

Pareto et incendie

Jérémie Barde

École d'actuariat
Université Laval, Québec, Canada

8 août 2024



Faculté des
sciences et de génie
École d'actuariat

CIMMUL

Quantact



- 1 Introduction
- 2 Loi Pareto
- 3 Notions sur la théorie des valeurs extrêmes
- 4 Loi Pareto généralisée
- 5 Modélisation des pertes pour le péril incendie (sévérité)
- 6 Caractéristique du modèle GB2-PaG

Première apparition de la distribution Pareto :

- Survenue il y a plus d'un siècle
- Vilfredo Pareto
- [Pareto, 1897]

Utilisation :

- Il remarque que le nombre de personnes avec un revenu plus grand que x peut être approximé par une distribution *power-law*

$$Cx^{-\alpha}$$

où $C > 0$ et $\alpha > 0$

- Il sera constaté que cette approximation fonctionne bien pour de grandes valeurs de x , on appellera ce phénomène *Paretian tail behavior*

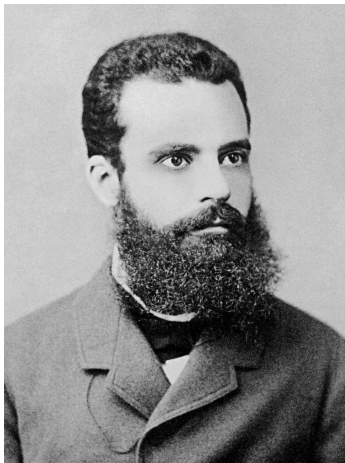
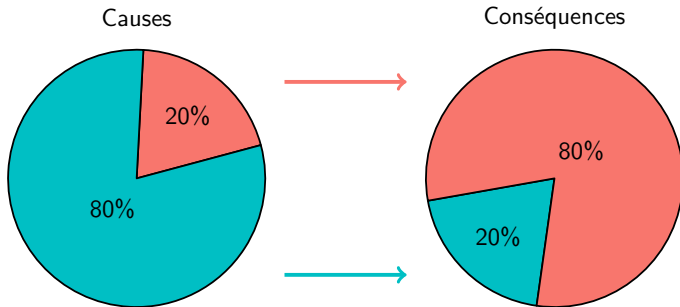


Illustration – Vilfredo Pareto, polymathe italien

Principe Pareto :

- Il constate que 80% des richesses appartiennent à environ 20% de la population
- On donnera son nom au principe Pareto qui stipule que 80% des conséquences proviennent de 20% de la cause



La loi Pareto de type I est une distribution de la loi de puissance continue qui est basée sur les observations de Pareto

La fonction de répartition est donnée par

$$F_X(x) = 1 - \left(\frac{x_0}{x} \right)^\alpha, \quad x > x_0$$

Caractéristiques de la loi Pareto de type I :

$$E[X] = x_0 \frac{\alpha}{\alpha - 1}, \quad \alpha > 1, \quad \text{Var}(X) = x_0^2 \frac{\alpha}{(\alpha - 1)^2 (\alpha - 2)}, \quad \alpha > 2$$

Remarque : Si $\alpha < 2$ la variance n'existe pas et si $\alpha < 1$ l'espérance aussi n'existe pas

La loi Pareto Lomax est essentiellement une loi Pareto de type I translaté pour que le support commence à 0. Si on a $X \sim \text{PaI}(x_0 = \lambda, \alpha)$, alors $X - x_0 \sim \text{PaLomax}(\alpha, \lambda)$.

La fonction de répartition est donnée par

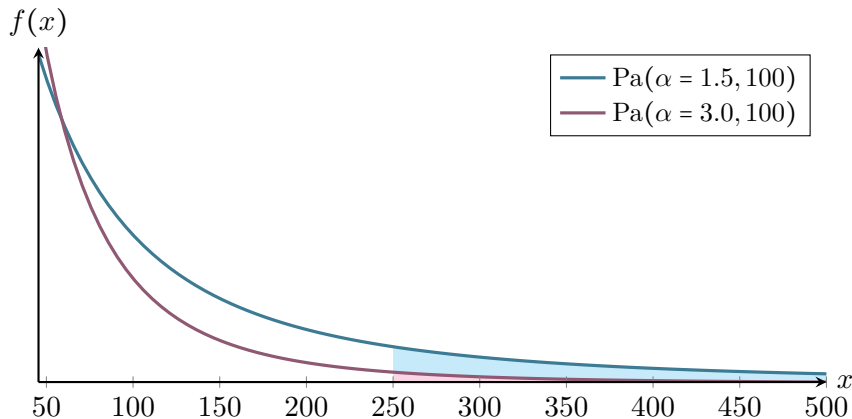
$$F_X(x) = 1 - \left(\frac{\lambda}{\lambda + x} \right)^\alpha, \quad x > 0.$$

Caractéristiques de la loi Pareto Lomax :

$$E[X] = \frac{\lambda}{\alpha - 1}, \quad \alpha > 1, \quad \text{Var}(X) = \frac{\alpha \lambda^2}{(\alpha - 1)^2 (\alpha - 2)}, \quad \alpha > 2.$$

Remarque : Si $\alpha < 2$ la variance n'existe pas et si $\alpha < 1$ l'espérance aussi n'existe pas.

La paramètre de forme α donne l'indication du niveau de dangerosité de la distribution.



La tableau ci-dessous montre la $Pr(X > 400)$ selon la pramètre α avec $\lambda = 100$

α	$Pr(X > 400)$
3.0	0.016
2.0	0.063
1.5	0.125
0.7	0.379

Remarque : lorsque α passe de 3 à 1.5 il y a 8 fois plus de chance d'avoir un sinistre supérieur à 400 et lorsqu'il passe de 3 à 0.7 il y a 24 fois plus de chance

Notions sur la théorie des valeurs extrêmes

Distribution généralisée de la valeur extrême

Les actuaires ont un intérêt particulier à étudier le comportement des pertes importantes.

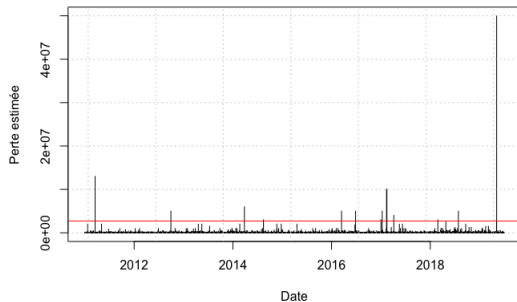
- Risque incendie : ces données ne sont pas aberrantes.

Pour cela on s'intéressera à la distribution généralisée de la valeur extrême (GVE), qui est justement utilisée pour décrire le comportement du maximum d'un échantillon de v.a. i.i.d.

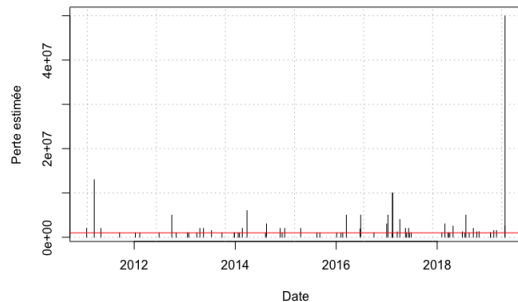
La fonction de répartition de cette distribution est donné par

$$H_{\xi, \sigma} = \begin{cases} \exp\left\{-\left(\frac{1}{1+\frac{\xi}{\sigma}x}\right)^{\frac{1}{\xi}}\right\} & , \xi < 0, 0 < x < -\frac{\sigma}{\xi} \\ \exp\left\{-\exp\left\{-\frac{x}{\sigma}\right\}\right\} & , \xi = 0, x > 0 \\ \exp\left\{-\left(\frac{1}{1+\frac{\xi}{\sigma}x}\right)^{\frac{1}{\xi}}\right\} & , \xi > 0, x > 0. \end{cases}$$

Pertes estimées selon la date du sinistre



Pertes estimées supérieures à 1 million selon la date du sinistre



Pourquoi il ne faut pas considérer ces données comme aberrantes ?

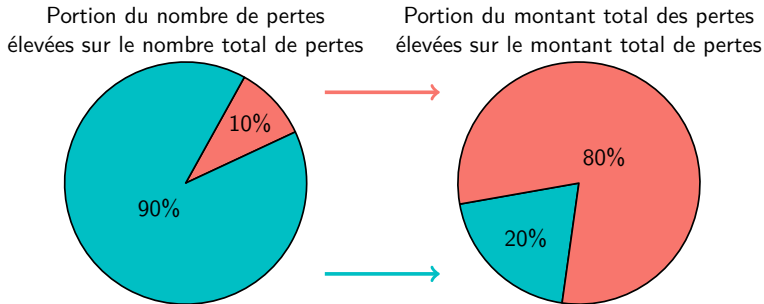


Illustration – Pour la base de données incendies de la ville de Toronto

Essentiellement, on considère une suite de v.a. i.i.d X_1, X_2, \dots

On définit la maximum des n premières v.a. par

$$M_n = \max\{X_1, \dots, X_n\}.$$

Soit $x_F = \sup\{x; F(x) < 1\}$, le point extrême droite de F . Il est possible de montrer que

$$\lim_{n \rightarrow \infty} M_n = x_F$$

Pour décrire le comportement asymptotique de la v.a. M_n on suppose qu'il existe des entiers normalisants $a_n > 0$ et b_n tels que

$$\Pr\left(\frac{M_n - b_n}{a_n} \leq x\right) = F(a_n x + b_n)^n \rightarrow H(x), \text{ quand } n \rightarrow \infty$$

où $H(x)$ est une fonction de répartition non dégénérée.

Lorsque cette condition est respectée, alors on dit que F_X est un élément du domaine d'attraction du maximum de la distribution H , qu'on note

$$F_X \in DAM(H).$$

Le théorème de Fisher-Tippett, [Fisher and Tippett, 1928], indique que la seule fonction de répartition non dégénérée H est la fonction de répartition de H_ξ .

Il y a trois classes de distributions $F_X \in DAM(H_\xi)$ selon que $\xi < 0$, $\xi = 0$ et $\xi > 0$.

On s'intéresse à la classe où $\xi > 0$, ce qui correspond aux distributions à queue lourde, par exemple, loi Pareto et Burr.

Remarque : Le Théorème de Fisher-Tippett est bien expliqué dans [McNeil, 1997].

Pour obtenir la loi Pareto généralisée, on applique la transformation $\alpha = \frac{1}{\xi}$ et $\lambda = \frac{\sigma}{\xi}$ à la loi Pareto Lomax. La fonction de répartition est donnée par

$$GP_{\xi,\sigma}(x) = 1 - \left(1 + \frac{\xi}{\sigma}x\right)^{-\frac{1}{\xi}}, \quad \xi > 0, x > 0$$

où ξ est le paramètre de forme et σ le paramètre d'échelle.

Il est possible d'introduire un paramètre de location θ , alors

$$GP_{\xi,\sigma,\theta}(x) = 1 - \left(1 + \frac{\xi}{\sigma}(x - \theta)\right)^{-\frac{1}{\xi}}, \quad \xi > 0, x > \theta.$$

Caractéristique de la loi Pareto généralisée :

$$E[X] = \frac{\sigma}{1 - \xi} + \theta, \quad \xi < 1 \quad \text{Var}(X) = \frac{\sigma^2}{(1 - \xi)^2(1 - 2\xi)}, \quad \xi < \frac{1}{2}$$

Loi Pareto généralisée

Lien entre la loi Pareto généralisée et la distribution généralisée des valeurs extrêmes

Loi Pareto généralisée et GVE

- Distribution de l'excédent
- Théorème de Pickands-Balkema-de Haan
- [Pickands III, 1975]
- [Balkema and De Haan, 1974]

Définition 1

Soit la v.a. X et la v.a. W_u représentant l'excédent d'un sinistre par rapport à un seuil θ . La v.a. W_θ est défini comme

$$W_\theta = (X - \theta | X > \theta).$$

La fonction de répartition de la v.a. W_u est donnée par

$$F_{W_\theta}(X) = \frac{F_X(x + \theta) - F_X(\theta)}{1 - F_X(\theta)}, \quad x > 0.$$

Loi Pareto généralisée

Lien entre la loi Pareto généralisée et la distribution généralisée des valeurs extrêmes

Le théorème de Pickhands-Balkema-de Haan :

- Sous la condition du $DAM(H_\xi)$ la loi Pareto généralisée est la distribution limite à droite de la distribution de l'excédent.
- Alors on peut trouver une fonction positive $\sigma(\theta)$ pour que

$$\lim_{\theta \rightarrow \infty} \sup_{x \leq 0} |F_{W_\theta}(x) - GP_{\xi, \sigma(\theta)}(x)| = 0$$

- Bases théoriques suffisante pour dire que pour un seuil θ assez élevé on peut approximer la distribution de l'excédent par une loi Pareto généralisée

Modélisation des pertes pour le péril incendie (sévérité)

Définition des mesure de risque VaR et $TVaR$

La mesure VaR est définie avec la fonction quantile de la v.a X par

$$VaR_{\kappa}(X) = \inf\{x \in \mathbb{R} : F_X(x) \geq \kappa\}, \kappa \in [0,1)$$

Cependant, elle ne donne pas d'information sur des valeurs supérieures à $VaR_{\kappa}(X)$.

La mesure $TVaR$ est donnée par

$$TVaR_{\kappa}(X) = \frac{1}{1 - \kappa} \int_{\kappa}^1 VaR_u(X) du, u \in [0,1)$$

La mesure de risque $TVaR_{\kappa}(X)$ peut être considérée comme la moyenne arithmétique des mesures de risque $VaR_{\kappa}(X)$ pour des valeurs de u supérieures à κ .

Modélisation des pertes pour le péril incendie (sévérité)

Méthode *Peak-over-Threshold*

Pour modéliser les pertes importantes, on utilise la méthode *Peak-over-Threshold* (POT)

- Proposée par les hydrologistes et les ingénieurs civils
- [Embrechts et al., 2013]

Pour appliquer cette méthode, on utilise la loi Pareto généralisée.

On voudra calculer les mesures VaR et $TVaR$, dont les formes analytiques, sont donné par

$$VaR_{\kappa}(X) = \frac{\sigma}{\xi} \left((1 - \kappa)^{-\xi} - 1 \right) + \theta, \quad TVaR_{\kappa}(X) = \frac{\sigma}{\xi} \left(\frac{1}{1 - \xi} (1 - \kappa)^{-\xi} - 1 \right) + \theta.$$

Modélisation des pertes pour le péril incendie (sévérité)

Méthode *Peak-over-Threshold*(POT)

Défi et enjeu de l'utilisation de la méthode POT

- Défi : identifier un seuil θ convenable
- Enjeu : avoir des résultats biaisés

Méthodes pour identifier le seuil :

- Analyse graphique : *QQplot* Pareto
- Détection automatique *Power-Law FIT* et double *bootstrap*

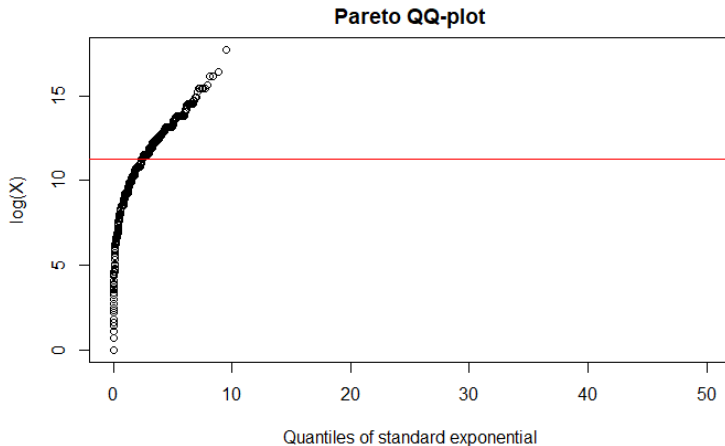
Description des données utilisées

	Toronto	San Francisco	NFID
Nombre de sinistres	17 536	628 147	467 929
Données utilisables	13 571	20 659	102 653
Pertes moyennes (\$)	40 296	47 077	75 800
Pertes maximales (millions \$)	50	100	200

Tableau – Données de périls incendies provenant de la ville de Toronto et San Francisco et les données *National Fire Information Database (NFID)*

Remarque : La perte maximale est extrêmement éloignée de la perte moyenne.

Dans un travail préalable, on avait déterminé un seuil à 75 000



On utilise la fonction `gpdFit` du package `tea` de R pour estimer les paramètres de la loi Pareto généralisée

	$\hat{\alpha}$	$\hat{\lambda}$	θ
Toronto	1.41	117 038	75 000
San Francisco	1.44	385 949	150 000
NFID	1.40	333 537	350 000

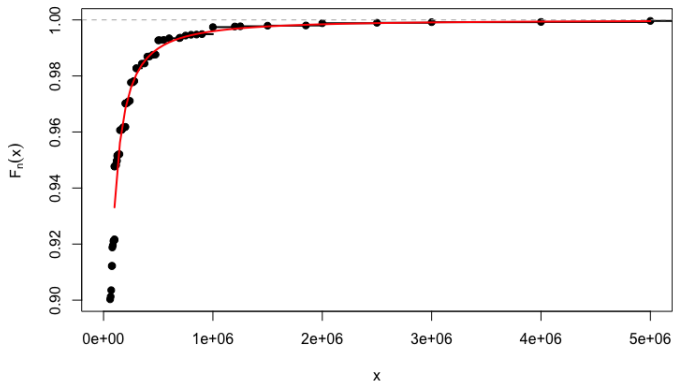
- On observe fréquemment une valeur de α entre 1 et 2.5 à partir de données pour le péril incendie
- Pour les compagnies industrielles on s'attend à avoir un alpha entre 1.2 et 1.8

Source : [Schmutz and Doerr, 1998], [Antal and Re, 2007]

Modélisation des pertes pour le péril incendie (sévérité)

Données en provenance de la ville de Toronto : fonction de répartition

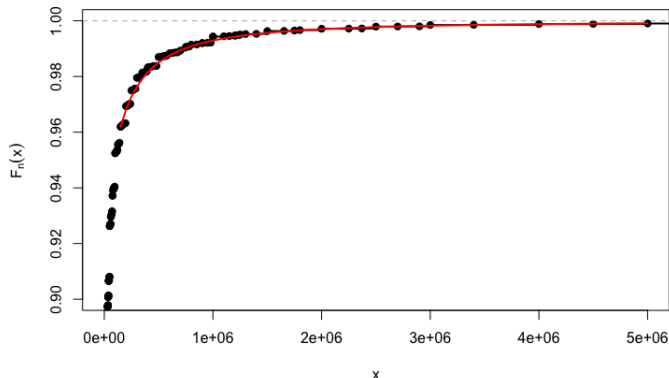
Comparaison de la fonction de répartition empirique avec la fonction de répartition de loi la Pareto généralisée pour les données en provenance de Toronto



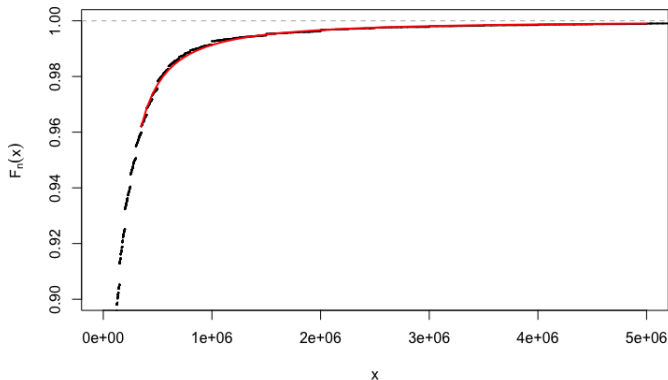
Modélisation des pertes pour le péril incendie (sévérité)

Données en provenance de la ville de San Francisco : fonction de répartition

Comparaison de la fonction de répartition empirique avec la fonction de répartition de loi la Pareto généralisée pour les données en provenance de San Francisco



Comparaison de la fonction de répartition empirique avec la fonction de répartition de loi la Pareto généralisée pour les données NFID



Modélisation des pertes pour le péril incendie (sévérité)

Raccordement de lois

Motivation :

- Ne pas utiliser la portion empirique pour les données sous le seuil
- Avoir une meilleure adéquation à celle que l'on obtient avec une seule loi

Définition 2

Soit X une variable continue qui obéit à une loi composite et soit $\{X_j, j = 1, 2\}$ des variables non identiquement distribuées. Alors,

$$f_X(x) = \begin{cases} w_1 \frac{f_{X_1}(x)}{F_{X_1}(\theta)} & , x \leq \theta \\ w_2 \frac{f_{X_2}(x)}{1-F_{X_2}(\theta)} & , x > \theta \end{cases}$$

où $w_i > 0$ et $\sum_{i=1}^2 w_i = 1$.

Modélisation des pertes pour le péril incendie (sévérité)

Raccordement de lois

Lois testées :

- Lognormale-Pareto généralisée (LN-PaG)
- Weibull-Pareto généralisée (We-PaG)
- Coxienne-2-Pareto généralisée (Cox2-PaG)
- Bêta généralisée de type 2-Pareto généralisée (GB2-PaG)

Dans [Brazauskas and Kleefeld, 2016], les auteurs ont montré qu'il était possible d'assurer la continuité et la dérivabilité au point θ en résolvant ce système d'équations :

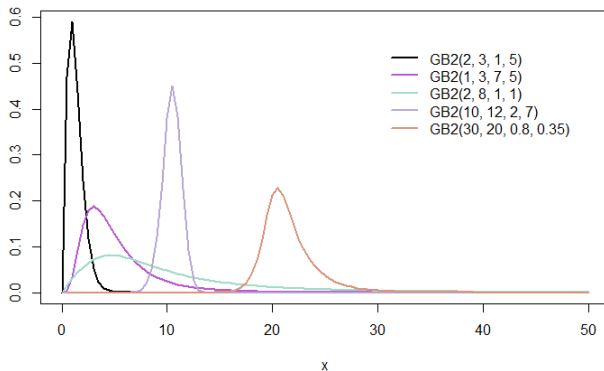
$$\begin{cases} w \frac{f_{X_1}(\theta)}{F_{X_1}(\theta)} = (1-w) \frac{f_{X_2}(\theta)}{F_{X_2}(\theta)} & \text{(continuité)} \\ w \frac{f'_{X_1}(\theta)}{F_{X_1}(\theta)} = (1-w) \frac{f'_{X_2}(\theta)}{F_{X_2}(\theta)} & \text{(dérivabilité)} \end{cases}$$

Modélisation des pertes pour le péril incendie (sévérité)

Loi bêta généralisée de type 2-Pareto généralisée

Loi : GB2-PaG de paramètres $(a, b, p, q, \theta, \xi, \sigma)$

Motivation : Avec ses 4 paramètres, la loi GB2 a l'avantage d'être très flexible



Modélisation des pertes pour le péril incendie (sévérité)

Loi bêta généralisée de type 2-Pareto généralisée

En maximisant numériquement la fonction de log-vraisemblance, on obtient les estimés suivants

	\hat{a}	\hat{b}	\hat{p}	\hat{q}	$\hat{\alpha}$	$\hat{\lambda}$	\hat{w}
Toronto	0.47	19 990	1.57	2.89	1.42	120 149	0.91
San Francisco	0.14	1 693	10.43	11.35	1.45	394 275	0.96
NFID	0.24	25 244	4.76	6.57	1.40	343 089	0.96

Remarque : Le paramètre α de ces 3 modèles sont relativement proche.

■ Espérance :

$$E[X] = \frac{w}{B\left(\frac{(\theta/b)^a}{1+(\theta/b)^a}; p, q\right)} \frac{bI(p + \frac{1}{a}, q - \frac{1}{a})}{I(p, q)} B\left(\frac{(\frac{\theta}{b})^a}{1 + (\frac{\theta}{b})^a}; p + \frac{1}{a}, q - \frac{1}{a}\right) + (1-w) \left(\frac{\sigma}{1-\xi} + \theta\right)$$

■ Espérance tronquée :

$$E[X \times 1_{\{X \leq d\}}] = \begin{cases} \frac{w}{F_{X_1}(\theta)} E[X_1 \times 1_{\{X_1 \leq \theta\}}] & , d \leq \theta \\ \frac{w}{F_{X_1}(\theta)} E[X_1 \times 1_{\{X_1 \leq \theta\}}] + (1-w) E[X_2 \times 1_{\{X_2 \leq d\}}] & , d > \theta \end{cases}$$

- Mesure VaR : outil d'optimisation
- Mesure $TVaR$:

$$TVaR_{\kappa}(X) = \begin{cases} \frac{w}{F_{X_1}(\theta)(1-\kappa)} (E[X_1 \times 1_{\{X_1 \leq \theta\}}] - E[X_1 \times 1_{\{X_1 \leq c\}}]) \\ \quad + \frac{1-w}{(1-\kappa)} E[X_2 \times 1_{\{X_{Pg} > \theta\}}] & , c \leq \theta \\ \frac{1-w}{1-\kappa} E[X_2 \times 1_{\{X_2 > c\}}] = (1-w)TVaR_{\kappa}(X_2) & , c > \theta \end{cases}$$

où $c = VaR_{\kappa}(X)$

Caractéristique du modèle GB2-PaG

Informations sur les loi GB2-PaG

Loi	$E[X]$	$E[X \times 1_{\{X > \theta\}}] / E[X]$
empirique	40296	77%
GB2-PaG	41077	78%

(a) Toronto

Loi	$E[X]$	$E[X \times 1_{\{X > \theta\}}] / E[X]$
empirique	47077	83%
GB2-PaG	47135	83%

(b) San Francisco

Loi	$E[X]$	$E[X \times 1_{\{X > \theta\}}] / E[X]$
empirique	75800	64%
GB2-PaG	73187	63%

(c) NFID

Remarque : Comme le paramètre α est inférieur à 2 pour tous ces modèles, la variance n'existe pas.

Voici les valeurs numériques pour la mesure VaR et $TVaR$:





Loi \ κ	0.90	0.95	0.99	0.995	0.999
empirique	130000	290000	855895	1500000	5000000
GB2-PaG	125968	273232	901700	1476447	4657006

Tableau – Mesure VaR

Loi \ κ	0.90	0.95	0.99	0.995	0.999
empirique	619893	1042588	3378455	5881809	18180293
GB2-PaG	592403	999682	3154707	5176620	16365563

Tableau – Mesure $TVaR$

α	$E[X \times 1_{\{4 \leq x < 8\}}]$
1.20	40450
1.42	35992
2.00	29057
2.50	26215

-  Antal, D. P. and Re, S. (2007).
Mathematical methods in reinsurance.
Lecture Notes, ETH Zurich.
-  Balkema, A. A. and De Haan, L. (1974).
Residual life time at great age.
The Annals of probability, 2(5) :792–804.
-  Brazauskas, V. and Kleefeld, A. (2016).
Modeling severity and measuring tail risk of norwegian fire claims.
North American Actuarial Journal, 20(1) :1–16.
-  Embrechts, P., Klüppelberg, C., and Mikosch, T. (2013).
Modelling extremal events : for insurance and finance, volume 33.
Springer Science & Business Media.



Fisher, R. A. and Tippett, L. H. C. (1928).

Limiting forms of the frequency distribution of the largest or smallest member of a sample.

In *Mathematical proceedings of the Cambridge philosophical society*, volume 24, pages 180–190. Cambridge University Press.



McNeil, A. J. (1997).



Estimating the tails of loss severity distributions using extreme value theory.

ASTIN Bulletin : The Journal of the IAA, 27(1) :117–137.



Pareto, V. (1897).

Cours d'economie politique, vol. 2, paris, f.

-  Pickands III, J. (1975).
Statistical inference using extreme order statistics.
the Annals of Statistics, pages 119–131.
-  Schmutz, M. and Doerr, R. R. (1998).
The Pareto model in property reinsurance : formulas and applications.
Swiss Reinsurance Company.