

## Actuariat de l'Assurance Non-Vie # 8

A. Charpentier (UQAM & Université de Rennes 1)



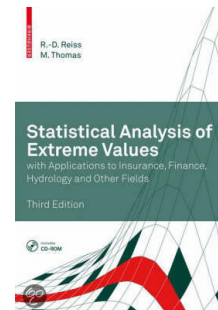
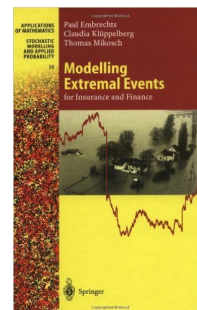
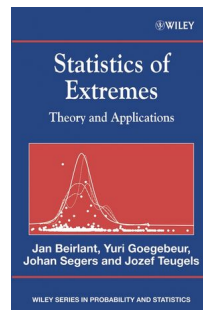
ENSAE ParisTech, Octobre / Décembre 2016.

<http://freakonometrics.hypotheses.org>

## Les grands risques

On avait vu dans la section #7 l'importance de l'écrêtement des grands risques.  
On va revenir ici sur la modélisation de ces coûts importants.

**Références:** Beirlant *et al.* (2004), Embrechts *et al.* (2013) ou Reiss & Thomas (2007)



## Le théorème central limite des extrêmes

**Théorème** Supposons qu'il existe des constantes de normalisation  $a_n \in \mathbb{R}$  et  $b_n > 0$ , et une loi non-dégénérée  $H$  telles que

$$b_n^{-1} \{X_{n:n} - a_n\} \xrightarrow{\mathcal{L}} H.$$

Alors  $H$  est du même type qu'une trois lois suivantes (données par leur fonction de répartition),

1. Loi de **Fréchet**,  $\Phi_\xi(x) = \exp(-x^{-\xi}) \mathbb{I}(x > 0)$ ,  $\xi > 0$ ,
2. Loi de **Weibull**,  $\Psi_\xi(x) = \exp(-x^{-\xi})$  si  $x \leq 0$ , et 1 sinon,  $\xi > 0$ ,
3. Loi de **Gumbel**,  $\Lambda(x) = \exp(-\exp(-x))$ .

On notera que

$$X \sim \Phi_\xi \iff \log X^\xi \sim \Lambda \iff -1/X \sim \Psi_\xi.$$

## Les trois lois limites

Ces trois lois sont en fait les trois cas particulier de la distribution **GEV - Generalized Extreme Value** (représentation de Jenkinson-von Mises)

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

dès lors que  $\mu + \xi x/\sigma > 0$ .

**Remarque** S'il existe des constantes de normalisation  $a_n \in \mathbb{R}$  et  $b_n > 0$ , et une loi non-dégénérée  $\text{GEV}(\xi)$  telles que

$$b_n^{-1} \{X_{n:n} - a_n\} \xrightarrow{\mathcal{L}} \text{GEV}(\xi).$$

on dira que  $F_X$  appartient au **max-domain d'attraction** (MDA) de  $\text{GEV}(\xi)$ .

## Les grands risques

On peut définir des risques **sous-exponentiels**, si  $X_i$  est une suite de variables i.i.d.

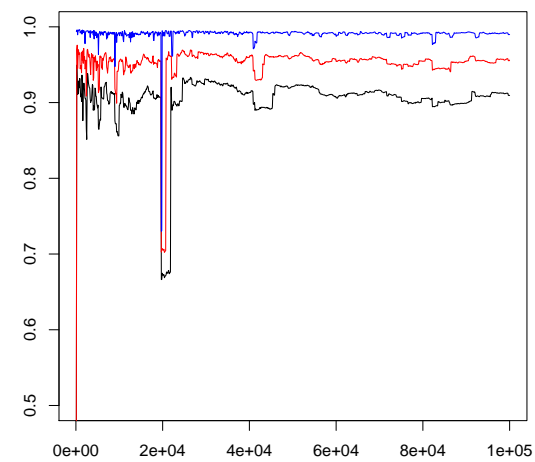
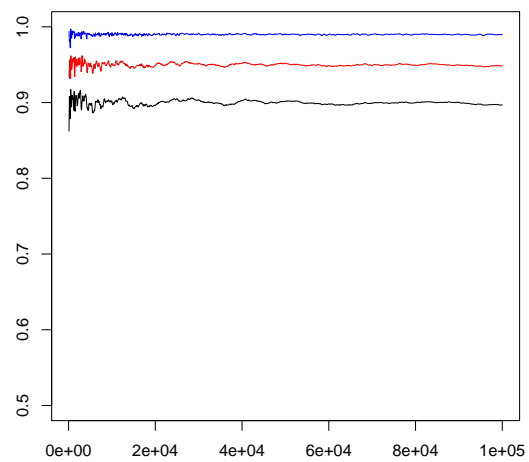
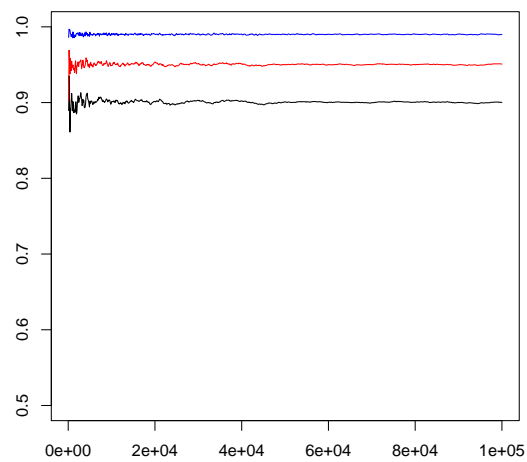
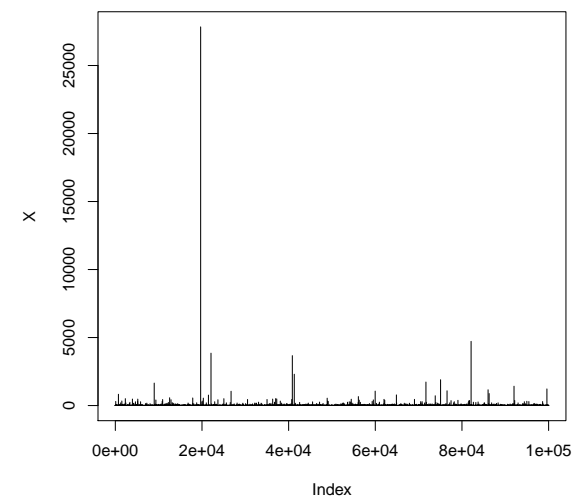
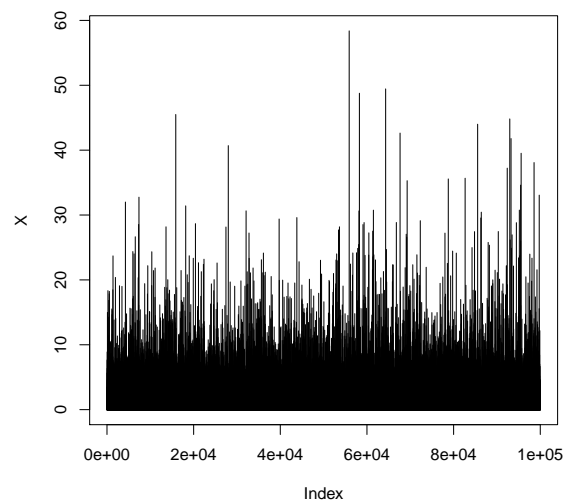
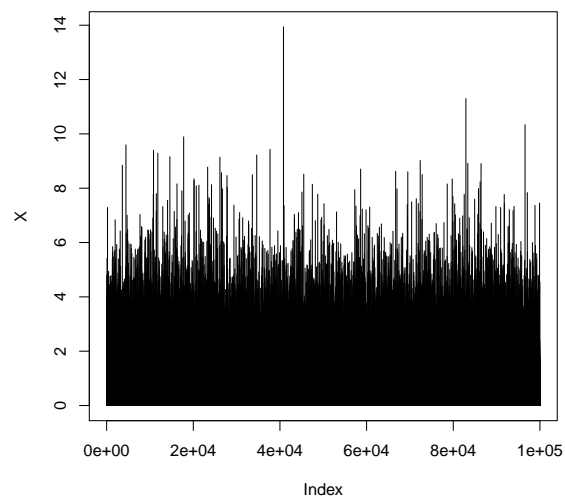
$$\lim_{x \rightarrow \infty} \frac{\mathbb{P}[X_1 + \dots + X_n > x]}{\mathbb{P}[\max\{X_1, \dots, X_n\} > x]} = 1, n \geq 1.$$

**Définition** Si  $F_X$  est une fonction de répartition continue d'espérance  $\mathbb{E}[X]$ , on définit l'indice de grands risques par

$$D_{F_X}(p) = \frac{1}{\mathbb{E}[X]} \int_{1-p}^1 F_X^{-1}(t) dt \text{ pour } p \in [0, 1].$$

La version empirique est alors  $T_n(p)$  la proportion des  $[np]$  plus gros sinistres par rapport à la somme totale, i.e.

$$D_n(p) = \frac{X_{1:n} + X_{2:n} + \dots + X_{[np]:n}}{X_1 + \dots + X_n} \text{ où } \frac{1}{n} < p \leq 1.$$



**Proposition**(théorème de Pickands-Balkema-de Haan) soit  $\xi \in \mathbb{R}$ , les deux résultats suivants sont équivalents,

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

$$\lim_{n \rightarrow \infty} \mathbb{P}(X_{n:n} \leq a_n x + b_n) = H_\xi(x), \quad x \in \mathbb{R}.$$

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

$$\begin{aligned} \lim_{u \rightarrow \infty} \frac{\overline{F}(u + xa(u))}{\overline{F}(u)} &= \lim_{u \rightarrow \infty} \mathbb{P}\left(\frac{X - u}{a(u)} > x \mid X > u\right) \\ &= \begin{cases} (1 + \xi x)^{-1/\xi} & \text{if } \xi \neq 0, \\ \exp(-x) & \text{if } \xi = 0. \end{cases} \end{aligned}$$

Aussi,  $F \in MDA(G_\xi)$  si et seulement si il existe  $\sigma(\cdot) > 0$  telle que

$$\lim_{u \rightarrow x_F} \sup_{0 < x < x_F} \{|\mathbb{P}(X - u \leq x | X > u) - H_{\xi, \sigma(u)}(\leq x)|\} = 0$$

## Loi de type Pareto

La loi de  $X$  est à variation régulière d'indice  $\alpha \in (0, +\infty)$  si pour tout  $x$

$$\lim_{t \rightarrow \infty} \frac{\mathbb{P}[X > tx]}{\mathbb{P}[X > t]} = \lim_{t \rightarrow \infty} \frac{\overline{F}(tx)}{\overline{F}(t)} = x^{-\alpha}$$

ou encore

$$\mathbb{P}[X > x] = x^{-\alpha} \mathcal{L}(x) \text{ où } \mathcal{L} \text{ est à variation lente, i.e. } \lim_{t \rightarrow \infty} \frac{\mathcal{L}(tx)}{\mathcal{L}(t)} = 1.$$



## Estimateur de Hill et *Pareto Plot*

Si  $u$  est suffisamment grand

$$\mathbb{P}(X \leq x | X > u) \approx 1 - \left(1 + \xi \frac{x + u}{\sigma}\right)^{-1/\xi},$$

pour un  $\sigma > 0$ , soit

$$\log \left(1 - \frac{F(x)}{1 - F(u)}\right) \approx -\frac{1}{\xi} \log \left(1 + \xi \frac{x + u}{\sigma}\right) \approx -\frac{1}{\xi} \log x + \text{constant}.$$

Un estimateur naturel de  $\xi$  est alors la pente de la droite passant au mieux par le nuage des  $(\log X_i, \log(1 - \hat{F}(X_i)))$ .

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

## Quelques estimateurs classiques de $\xi$

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

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

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

où

$$\xi_{n,k}^{H(r)} = \frac{1}{k} \sum_{i=1}^{k-1} [\log X_{n-i:n} - \log X_{n-k:n}]^r, \quad r = 1, 2, \dots$$

## Lois à queue fine

- loi **exponentielle**,  $f(x) = \beta \exp(-\beta x)$ , pour  $\beta > 0$ ,
- loi **Gamma**,  $f(x) = \beta^\alpha \Gamma(\alpha)^{-1} x^{\alpha-1} \exp(-\beta x)$  pour  $\alpha, \beta > 0$ ,
- loi de **Weibull**,  $f(x) = c\beta x^{\beta-1} \exp(-cx^\beta)$  pour  $c > 0$  et  $\beta \geq 1$ ,

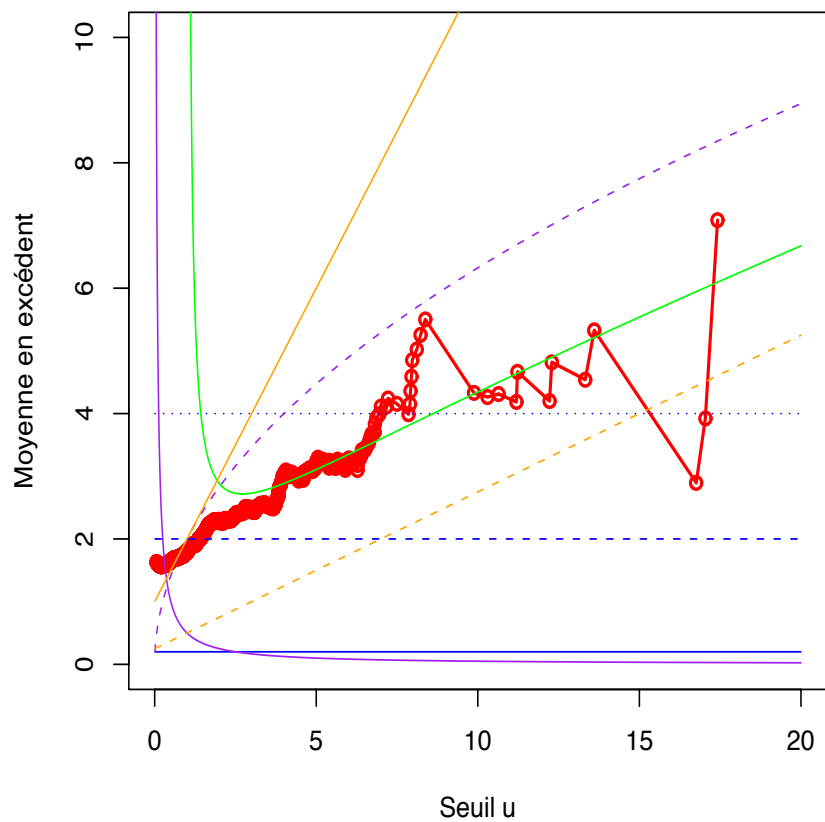
## Lois à queue épaisse

- loi de **Weibull**,  $f(x) = c\beta x^{\beta-1} \exp(-cx^\beta)$  pour  $c > 0$  et  $\beta < 1$ ,
- loi **lognormale**,  $f(x) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\left(-\frac{(\log x - \mu)^2}{2\sigma^2}\right)$  pour  $\sigma > 0$
- loi de **Pareto**,  $f(x) = \frac{\alpha}{\theta+x} \left(\frac{\theta}{\theta+x}\right)^\alpha$  pour  $\alpha, \theta > 0$
- loi de **Burr**,  $f(x) = (\theta + x^\gamma)^{-(\alpha+1)} \alpha \gamma \theta^\alpha x^{\gamma-1}$  pour  $\alpha, \gamma, \theta > 0$ .

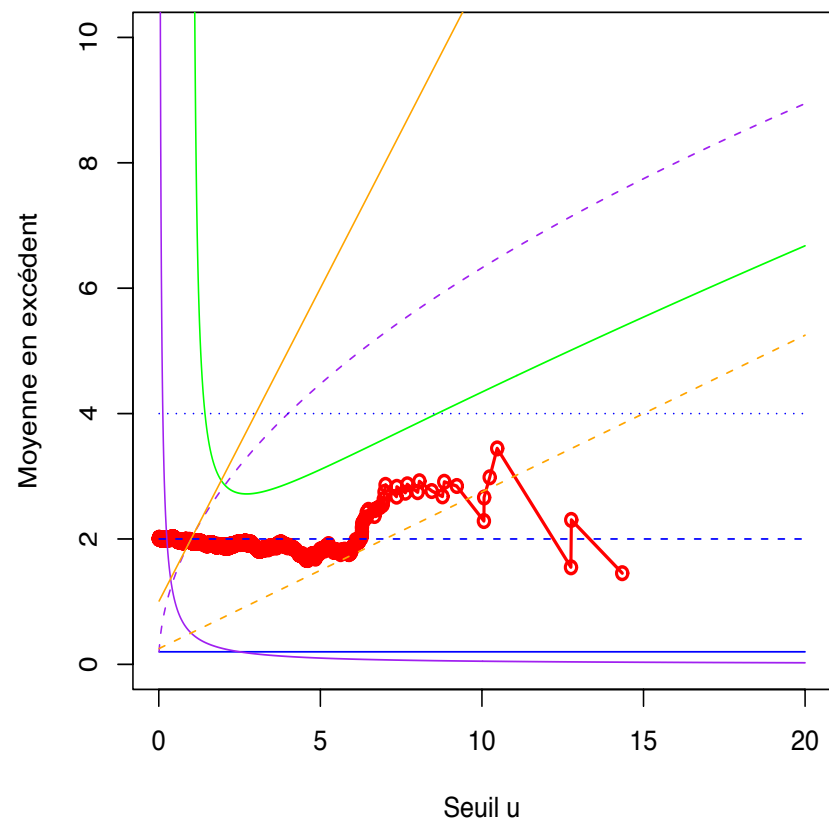
## La fonction de moyenne en excès

	Fonction de répartition	Fonction en excédent
Loi lognormale	$\int_0^x \frac{1}{\sqrt{2\pi}\sigma t} \exp\left(-\frac{(\log t - \mu)^2}{2\sigma^2}\right)$	$\sim \frac{\sigma^2 u}{\log u - \mu}$
Exponentielle	$1 - \exp(-\lambda x)$	$\frac{1}{\lambda}$
Pareto	$1 - \left(\frac{\theta}{\theta+x}\right)^\alpha$	$\frac{\theta+x}{\alpha-1}$
Gamma	$\int_0^x \beta (\beta t)^{\alpha-1} \frac{\exp(-\beta t)}{\Gamma(\alpha)} dt$	$\sim \frac{1}{\beta}$
Weibull	$1 - \exp(-\lambda x^\beta)$	$\sim \frac{u^{1-\beta}}{\lambda\beta}$

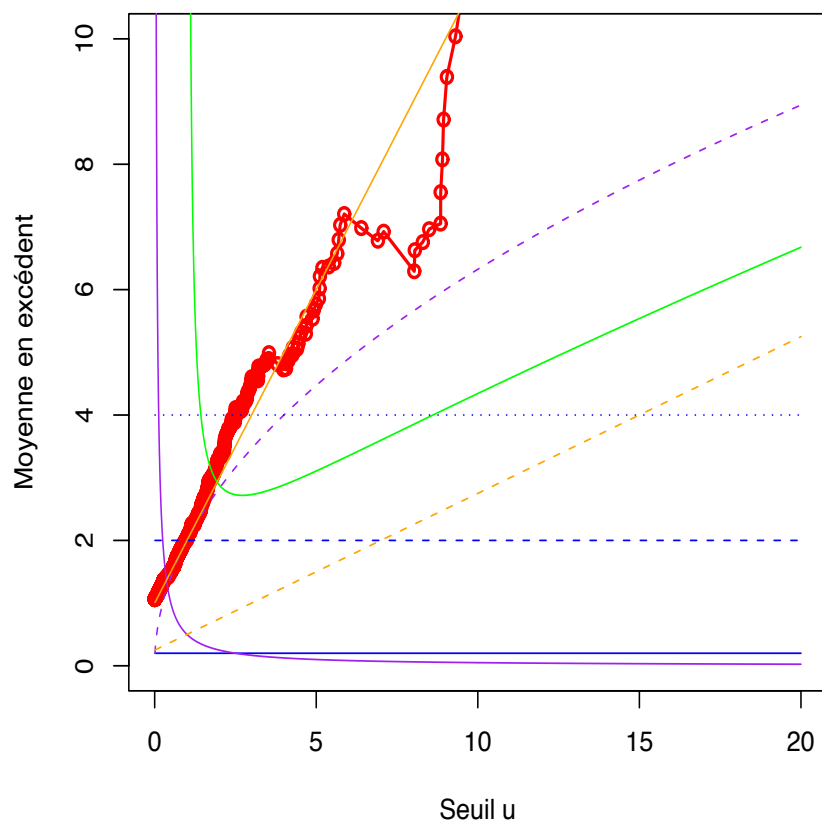
Fonction de moyenne en excès – loi log-normale



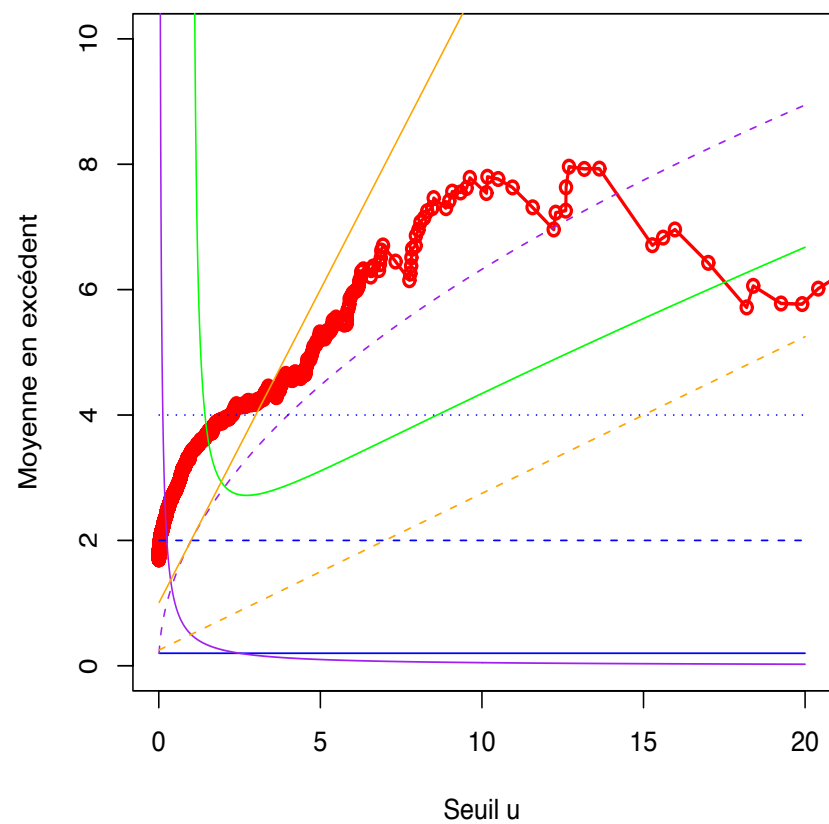
Fonction de moyenne en excès – loi exponentielle



Fonction de moyenne en excès – loi Pareto



Fonction de moyenne en excès – loi Weibull



## Estimation de quantiles extrêmes

Le **quantile** d'ordre  $p \in ]0, 1[$  - correspondant à la **Value-at-Risk** - associé à  $X$ , de fonction de répartition  $F_X$ , se définit par

$$VaR(X, p) = x_p = F^{-1}(p) = \sup\{x \in \mathbb{R}, F(x) \geq p\}.$$

Notons que les assureurs parlent plutôt de **période de retour**. Si  $T$  est la première année où on observe un phénomène qui se produit annuellement avec probabilité  $p$ . Alors

$$\mathbb{P}[T = k] = [1 - p]^{k-1}p \text{ de telle sorte que } \mathbb{E}[T] = \frac{1}{p}$$

Parmi les autres mesures pertinentes (et intéressantes en réassurance), on retiendra la  $TVaR$  - ou  $ES$  - correspondant à la valeur moyenne sachant que la  $VaR$  a été dépassée,

$$TVaR(X, p) = \mathbb{E}(X | X > VaR(X, p)).$$

## Utilisation de l'approximation GPD pour estimer $Var(X, p)$

La première méthode ([Smith \(1987\)](#)) repose sur la modélisation de la queue de la distribution, au delà d'un seuil  $u$ . Notons

$$N_u = \sum_{i=1}^n \mathbb{I}(X_i > u),$$

le nombre de dépassement de  $u$  dans un échantillon  $X_1, \dots, X_n$ . Pour  $x > u$ ,

$$\bar{F}(x) = \mathbb{P}(X > x) = \mathbb{P}(X > u) \mathbb{P}(X > x | X > u) = \bar{F}(u) \mathbb{P}(X > x | X > u),$$

où

$$\mathbb{P}(X > x | X > u) = \bar{F}(x)_u(x - u)$$

avec  $F_u(t) = \mathbb{P}(X - u \leq t | X > u) \sim H_{\xi, \beta}(t)$ , pour des valeurs  $\xi$  et  $\beta$  ( $H_{\xi, \beta}$  désignant la loi GPD).



## Utilisation de l'approximation GPD pour estimer $VaR(X, p)$

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.

Aussi, un estimateur naturel de la  $VaR(X, p)$  est  $\widehat{Q}_p$  défini par

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

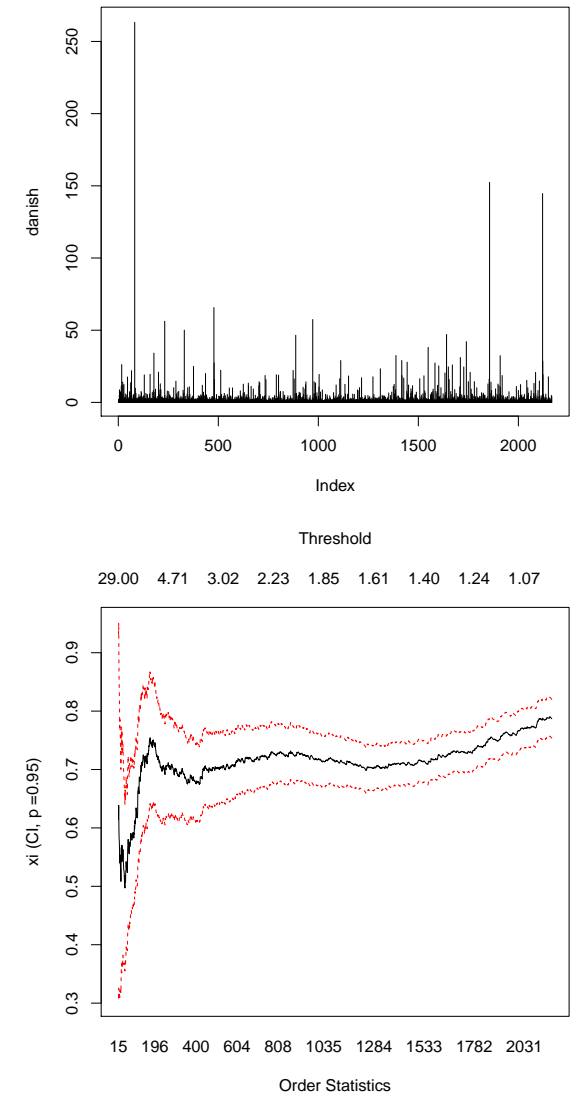
Notons qu'un intervalle de confiance asymptotique peut être obtenu en considérant la méthode de vraisemblance profilée (*profile likelihood*).

## Mise en oeuvre pratique, quantile bi-centenaire

```
1 > library(evir)
2 > data(danish)
3 > plot(danish, type="h")
```

on peut estimer  $\xi$  avec l'estimateur de Hill  $\hat{\xi}_{n,k}^{Hill}$ ,

```
1 > hill(danish, option="xi")
```



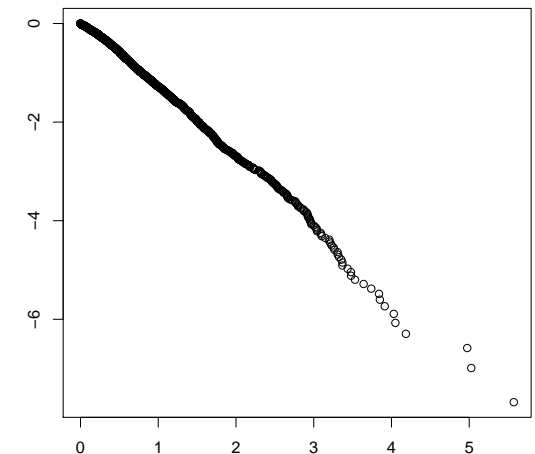
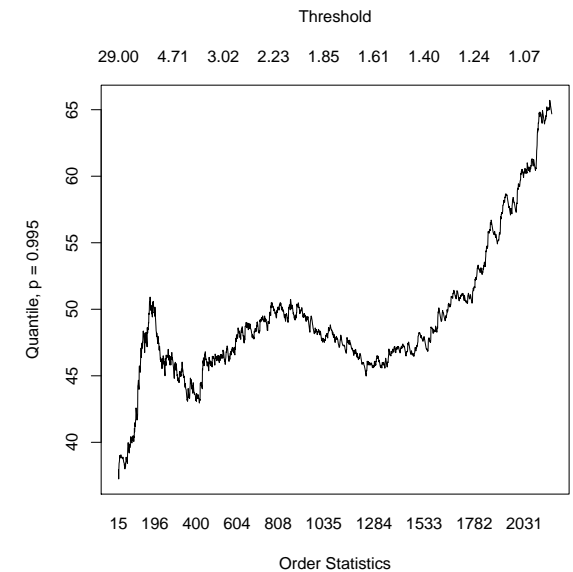
## Mise en oeuvre pratique, quantile bi-centenaire

ou directement un quantile  $\hat{Q}_p$  pour une probabilité élevée, e.g. = 99.5%

```
1 > hill(danish, option="quantile", p=.995)
```

Sinon, on peut utiliser l'approximation GPD. On effet, le **Pareto plot** semble confirmer une loi de Pareto

```
1 > n=length(danish)
2 > Xs=sort(danish, decreasing=TRUE)
3 > plot(log(Xs), log((1:n)/(n+1)))
```



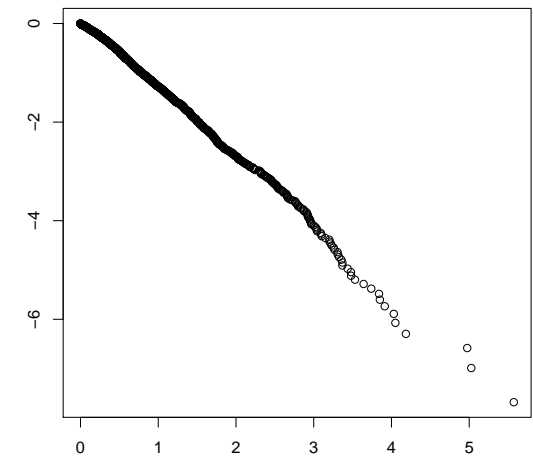
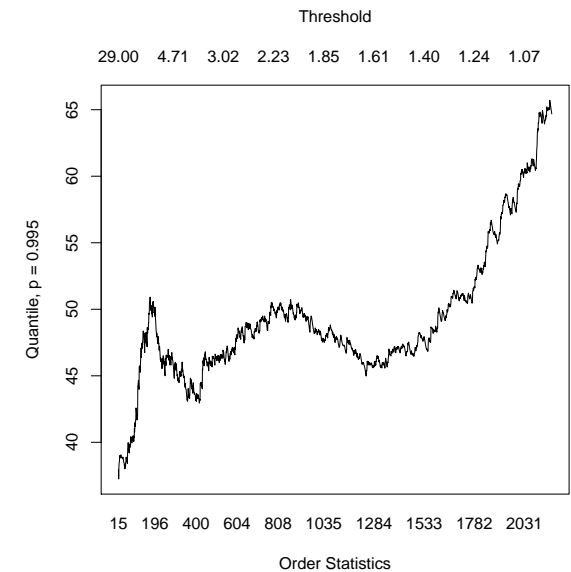
## Mise en oeuvre pratique, quantile bi-centenaire

L'estimation de la loi GPD donne

```

1 > u=5
2 > nu=sum(Xs>u)
3 > p=.995
4 > gpd(danish,5)
5 $par.ests
6           xi          beta
7 0.6320499  3.8074817
8
9 $par.ses
10          xi          beta
11 0.1117143  0.4637269

```



## Mise en oeuvre pratique, quantile bi-centenaire

La densité de la loi de Pareto, en fonction de  $\xi$  et  $\sigma$  est

$$g_{\xi,\sigma}(x) = \frac{1}{\sigma} \left( 1 + \frac{\xi x}{\sigma} \right)^{-\frac{1}{\xi}-1}$$

que l'on peut écrire en fonction de  $\xi$  et  $Q_p$ , si on considère une loi de Pareto au delà d'un seuil  $u$ ,

$$g_{\xi,Q_p}(x) = \frac{\left( \frac{n}{N_u}(1-p) \right)^{-\xi} - 1}{\xi[Q_p - u]} \left( 1 + \frac{\left( \frac{n}{N_u}(1-p) \right)^{-\xi} - 1}{[Q_p - u]} \cdot x \right)^{-\frac{1}{\xi}-1}$$

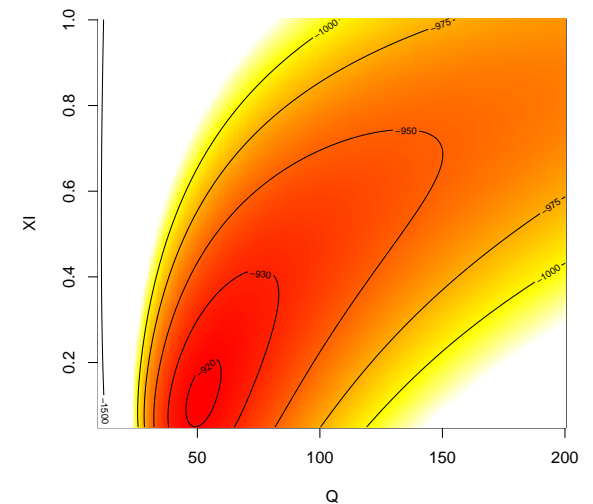
```
1 > gq=function(x,xi,q){
2 + ( (n/nu*(1-p) )^(-xi)-1)/(xi*(q-u))*(1+((n/nu*(1-p))^(-xi)-1)/(q-u)
    *x)^(-1/xi-1)}
```

## Mise en oeuvre pratique, quantile bi-centenaire

On cherche le maximum de vraisemblance

$$(\hat{\xi}, \hat{Q}_p) = \operatorname{argmax} \left\{ \underbrace{\sum_{i=0}^{N_u-1} \log[g_{\xi, Q_p}(x_{n-i:n})]}_{\log[\mathcal{L}(\xi, Q_p)]} \right\}$$

```
1 > loglik=function(param){
2 +   xi=param[2];q=param[1]
3 +   lg=function(i) log(gq(Xs[i],xi,q))
4 +   return(-sum(Vectorize(lg)(1:nu)))}
```



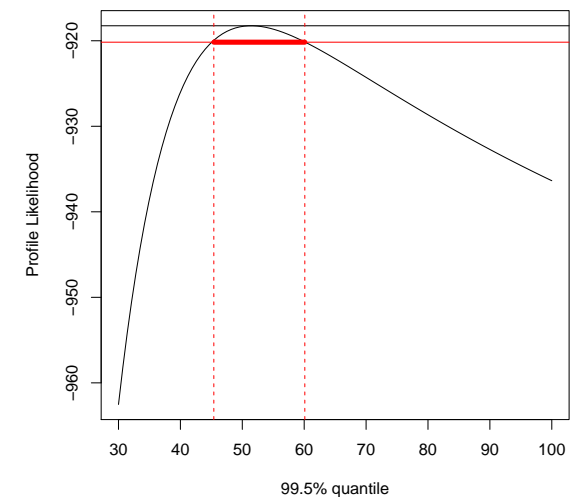
## Mise en oeuvre pratique, quantile bi-centenaire

ou mieux, le maximum de la vraisemblance profilée (seul  $Q_p$  nous intéresse),

$$\hat{Q}_p = \operatorname{argmax}\left\{ \max_{\xi} \{\log[\mathcal{L}(\xi, Q_p)]\} \right\}$$

```
1 > PL=function(Q){
2 +   profilelikelihood=function(xi){
3 +     loglik(c(Q,xi))}
4 +   return(optim(par=.6,fn=profilelikelihood)$
      value)}
```

```
1 > (OPT=optimize(f=PL,interval=c(10,100)))
2 $minimum
3 [1] 51.42182
4
5 $objective
6 [1] 918.2469
```



## Mise en oeuvre pratique, quantile bi-centenaire

On peut utiliser le test de rapport de vraisemblance pour obtenir un intervalle de confiance, puisque

$$2(\log[\mathcal{L}_p(\hat{Q}_p)] - \log[\mathcal{L}_p(Q_p)]) \xrightarrow{\mathcal{L}} \chi^2(1), \text{ lorsque } n \rightarrow \infty.$$

```

1 > XQ=seq(30,100,length=101)
2 > L=Vectorize(PL)(XQ)
3 > up=OPT$objective
4 > I=which(-L>=-up-qchisq(p=.95,df=1)/2)
5 > range(XQ[I])
6 [1] 45.4 60.1

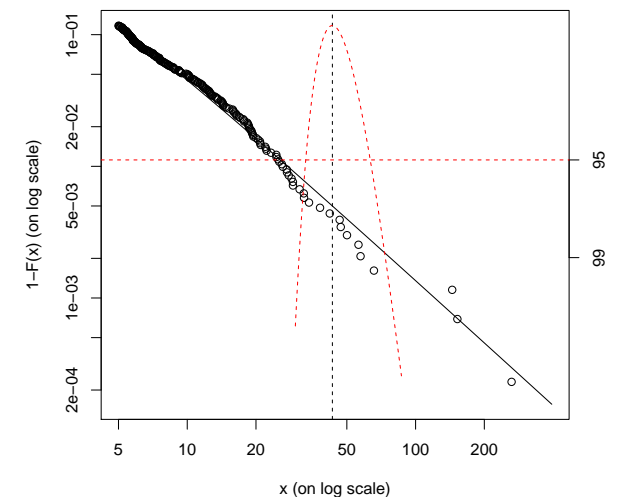
```

On peut utiliser le code suivant

```

1 > gpd.q(tailplot(gpd(Xs,u)),.995)

```





## Choix du seuil $u$

Supposons que l'on choisisse un seuil  $u$  tel que  $u = X_{n-k:n}$ , alors

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

pour  $k > n(1-p)$ . On notera que si  $k = n(1-p)$ ,  $\hat{x}_{p,k}$  coïncide avec l'estimateur empirique du quantile, i.e.  $X_{[np+1]:n}$ . Notons de plus que

$$MSE(\hat{x}_p) = Var(\hat{x}_p) + \mathbb{E}(\hat{x}_p - VaR(X, p))^2.$$

## Estimation de la période de retour

Pour rappel, le niveau  $P_m$  période de retour associée à une période de  $m$  années (niveau qui sera excédé, en moyenne, toutes les  $m$  observations) est la solution de

$$\mathbb{P}(X > u) \left( 1 + \xi \left( \frac{P_m - u}{\sigma} \right) \right)^{1-\xi} = \frac{1}{m},$$

si  $X|X > u$  peut être modélisé par une loi de Pareto généralisée, soit

$$P_m = u + \frac{\sigma}{\xi} ([m\mathbb{P}(X > u)]^\xi - 1),$$

sous l'hypothèse où  $\xi \neq 0$ , et

$$P_m = u + \sigma \log(m\mathbb{P}(X > u)),$$

si  $\xi = 0$ .

## Utilisation de l'estimateur de Hill

Rappelons que l'estimateur de Hill basé sur la statistique d'ordre est

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

qui est particulièrement intéressante si  $\xi > 0$ .

Rappelons que  $\overline{F}(x) = \mathbb{P}(X > x) = x^{-1/\xi} \mathcal{L}(x)$ , et donc, pour  $x \geq X_{n-k:n}$  et  $k$  suffisamment faible,

$$\frac{\overline{F}(x)}{\overline{F}(X_{n-k:n})} = \frac{\mathcal{L}(x)}{\mathcal{L}(X_{n-k:n})} \left( \frac{x}{X_{n-k:n}} \right)^{-1/\xi}.$$

Si l'on suppose que le ration des fonctions à variation lente peut être supposé comme négligeable, alors

$$\overline{F}(x) \sim \overline{F}(X_{n-k:n}) \left( \frac{x}{X_{n-k:n}} \right)^{-1/\xi}$$

Aussi, un estimateur naturel de la fonction de répartition est

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

En considérant l'inverse de cette fonction, on en déduit l'estimation naturelle de la *VaR* suivante,

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

pour des quantiles  $x_p$  tels que  $p > 1 - k/n$ . 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,

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

## Dans quels cas peut-on utiliser cette méthode ?

Rappelons que l'estimateur de Hill n'est pertinent que si  $\xi > 0$  (et donc pas si  $\xi = 0$ ). Dans ce dernier cas, des mauvaises conclusions peuvent être déduites.

De plus les propriétés de cet estimateur à distance finie (en particulier pour des petits échantillons) sont relativement décevantes. Néanmoins, les propriétés de l'estimateur de Hill ont été beaucoup plus étudiées dans la littérature que celui du maximum de vraisemblance.

## La tarification en réassurance

Le principe de base est le calcul de la **prime pure**, correspondant à l'espérance mathématique des paiements, avec franchise.

Pour un traité **stop-loss**, de portée infinie, le montant de l'indemnité versée par le réassureur est  $(X - d)_+$ , et la prime pure est donc

$$\mathbb{E}[(X - d)_+] = \mathbb{E}[(X - d)_+ | X \leq d] \cdot \mathbb{P}[X \leq d] + \mathbb{E}[(X - d)_+ | X > d] \cdot \mathbb{P}[X > d],$$

ce qui se simplifie simplement en

$$\pi_d = \mathbb{E}[(X - d)_+] = \mathbb{E}[X - d | X > d] \cdot \mathbb{P}[X > d].$$

La quantité  $\mathbb{E}[X - d | X > d]$  correspond à la **fonction de moyenne en excès**, notée  $e(d)$ ,

$$e(x) = \frac{1}{\mathbb{P}[X > x]} \int_x^\infty \mathbb{P}[X > t] dt.$$

## La méthode du *burning-cost*

On utilise ici tout simple un estimateur empirique (non paramétrique) de la prime pure,

$$\hat{\pi}_d = \frac{1}{n} \sum_{i=1}^n (X_i - d)_+$$

Par très robuste, impossible de valoriser une tranche de niveau élevé.

## L'approche paramétrique et la loi de Pareto

Rappelons que les sinistres suivent ici une loi de Pareto, i.e. la densité s'écrit

$$f(x) = \alpha \cdot \theta^\alpha \cdot x^{-\alpha-1}, x \geq \theta,$$

et la fonction de répartition

$$F(x) = \mathbb{P}(X \leq x) = 1 - \left(\frac{\theta}{x}\right)^\alpha, x \geq \theta.$$

( $\theta$  est parfois noté *OP* la **point d'observation** (*observation point*)).

**Remarque** L'estimateur du maximum de vraisemblance de  $\alpha$  vérifie

$$\frac{1}{\hat{\alpha}} = \frac{1}{n} \sum_{i=1}^n \log \left( \frac{X_i}{\theta} \right),$$

si  $\theta$  est connu (sinon, on considère  $\hat{\theta} = \min\{X_1, \dots, X_n\}$ ).

On décompose alors le calcul de la prime pure de façon usuelle

$$\text{espérance} = \text{fréquence} \times \text{coût moyen}$$



## Coût à charge de la tranche, *expected excess loss*

Pour les sinistres dont le montant est compris entre  $DE$  (la franchise) et  $EP$  (le maximum), le montant de la perte en excès s'écrit  $\min\{CO, \max\{X - DE, 0\}\}$ , et donc

$$EL = \int_{DE}^{EP} (x - DE) f(x) dx + \int_{EP}^{\infty} CO f(x) dx = \frac{DE}{1 - \alpha} (RL^{1-\alpha} - 1) \text{ si } \alpha \neq 1$$

(et  $EL = DE \cdot \log(RL)$  si  $\alpha = 1$ ), où pour rappel,  $EL = (CO + DE)/CO$  désigne la taille relative de la tranche.

## La fréquence

On notera également que la fréquence associée à un niveau  $x \geq \theta$  est

$$FQ(x) = FQ(\theta) \cdot \left(\frac{\theta}{x}\right)^{\alpha}$$

## La prime pure

Enfin, la prime est alors donnée par

$$\pi = FQ(DE) \cdot EL = FQ(\theta) \cdot \left( \frac{\theta}{DE} \right)^\alpha \cdot \frac{DE}{1 - \alpha} (RL^{1-\alpha} - 1) \text{ si } \alpha \neq 1.$$

Le rate-on-line est alors donnée, pour  $\alpha \neq 1$ ,

$$ROL = \frac{\pi}{CO} = \frac{FQ(\theta) \cdot \theta^\alpha \cdot (RL^{1-\alpha} - 1)}{DE^\alpha \cdot (RL - 1) \cdot (1 - \alpha)}$$

## Peut-on toujours utiliser la loi de Pareto ?

Cette approche a été proposée, étudiée et implémentée depuis longtemps ([Swiss Re \(2001\)](#), [Huyghues-Beaufond \(1991\)](#)).

Cette loi peut effectivement être souvent utile, elle est simple d'utilisation, et facile à ajuster (un seul paramètre si on suppose connu  $\theta$ ).

## La tarification par exposition

Il s'agit ici de calculer la prime pour des traités non-proportionnels.

Soit  $M$  la somme assurée. On notera  $x_i = X_i/M$  les montants de sinistres (i.i.d.) et  $d = D/M$  où  $D$  désigne la priorité du traité en excédent de sinistre. La prime pure, du point de vue de l'assureur est

$$\mathbb{E} [\min \{X, D\}] \mathbb{E} [N] = M \mathbb{E} [\min \{x, d\}] \mathbb{E} [N]$$

et la prime pure dite de base sera

$$\mathbb{E} [X] \mathbb{E} [N] = M \mathbb{E} [\min \{x, 1\}] \mathbb{E} [N].$$

La courbe d'exposition  $G$  est le rapport entre la prime de l'assureur et la prime de base,  $G : [0, 1] \rightarrow [0, 1]$

$$G(d) = \frac{\mathbb{E} [\min \{x, d\}]}{\mathbb{E} [\min \{x, 1\}]} \text{ pour } d \in [0, 1].$$

## La courbe d'exposition

La courbe d'exposition est définie simplement par

$$G(d) = \frac{\mathbb{E}[\min\{X, d\}]}{\mathbb{E}[X]} = \frac{1}{\mathbb{E}[X]} \int_0^d (1 - F(t)) dt.$$

On notera que cette fonction  $G$  vérifie  $G(0) = 0$ ,  $G(1) = 1$ , qu'elle est croissante et concave. En effet,

$$\frac{\partial G(d)}{\partial d} = \frac{1 - F(d)}{f(d)} \geq 0,$$

et

$$\frac{\partial^2 G(d)}{\partial d^2} = \frac{-f(d)}{f(d)} \leq 0,$$

La concavité reflète la part des grands risques: plus la courbe est proche de la diagonale, plus la part des grands risques dans la charge totale tend à être négligeable.

**Example** Pour une loi de Pareto  $\mathcal{P}(\theta, \alpha)$ ,

$$G(d) = \frac{1 - \theta^{\alpha-1}(\theta + d)^{1-\alpha}}{1 - \theta^{\alpha-1}(\theta + 1)^{1-\alpha}}.$$

**Example** [Bernegger \(1997\)](#) a suggéré une modélisation de  $G$  à l'aide de la fonction suivante à deux paramètres,

$$G_{a,b}(x) = \frac{\log(a + b^x) - \log(a + 1)}{\log(a + b) - \log(a + 1)},$$

correspond au modèle dit MBBEFD

(Maxwell-Boltzman-Bose-Einstein-Fermi-Dirac), issu de la littérature de physique statistique. La distribution  $X$  sous-jacente est alors

$$F_X(x) = \begin{cases} 1 - \frac{(a+1)b^t}{a+b^t} & \text{pour } 0 \leq t < 1 \\ 1 & \text{pour } t = 1 \end{cases}$$

## Réassurance, Exemple des Ouragans

On utilise la base *Normalized Hurricane Damages in the United States* pour la période 1900-2005, à partir de [Pielke et al. \(2008\)](#).

```

1 > library(gdata)
2 > db=read.xls(
3 + "http://sciencepolicy.colorado.edu/publications/special/public_data
   _may_2007.xls",sheet=1)
4 > tail(db)

```

	Year	Hurricane.Description	State	Category	Base.Economic.Damage
202	2005	Cindy	LA	1	320,000,000
203	2005	Dennis	FL	3	2,230,000,000
204	2005	Katrina	LA,MS	3	81,000,000,000
205	2005	Ophelia	NC	1	1,600,000,000
206	2005	Rita	TX	3	10,000,000,000
207	2005	Wilma	FL	3	20,600,000,000

## Réassurance, Exemple des Ouragans

	Normalized.PL05	Normalized.CL05
202	320,000,000	320,000,000
203	2,230,000,000	2,230,000,000
204	81,000,000,000	81,000,000,000
205	1,600,000,000	1,600,000,000
206	10,000,000,000	10,000,000,000
207	20,600,000,000	20,600,000,000

La lecture des données est un peu technique

```

1 > stupidcomma = function(x){
2 + x=as.character(x)
3 + for(i in 1:10){x=sub(",","",as.character(x))}
4 + return(as.numeric(x))}

```

## Réassurance, Exemple des Ouragans

et on obtient alors des chiffres

```
1 > base=db[,1:4]
2 > base$Base.Economic.Damage=Vectorize(stupidcomma)(db$Base.Economic.
    Damage)
3 > base$Normalized.PL05=Vectorize(stupidcomma)(db$Normalized.PL05)
4 > base$Normalized.CL05=Vectorize(stupidcomma)(db$Normalized.CL05)
```

et on obtient la base suivante

```
1 > tail(base)
```

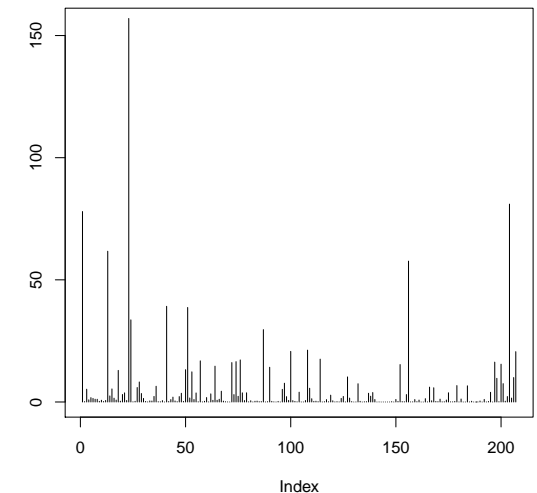
	Year	Hurricane.Description	State	Category	Base.Economic.Damage	
3	202	2005	Cindy	LA	1	3.20e+08
4	203	2005	Dennis	FL	3	2.23e+09
5	204	2005	Katrina	LA,MS	3	8.10e+10
6	205	2005	Ophelia	NC	1	1.60e+09
7	206	2005	Rita	TX	3	1.00e+10
8	207	2005	Wilma	FL	3	2.06e+10



## Réassurance, Exemple des Ouragans

La base historique de coûts d'ouragans est

```
1 > plot(base$Normalized.PL05/1e9, type="h", ylim=
      c(0,155))
```



L'analyse se fait (comme en assurance non-vie classique) en deux temps : on va modéliser la fréquence annuelle, et les coûts individuels. La base des fréquence s'obtient avec

```
1 > TB <- table(base$Year)
2 > years <- as.numeric(names(TB))
3 > counts <- as.numeric(TB)
4 > years0=(1900:2005)[which(!(1900:2005)%in%years)]
5 > db <- data.frame(years=c(years, years0),
```

```

6 + counts=c(counts,rep(0,length(years0)))
7 > db[88:93,]
8   years counts
9 88 2003      3
10 89 2004      6
11 90 2005      6
12 91 1902      0
13 92 1905      0
14 93 1907      0

```

On a, en moyenne, deux ouragans par an,

```

1 > mean(db$counts)\begin{lstlisting}[escapechar=
    ,style=mystyle,firstnumber=1][1]
    1.95283

```

On peut aussi tenter un modèle de régression, linéaire

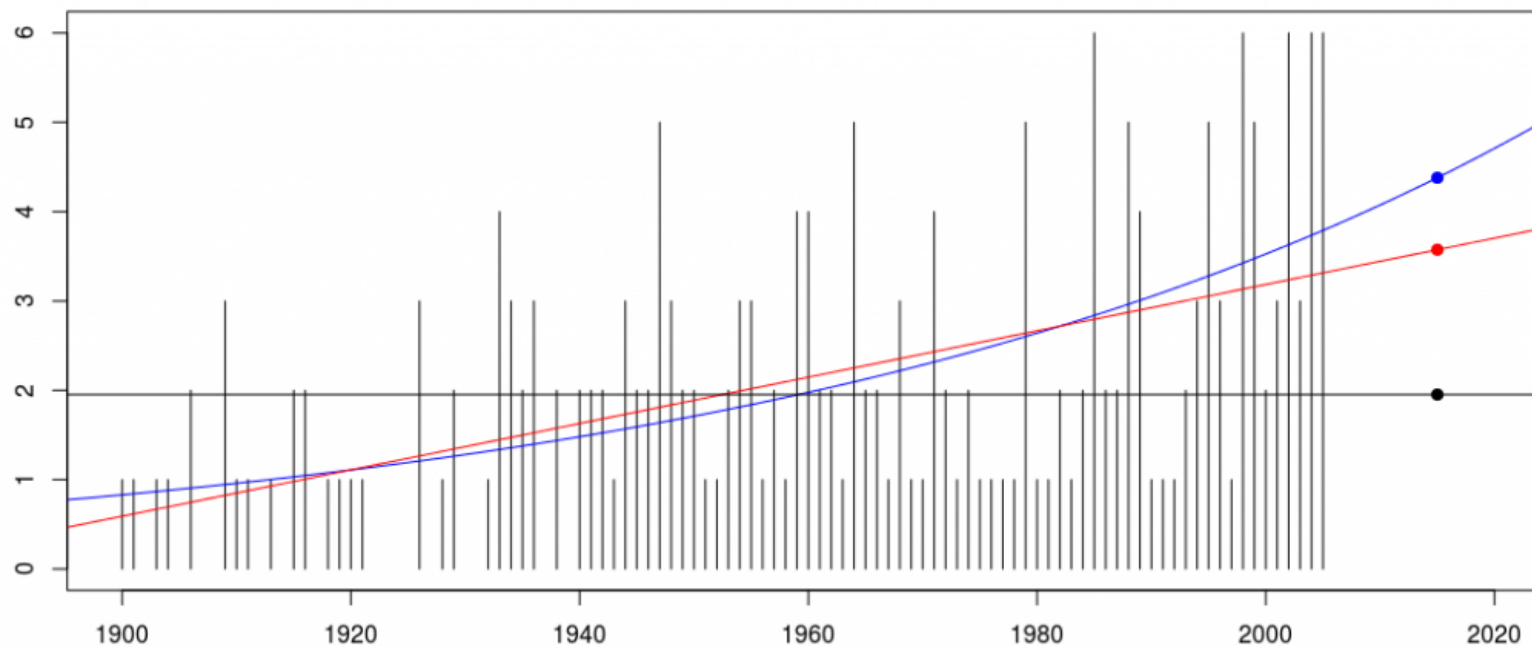
```

1 > reg0 <- glm(counts~years,data=db,family=poisson(link="identity"),
2 + start=lm(counts~years,data=db)$coefficients)

```

ou exponentiel

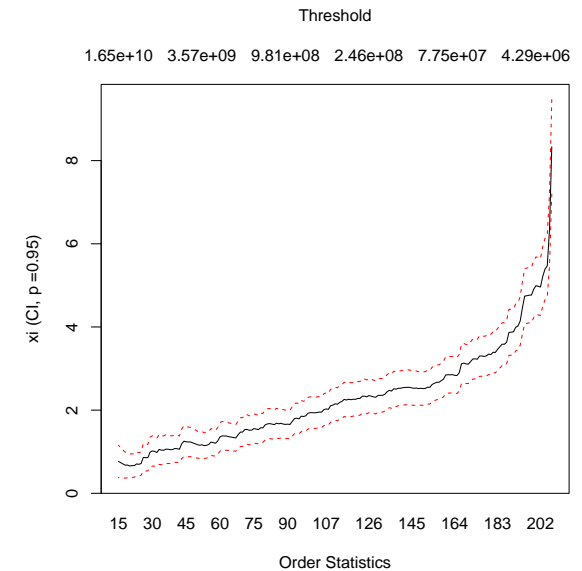
```
1 > reg1 <- glm(counts~years ,data=db ,family=poisson(link="log"))
```



```
1 > (predictions=cbind(constant=mean(db$counts),linear= cpred0[126],
  exponential=cpred1[126]))
2   constant   linear exponential
3 126   1.95283  3.573999    4.379822
```

Pour modéliser les coûts individuels des ouragans

```
1 > library(evir)
2 > hill(base$Normalized.PL05)
```



Considérons une compagnie qui a 5% de part de marché, et qui veut valoriser un traité de réassurance, de franchise 2 milliards, avec une limite de 4 milliards.

Considérons une loi de Pareto, au delà de  $u = 500$  millions,

```
1 > threshold=.5
2 > (gpd.PL <- gpd(base$Normalized.PL05/1e9/20,threshold)$par.ests)
3           xi           beta
4 0.4424669 0.6705315
```

Pour rappel 1 ouragan sur 8 atteint ce seuil  $u$ ,

```
1 > mean(base$Normalized.CL05/1e9/20>.5)
2 [1] 0.1256039
```

On peut alors valoriser un contrat de réassurance To compute it we can use

```
1 > E <- function(yinf,ysup,xi,beta){
2 +   as.numeric(integrate(function(x) (x-yinf)*dgp(x,xi,mu=threshold,
      beta), lower=yinf,upper=ysup)$value+(1-pgp(ysup,xi,mu=threshold,
      beta))*(ysup-yinf)) }
```

Si on espère être touché par 2 ouragans, une année donnée

```
1 > predictions[1]
2 [1] 1.95283
```

que chaque ouragan a 12.5% de chances de coûter plus de 500 million

```
1 > mean(base$Normalized.PL05/1e9/20>.5)
2 [1] 0.1256039
```

et que si un ouragan dépasse 500 million, le coût moyen en excès (en millions) est

```
1 > E(2,6,gpd.PL[1],gpd.PL[2])*1e3  
2 [1] 330.9865
```

de telle sorte que la prime pure d'un contrat de réassurance serait

```
1 > predictions[1]*mean(base$Normalized.PL05/1e9/20>.5)*  
2 + E(2,6,gpd.PL[1],gpd.PL[2])*1e3  
3 [1] 81.18538
```

en millions d'euros, pour une couverture de 4 milliards, en excès de 2.

## Réassurance, Exemple de la Perte d'Exploitation

Considérons les données suivantes, de perte d'exploitation,

```
1 > library(gdata)
2 > db=read.xls("http://perso.univ-rennes1.fr/arthur.charpentier/SIN_
    1985_2000-PE.xls", sheet=1)
```

Comme pour tout contrat d'assurance, on distingue

- le nombre moyen de sinistres
- le coût moyen des sinistres

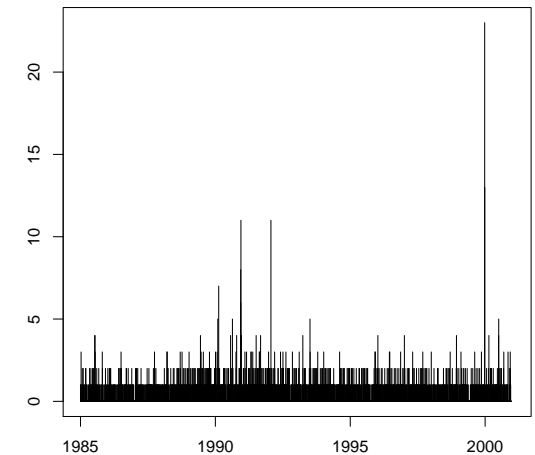
La encore (pour l'instant), on n'utilise pas de covariables.

## Fréquence des sinistres en Perte d'Exploitation

```

1 > date=db$DSUR
2 > D=as.Date(as.character(date),format="%Y%m%d"
  )
3 > vD=seq(min(D),max(D),by=1)
4 > sD=table(D)
5 > d1=as.Date(names(sD))
6 > d2=vD[-which(vD%in%d1)]
7 > vecteur.date=c(d1,d2)
8 > vecteur.cpte=c(as.numeric(sD),rep(0,length(
  d2)))
9 > base=data.frame(date=vecteur.date,cpte=
  vecteur.cpte)

```



Utilisons une régression de Poisson pour modéliser la fréquence journalière de sinistres.

```

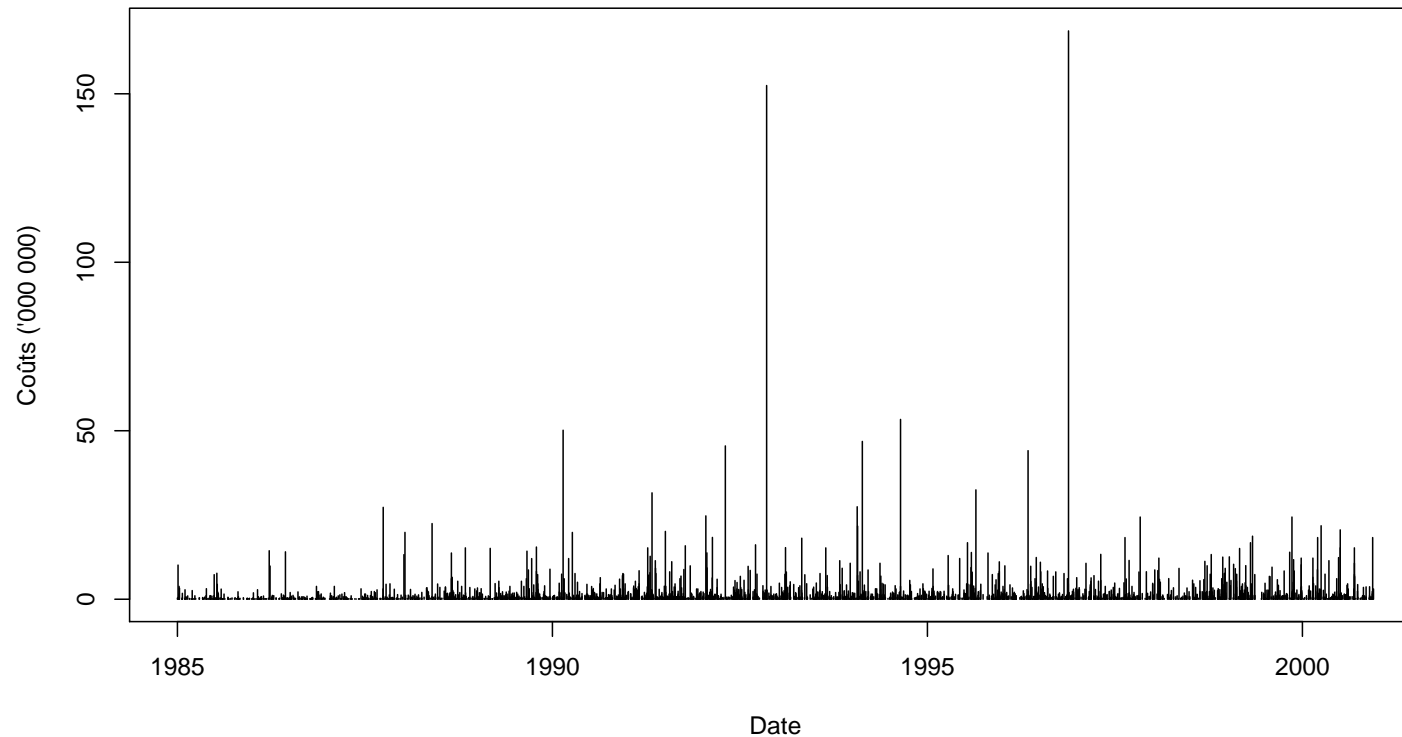
1 > regdate=glm(cpte~date,data=base,family=poisson(link="log"))

```



```
2 > nd2016=data.frame(date=seq(as.Date(as.character(20160101),format="
    %Y%m%d"),
3 +                               as.Date(as.character(20161231),format="%Y%m
    %d"),by=1))
4 > pred2016 =predict(regdate,newdata=nd2016,type="response")
5 > sum(pred2016)
6 [1] 163.4696
```

Il y a environ 165 sinistres, par an.



Sur 16 ans, le 32ème plus gros sinistre (2 par an) était de l'ordre de 15 million.

```
1 > quantile(db$COUTSIN, 1-32/2400) / 1e6  
2 98.66667%  
3 15.34579
```

## Réassurance, Exemple de la Perte d'Exploitation

Considérons un contrat avec une franchise de 15 millions, pour une couverture totale de 35 millions. Pour la compagnie de réassurance, le coût espéré est  $\mathbb{E}(g(X))$  avec

$$g(x) = \min\{35, \max\{x - 15, 0\}\}$$

La fonction d'indemnité est

```
1 > indemn=function(x) pmin((x-15)*(x>15), 50-15)
```

Sur 16 ans, l'indemnité moyenne aura été de

```
1 > mean(indemn(db$COUTSIN/1e6))
```

```
2 [1] 0.1624292
```

## Réassurance, Exemple de la Perte d'Exploitation

Par sinistre, la compagnie de réassurance paie, en moyenne, 162,430 euros. Donc avec 160 sinistres par an, la prime pure - burning cost - est de l'ordre de 26 millions.

```
1 > mean(indemn(db$COUTSIN/1e6))*160  
2 [1] 25.98867
```

On peut aussi tenter un modèle paramétrique, de type Pareto. Les trois paramètres sont

- le **seuil**  $\mu$  (considéré comme fixe)
- le paramètre d'**échelle** (scale)  $\sigma$  (noté aussi  $\beta$ )
- the **tail index**  $\xi$

## Réassurance, Exemple de la Perte d'Exploitation

Prenons un seuil  $u$  plus faible que la franchise, e.g. 12 million.

```
1 > gpd.PL <- gpd(db$COUTSIN,12e6)$par.ests
2 > gpd.PL
3           xi           beta
4 7.004147e-01 4.400115e+06
```

et on pose

```
1 > E <- function(yinf,ysup,xi,beta,threshold){
2 +   as.numeric(integrate(function(x) (x-yinf)*dgpd(x,xi,mu=threshold,
3 +   beta),lower=yinf,upper=ysup)$value+(1-pgpd(ysup,xi,mu=threshold,
4 +   beta))*(ysup-yinf))
5 }
```

## Réassurance, Exemple de la Perte d'Exploitation

Sachant qu'un sinistre dépasse 12 million, le remboursement moyen est de l'ordre de 6 millions

```
1 > E(15e6, 50e6, gpd.PL[1], gpd.PL[2], 12e6)
2 [1] 6058125
```

La probabilité qu'un sinistre dépasse  $u$  est

```
1 > mean(db$COUTSIN > 12e6)
2 [1] 0.02639296
```

Aussi, avec 160 sinistres par an,

```
1 > p
2 [1] 159.4757
```

et 2.6% des sinistres qui dépassent 12 millions,

```
1 > mean(db$COUTSIN > 12e6)
2 [1] 0.02639296
```

## Réassurance, Exemple de la Perte d'Exploitation

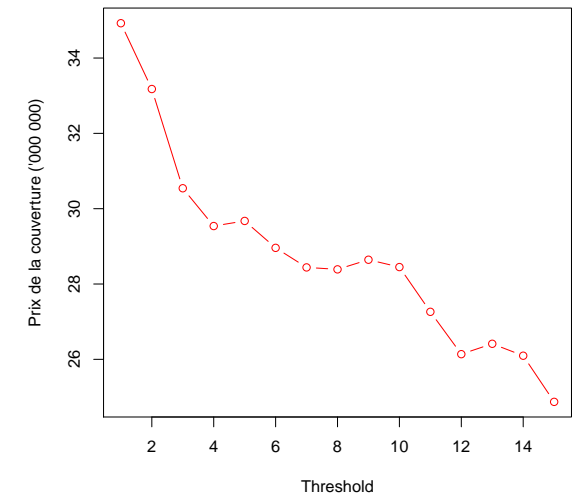
Finalement, la prime pure est

```
1 > p*mean(db$COUTSIN>12e6)*E(15e6,50e6,gpd.PL[1],gpd.PL[2],12e6)
2 [1] 25498867
```

qui est proche de la valeur empirique.

On peut regarder la robustesse au choix de  $u$ ,

```
1 > esp=function(threshold=12e6,p=sum(pred2010)){
2 +   (gpd.PL <- gpd(db$COUTSIN,threshold)$par.
      ests)
3 +   return(p*mean(db$COUTSIN>threshold)*E(15e6
      ,50e6,gpd.PL[1],gpd.PL[2],threshold))}
```



## Retour aux données RC

On peut utiliser un modèle avec comme loi pour les coûts une loi de Pareto

```
1 > ?gamlss.family
```

```
2 > ?PARETO2
```

La loi de Pareto a ici pour densité

$$f(y|\mu, \sigma) = \frac{1}{\sigma} \mu^{\frac{1}{\sigma}} [y + \mu]^{-\frac{1}{\sigma} - 1}$$

et sa moyenne est alors

$$\mathbb{E}[Y] = \frac{\mu\sigma}{1 - \sigma}$$

avec ici

$$\mathbb{E}[Y|\mathbf{X} = \mathbf{x}] = \frac{e^{\mathbf{x}^\top \boldsymbol{\alpha}} e^{\mathbf{x}^\top \boldsymbol{\beta}}}{1 - e^{\mathbf{x}^\top \boldsymbol{\beta}}}$$



## Retour aux données RC

```

1 > seuil=1500
2 > regpareto=gamlss(cout~ageconducteur, sigma.formula= ~ageconducteur
, data=base_RC[base_RC$cout>seuil,], family=PARETO2(mu.link = "log",
sigma.link = "log"))
3 -----
4 Mu link function: log
5 Mu Coefficients:
6
7 Estimate Std. Error t value Pr(>|t|)
8 (Intercept) 9.166591 0.212249 43.188 <2e-16 ***
9 ageconducteur 0.007024 0.004420 1.589 0.113
10 -----
11 Sigma link function: log
12 Sigma Coefficients:
13
14 Estimate Std. Error t value Pr(>|t|)
15 (Intercept) -0.793548 0.166166 -4.776 2.58e-06 ***
ageconducteur -0.009316 0.003513 -2.652 0.00835 **

```