

Classification à partir de comportements extrêmes : exemples en cyber-risque et risques naturels

Olivier Lopez et Maud Thomas

ISUP, Sorbonne Université



La recherche à l'ISUP

- Les principaux laboratoires de Sorbonne Université en lien avec l'ISUP :
 - Mathématiques : Laboratoire de Probabilités Statistique et Modélisation (LPSM), Laboratoire Jacques-Louis Lions (LJLL)
 - Informatique : LIP6
 - Santé Publique : Institut Pierre Louis d'Épidémiologie et Santé Publique
 - Géosciences : Institut de Sciences de la Terre de Paris (ISTeP)
 - Géographie : Laboratoire de Géographie Physique (LGP)
 - ...
- **Laboratoire de Probabilités Statistique et Modélisation (UMR CNRS 8001) :**
 - Création récente (fusion de deux laboratoires en 2018)
 - Laboratoire Sorbonne Université - Université de Paris
 - Environ 200 permanents sur deux sites
 - 6 équipes (notamment "Mathématiques financières et actuarielles, probabilités numériques", "Statistique, données, algorithmes", "Modélisation aléatoire du vivant").

Quelques thèmes de recherche

- **La recherche en actuariat à l'ISUP :**

- Intelligence artificielle (machine learning) en assurance
- Finance, gestion ressources naturelles
- Comportements des clients (résiliations, rachats...)
- Risques naturels
- Risques biométriques
- Cyber assurance (Initiative de recherche commune avec Ensaë)
- Nouveaux risques

- **Interdisciplinarité :** l'un des axes stratégiques de l'ISUP est de faire émerger des collaborations de recherche entre diverses équipes de recherche de Sorbonne Université.

- Groupe de travail ARC (Actuariat et Risques Contemporains) (co-organisé avec nos collègues de l'Ensaë et Paris Dauphine) le vendredi après-midi (accès libre).

Des exemples de risques extrêmes



- Gestion du risque
- Événement extrême : observations dépassant un seuil élevé
- Manque de données et/ou d'informations historiques
- Présence d'une hétérogénéité



⇒ Évaluer le coût potentiel d'un événement extrême peut être un véritable défi

La question de l'assurabilité

- Risques extrêmes : un seul sinistre peut détruire l'équilibre d'un portefeuille.
- Risques systémiques : défaillance simultanée de nombreux assurés met à mal le mécanisme de mutualisation.
- Montée en puissance de certaines garanties : le nombre d'assurés peut être insuffisant pour assurer la mutualisation.
- Ces problématiques pose la question de **l'assurabilité** des risques.
- Nous nous intéressons à l'étude de cette question via deux angles principaux :
 - une meilleure connaissance du risque : les techniques statistiques permettent une meilleure connaissance du risque, et donc une réduction de l'incertitude.
 - la modélisation des comportements des différents acteurs.

Plan du webinaire

Théorie des valeurs des extrêmes

- Principaux résultats

- Extrêmes et assurabilité

- Mélanges d'extrêmes

Zoom sur deux risques : cyber et catastrophes naturelles

- Cyber-assurance

- Catastrophes naturelles

Tree-based methods and extreme value theory

- Regression trees

- GPD CART

- Illustrations

Conclusion, perspectives

Plan du webinaire

Théorie des valeurs des extrêmes

- Principaux résultats

- Extrêmes et assurabilité

- Mélanges d'extrêmes

Zoom sur deux risques : cyber et catastrophes naturelles

- Cyber-assurance

- Catastrophes naturelles

Tree-based methods and extreme value theory

- Regression trees

- GPD CART

- Illustrations

Conclusion, perspectives

Théorie des valeurs extrêmes

Un peu d'histoire



- À Delft (Pays-Bas) en 1953, une tempête fût responsable de milliers de morts et la destruction de centaines de maisons.
- Décision du gouvernement
→ Construire une digue telle qu'il n'y ait qu'une inondation tous les 10 000 ans
- MAIS avec uniquement un historique de 100 ans

Théorie des valeurs extrêmes

Un peu d'histoire



- À Delft (Pays-Bas) en 1953, une tempête fût responsable de milliers de morts et la destruction de centaines de maisons.
- Décision du gouvernement
→ Construire une digue telle qu'il n'y ait qu'une inondation tous les 10 000 ans
- MAIS avec uniquement un historique de 100 ans

⇒ Comment déterminer la hauteur de la digue ?

Théorie des valeurs extrêmes

Un peu d'histoire



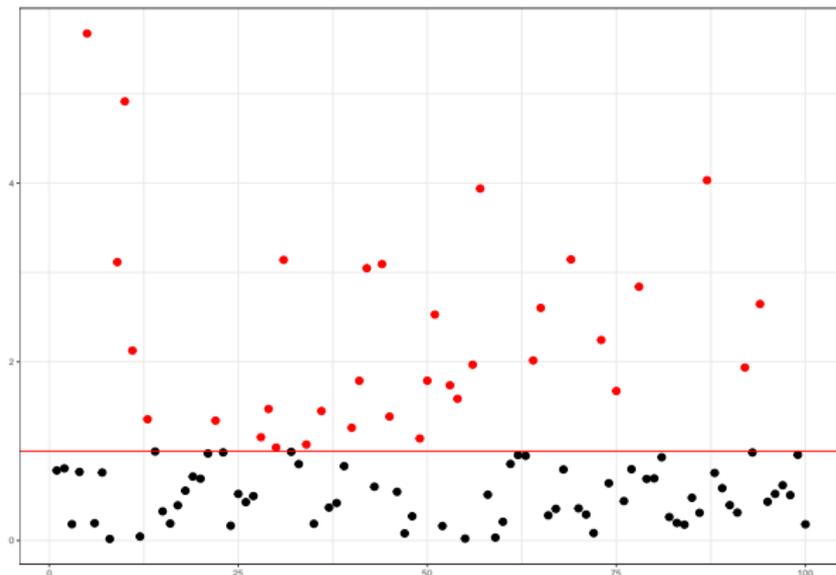
Buts de la théorie des valeurs extrêmes

1. Estimer des quantiles extrêmes (ordre très grand ou très petit)
2. Estimer la probabilité d'occurrence d'un événement plus extrêmes que ceux déjà observés

⇒ Inférence au-delà du support de l'échantillon

Méthode "Peaks-over-Threshold"

- Y_1, Y_2, \dots une suite de variables aléatoires i.i.d.
- Fixons un seuil (élevé) u
- Événement extrême = Y_i dépasse u
 - Sachant que $Y_i > u$, un excès est défini par $Z_i = Y_i - u$



Méthode "Peaks-over-Threshold"

- Y_1, Y_2, \dots une suite de variables aléatoires i.i.d.
- Fixons un seuil (élevé) u
- **Événement extrême** = Y_i dépasse u
→ Sachant que $Y_i > u$, un **excès** est défini par $Z_i = Y_i - u$

Balkema et de Haan (1974), Pickands (1975)

Sous certaines conditions, la loi des excès F_u

$$F_u(z) = \mathbb{P}[Y_i - u \geq z \mid Y_i > u]$$

converge vers une loi de Pareto généralisée (GPD) dont la fonction de répartition est

$$H_{\sigma, \gamma}(z) = \begin{cases} 1 - \left(1 + \frac{\gamma}{\sigma} z\right)^{-1/\gamma} & \text{si } \gamma \neq 0 \\ 1 - \exp\left(-\frac{z}{\sigma}\right) & \text{si } \gamma = 0 \end{cases}$$

- Familles des lois possibles pour les excès = famille paramétrique
→ **loi de Pareto généralisée (GPD)**

Generalized Pareto Distributions

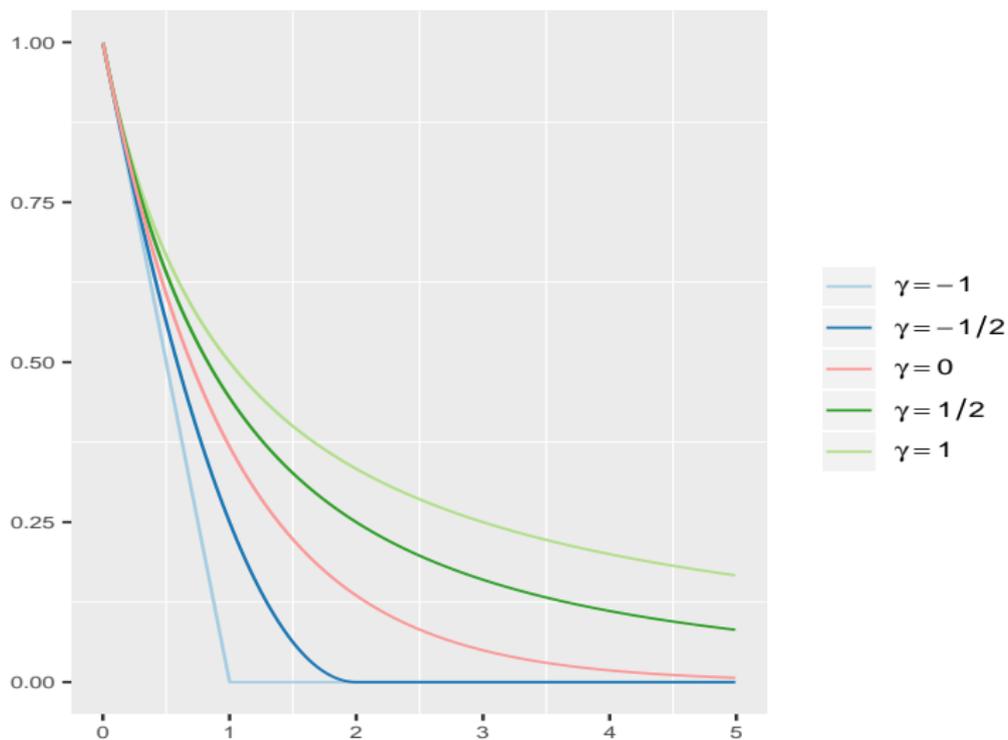


Figure: Fonctions de survie des GPD

3 domaines d'attraction

1. Domaine de Fréchet ($\gamma > 0$): **lois à queue lourde**

$$1 - G_\gamma(z) \underset{+\infty}{\sim} \gamma^{-1/\gamma} z^{-1/\gamma}$$

Exemples: Cauchy, Log-gamma, Student

2. Domaine de Gumbel ($\gamma = 0$): **lois à queue légère**

$$1 - G_0(z) \underset{+\infty}{\sim} \exp(-z)$$

Exemples: Normale, Gamma, Exponentielle

3. Domaine de Weibull ($\gamma < 0$): **lois à queue finie**

$$1 - G_\gamma\left(\frac{1}{\gamma} - z\right) \underset{0}{\sim} (-\gamma)^{-1/\gamma} z^{-1/\gamma}$$

Exemples: Uniforme, Bêta

Théorie des Valeurs Extrêmes et Assurabilité

- N = nombre de sinistres ($E[N]$ = **Fréquence**)
- Y coût moyen d'un sinistre ($E[Y]$ = **Sévérité**)

Tarification = équilibre (en moyenne) entre le coût d'un assuré et les engagements de l'assureur

$$\pi = E[N]E[Y]$$

- π = prime du contrat d'un assuré
- Hypothèse classique: Y et N sont indépendantes
- **Mutualisation** : si on dispose de n assurés identiques, la perte de l'assureur devrait être environ $n\pi$.

Théorie des Valeurs Extrêmes et Assurabilité

- $Y =$ coût d'un sinistre.
- $\gamma > 0$ (loi à queue lourde).
- Si $Z = Y - u | Y \geq u$ suit une GPD de paramètres γ et σ ,

$$\mathbb{E}[Z | Z \geq 0] = \begin{cases} \frac{\sigma}{1-\gamma} & \text{si } \gamma < 1. \\ \infty & \text{si } \gamma \geq 1. \end{cases}$$

- Dans le cas $\gamma \geq 1$, le sinistre n'est "pas assurable". L'assureur :
 - peut **exclure** le risque.
 - peut introduire des **limites** de garanties (d'autant plus basses que γ est grand).

Risques extrêmes et limites de police

- Si on met une limite, γ devient négatif.
- Néanmoins, plus le γ "d'origine" était important, plus la limite de la police doit être **basse** pour permettre d'avoir un tarif raisonnable, et une provision viable.
- **Idée générale** : la valeur de ce paramètre γ va conditionner la qualité de la couverture. Plus γ est grand, plus des restrictions sont à prévoir, limitant l'efficacité de la police dans les cas dits "extrêmes".

γ et population assurée hétérogène

- Exemple : on considère deux types d'incidents :
 - type 1 : peut engendrer des pertes sévères, mais reste assurable ($\gamma_1 < 1$).
 - type 2 : peut engendrer des pertes sévères, très aléatoires, avec $\gamma_2 > 1$.
- Un(e) statisticien(ne) se trouve face à un jeu de données recensant des coûts de tels sinistres **sans savoir s'il s'agit d'un incident de type 1 ou de type 2.**
- Conséquence : il/elle va estimer un γ proche de γ_2 .
- Alternative : si on a pu retracer la différence entre les deux sévérités, on peut proposer une **limite moins basse** pour les incidents de type 1, sans forcément un gros impact sur le prix.
- D'où l'idée de comprendre comment des variables décrivant les caractéristiques de l'événement, le comportement des individus, influent sur le comportement de queue de distribution.

γ et population assurée hétérogène

- Exemple : on considère deux types d'incidents :
 - type 1 : peut engendrer des pertes sévères, mais reste assurable ($\gamma_1 < 1$).
 - type 2 : peut engendrer des pertes sévères, très aléatoires, avec $\gamma_2 > 1$.
- Un(e) statisticien(ne) se trouve face à un jeu de données recensant des coûts de tels sinistres **sans savoir s'il s'agit d'un incident de type 1 ou de type 2**.
- Conséquence : il/elle va estimer un γ proche de γ_2 .
- Alternative : si on a pu retracer la différence entre les deux sévérités, on peut proposer une **limite moins basse** pour les incidents de type 1, sans forcément un gros impact sur le prix.
- D'où l'idée de comprendre comment des variables décrivant les caractéristiques de l'événement, le comportement des individus, influent sur le comportement de queue de distribution.

γ et population assurée hétérogène

- Exemple : on considère deux types d'incidents :
 - type 1 : peut engendrer des pertes sévères, mais reste assurable ($\gamma_1 < 1$).
 - type 2 : peut engendrer des pertes sévères, très aléatoires, avec $\gamma_2 > 1$.
- Un(e) statisticien(ne) se trouve face à un jeu de données recensant des coûts de tels sinistres **sans savoir s'il s'agit d'un incident de type 1 ou de type 2**.
- Conséquence : il/elle va estimer un γ proche de γ_2 .
- Alternative : si on a pu retracer la différence entre les deux sévérités, on peut proposer une **limite moins basse** pour les incidents de type 1, sans forcément un gros impact sur le prix.
- D'où l'idée de comprendre comment des variables décrivant les caractéristiques de l'événement, le comportement des individus, influent sur le comportement de queue de distribution.

γ et population assurée hétérogène

- Exemple : on considère deux types d'incidents :
 - type 1 : peut engendrer des pertes sévères, mais reste assurable ($\gamma_1 < 1$).
 - type 2 : peut engendrer des pertes sévères, très aléatoires, avec $\gamma_2 > 1$.
- Un(e) statisticien(ne) se trouve face à un jeu de données recensant des coûts de tels sinistres **sans savoir s'il s'agit d'un incident de type 1 ou de type 2**.
- Conséquence : il/elle va estimer un γ proche de γ_2 .
- Alternative : si on a pu retracer la différence entre les deux sévérités, on peut proposer une **limite moins basse** pour les incidents de type 1, sans forcément un gros impact sur le prix.
- D'où l'idée de comprendre comment des variables décrivant les caractéristiques de l'événement, le comportement des individus, influent sur le comportement de queue de distribution.

Théorie des valeurs extrêmes et régression

- Cadre de la régression
 - Considère une observation de caractéristiques X
 - Suppose que la loi de $Y | X$ est à queue lourde = la loi des excès $Z | X$ converge vers une GPD de paramètre $\gamma(X) > 0$ et $\sigma(X)$

$$H_{\sigma, \gamma}(z) = 1 - \left(1 + \frac{\gamma(X)}{\sigma(X)} z \right)^{-1/\gamma(X)}$$

- But : estimer $\gamma(X)$
- Espoir : pour certaines valeurs de X , $\gamma(X) < 1$.
- Méthodes :
 - Approches semi-paramétriques :
 - Exponential regression model (Beirlant et al., 2003)
 - Smoothing splines (Chavez-Demoulin et al., 2015)
 - Approches non paramétriques (Beirlant and Goegebeur, 2004) :
 - Local polynomial maximum likelihood
 - Only for continuous covariates

Plan du webinaire

Théorie des valeurs des extrêmes

Principaux résultats

Extrêmes et assurabilité

Mélanges d'extrêmes

Zoom sur deux risques : cyber et catastrophes naturelles

Cyber-assurance

Catastrophes naturelles

Tree-based methods and extreme value theory

Regression trees

GPD CART

Illustrations

Conclusion, perspectives

Cyber-risque

- Cyber-risque: usage inapproprié d'outils numériques et de systèmes d'informations.
- Un cyber incident peut être volontaire (attaque cyber) ou non (accidentel).
- Dans le cas du piratages, les hackers utilisent des vulnérabilités dans les systèmes d'information, mais aussi des vulnérabilités humaines dans les organisations.
- Différents types d'attaques (ransomware, phishing, fraudes classiques utilisant le vecteur numérique...)
- Frappe les états, les industries, les individus.

Wannacry



- Rançongiciel Wannacry : attaque cyber mondiale en Mai 2017.
- Utilise la vulnérabilité "EternalBlue".
- Approximativement 300 000 postes infectés dans 150 pays sur une semaine approximativement.
- Estimation du coût : en centaines de millions de dollars voire milliards. (\$100 millions pour le NHS seul).

Colonial Pipeline

The Colonial Pipeline Cyberattack | What to Know | Pipeline Resumes | \$5 Million Ransom | DarkSide Says Shutting Down | Weaknesses in U.S. Cybersecurity

Cyberattack Forces a Shutdown of a Top U.S. Pipeline

The operator, Colonial Pipeline, said it had halted systems for its 5,500 miles of pipeline after being hit by a ransomware attack.



- Augmentation de 4.2% du WTI et du Brent.
- "Double extorsion": rançongiciel combiné avec chantage à la divulgation de données.
- Auteurs: le groupe de hackers "Darkside" (Ransomware as a service).

"Lucy"

- L'AMRAE ("Association pour le Management des Risques et des Assurances de l'Entreprise") a publié en 2021 une étude intitulée "LUMière sur la CYberassurance".
- 87% des grandes entreprises ont souscrit une assurance cyber, seulement 8% des compagnies de tailles intermédiaires.
- Volume de primes qui a augmenté de 49% entre 2019 et 2020.
- Le nombre de sinistres a été multiplié par 3.
- Le S/P est de 167% en 2020, 84% en 2019.

Régime CatNat

- Catastrophes naturelles assurées par un partenariat privé-public
- Couvre un grand nombre de risques: inondations, glissements de terrain, sécheresse, tremblements de terre,...
- **Quelques chiffres**
 - entre 1988-2013 :
 - 48,3M€ ont été payés par les assurances pour des catastrophes naturelles
 - 431 sinistres par an
 - 1,9M€ par an
 - entre 2014 - 2039 :
 - 92M€ devraient être payés par les assurances pour des catastrophes naturelles
 - +114% sur le coût des inondations, +39% sur le coût des tempêtes
 - 1,9M€ par an
- **Spécificité** : pour recevoir une indemnité la ville de l'assuré doit faire une demande auprès d'un comité pour être reconnue en état de Catastrophe Naturelle

Objectifs

→ Est-il possible de prédire son coût peu de temps après son occurrence et pour l'ensemble du marché français ?

- Projet R&D "CatClimData" de la Mission Risques Naturels (MRN), dont les objectifs sont
 1. de contribuer au processus d'évaluation des événements CatNat en France, peu de temps après leur occurrence
 2. de mieux comprendre les dommages dus aux événements de catastrophes naturelles, peu de temps après leur occurrence
 3. de contribuer à l'intérêt général de l'assurance en France
- Pour cela, besoin d'une méthode qui nous permette de
 - mieux comprendre l'hétérogénéité de ces différents événements
 - mieux comprendre l'impact de leurs caractéristiques sur leur sévérité

Plan du webinaire

Théorie des valeurs des extrêmes

- Principaux résultats

- Extrêmes et assurabilité

- Mélanges d'extrêmes

Zoom sur deux risques : cyber et catastrophes naturelles

- Cyber-assurance

- Catastrophes naturelles

Tree-based methods and extreme value theory

- Regression trees**

- GPD CART**

- Illustrations**

Conclusion, perspectives

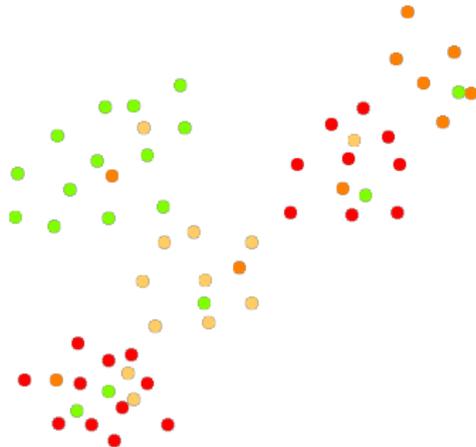
Classification And Regression Trees (CART)

Regression tree (Breiman et al., 1984)

$$m^* = \arg \min_{m \in \mathcal{M}} \mathbb{E}[\phi(Y, m(\mathbf{X}))],$$

- Y est une variable à expliquer (le coût d'un sinistre dans notre cas)
- $\mathbf{X} \in \mathcal{X} \subset \mathbb{R}^d$ est un ensemble de variables explicatives
- \mathcal{M} est une classe de fonctions cibles sur \mathbb{R}^d
- ϕ est une fonction de perte qui dépend de la quantité que l'on souhaite estimer

Croissance de l'arbre



CART : Step 0

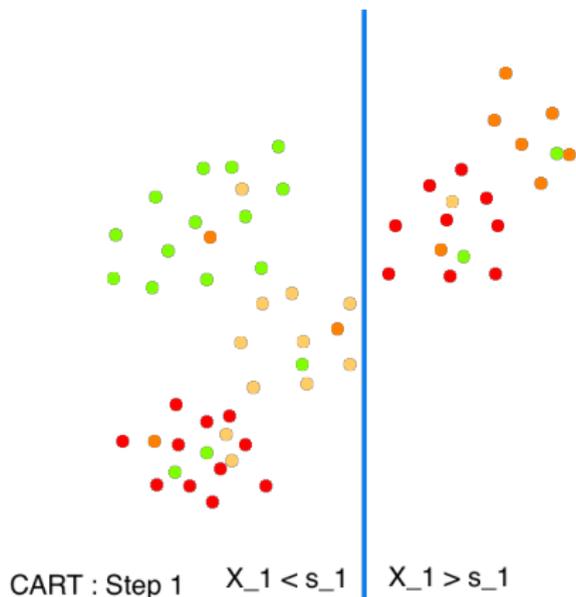
Croissance de l'arbre

Splitting rules

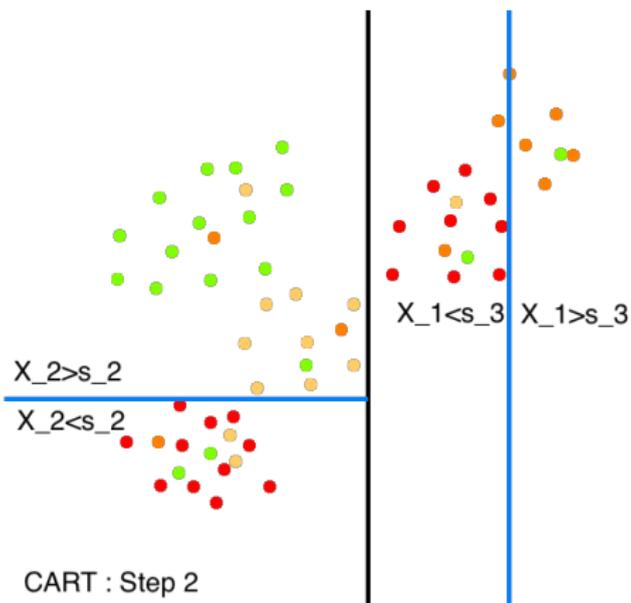
$$\mathbf{x} = (x^{(1)}, \dots, x^{(d)}) \longrightarrow R_j(\mathbf{x})$$

with

$$\begin{cases} R_j(\mathbf{x}) & = 0 \text{ ou } 1 \\ R_j(\mathbf{x}) R_{j'}(\mathbf{x}) & = 0 \text{ for } j \neq j' \\ \sum_j R_j(\mathbf{x}) & = 1 \end{cases}$$



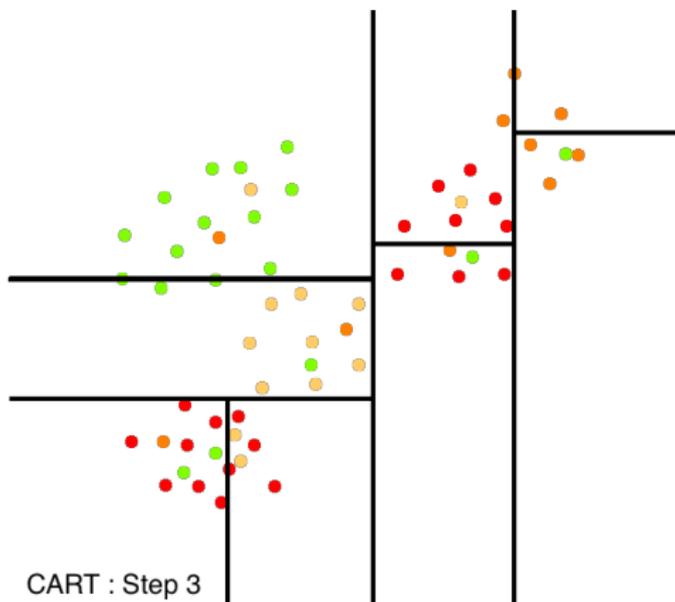
Croissance de l'arbre



Croissance de l'arbre

Estimateur de la fonction de régression $\hat{m}^{\mathcal{R}}(\mathbf{x})$ donné par

$$\hat{m}^{\mathcal{R}}(\mathbf{x}) = \sum_{j=1}^s \hat{m}(R_j) R_j(\mathbf{x}) \text{ where } \hat{m}(R_j) = \arg \min_{m \in \mathcal{M}} \sum_{i=1}^n \phi(Y_i, \mathbf{X}_i) R_j(\mathbf{X}_i)$$



Fonction de perte

- Perte **quadratique** → "Mean regression" (espérance conditionnelle)

$$\phi(y, m(\mathbf{x})) = (y - m(\mathbf{x}))^2$$

$$\hookrightarrow m^*(\mathbf{x}) = \mathbb{E}[Y | \mathbf{X} = \mathbf{x}]$$

- Perte **absolue** → "Median regression" (médiane conditionnelle)

$$\phi(y, m(\mathbf{x})) = |y - m(\mathbf{x})|$$

$$\hookrightarrow m^*(\mathbf{x}) = \text{médiane conditionnelle}$$

- Perte liée à la **log-vraisemblance**, ici GPD

$$\phi(y, m(\mathbf{x})) = -\log(\sigma(\mathbf{x})) - \left(\frac{1}{\gamma(\mathbf{x})} + 1 \right) \log \left(1 + \frac{y\gamma(\mathbf{x})}{\sigma(\mathbf{x})} \right),$$

$$\hookrightarrow m^*(\mathbf{x}) = (\sigma(\mathbf{x}), \gamma(\mathbf{x}))$$

Fonction de perte

- Perte **quadratique** → "Mean regression" (espérance conditionnelle)

$$\phi(y, m(\mathbf{x})) = (y - m(\mathbf{x}))^2$$

$$\hookrightarrow m^*(\mathbf{x}) = \mathbb{E}[Y | \mathbf{X} = \mathbf{x}]$$

- Perte **absolue** → "Median regression" (médiane conditionnelle)

$$\phi(y, m(\mathbf{x})) = |y - m(\mathbf{x})|$$

$$\hookrightarrow m^*(\mathbf{x}) = \text{médiane conditionnelle}$$

- Perte liée à la **log-vraisemblance**, ici GPD

$$\phi(y, m(\mathbf{x})) = \left\{ -\log(\sigma(\mathbf{x})) - \left(\frac{1}{\gamma(\mathbf{x})} + 1 \right) \log \left(1 + \frac{(y - u)\gamma(\mathbf{x})}{\sigma(\mathbf{x})} \right) \right\} \mathbf{1}_{y \geq u},$$

$$\hookrightarrow m^*(\mathbf{x}) = (\sigma(\mathbf{x}), \gamma(\mathbf{x}))$$

Élagage : sélection de modèle

- Soit T_{\max} l'arbre maximal obtenu dans la première phase et K_{\max} le nombre de ses feuilles
- Consiste à extraire de T_{\max} un sous-arbre.
- Critère pénalisé (n_T nombre de feuilles de l'arbre T)

$$C_\alpha(T) = \sum_{i=1}^n \phi(Y_i, m^{\mathcal{R}^T}(\mathbf{X}_i)) + \alpha n_T$$

- $\alpha > 0$ est choisi par cross-validation
- On désigne par \hat{T}_K le meilleur arbre à K feuilles d'après ce critère.
- \hat{T} l'arbre minimisant le critère pénalisé, \hat{K} son nombre de feuilles ($\hat{T} = \hat{T}_{\hat{K}}$).

Privacy Rights Clearinghouse (nonprofit association)

- Fondée en 1992
- Publique
- Référence pour les travaux académiques d'analyse des événements cyber liés à la fuite de données
- But : faire prendre conscience des problématiques de respect de la vie privée.
- Chronologie des fuites de données maintenue depuis 2005.
- Contient des informations sur des événements de multiples sources:
 - Agences gouvernementales US (Federal level–HIPAA).
 - Agences gouvernementales US (State level)
 - Média
 - Autres organisations
- 8860 événements

PRC database: variables

- Variables liées à l'exposition : informations sur la victime.
- Variables liées à l'événement : informations sur la fuite de données.

Exposition variables	Name of the organisation Type of organisation Localisation of the organisation
Event variables	Source of the release Date of the release Type of breaches Number of affected records Description of the event

Lien entre "Number of Records" et coût

- On note :
 - Y = number of records;
 - L = perte associée.
- Jacobs (2014) estime, à partir de données de l'institut Ponemon (non publiques), le lien entre L et Y de la façon suivante :

$$\log L = 7.68 + 0.76 \log Y.$$

- Dans notre travail, nous proposons une version actualisée, compatible avec des "mega-breaches" apparues en 2018 :

$$\log L = 9.59 + 0.57 \log Y.$$

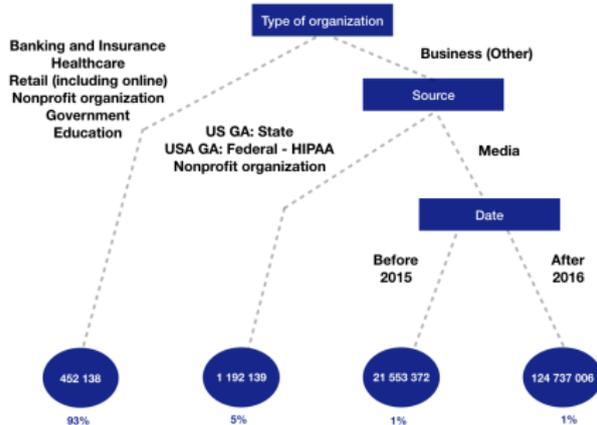
- Il y a un lien entre le tail index de L (γ_L) et celui de Y (γ_Y) :

$$\gamma_L = 0.57 \gamma_Y.$$

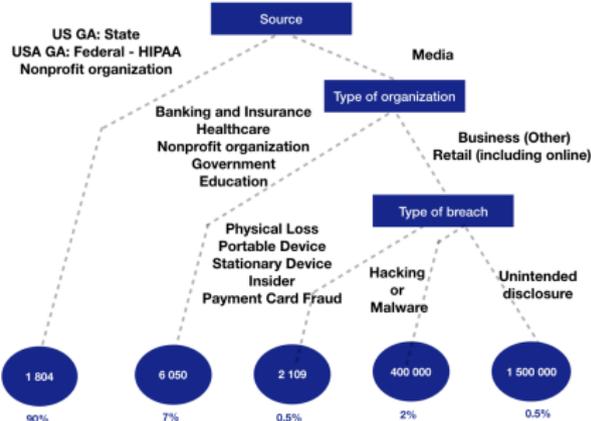
Application à la PRC database

Mean et median regression trees

Analysis of risk heterogeneity by a mean tree with quadratic loss



Analysis of risk heterogeneity by a median tree with absolute loss

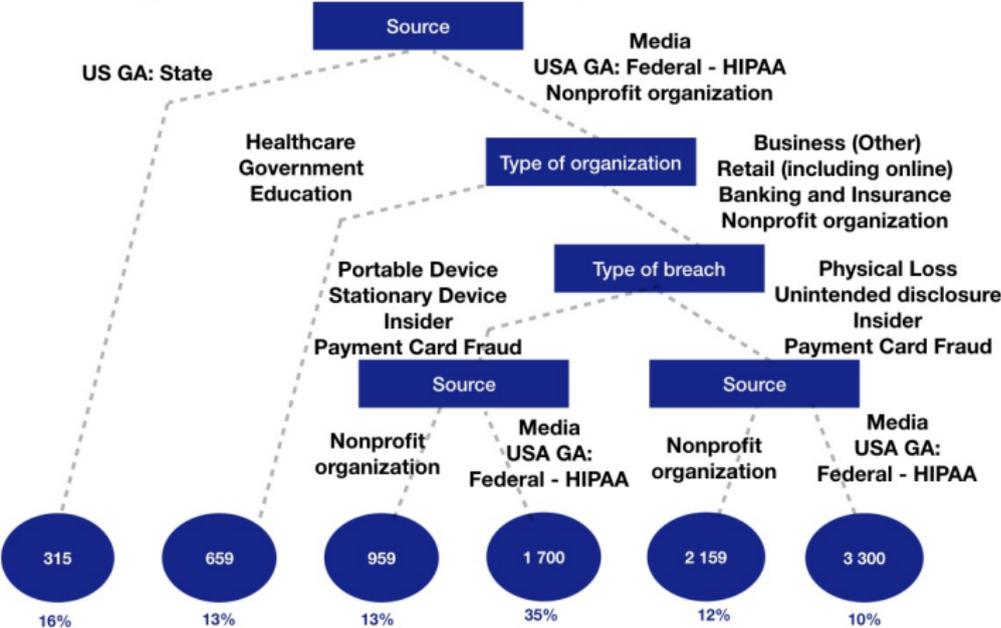


Application à la base PRC

- Sévérité caractérisée par le nombre de records
- On estime toute la distribution du nombre de records
- On a besoin d'analyser séparément la partie centrale de la distribution et sa queue
- On choisit un seuil u
 - ↪ Fixé ici à $u = 27000$
- **Partie centrale** de la distribution
 - On ne considère que les cas où le nombre de records est plus petit que 27 000
 - On ajuste un CART "médiante"
 - Sur chaque feuille, on peut ajuster une distribution (par exemple log-normal)
- **Tail** of the distribution
 - On considère les observations telles que le nombre de records est au-dessus de 27 000
 - On ajuste un GPD CART

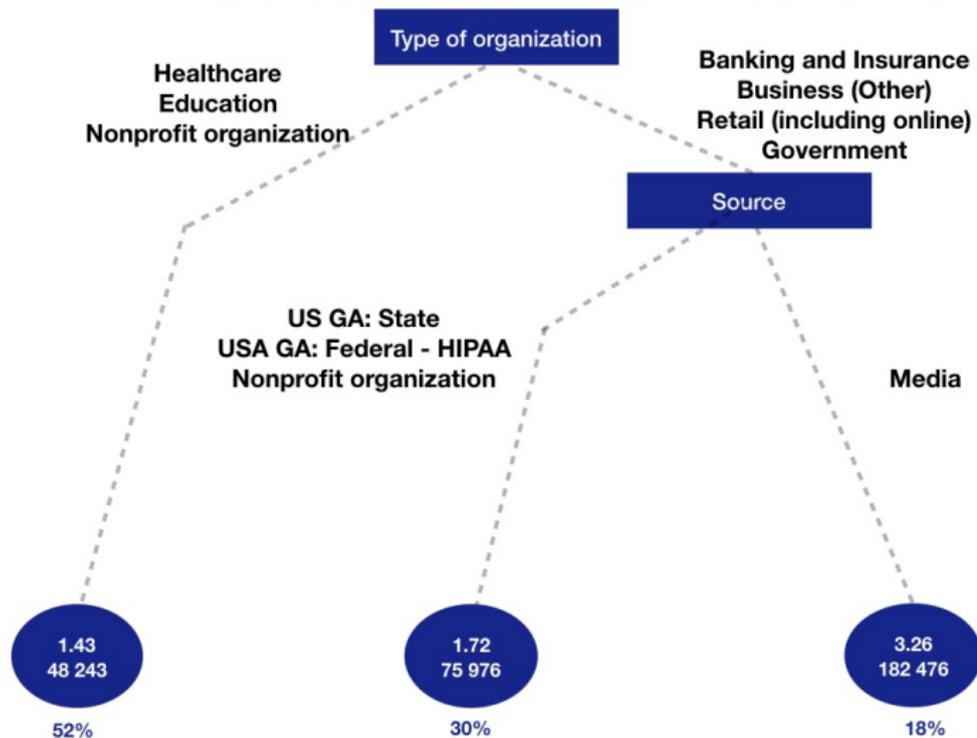
Partie centrale de la distribution: Median tree

Analysis of the central part of the distribution by a median tree with absolute loss



Queue de distribution: GPD tree

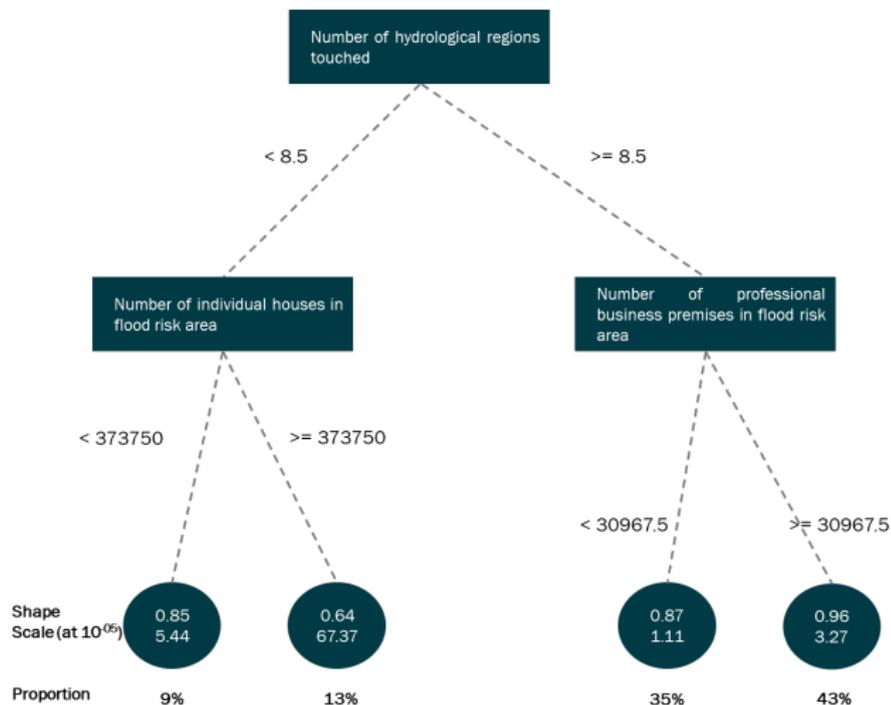
Analysis of the tail part of the distribution by a Generalized Pareto Regression tree



Application aux événements inondations

- Base de données SILECC
 - Partenariat avec la MRN
 - Constituée des sinistres des plus grandes compagnies d'assurance (70% du marché français)
 - 700 000 sinistres de 1990 à 2019 dont **3 147 événements inondations**
- Covariables (disponibles peu de temps après l'occurrence de l'événement)
 - la région météorologique
 - la saison
 - le type d'inondations
 - le nombre de régions hydrologiques touchées
 - le nombre de maisons de maisons individuelles
 - le nombre de locaux professionnels dans la zone inondable
- Le seuil u a été choisi égal à 100 000€, ce qui correspond à 1 083 événements

Application aux événements inondations



Plan du webinaire

Théorie des valeurs des extrêmes

- Principaux résultats

- Extrêmes et assurabilité

- Mélanges d'extrêmes

Zoom sur deux risques : cyber et catastrophes naturelles

- Cyber-assurance

- Catastrophes naturelles

Tree-based methods and extreme value theory

- Regression trees

- GPD CART

- Illustrations

Conclusion, perspectives

Conclusion

- Classification des comportements extrêmes via des méthodes d'arbres :
 - permet de considérer des nonlinéarités dans cette dépendance
 - adapté aux variables X qui sont discrètes comme continues (intéressant notamment pour l'étude des comportements car de nombreuses variables sont qualitatives)
 - permet une classification
 - mais peut parfois être instable
- Des extensions naturelles, moins intelligibles mais plus précises (gradient boosting, forêts aléatoires) peuvent être utilisées
- De façon générale, outil d'aide à la décision pour tracer la ligne entre ce qui est "assurable" et ne l'est pas.

Merci de votre attention !