UNIVERSITE DE LIEGE

Faculté des Sciences

Distributions des valeurs extrêmes

Mémoire présenté par
Anne-Françoise Donneau
En vue de l'obtention du grade de
Licenciée en Sciences Mathématiques

Année académique 2001-2002

Introduction

Le développement de la théorie des valeurs extrêmes est justifié par le fait que les parties extrêmes d'un échantillon peuvent être porteuses d'une information d'une importance exceptionnelle. Elles peuvent nous permettre de mettre en évidence le risque potentiel de phénomènes aléatoires tels que les innondations, les ouragans, le taux nocif de pollution de l'air,...

L'évaluation correcte de la probabilité d'une future catastrophe peut aider les décideurs à mener des actions visant à se prémunir d'un grand désastre.
Moins spectaculaire mais tout de même important, la statistique des extrêmes peut jouer un rôle décisif dans la vie économique de tous les jours ou dans la recherche de solutions à des problèmes écologiques ou techniques.

Les fondateurs du calcul des probabilités étaient certainement trop occupés par le comportement général des statistiques décrivant les valeurs centrales des distributions pour s'intéresser aux extrêmes.
Cependant, dès 1709 Nicolaus Bernoulli considéra le problème actuariel suivant :
- Quelle est en moyenne la durée de vie du dernier survivant parmi n personnes de même âge ?
Il transposa ce problème et en vint à se poser la question suivante :
- Si l'on considère n points au hasard sur un segment de droite, que vaut en moyenne la distance à l'origine du segment du point le plus éloigné ?

La théorie des valeurs extrêmes venait de voir le jour.
Par la suite, elle ne cessa de se développer, de susciter l'intérêt de chercheurs et de mathématiciens parmi lesquels nous pouvons citer R.Van Misses, L.H.C.Tippett, M.Fréchet, E.J.Gumbel,...

Cependant ce n'est qu'à partir de 1950 qu'une méthodologie a été proposée de manière sérieuse pour la modélisation de véritables phénomènes physiques. Ce n'est donc pas une coïncidence si les premières applications des modèles des valeurs extrêmes relevaient du génie civil. En effet, les ingénieurs ont toujours eu besoin de réaliser des plans des structures qu'il leur faut réaliser afin que ces dernières soient capables de résister aux forces qui leurs seront appliquées par la suite.
La théorie des valeurs extrêmes leur fournit un renseignement précieux permettant d'estimer, à partir de données historiques, les contraintes extérieures auxquelles devront résister ces structures.

Introduisons ici le concept des valeurs extrêmes à l'aide d'un exemple. Supposons être intéressé par l'étude des variations du niveau des précipitations mesurées jour après jour en un lieu donné. La suite des niveaux des précipitations observés de manière journalière

constitue les données. Le rôle de la statistique n'est pas tant de résumer les événements qui ont déjà eu lieu mais de faire des déductions à partir du caractère aléatoire du processus qui génère les données. Par exemple, si quarante pour-cent des observations correspondent à des jours de sécheresse, quelle chance avons-nous pour que demain soit un jour sans pluie ? Et si la plus grande valeur observée du niveau des précipitations est de six centimètres, quelles conclusions pouvons-nous émettre sur les niveaux des précipitations qui seront enregistrés lors des pluies battantes qui peuvent survenir dans le futur ?
La statistique des extrêmes se propose de donner une réponse à toutes ces questions.

Ce travail de fin d'étude est composé de quatre chapitres.
Dans le premier chapitre, après avoir rappelé les notions de base et fixé les notations, on se propose d'introduire et de développer les notions de dépassements, de maxima et de minima au sein d'un échantillon. On y explicite également les problèmes liés au fait que la distribution parent F est en pratique inconnue.

Ces problèmes seront résolus dans les chapitres II et III, grâce à l'introduction de différentes familles de modèles. Ce premier chapitre va donc nous servir de tremplin pour les suivants.
La première famille que l'on rencontre est la distribution généralisée des valeurs extrêmes. Celle-ci va nous servir à modéliser des blocs de données. C'est ce qui est expliqué au chapitre II, où après avoir introduit les lois de Gumbel, Fréchet et de Weibul, on présente l'un des théorèmes les plus importants de la théorie des valeurs extrêmes. Il est dû à Fisher-Tippett. Ce théorème donne la distribution asymptotique des maxima quelque soit la distribution de la variable considérée. Il spécifie que cette distribution asymptotique doit être l'une des lois de Gumbel, Fréchet ou de Weibull.

La méthode du maximum de vraisemblance est utilisée afin d'estimer les paramètres attachés à la distribution généralisée des valeurs extrêmes. Nous présentons ensuite les méthodes d'estimation des niveaux de retour sur une période donnée. Enfin, pour vérifier que le modèle choisi est bien adapté aux données, nous présentons différentes méthodes graphiques permettant de juger de la validité de l'ajustement réalisé.

Le troisième chapitre reprend la notion de dépassement d'un seuil.
On y trouve la présentation de la distribution généralisée de Pareto qui décrit la distribution des dépassements d'un seuil suffisamment grand, ainsi que deux méthodes qui vont nous permettre de choisir un seuil. La méthode du maximum de vraisemblance est aussi utilisée pour estimer les paramètres de la distribution de Pareto.

Nous clôturons ce chapitre en adaptant les techniques graphiques introduites au chapitre II à la distribution généralisée de Pareto afin de vérifier l'adéquation du modèle à nos données.

Dans le but de mettre en pratique les différentes méthodes présentées, le quatrième chapitre de ce travail est consacré à une application basée sur des observations météorologiques. Dans cette application, nous recherchons la distribution la mieux adaptée à la modélisation des températures maximales annuelles relevées à l'observatoire de Paris depuis 1700 jusque 1886. Nous en déduirons des estimations des niveaux de retour décennal et séculaire.

Table des matières

1 Seuil, dépassement et maximum **3**

1.1 Nombre de dépassements 4

1.2 Approximation de la loi Binomiale par la loi de Poisson 4

1.3 Maximum 5

1.4 Minimum 7

1.5 Amplitude des dépassements et statistiques d'ordre supérieur 7

1.6 T-seuil 9

1.7 Les dates de dépassement et le niveau de retour en T 10

2 Distributions asymptotiques des valeurs extrêmes **15**

2.1 Modèle asymptotique 15

2.1.1 La paramétrisation α 15

2.1.2 La paramétrisation γ 20

2.1.3 Distribution limite du maximum 22

2.2 Modèle asymptotique dans le cas de minima 34

2.3 Niveau de retour 36

2.4 Inférence sur la distribution généralisée des valeurs extrêmes 37

2.4.1 Considérations générales 37

2.4.2 Estimation par la méthode du maximum de vraisemblance 39

2.4.3 Inférence sur le niveau de retour 43

2.4.4 Profil de vraisemblance 44

2.4.5 Vérification du modèle 46

3 Les modèles du seuil **49**

3.1 Distribution généralisée de Pareto 49

3.2 Modélisation du seuil de dépassement 52

3.2.1 Choix d'un seuil 52

3.2.2 Estimation des paramètres 56

3.2.3 Les niveaux de retour 57

3.2.4 Bref retour sur le choix du seuil 60

3.2.5 Vérification du modèle 61

4 Application **63**

4.1 Estimation des paramètres 63

4.2 Intervalles de confiance 65

4.3 Niveau de retour 67

4.4 Vérification du modèle 71
4.5 Lesquels des trois modèles? 73

A Calcul des moments **76**