+1)/\xi > 0. \end{aligned}$$

Dès lors que l'on se place dans le contexte de distributions à queue épaisse ($\xi > 0$), l'existence des moments aux différents ordres n'est plus assurée. Précisément, on a :

Propriété 2. La variable aléatoire Y admet des moments jusqu'à l'ordre $\lceil \xi^{-1} \rceil$ et l'on a

¹ En pratique, les distributions appartenant à l'un des trois domaines d'attraction classiques des valeurs extrêmes.

$$E(X^r) = \frac{\beta^r \Gamma(\xi^{-1} - r)}{\xi^{r+1} \Gamma(\xi^{-1} + 1)}, \text{ pour } r \leq [\xi^{-1}].$$

$$\text{En particulier } E(X) = \frac{\beta}{1 - \xi}.$$

Propriété 3. (Stabilité) La variable aléatoire $X_u = [X - u | X > u]$ est distribuée selon une GPD de paramètre $(\xi, \beta + \xi u)$. On en déduit que si $\xi < 1$, alors pour tout $u < \omega_F$,

$$E[X - u | X > u] = \frac{\beta + \xi u}{1 - \xi}, \text{ pour } \beta + \xi u > 0.$$

Si $\xi \geq 1$ l'espérance n'existe plus.

2.2. Estimation des paramètres de la loi GPD

L'estimation des paramètres peut être effectuée par le maximum de vraisemblance. La densité de la distribution GPD s'écrit en effet :

$$g_{\xi, \beta}(x) = \begin{cases} \frac{1}{\beta^\xi} (\beta + \xi x)^{-\frac{1}{\xi} - 1} & \text{si } \xi \neq 0 \\ \beta^{-1} \exp\left(-\frac{x}{\beta}\right) & \text{si } \xi = 0 \end{cases}$$

et on en déduit la log-vraisemblance :

$$\ln L(\xi, \beta; X_1, \dots, X_n) = -n \ln \beta - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^n \ln \left(1 + \frac{\xi}{\beta} X_i\right).$$

En utilisant la reparamétrisation $\tau = \xi/\beta$, l'annulation des dérivées partielles de la log-vraisemblance conduit au système :

$$\begin{cases} \xi = \frac{1}{n} \sum_{i=1}^n \ln(1 + \tau X_i) =: \hat{\xi}(\tau), \\ \frac{1}{\tau} = \frac{1}{n} \left(\frac{1}{\xi} + 1\right) \sum_{i=1}^n \frac{X_i}{1 + \tau X_i}. \end{cases}$$

L'estimateur du maximum de vraisemblance de (ξ, τ) est $(\hat{\xi} = \hat{\xi}(\hat{\tau}), \hat{\tau})$ où $\hat{\tau}$ est solution de

$$\frac{1}{\tau} = \frac{1}{n} \left(\frac{1}{\hat{\xi}(\tau)} + 1\right) \sum_{i=1}^n \frac{X_i}{1 + \tau X_i}.$$

Cette dernière équation se résout numériquement de manière itérative pour autant que l'on dispose d'une valeur initiale τ_0 pas trop éloigné de τ . En pratique, cette valeur initiale pourra être obtenue par la méthode des moments (pour autant que ceux-ci existent jusqu'à l'ordre 2) ou par la méthode des quantiles. Une fois obtenue l'estimateur $\hat{\tau}$ on en déduit immédiatement $\hat{\xi} = \hat{\xi}(\hat{\tau})$. Lorsque $\xi > -1/2$, Hosking et Wallis (1987) ont montré la normalité asymptotique des estimateurs du maximum de vraisemblance :

$$n^{1/2} \left(\hat{\xi}_n - \xi, \frac{\hat{\beta}_n}{\beta} - 1 \right) \xrightarrow[n \rightarrow \infty]{L} N \left[0, (1 + \xi) \begin{pmatrix} 1 + \xi & -1 \\ -1 & 2 \end{pmatrix} \right].$$

Ce résultat permet en particulier de calculer les erreurs approximatives d'estimation commises par les estimateurs du maximum de vraisemblance. On peut aussi remarquer que la fonction quantile a une expression explicite, lorsque $\xi > 0$:

$$x_p = G_{\xi, \beta}^{-1}(p) = \frac{\beta}{\xi} \left((1-p)^{-\xi} - 1 \right).$$

Cela permet de proposer un estimateur alternatif de ξ par la méthode des quantiles en remarquant que :

$$\frac{x_{p_1}}{x_{p_2}} = \frac{(1-p_1)^{-\xi} - 1}{(1-p_2)^{-\xi} - 1}.$$

Une fois l'estimation terminée, on peut vérifier graphiquement la pertinence des estimations en comparant la distribution GPD estimée avec la distribution empiriquement observée des observations au-delà du seuil. D'un point de vue pratique l'estimation des paramètres (ξ, β) pour un seuil u fixé conduit à l'estimation suivante de la distribution au-delà du seuil : $\hat{S}_u(x) \approx S_{\hat{\xi}, \hat{\beta}(u)}(x)$. On notera que dans l'approximation proposée par cette formule le paramètre β dépend du seuil u choisi.

3. Application : l'espérance au-delà d'un seuil

L'espérance au-delà du seuil u (ou « moyenne en excédent ») est définie pour une distribution F par :

$$e_X(u) = E(X - u | X > u).$$

Il s'agit donc de l'espérance de la distribution conditionnelle introduite ci-dessus :

$$e_X(u) = \int_0^{\omega_F} S_u(x) dx.$$

Dans le cas particulier où X est distribué selon une loi GPD de paramètres (ξ, β) , on a vu dans le paragraphe 2.1 que $e_X(u) = \frac{\beta + \xi u}{1 - \xi}$. Cette espérance n'existe que lorsque $\xi < 1$, on supposera donc dans la suite que cette condition est satisfaite.

3.1. Estimateurs de l'espérance au-delà du seuil

En pratique, si N_u est le nombre d'observations au-dessus du seuil u , on a l'estimateur empirique suivant de $e(u)$:

$$\hat{e}_1(u) = \frac{1}{N_u} \sum_{j=1}^n (x_j - u) 1_{\{x > u\}}(x_j).$$

On peut également proposer comme estimateur :

$$\hat{e}_2(u) = \frac{\hat{\beta}(u)}{1 - \hat{\xi}},$$

avec les paramètres (ξ, β) estimés par le maximum de vraisemblance (cf. le paragraphe 2.2). Enfin, une troisième méthode permet de se ramener à l'estimation du seul paramètre ξ , dans le cas particulier où $\xi > 0$. Pour cela on observe que la fonction de survie de la distribution peut se mettre sous la forme

$$S(x) = 1 - F(x) = x^{-1/\xi} L(x),$$

où L est une fonction « à variation lente » à l'infini au sens où

$$\lim_{t \rightarrow \infty} \frac{L(tx)}{L(t)} = 1.$$

Avec le seuil u , on peut donc écrire que

$$S_u(x) = \frac{S(x)}{S(u)} = \frac{L(x)}{L(u)} \left(\frac{x}{u}\right)^{-1/\xi}$$

et en considérant que le rapport des fonctions à variation lente est proche de 1 on trouve :

$$S_u(x) \approx \left(\frac{x}{u}\right)^{-1/\xi}.$$

Dès lors que $\xi < 1$ on trouve donc que $e_x(u) = \int_0^{+\infty} S_u(u+t) dt \approx \frac{\xi u}{1-\xi}$. Il suffit donc de déterminer un estimateur de ξ , par exemple l'estimateur de Hill, pour proposer comme estimateur de l'espérance au-delà de u :

$$\hat{e}_3(u) = \frac{\hat{\xi} u}{1-\hat{\xi}}.$$

$\hat{\xi}$ peut être choisi estimé pour un seuil $u_0 \leq u$ fixé, ou ré-estimé pour chaque valeur de u , ce qui conduit alors à $\hat{\xi} = \hat{\xi}(u)$.

3.2. Illustration : cas de la loi de Pareto

On illustre la mise en œuvre de ces estimateurs dans le cas d'une loi de Pareto de fonction de répartition $F(x) = 1 - \left(\frac{x_m}{x}\right)^\alpha$ pour $x \geq x_m$. Dans cette configuration l'approximation GDP est en fait exacte :

$$S_u(x-u) = \Pr(X > x | X > u) = \left(\frac{x}{u}\right)^{-\alpha} = \left(1 + \frac{\xi}{\beta}(x-u)\right)^{-1/\xi}$$

en posant $\xi = \frac{1}{\alpha}$ et $\beta(u) = \xi u$. Dans ce cas particulier, du fait de l'invariance par changement de seuil de la loi de Pareto, l'estimation est en fait ramenée à l'estimation du seul paramètre $\xi = \frac{1}{\alpha}$.

3.3. Détermination du seuil

On déduit de la propriété 3 que, pour un seuil u fixé tel que $S_u(x) \approx S_{\xi, \beta(u)}(x)$ et si l'on considère $v \geq u$, alors en notant $Y = [X - u | X > u]$, on a l'égalité en loi

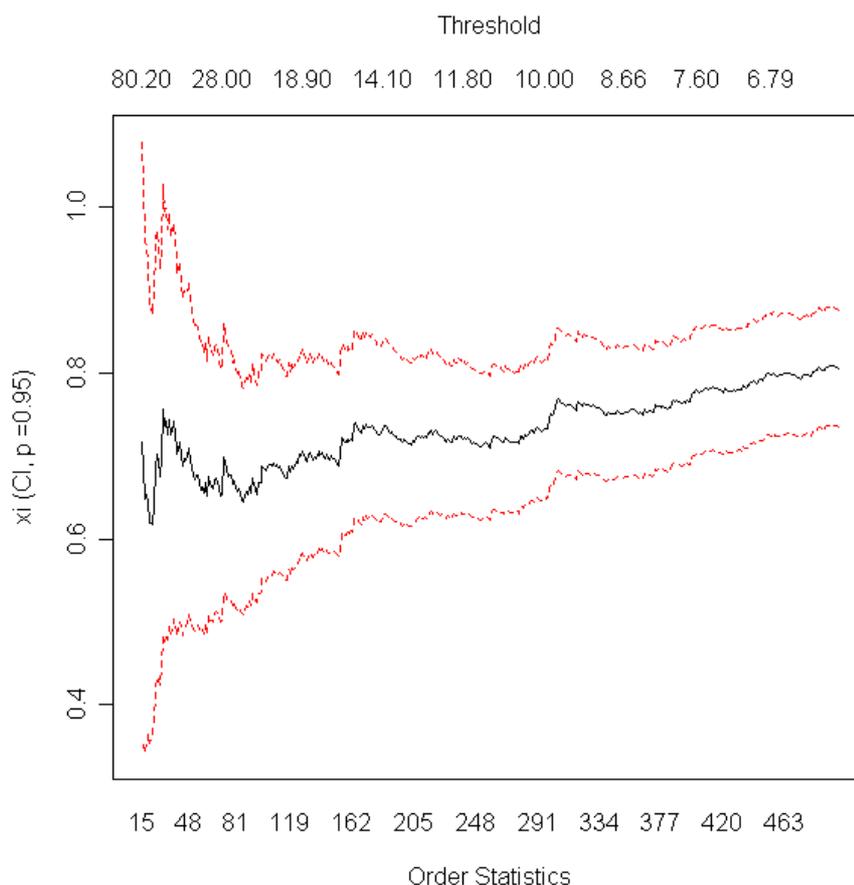
$$[X - v | X > v] = [Y - (v-u) | Y > v-u],$$

et comme Y suit une loi de Pareto généralisée $(\alpha, \beta(u))$:

$$E[X - v | X > v] = \frac{\beta(u) + \xi \times (v-u)}{1-\xi}.$$

Ainsi, l'espérance conditionnelle des excès est une fonction affine de v , lorsque v est supérieur au seuil de référence u . Cela donne un moyen empirique de tester si le seuil u choisi pour les calculs est suffisamment élevé : les espérances des excès au-delà des seuils plus élevés doivent être alignées sur une droite de pente $\frac{\xi}{1-\xi}$.

Une approche alternative consiste à calculer l'estimateur de Hill de l'indice de queue pour différents seuils et à chercher à partir de quel seuil cet estimateur est approximativement constant :



4. Applications numériques

4.1. Méthodologie

Les applications présentées ici sont réalisées avec le logiciel R². L'objectif est de comparer les trois estimateurs proposés :

$$\hat{e}_1(u) = \frac{1}{N_u} \sum_{j=1}^n (x_j - u) 1_{\{x_j > u\}}(x_j),$$

$$\hat{e}_2(u) = \frac{\hat{\beta}_{GPD}(u)}{1 - \hat{\xi}_{GPD}}, \text{ et}$$

$$\hat{e}_3(u) = \frac{\hat{\xi}_{HILL}}{1 - \hat{\xi}_{HILL}} u.$$

L'approche utilisée consiste à simuler un échantillon de données selon une loi connue, puis à comparer, pour différents seuils, la valeur théorique $e(u)$ avec les résultats des estimations. En pratique cette technique sera appliquée dans deux contextes :

- lorsque la loi sous-jacente appartient au domaine de Fréchet : $\xi > 0$ et les estimateurs \hat{e}_2 et \hat{e}_3 sont justifiés théoriquement ;
- lorsque la loi sous-jacente appartient au domaine de Gumbel : $\xi = 0$, et dans ce cas les estimateurs \hat{e}_2 et \hat{e}_3 n'ont pas de sens ; ils seront calculés afin de porter un avis sur la robustesse de ces techniques.

² Le code R est disponible sur le site <http://www.isfa.info/r>.

Le choix de $\xi > 0$ ou $\xi = 0$ est complexe et nécessite des investigations dépassant le cadre de du présente article. Une attention particulière est portée à la loi log-normale, celle-ci étant à la fois à variation lente et sous-exponentielle, ce qui lui donne un comportement intermédiaire entre les lois « à queue légère » et celles « à queue lourde ».

4.2. Implémentation dans le logiciel R

Les fonctions R utilisées sont détaillées ci-après. Ces fonctions supposent qu'un échantillon X est disponible. Les calculs sont effectués pour tous les seuils associés aux quantiles de 90 % à 99,5 %, avec un pas de 0,5 %. Ces seuils sont obtenus *via* la fonction R suivante :

```
seuils=function(X){
# Renvoie un tableau avec les 20 quantiles de 90% à 99,5%
  seuils=as.vector(1:20)
  for (i in 1:20){
    seuils[i]=quantile(X,.895+i/200)
  }
  seuils
}
```

Ce sont donc des quantiles empiriques.

4.2.1. Estimateur empirique

```
e_emp=function(X,seuils){
# Ajustement empirique de l'espérance au-delà d'un seuil
# Retourne un tableau avec l'estimation du mean XS pour les quantiles
définis dans "seuils"

  e_est=as.vector(1:length(seuils))
  for (i in 1:length(seuils)){
    e_est[i]=mean(X[X>seuils[i]])
  }
  plot(seuils, e_est)
  res=data.frame(e_est)
  res
}
```

4.2.2. Estimateur GPD

```
e_gpd=function(X,seuils){
# Ajustement GPD de l'espérance au-delà d'un seuil
# Retourne un tableau avec l'estimation du mean XS pour les quantiles
définis dans "seuils"

  xi_est=as.vector(1:length(seuils))
  beta_est=as.vector(1:length(seuils))
  e_gpd=as.vector(1:length(seuils))
  library(evir)
  for (i in 1:length(seuils)){
    w=gpd(X,seuils[i])
    fitgpd=w$par.ests
    xi_est[i]=fitgpd[1]
    beta_est[i]=fitgpd[2]
  }

  e_est=beta_est/(1-xi_est)
  plot(seuils,xi_est)
```

```

    res=data.frame(e_est,xi_est,beta_est)
  res
}

```

4.2.3. Estimateur de Hill

```

e_hill=function(X,seuils){
# Ajustement Hill de l'espérance au-delà d'un seuil
# Retourne un tableau avec l'estimation du mean XS pour les quantiles
définis dans "seuils"

  xi_est=as.vector(1:length(seuils))
  e_est=as.vector(1:length(seuils))
  library(evir)
  #On ne travaille que sur les 10% de valeurs les plus élevées de
  X
  p=.1
  u=quantile(X,1-p)
  Xu=X[X>u]
  Xs=rev(sort(Xu))
  esthill=hill(Xs,"xi")
  for (i in 1:length(seuils)){
    k=trunc((.895+i/(length(seuils)*10))*length(Xs))
    xi_est[i]=esthill$y[esthill$x==k]
  }
  e_est=xi_est*seuils/(1-xi_est)
  plot(seuils,xi_est)
  res=data.frame(e_est,xi_est)
  res
}

```

Dans le code ci-dessus, on limite l'estimation au sous-échantillon composé des 10 % de valeurs les plus élevées.

4.2.4. Valeurs théoriques

Les valeurs théoriques utilisées comme référence sont fournies par la fonction suivante :

```

e_th=function(dist,param,seuils){

  #Loi exponentielle
  if (dist=="dexp"){
    e_th=1/param[1]*(seuils/seuils)
  }
  #Loi normale
  if (dist=="dnorm"){
    e_th=1
  }
  #Loi log-normale
  if (dist=="dlnorm"){
    s=param[1]
    e_th=s^2*seuils/log(seuils)
  }
  #Loi de Weibull (Fréchet)
  if (dist=="dweibull"){
    xi=param[1]
    e_th=xi*seuils^(1-1/xi)
  }
  #Loi GPD
  if (dist=="dgpdp"){

```

```

        xi=param[1]
        beta=param[2]
        e_th=(beta+xi*seuils)/(1-xi)
    }
    res=data.frame(e_th)
    res
}

```

Les seuils passés en paramètre à cette fonction ne sont pas les seuils empiriques, mais les seuils théoriques, calculés à partir des fonctions *qlnorm*, *qgpd*, etc.

4.3. Résultats

Les résultats présentés ici ont vocation à illustrer la démarche ; il convient de réaliser des simulations complémentaires dans le contexte propre à chaque situation particulière (tant en termes de lois sous-jacentes que de taille d'échantillon) afin de retenir la méthode la mieux adaptée.

On utilise le code R suivant (pour le cas de la loi GPD, l'adaptation à d'autres lois est simple) :

```

library(evir)
xi_th=.75
beta_th=1
n_obs=10^6
X=rgpd(n_obs,xi_th,0,beta_th)
source("C:\\Logiciel R\\EVT\\seuils.r")
seuils=seuils(X)
source("C:\\Logiciel R\\EVT\\e_gpd.r")
res=e_gpd(X,seuils)
source("C:\\Logiciel R\\EVT\\e_hill.r")
res2=e_hill(X,seuils)
source("C:\\Logiciel R\\EVT\\e_emp.r")
res3=e_emp(X,seuils)
source("C:\\Logiciel R\\EVT\\e_th.r")
p=as.vector(1:2)
p[1]=xi_th
p[2]=beta_th
for (i in 1:20){
    seuils[i]=.895+i/200
}
seuils=qgpd(seuils,xi_th,0,beta_th)
res4=e_th("dgpdp",p,seuils)

```

Deux séries de simulations sont présentées ci-après : l'une avec 10^6 réalisations, effectif non disponible en pratique, et l'autre avec 5 000 réalisations, ce qui s'avère proche de la taille des échantillons de données disponibles.

4.3.1. Echantillons issus d'une loi GPD

Tableau 1 - Résultats basés sur un échantillon de 10^6 tirages de la loi GPD

Quantile	Estimateur GPD			Estimateur Hill		Estimateur empirique	Valeur théorique
	e_est	xi_est	beta_est	e_est	xi_est	e_est	e_th
90,0%	22,82	0,75	5,66	30,55	0,83	26,79	22,49
90,5%	23,70	0,75	5,88	32,03	0,83	27,87	23,38
91,0%	24,77	0,75	6,11	33,68	0,83	29,05	24,34
91,5%	25,90	0,75	6,37	35,48	0,83	30,35	25,41
92,0%	27,08	0,75	6,68	37,56	0,83	31,79	26,59
92,5%	28,49	0,75	7,00	39,95	0,83	33,39	27,91
93,0%	29,91	0,75	7,40	42,64	0,83	35,19	29,39
93,5%	32,02	0,76	7,76	45,79	0,83	37,22	31,07
94,0%	34,30	0,76	8,21	49,22	0,84	39,54	32,99
94,5%	36,22	0,76	8,82	53,10	0,84	42,21	35,22
95,0%	38,65	0,75	9,52	57,71	0,84	45,35	37,83
95,5%	41,45	0,75	10,37	63,08	0,84	49,08	40,94
96,0%	44,97	0,75	11,38	69,59	0,84	53,59	44,72
96,5%	49,88	0,75	12,55	78,22	0,84	59,20	49,43
97,0%	55,57	0,74	14,21	89,10	0,84	66,37	55,49
97,5%	63,05	0,74	16,29	103,73	0,84	75,94	63,62
98,0%	73,62	0,74	19,32	124,31	0,84	89,47	75,21
98,5%	90,40	0,73	24,02	156,94	0,84	110,38	93,32
99,0%	119,43	0,73	32,73	216,18	0,84	148,05	126,49
99,5%	200,26	0,73	53,76	371,16	0,84	242,86	212,73

Tableau 2 - Résultats basés sur un échantillon de 5 000 tirages de la loi GPD

Quantile	Estimateur GPD			Estimateur Hill		Estimateur empirique	Valeur théorique
	e_est	xi_est	beta_est	e_est	xi_est	e_est	e_th
90,0%	21,33	0,78	4,63	20,56	0,78	29,46	22,49
90,5%	23,60	0,80	4,66	21,70	0,78	30,70	23,38
91,0%	24,83	0,80	4,85	22,18	0,78	32,06	24,34
91,5%	32,37	0,86	4,66	23,54	0,78	33,56	25,41
92,0%	41,01	0,89	4,68	25,04	0,78	35,23	26,59
92,5%	39,54	0,87	5,07	25,95	0,78	37,09	27,91
93,0%	56,79	0,91	5,06	27,68	0,78	39,20	29,39
93,5%	101,55	0,95	5,09	29,85	0,78	41,58	31,07
94,0%	87,50	0,94	5,62	31,26	0,78	44,33	32,99
94,5%	85,19	0,93	6,16	33,58	0,78	47,55	35,22
95,0%	-707,40	1,01	5,96	36,21	0,78	51,33	37,83
95,5%	207,88	0,97	7,07	38,17	0,78	55,88	40,94
96,0%	-188,70	1,04	7,12	42,67	0,78	61,47	44,72
96,5%	-47,04	1,15	7,02	46,96	0,78	68,47	49,43
97,0%	-80,05	1,11	8,83	52,71	0,79	77,63	55,49
97,5%	455,07	0,97	13,12	58,45	0,79	90,23	63,62
98,0%	-237,98	1,06	14,13	74,28	0,79	108,57	75,21
98,5%	-133,02	1,13	17,46	94,40	0,80	137,47	93,32
99,0%	-234,57	1,12	28,61	125,14	0,79	192,11	126,49
99,5%	582,39	0,84	90,45	220,16	0,79	341,64	212,73

On note tout d'abord la mauvaise qualité de l'estimateur empirique, et ce d'autant plus que la taille de l'échantillon est faible. En effet, cet estimateur est beaucoup trop volatile et sensible aux grandes valeurs de l'échantillon. Les résultats obtenus pour les estimateurs GPD et Hill sont peu intuitifs :

- dans le cas d'un gros échantillon, l'estimateur GPD domine l'estimateur de Hill,
- la situation est inverse dans le cas d'un « petit » échantillon.

On peut également relever que l'estimation du paramètre de queue s'avère particulièrement délicate et que l'on peut avoir une estimation correcte de l'espérance au-delà du seuil avec une mauvaise estimation de ξ .

4.3.2. Echantillons issus d'une loi log-normale

Dans ce cas, en toute rigueur, les estimateurs GPD et Hill ne s'appliquent pas, puisqu'on a $\xi = 0$; on utilise toutefois mécaniquement les algorithmes proposés pour tester leur robustesse. On obtient les résultats suivants :

Tableau 3 - Résultats basés sur un échantillon de 10⁶ tirages de la loi log-normale

Quantile	Estimateur GPD			Estimateur Hill		Estimateur empirique	Valeur théorique
	e_est	xi_est	beta_est	e_est	xi_est	e_est	e_th
90,0%	2,82	0,26	2,09	3,13	0,46	6,42	2,81
90,5%	2,87	0,26	2,13	3,23	0,47	6,57	2,83
91,0%	2,91	0,26	2,17	3,33	0,47	6,72	2,85
91,5%	2,95	0,25	2,20	3,45	0,47	6,89	2,87
92,0%	3,00	0,25	2,24	3,57	0,47	7,07	2,90
92,5%	3,05	0,25	2,28	3,70	0,47	7,27	2,93
93,0%	3,11	0,25	2,33	3,85	0,47	7,48	2,96
93,5%	3,16	0,25	2,37	4,01	0,47	7,71	3,00
94,0%	3,23	0,25	2,43	4,19	0,47	7,96	3,04
94,5%	3,30	0,25	2,48	4,38	0,47	8,25	3,09
95,0%	3,38	0,25	2,55	4,59	0,47	8,57	3,15
95,5%	3,48	0,24	2,62	4,84	0,47	8,93	3,21
96,0%	3,58	0,24	2,72	5,12	0,47	9,34	3,29
96,5%	3,70	0,24	2,80	5,46	0,47	9,83	3,38
97,0%	3,84	0,24	2,92	5,85	0,47	10,41	3,49
97,5%	4,01	0,24	3,03	6,36	0,47	11,12	3,62
98,0%	4,24	0,24	3,24	6,98	0,47	12,04	3,80
98,5%	4,54	0,24	3,45	7,86	0,47	13,30	4,04
99,0%	5,00	0,23	3,83	9,19	0,47	15,23	4,40
99,5%	5,92	0,22	4,62	11,76	0,47	18,98	5,10

Tableau 4 - Résultats basés sur un échantillon de 5 000 tirages de la loi log-normale

Quantile	Estimateur GPD			Estimateur Hill		Estimateur empirique	Valeur théorique
	e_est	xi_est	beta_est	e_est	xi_est	e_est	e_th
90,0%	2,52	0,20	2,02	2,78	0,44	6,09	2,81
90,5%	2,54	0,20	2,05	2,85	0,44	6,22	2,83
91,0%	2,58	0,19	2,09	2,91	0,44	6,36	2,85
91,5%	2,63	0,17	2,18	2,97	0,43	6,51	2,87
92,0%	2,66	0,17	2,21	3,09	0,44	6,67	2,90
92,5%	2,68	0,18	2,21	3,24	0,44	6,84	2,93
93,0%	2,70	0,19	2,19	3,38	0,44	7,02	2,96
93,5%	2,77	0,16	2,31	3,48	0,44	7,23	3,00
94,0%	2,81	0,16	2,35	3,67	0,44	7,45	3,04
94,5%	2,86	0,15	2,42	3,80	0,44	7,70	3,09
95,0%	2,85	0,19	2,32	4,04	0,44	7,97	3,15
95,5%	2,87	0,21	2,27	4,25	0,44	8,27	3,21
96,0%	2,96	0,20	2,35	4,43	0,44	8,61	3,29
96,5%	3,06	0,19	2,47	4,71	0,44	9,01	3,38
97,0%	3,19	0,16	2,66	4,99	0,44	9,49	3,49
97,5%	3,40	0,11	3,03	5,31	0,44	10,10	3,62
98,0%	3,44	0,13	2,99	5,91	0,44	10,85	3,80
98,5%	3,46	0,18	2,84	6,67	0,44	11,85	4,04
99,0%	3,68	0,20	2,93	7,63	0,44	13,30	4,40
99,5%	4,29	0,49	2,19	9,68	0,44	15,93	5,10

On retrouve les très mauvais résultats de l'estimateur empirique. De manière assez surprenante, l'estimateur GPD donne de bons résultats pour les 2 tailles d'échantillons, les résultats sont peu convaincants pour l'estimateur de Hill.

Au global, les résultats obtenus sur un échantillon de 5 000 valeurs peuvent être assez variables, mais tendent toutefois à montrer que l'estimateur GPD est plus robuste. L'estimateur de Hill s'avère de son côté délicat à utiliser.

5. Conclusion

L'estimation robuste de l'espérance au-delà d'un seuil s'avère un exercice délicat, à tout le moins lorsque le volume de données disponibles est modeste. L'utilisation d'un estimateur empirique doit être écartée, du fait de l'imprécision et de la volatilité de celui-ci.

Dans un cadre non paramétrique, l'estimateur GPD fournit des résultats globalement satisfaisants.

Mais, quel que soit l'estimateur retenu (GPD ou Hill), son utilisation comporte un risque relativement important sur un échantillon de l'ordre de 5 000 valeurs. Dans ce contexte, on peut envisager de se tourner vers une modélisation paramétrique de la distribution globale, en utilisant une expression explicite de l'espérance au-delà du seuil, que l'on peut alors calculer (au moins de manière approchée).

Les distributions log-normales, Benktender, ou des lois mélangées de type log-normale / Pareto constituent des modélisations intéressantes à ce titre. Au surplus, le modèle retenu peut être étalonné par comparaison avec les tarifs des traités de réassurance en excédent.

Bibliographie

- Danielson J., De Vries C.G. (1997) « Extreme returns, tail estimator and Value-at-Risk », Working paper, London School of Economics.
- De Haan L., Rotzen H. (1993) « On the estimation of high quantiles », *Journal of Statistical Planning and Inference* 35 (1), 1-13.
- Embrechts P., Kluppelberg C., Mikosch T. (1997) *Modelling extremal events*, Berlin: Springer Verlag.
- Garrido M. (2002) *Modélisation des évènements rares et des quantiles extrêmes, méthodes de sélection de modèle pour les queues de distribution*, Thèse de doctorat, Université J. Fourier.
- Hosking J. R., Wallis J. R. (1987) « Parameter and quantile estimation for the generalized pareto distribution », *Technometrics* 29, 339-49.
- Partrat C., Besson J.L. (2005) *Assurance non-vie, modélisation, simulation*, Paris : Economica.
- Pickands J. (1975) « Statistical inference using extreme order statistics », *Annals of Statistics* 3, 119-31.
- Planchet F., Thérond P.E. (2007) « Model risk and determination of economic capital in the Solvency 2 project », *Proceedings of the 17th AFIR colloquium*.
- Resnick S., Starica C. (1997) « Smoothing the Hill estimator », *Advanced Applied Probabilities* 29, 271-93
- Thérond P.E., Planchet F. (2007) « Provisions techniques et capital de solvabilité d'une compagnie d'assurance : méthodologie d'utilisation de Value-at-Risk », *Assurances et gestion des risques* 74 (4), 533-63.