Illustration et limites de la théorie des valeurs extrêmes
pour des cumuls pluviométriques exceptionnels

Liliane Bel

Séminaire d'Alembert

17 novembre 2010

Situations à risque

Situations pour lesquelles on a besoin d'évaluer un risque

- risque climatique : inondations, crues, avalanches
- risque géologique : séisme, irruption volcanique
- risque environnemental : pollution de l'air, de l'eau
- risque alimentaire
- risque médical
- risque financier
- ...

Protection contre les risques

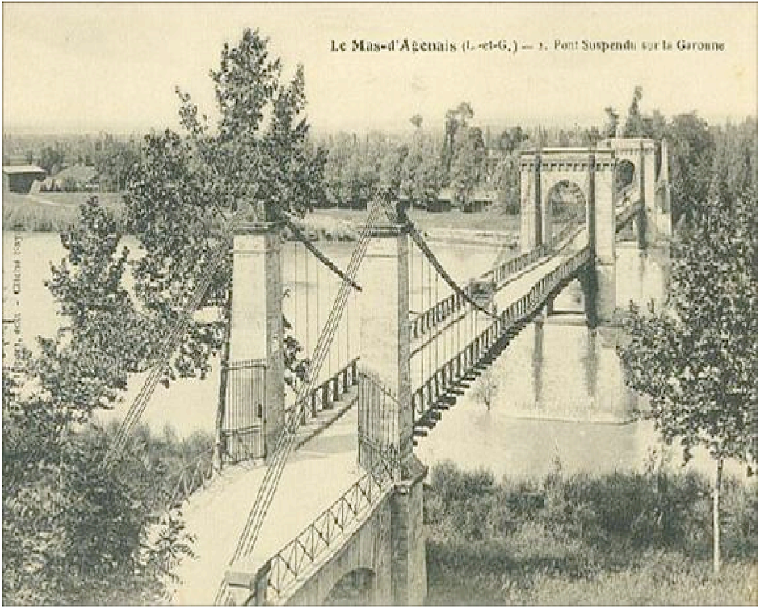
Dans chaque situation on peut anticiper et essayer de prévenir le risque

- digues, ouvrages paravalanches
- réglementation: POS, normes de construction
- alertes à la pollution
- controles sanitaires

Il faut pouvoir quantifier le risque pour y adapter la prévention.

La Garonne en colère

La Garonne



Problème décisionnel

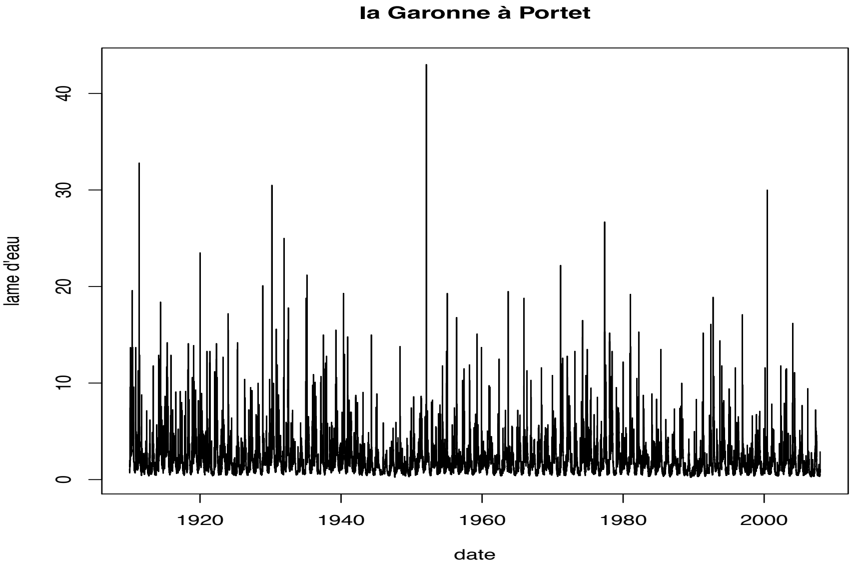
Pour se prémunir des dégats causés par les crues de la Garonne on voudrait construire une digue.

- Quel type de digue (en béton, en terre)?
- Quelle hauteur de digue?

Compromis entre

- le cout de construction de la digue
- le cout des dommages potentiels suite à une crue.

Problème décisionnel



Problème décisionnel

Première idée

construire la digue pour se prémunir contre la hauteur maximale observée.

Problème décisionnel

Première idée

construire la digue pour se prémunir contre la hauteur maximale observée.

mauvaise idée

Problème décisionnel

Première idée

construire la digue pour se prémunir contre la hauteur maximale observée.

mauvaise idée

Deuxième idée

hauteur maximale + hauteur de sécurité

Problème décisionnel

Première idée

construire la digue pour se prémunir contre la hauteur maximale observée.

mauvaise idée

Deuxième idée

hauteur maximale + hauteur de sécurité

Comment évaluer cette hauteur de sécurité?

Problème statistique:

calculer la probabilité que la crue dépasse un niveau donné

Problème décisionnel

Première idée

construire la digue pour se prémunir contre la hauteur maximale observée.

mauvaise idée

Deuxième idée

hauteur maximale + hauteur de sécurité

Comment évaluer cette hauteur de sécurité?

Problème statistique:

calculer la probabilité que la crue dépasse un niveau donné

Difficulté: ce niveau n'a jamais été observé.

Théorie des valeurs extrêmes

Objectif estimer les queues de distribution à partir de l'observation d'un échantillon

Distribution de probabilité: probabilité d'observer un niveau inférieur à un ensemble de valeurs

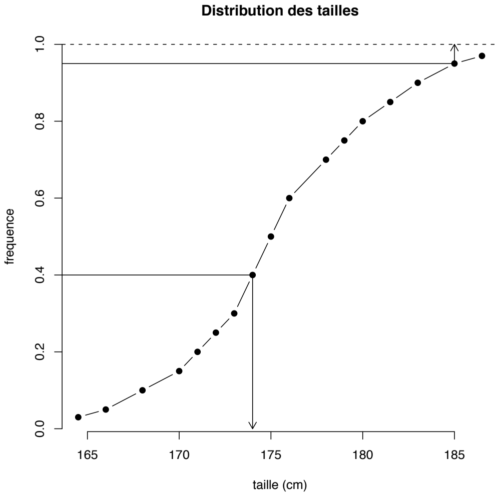
X variable aléatoire

$$F_X(x) = Proba(X < x) \qquad x \in \mathbb{R}$$

Exemple: taille des individus de sexe masculin

Distribution de probabilité

seuil	fréquence
164.5 cm	3 %
166.0 cm	5 %
168.0 cm	10 %
170.0 cm	15 %
171.0 cm	20 %
172.0 cm	25 %
173.0 cm	30 %
174.0 cm	40 %
175.0 cm	50 %
176.0 cm	60 %
178.0 cm	70 %
179.0 cm	75 %
180.0 cm	80 %
181.5 cm	85 %
183.0 cm	90 %
185.0 cm	95 %
186.5 cm	97 %



Loi des dépassements

Soit X la variable aléatoire, u un seuil.

Thèorème de Pickands (1975)
Si u est assez grand, la probabilité que $X - u$ ne dépasse pas la valeur x sachant que X dépasse u

$$P(X - u < x | X > u) \to G_{\xi,\beta}(x) = 1 - \left(1 + \frac{\xi x}{\beta}\right)^{-1/\xi}$$

Il suffit de fixer u et de connaitre les deux paramètres ξ et β pour calculer

$$P(X > x + u | X > u)$$

Période et niveau de retour

Soit $p_{x+u} = P(X > x+u)$ la probabilité de dépasser le niveau $x+u$.
$\dfrac{1}{p_{x+u}}$ est la période de retour associée au niveau $x+u$.

La crue centennale est la crue qui est observée en moyenne une fois tous les 100 ans.
Elle a une probabilité d'être observée égale à 1/100 si on a des données annuelles, ou 1/(100*365) si on a des mesures journalières.

Calcul du niveau $x+u$

$$\begin{aligned} P(X > x+u) &= P(X-u > x \mid X > u)P(X > u) \\ &= (1 - G_{\xi\beta}(x))P(X > u) \\ x &= G_{\xi\beta}^{-1}\left(1 - p_{x+u}\frac{N_u}{N}\right) \end{aligned}$$

Estimation des paramètres

Deux grandes méthodes d'estimation

Maximum de vraisemblance
: on écrit la probabilité d'observer les données en fonction de la distribution. Les estimations des paramètres sont les valeurs qui maximisent cette probabilité (la vraisemblance).

Méthodes des moments
: on écrit les moments de la distribution (moyenne, variance, ...) en fonction des paramètres et on inverse la relation. Les paramètres sont fonction des moments empiriques.

Les paramètres obtenus le sont avec une certaine précision (intervalle de confiance).
Le niveau de retour est aussi connu avec une certaine précision.

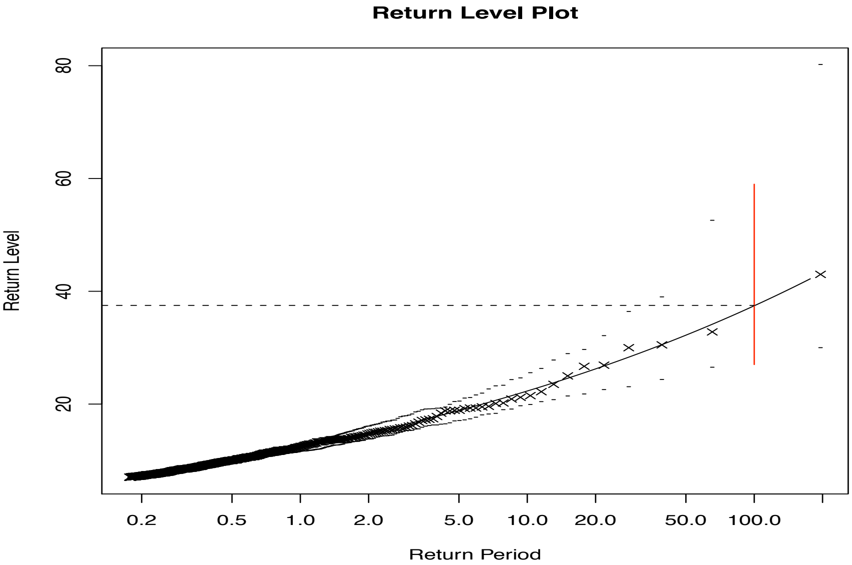
Estimation des paramètres

```
Threshold: 7
Number Above: 555
Proportion Above: 0.0155

Estimates
 scale   shape
2.5988  0.1772

Standard Errors
  scale    shape
0.17306  0.05186
```

Estimation du niveau de retour centennal



Stationnarité

Le théorème de Pickands est valide pour des suites de variables $(X_t)_t$ indépendantes.
Il peut s'étendre aux suites de variables stationnaires:
la distribution de probabilité reste la même au cours du temps i.e.

- l'espérance, la variance de X_t
- la corrélation entre X_t et X_{t+h}
- la probabilité de dépasser un niveau

ne changent pas au cours du temps.
Or

- l'environnement
- le climat

changent.

Si on pense que la suite des variables aléatoires n'est pas stationnaire, il faut isoler les causes de variations structurelles, des variations aléatoires.

Avalanche de Montroc (1999)

12 morts, 15 chalets détruits, 5 chalets endommagés.

Avalanche de Montroc (1999)

Le site de Montroc est en zone blanche (risque nul) ou en zone bleue (risque modéré).

- À Montroc, la dernière avalanche a été observée en 1945.
- Montroc aurait subi quatre autres avalanches :
    1. l'avalanche de 1843 citée dans un document établi en 1943,
    2. l'avalanche de 1908 dans les fiches des eaux et forêts,
    3. les deux autres avalanches reconnues ne sont pas datées.

Avalanche de Montroc (1999)

Modélisation Bayésienne

Ecrire un modèle qui contient de l'information a priori: expertise des guides de montagne par exemple.

Changement climatique

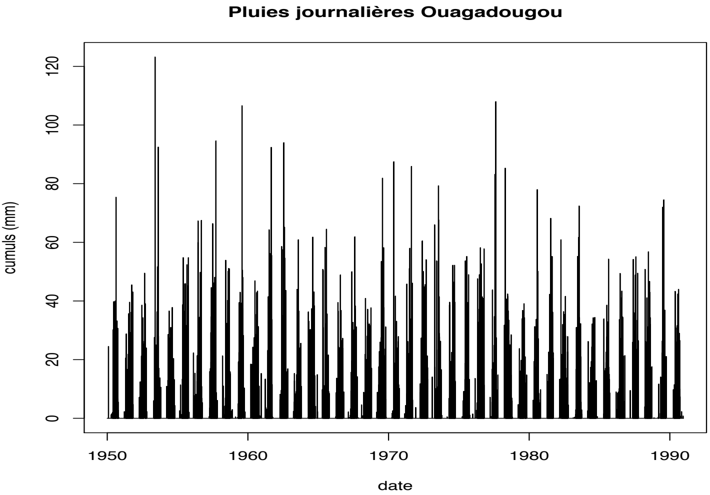
Les spécialistes s'entendent pour affirmer qu'il y a de moins en moins de neige, mais les avalanches observées sont beaucoup plus impressionnantes.

- la moyenne diminue
- la variance augmente

Inondations de Ouagadougou (2009)

5 morts et 150000 sans-abri.

Inondations de Ouagadougou (2009)



Niveau de pluviométrie observé 3/9/09: 263 mm (2 fois le max sur 1950-1990). Période de retour: 100000 ans!

Inondations de Ouagadougou(2009)

Changement climatique

- période 1990-2009 ?
- la moyenne diminue
- la variance augmente

Homogénéisation des données

- changement d'instruments de mesures
- changement de stations de mesure
- changement de protocole de mesure

Dépendance

La loi marginale du processus peut être très différente de la loi conditionnelle.

Exemple la probabilité pour un jeune de faire des études supérieures est de 50%,
la probabilité pour un jeune dont les parents sont professeurs de faire des études supérieures est de 80%.

La probabilité marginale de dépassement d'un niveau peut être très faible, mais elle peut augmenter considérablement si elle est conditionnée par un évènement favorable.

Tempête Xynthia (2010)

40 morts, des communes inondées, un million de foyers privés d'électricité.

Tempête Xynthia (2010)

Patrick Galois (Meteo France)

"Les tempêtes sont des phénomènes que l'on observe tous les cinq à dix ans en France en raison d'aléas climatiques. Si elle présente un caractère remarquable, Xynthia n'est pas pour autant le phénomène du siècle. Elle est ainsi moins exceptionnelle que celles de 1999 et ses vents sont moins intenses qu'en 2009. Mais son issue dramatique réside dans sa conjonction à un fort coefficient de marée sur la côte atlantique, au moment même de la marée haute."

Crise financière (2008)

Recipe for Disaster: The Formula That Killed Wall Street

By Felix Salmon 02.23.09

$$\Pr[T_A<1, T_B<1] = \phi_2(\phi^{-1}(F_A(1)), \phi^{-1}(F_B(1)), \gamma)$$

Here's what killed your 401(k) *David X. Li's Gaussian copula function as first published in 2000. Investors exploited it as a quick—and fatally flawed—way to assess risk. A shorter version appears on this month's cover of* Wired.

Modèles max-stables

Les modèles gaussiens couramment utilisés ne présentent pas de dépendance dans les extrêmes:
deux variables gaussiennes corrélées, n'ont pas leurs extrêmes corrélés.

Pour modéliser des phénomènes avec des extrêmes corrélés il faut faire appel à des modèles max-stables:
la distribution du maximum d'un échantillon est la même que la distribution de l'échantillon (à des constantes de normalisation près).

Les queues de distributions sont plus "épaisses": la probabilité d'observer une forte valeur est plus importante.

Modèles max-stables

Coefficient extrêmal:

$$\chi = \lim_{u \to u^*} P(X > u | Y > u) = \lim_{u \to u^*} \frac{P(X > u, Y > u)}{P(Y > u)}$$

Si (X, Y) couple max-stable $\chi > 0$, la probabilité d'observer une forte valeur pour X est plus élevée si on a observé une forte valeur pour Y.

Thème de recherche actif, avec des applications en particulier en finances, et en environnement pour des processus spatialisés.