TVE : Il est impossible que l’improbable n’arrive jamais…

Modélisation Mathématique

DELIOT Maxence
DESPORTES Kilian
JALADE Samuel 2B

Table des matières

1 Introduction

Mémoire réalisé dans le cadre du module **Modélisation Mathématique** en seconde année de DUT Informatique.
Le but de ce module est de travailler collaborativement sur des problèmes de mathématiques ayant des applications à l'informatique.
Nous allons traiter le sujet concernant **la Théorie des valeurs extrêmes**.
Nous présenterons le contexte général de la Théorie des valeurs extrêmes ainsi que quelques exemples d'utilisation. Nous expliquerons ensuite les principales lois de probabilités utilisées dans la Théorie des valeurs extrêmes, les lois de **Weibull**, **Gumbel** et **Fréchet**.

2 Utilisation d’une loi de probabilité continue, et introduction à la loi normale

2.1 Utilisation d’une loi continue (densité, fonction de répartition,…)

2.1.1 Définition d’une loi de probabilité

Une loi de probabilité est constituée d’une expérience aléatoire E associée à un univers Ω.
Soit une variable aléatoire X définie sur Ω
X(Ω) étant l’ensemble des valeurs prises par X.
La loi de probabilité X attribut à chaque valeur xi la probabilité Pi
X=xi

2.1.2 Loi de probabilité continue

Une loi de probabilité continue est une loi de probabilité ou l’univers contient une infinité d’issues (à l’inverse des lois de probabilité discrètes qui contiennent un univers fini d’issue).

Exemple : Distance d’un projectile à l’objectif dans un tir répété

La probabilité que X soit égal à 3 mètres vaut 0, car le nombre d’issue de X possibles est infini, la probabilité que X soit égal à un nombre vaut donc 1/infini = 0.

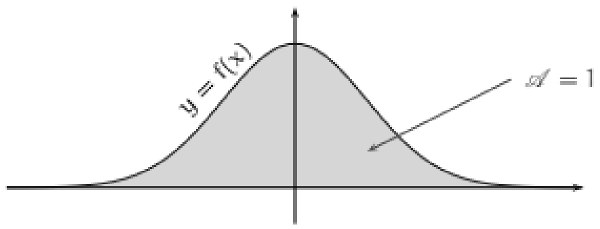
Pour exprimer les probabilités d’issues d’une telle loi, on donnera donc la définition de la fonction de répartition, si elle existe.

Soit $F(x)=Pr\{X<x\}$
On définit une fonction par $F(x) = p(X<x)$

Par exemple : donc $F(2)=p(X<2)$
$P(a<X<=b) = F(b) – F(a) = P(X<3.1) – P(X<2.9)$
$F(2) = P(X<2) = P(-inf <X <2)=F(2) – F(-inf) = F(2)$ car $F(-inf)= P(XS<-inf)=0$

On pourra également donné la fonction de densité d'une loi de probabilité, on dit que F est une fonction de densité si :
- F est une fonction positive sur R
- F est continue sur R (sauf éventuellement en un nombre fini de points)
- L'aire du domaine D définie par la courbe représentation de f et l'axe des abscisses est égal à 1.

La dérivée de la fonction de répartition est égale à la fonction de densité.
La fonction de répartition d'une variable aléatoire continue est la primitive de la densité qui s'annule en -∞.



L'égalité $\mathscr{A} = 1$ peut encore s'écrire $\int_{-\infty}^{+\infty} f(t)\,dt = 1$ ou $\int_{\mathbb{R}} f(t)\,dt = 1$ ou $\lim\limits_{x\to-\infty,\ y\to+\infty} \int_x^y f(t)\,dt = 1$.

Fonction de densité d'une fonction f(x)

2.1.3 Probabilité d'un événement. Fonction de répartition

Applications des lois de probabilités :
Les lois de probabilités sont principalement utilisées pour les jeux de hasard mais de nombreux domaines s'appuient ou se servent des probabilités. Entre autre :
- La statistique, vaste domaine qui s'appuie sur les probabilités pour le traitement et l'interprétation des données
- La théorie des jeux qui s'appuie fondamentalement sur la probabilité est utile en économie et plus précisément en micro-économie.
- L'estimation optimale par usage de la loi de Bayes(probabilités conditionnelles), qui sert dans les applications de décision automatique (imagerie médicale, astronomie, reconnaissance de caractère).
- En physique et en biologie moléculaire (étude du mouvement brownien)
- Mathématique financière (étude des cours de la bourse)
- Études probabilistes de sûreté (évaluation de la probabilité d'occurrence d'un événement indésirable)

La loi normale apparaît naturellement dans toutes ces applications, pas seulement lorsque l'on lance une pièce de monnaie mais dès qu'un phénomène aléatoire est une somme (en grand nombre) de VAR aléatoires indépendantes et de même loi, ce qui ne manque pas, d'où son importance, par exemple:

- quasiment tout ce qui est humain (taille, poids, pousse des cheveux, ongles, durées de sommeil, …)
- Au delà, quasiment tout le vivant (taille et poids des graines, vitesse de pousse, poids des animaux, …)
- La production industrielle de masse (poids, résistance à l’usure, .. d’un produit)
- La loi normale est aussi utilisée comme approximation de la loi Binomiale lorsque la loi de poisson convient moins bien (que les paramètres ne sont pas respectés)

2.2 Histoire de la loi normale

La loi normale est, en probabilité et en statistique, l'une des lois les plus adaptées pour modéliser des phénomènes naturels issus de plusieurs événements aléatoires.

C'est au XVIIe siècle que Galilée introduisit le premier concept qui donnera plus tard la loi normale, en se posant la question « Pourquoi la somme 10 semble se présenter plus fréquemment que 9 sur un lancé de 3 dés ? ». Il publia, en 1618, une solution en faisant un décompte des différents cas.

T_3	Dénombrement	Quantité	Probabilité	En %
3	1	1	1/216	0,463 %
4	2 1	3	3/216	1,389 %
5	3 2 1	6	6/216	2,788 %
6	4 3 2 1	10	10/216	4,630 %
7	5 4 3 2 1	15	15/216	6,944 %
8	6 5 4 3 2 1	21	21/216	9,722 %
9	5 6 5 4 3 2	25	25/216	11,574 %
10	4 5 6 5 4 3	27	27/216	12,500 %
11	3 4 5 6 5 4	27	27/216	12,500 %
12	2 3 4 5 6 5	25	25/216	11,574 %
13	1 2 3 4 5 6	21	21/216	9,722 %
14	1 2 3 4 5	15	15/216	9,644 %
15	1 2 3 4	10	10/216	4,630 %
16	1 2 3	6	6/216	2,778 %
17	1 2	3	3/216	1,389 %
18	1	1	1/216	0,463 %

Tableau de décompte des différents cas publié par Galilée en 1618.

En 1713, Jacques Bernoulli établira la Loi des grands nombres en se basant sur la probabilité que l'on obtienne pile ou face à un lancé de pièce. Il observe que lorsque l'échantillon (le nombre de lancés) augmente, la fiabilité par rapport aux statistiques augmente également.

On obtient ce résultat par la manipulation d'une loi binomiale, la somme de variable de Bernoulli.

$$\mathbb{P}(X=k)=\binom{n}{k}p^k(1-p)^{n-k}.$$

Loi Binomiale

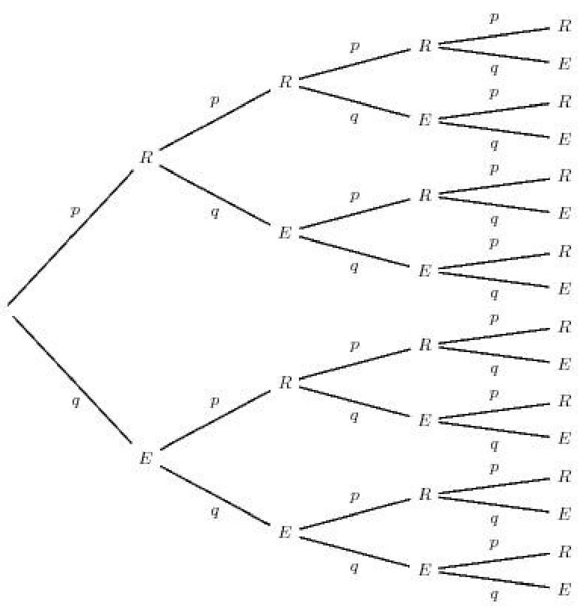


Schéma de Bernoulli

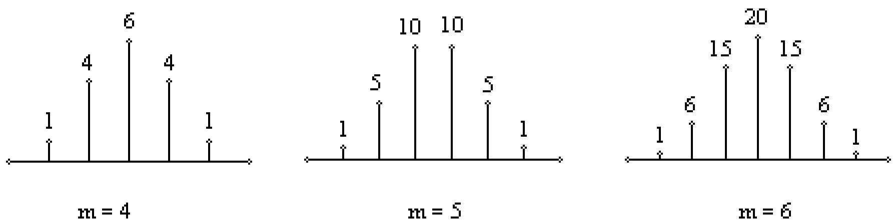
On observe notamment grâce à ce schéma de Bernoulli (représentation visible de la loi de Bernoulli) que la probabilité d'obtenir un même nombre de pile et de face (ici 2 piles, 2 faces) se présente de plus en plus de fois au fur et à mesure que l'on fait des lancés, et celle d'obtenir que des piles ou que des faces diminue.

Abraham de Moivre publiera en 1735 dans « The Doctrine of Chance » la première approche de la loi normale, la faisant apparaître comme la limite d'une loi binomiale et donc la limite d'une somme de variable de Bernoulli.

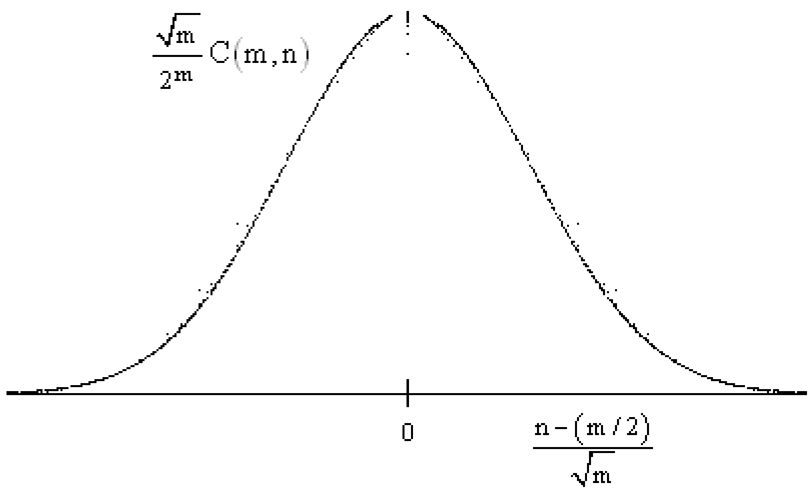
Il va utiliser la formule du binôme pour obtenir la courbe de la loi normale, suivant ce procédé :

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k,$$ Formule du Binôme

- Il va poser x=1 et y=x, la somme des coefficients du binôme vaudra donc 2^n.
- Il va ensuite diviser chaque coefficients par 2^n, la somme des coefficients vaudra donc 1 ($2^n/2^n$).
- On obtient donc des coefficients symétriques que l'on peut aligner sur la valeur centrale par le décalage de n par m/2.
- On augmente ensuite tous les coefficients par √m et l'indice par 1/√m.
- Tous les coefficients s'allongent donc sur une courbe unique.



La figure ci-contre montre les coefficients normalisés pour les valeurs de m valant 4,5 et 6.



de ce procédé, une courbe en cloche.

3 Théorème central limite

Le théorème central limite établit la convergence en loi de la somme d'une suite de variables aléatoires vers la **loi normale**. Intuitivement, ce résultat affirme que toutes sommes de variables aléatoires indépendantes et identiquement distribuées tendent vers une variable aléatoire gaussienne.
La démonstration de cette loi poursuit les travaux de De Moivre qui avait établi cette loi dans le cas où la loi de Bernoulli est de paramètre p=0,5.

Caractéristiques du théorème central limite :

X_1, X_2… suite de variable aléatoire réelles définies sur le même espace de probabilité, indépendantes et identiquement distribuées suivant une même loi.
On suppose que l'espérance μ et l'écart-type σ sont définis et que σ ≠ 0.

$S_n = X_1 + X_2 + \ldots + X_n$.

Alors l'espérance de S_n est **nμ** et l'écart-type vaut **σ√n**.

Quand n est assez grand, la loi normale N(nμ,nσ) est une bonne approximation de la loi de S_n.

Pour formuler cette approximation on peut centrer et réduire S_n pour que l'espérance soit égale à 0 et l'écart-type 1.

$$\overline{X}_n = \frac{S_n}{n} = \frac{(X_1 + X_2 + \ldots + X_n)}{n}$$

et

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}.$$

Le théorème central limite énonce alors que la suite de variables aléatoires $Z_1,Z_2,\ldots$ converge en loi vers une variable aléatoire Z, définie sur le même espace probabilité, et de loi normale centrée réduite N(0,1).
Cela signifie que si Φ est la fonction de répartition de N(0,1), alors pour tout réel z :

$$\lim_{n\to\infty} \mathrm{P}(\mathrm{Z}_n \leq z) = \Phi(z),$$

ou de façon équivalente :

$$\lim_{n\to\infty} \mathrm{P}\left(\frac{\overline{\mathrm{X}}_n - \mu}{\sigma/\sqrt{n}} \leq z\right) = \Phi(z)$$

Ainsi, ce théorème et ses généralisations offrent une explication de l'omniprésence de la loi normale dans la nature : de nombreux phénomènes sont dus à l'addition d'un grand nombre de petites perturbations aléatoires.

Le théorème central limite est souvent défini comme la loi de la moyenne d'une somme de variables aléatoires.

Qu'en est-il des lois du minimum et du maximum d'une somme de variable aléatoire .. ?

4 Les bases de la théorie des valeurs extrêmes

Est-il possible de prévoir des événements rares ? Peut-on prévoir des événements météorologiques majeurs tels que des ouragans, des inondations, des températures extrêmes, etc...? Prévoir l'improbable, c'est en cela qu'est l'enjeu de la théorie des valeurs extrêmes. Étudier des événements rares nous permettrait d'éviter de nombreux morts, dégâts matériels et conséquences financières et environnementales .

Mais un problème se pose, ce sont des événements rares, comment peut t-on étudier de façons pertinentes sans avoir beaucoup de données. C'est pour remédier à ce problème que des mathématiciens ont établi au bout de nombreuses années de travail, la théorie des valeurs extrêmes.

Le mathématicien allemand Emil Julius Gumbel en est le fondateur mais cette théorie doit beaucoup au travail de Maurice Fréchet, mathématicien français et Waloddi Weibull, mathématicien et ingénieur suédois. Qu'est-ce qu'est cette théorie comment fonctionne-t-elle, nous l'étudierons dans cette partie. Et nous verrons que son fonctionnement n'est pas très différent du théorème central limite.

4.1 Dans quel but utilise-t-on les TVE

La théorie des valeurs extrêmes est un ensemble de lois de probabilités capable de définir le comportement des valeurs extrêmes lors d'études d’événements.

Les lois composant cette théorie sont des lois de probabilités utilisées dans les branches des statistiques qui s'intéressent aux valeurs extrêmes des distributions de probabilité. Elles permettent de caractériser le comportement de très grandes valeurs qui sont rares et extrêmement difficiles à observer.

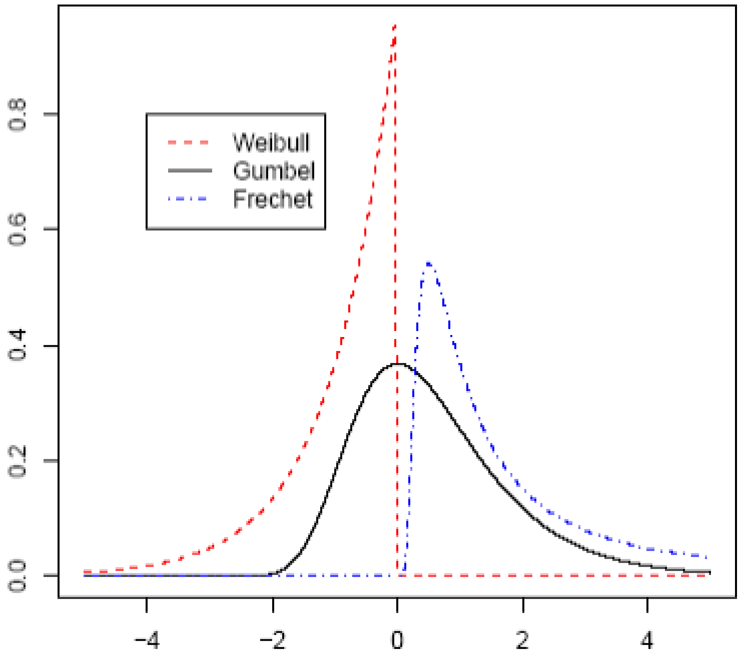
Attention, il ne faut pas confondre la théorie des valeurs extrêmes avec le théorème des valeurs extrêmes.

4.2 3 grands types de lois

Les lois de probabilités composant la théorie des valeurs extrêmes font partie de trois grandes familles de loi de probabilité (toutes ont hérité du nom de leur fondateur) :

- Les lois Gumbel, queue légère
- Les lois Frechet, queue lourde
- Les lois Weibull, queue finie

Ces lois ont pour différence le comportement de la distribution du maxima dans la zone éloignée de sa valeur centrale (côté droit), c'est aussi ceux que l'on appelle la queue supérieure.



4.3 Rapide exemples des domaines d'utilisation

Ces lois sont utilisées pour prévoir les crues en hydrologie, prévoir les grands sinistres en assurance, en finance et météorologie. L'étude des pluies extrêmes est indispensable pour le bon dimensionnement des ouvrages hydrauliques.

5 Les lois Gumbel, Frechet et Weibull

5.1 La loi Gumbel

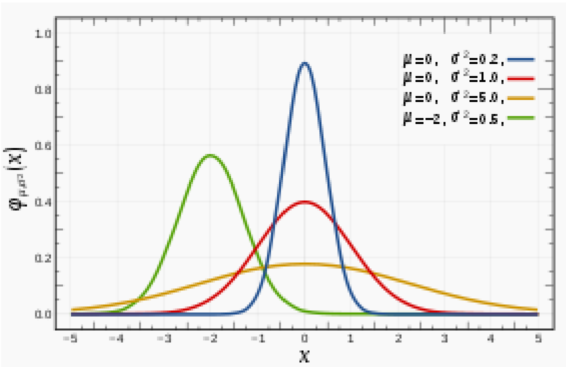
Définition : On dit que X suit la loi de Gumbel de paramètre µ et σ si sa fonction de répartition vaut **:**

$$F(x) = e^{-e^{-\frac{x-\mu}{\sigma}}}$$

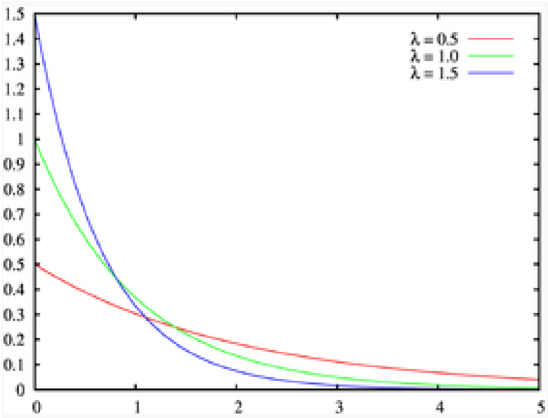
Avec deux paramétres : µ = position (éloignement avec l'abscisse) et σ = échelle (aplatissement)
Par dérivation, on obtient la formule de sa fonction de densité :

$$f(x) = e^{-\frac{x-\mu}{\sigma}} e^{-e^{-\frac{x-\mu}{\sigma}}}$$

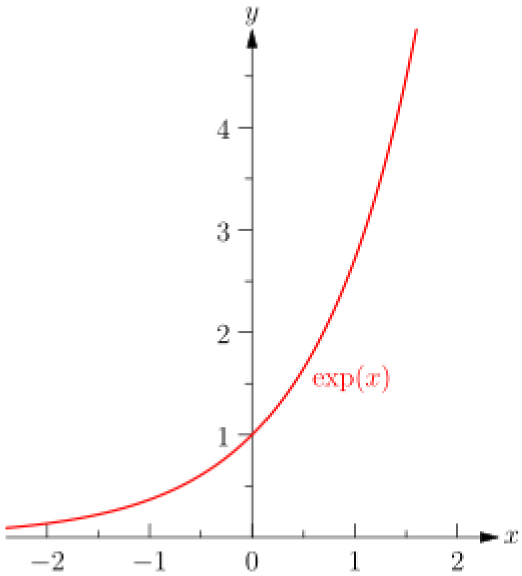
La loi de Gumbel est celle qui se rapproche le plus de la loi normale.
Elle permet d'approximer les maxima de loi dont les distributions sont non bornées et dont la queue supérieure de la distribution décroît de façon exponentielle. C'est-à-dire que pour de très grandes valeurs, la densité décroît très rapidement.
De plus, les observations permettent de constater, la distribution du maximum d'une loi normale, exponentielle ou gamma ont de grandes chances de suivre une loi de Gumbel.

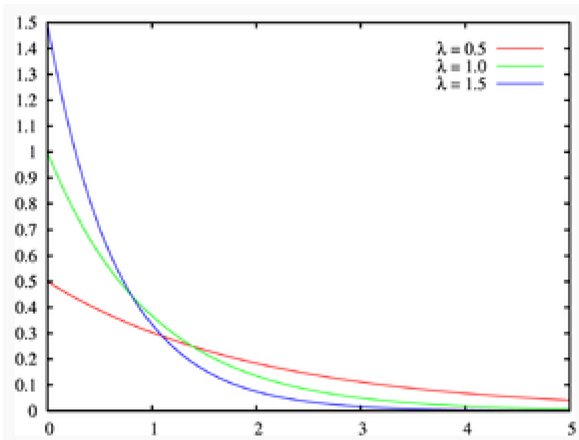


Loi normale



Loi Gamma

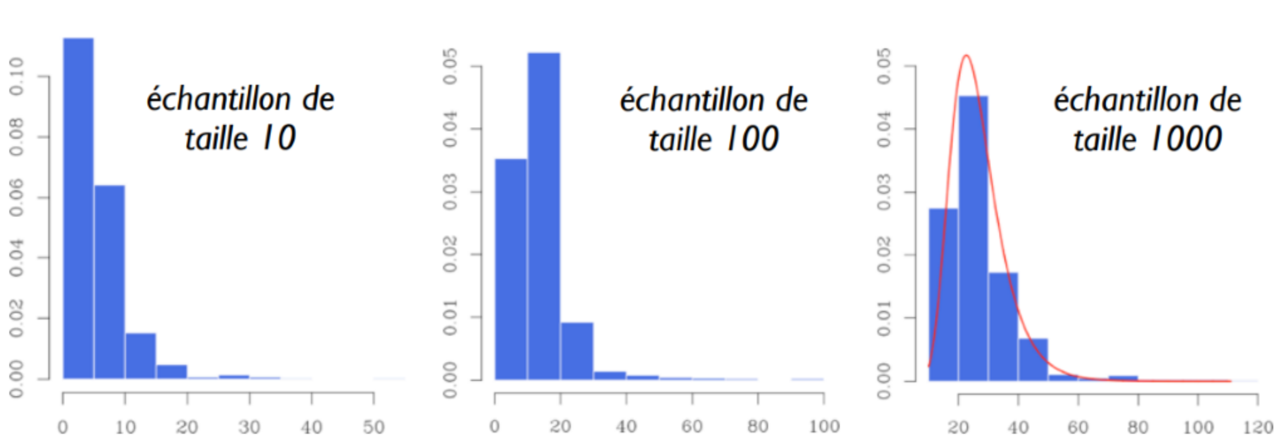




Loi exponentielle

Elle est très utilisée en hydrologie (débits de cours d'eau, précipitation) et en climatologie pour estimer les valeurs extrêmes de phénomènes. Mais elle possède quelques défauts pour des valeurs très fortes.

Exemple : Prenons l'exemple d'une loi Log-Normale. On retient que la valeur maximale d'un échantillon de taille 10. On répète cette action 10 000 fois.
Puis on recommence avec un échantillon de taille 100 puis 1 000.
On remarque alors que la distribution du maximum tend vers la loi de Gumbel.



5.2 La loi Fréchet

Définition : On dit que X suit la loi de Fréchet de paramètre α si sa fonction de répartition vaut :

$$F(x) = e^{-x^{-\alpha}}$$

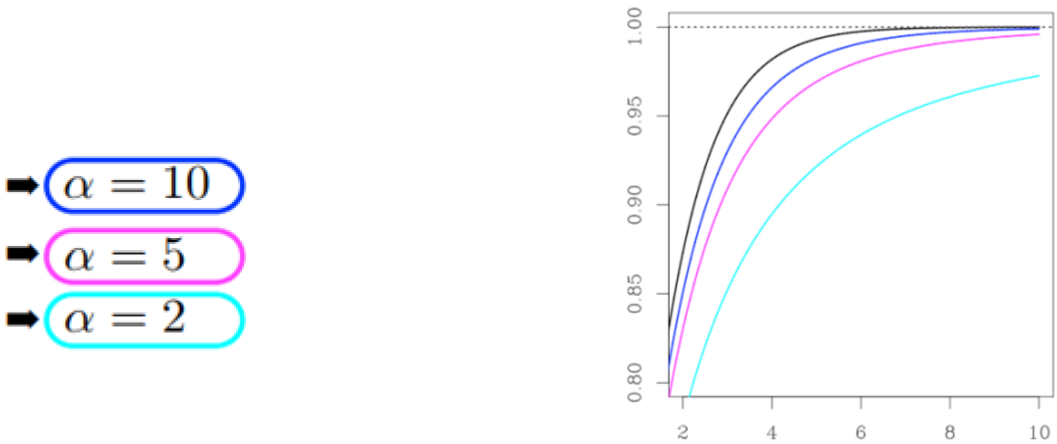
Si α>0 et x>0, 0 sinon.

Avec α= paramètre de forme (obligatoire), on peut y associer un paramètre d'échelle et de position pour obtenir la loi généralisé (par défaut ils sont respectivement à 1 et 0).

Par dérivation, on obtient la formule de sa fonction de densité :

$$f(x) = \alpha x^{-\alpha - 1} e^{-x^{-\alpha}}$$

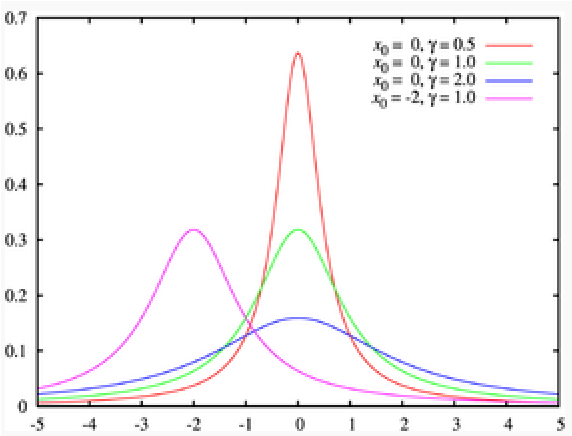
La loi Fréchet est un modèle pour les données ayant de très grandes valeurs. Elle permet de caractériser le comportement du maxima de loi dont les distributions sont non bornées et dont la queue supérieure décroît à vitesse polynomiale x à la puissance -α.



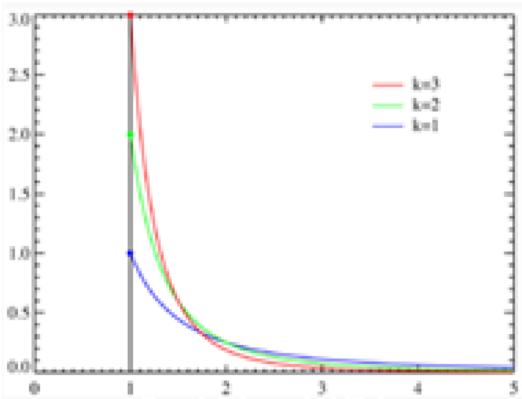
On remarque alors que :

- Plus α est petit, plus la croissance est lente et par conséquent plus des valeurs extrêmes sont susceptibles de se produire, car elle mettra plus de temps à atteindre 1.
- De même, plus α est grand, plus la croissance est forte et moins de valeurs extrêmes vont se présenter, étant donné qu'elle ira plus vite atteindre 1.

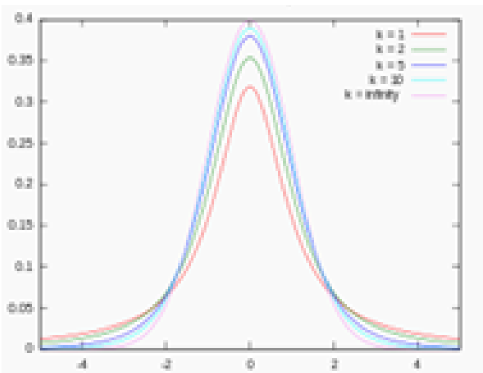
Elle est appropriée pour modéliser la distribution du maximum d'une loi de Pareto, Strudent ou Cauchy par exemple.



Loi de Gauchy



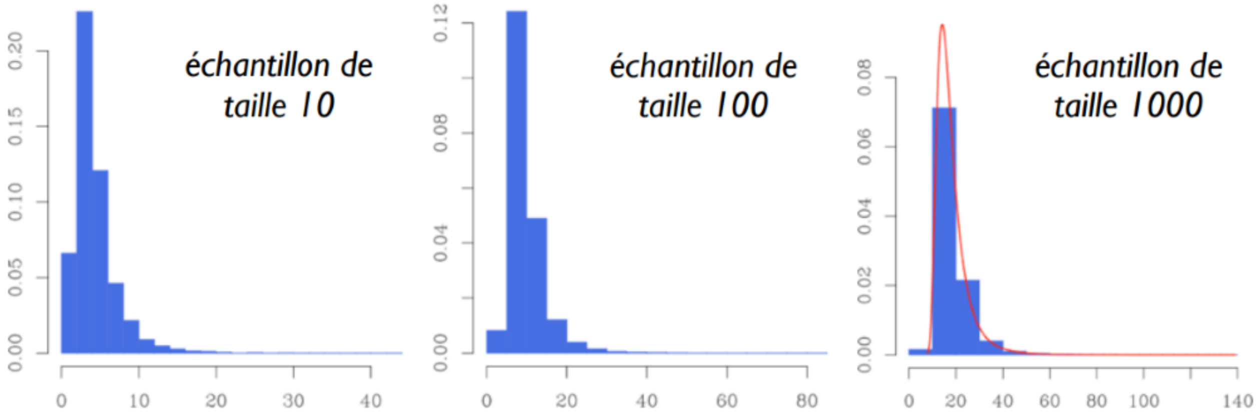
Loi de Pareto



Loi Student

Elle est utilisée pour prévoir des pics de pollution, des précipitations, des crues de rivières et souvent dans le domaine des finances et de l'économie.

Exemple : Effectuons le même raisonnement que dans le 1) mais cette fois avec la loi de Pareto. On remarque que la distribution du maxima de la loi de Pareto tend vers la loi de Fréchet.



5.3 Loi Weibull

Définition : On dit que X suit la loi de Weibull de paramètre α et **λ** si sa fonction de répartition vaut :

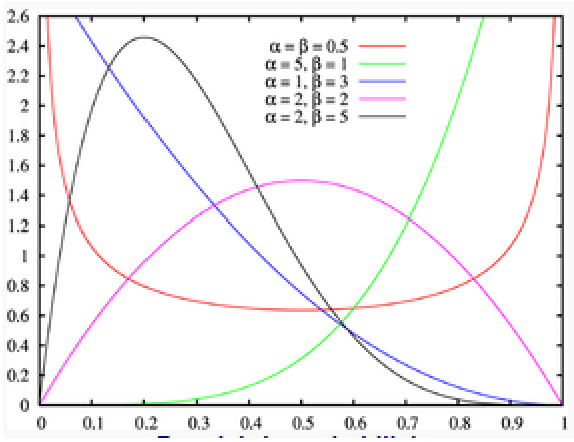
$$F(x) = 1 - e^{-(\lambda x)^{\alpha}}$$

Avec x>0, 0 sinon.
Avec α= paramètre de forme et **λ** = paramètre d'échelle.
La formule de sa fonction de densité est :

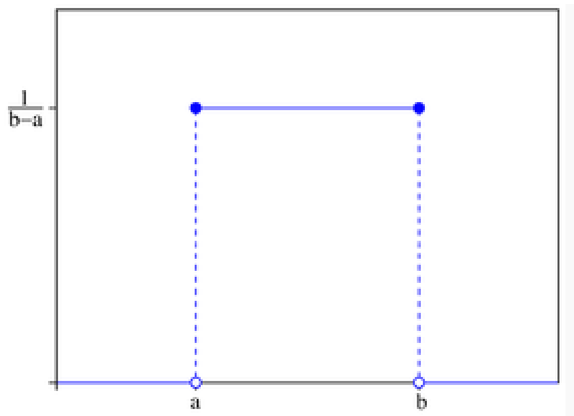
$$f(x) = \alpha\lambda(\lambda x)^{\alpha - 1} e^{-(\lambda x)^{\alpha}}$$

La loi Weibull est un modèle pour les données ayant une borne supérieure. Elle permet de caractériser le comportement du maxima de loi dont les distributions sont bornées et dont la queue supérieure est finie (La loi possède un maximum qu'elle ne dépassera pas).

On peut l'utiliser pour la distribution du maximum de la loi uniforme ou la loi Beta qui est définie sur [0,1] par exemple.



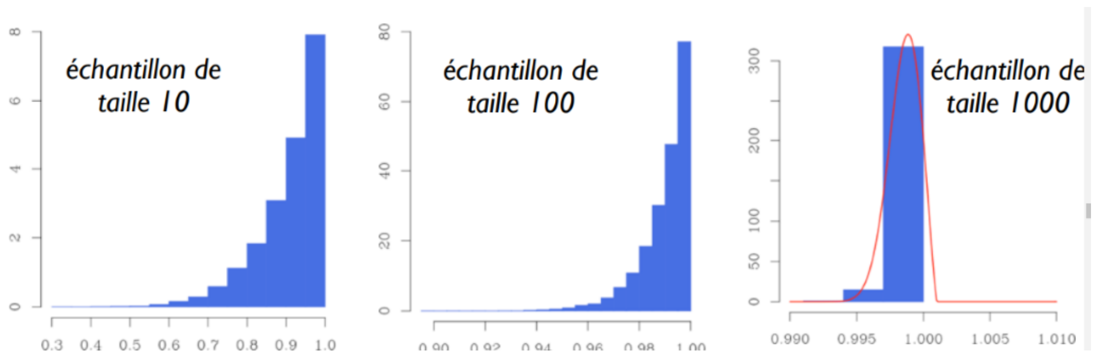
Loi Beta



Loi Uniforme

Elle est un modèle réaliste lorsque les données ne peuvent pas dépasser une valeur maximale, une barrière physique comme par exemple des données de température. Elle est utilisée dans la prévision de fatigue des matériaux, de températures.

Exemple : Effectuons une dernière fois le test avec une loi uniforme. On remarque alors on remarque que le comportement du maxima tend vers une loi de Weibull.



6 Théorème de la valeur extrême (théorème de Fisher-Tippet-Gnedenko)

6.1 Explication

L'objectif de l'utilisation de ce théorème est de modéliser les maxima. C'est-à-dire prévoir la convergence de ces maxima.

Le rôle de ce théorème pour le maximum est similaire à celui du théorème central limite pour les moyennes.

Ce théorème énonce que si la loi d'un maximum normalisé converge, alors sa limite ne peut être qu'un certain type de loi. Il y a trois et seulement trois distributions possibles pour les maxima de variables aléatoires.

Mais attention, ce théorème n'énonce pas que la loi d'un maximum normalisé converge, c'est un mauvais raccourci.

6.2 Relation avec la théorie des valeurs extrêmes

Énoncé :

Soit X1, X2,…,Xn plusieurs séquences de variables aléatoires idd (indépendantes et identiquement distribué) et Mn=max{X1,X2,…,Xn} (ensemble composé de la valeur maximum de chaque séquences). S'il existe des suites de constantes An et Bn avec An>0 telles que :

$$\lim_{n\to\infty} P\left(\frac{M_n - b_n}{a_n} \leq x\right) = F(x)$$

Alors la distribution de Mn suit l'une des trois lois :

Loi Gumbel : loi d'extremum de type I pour tout x

Loi Fréchet : loi d'extremum de type II pour x>=0

Loi Weibull : loi d'extremum de type III pour x>=0

On déterminera l'appartenance grâce à la queue supérieure de la distribution du maxima de l'événement testé.

6.3 Relation entre les trois lois

Il existe des relations à savoir entre ces lois :

- Si X suit une loi Fréchet alors ln(X) suit une loi Gumbel.
- Si X suit une loi Weibull alors 1/X suit une loi Fréchet.
- Si X suit une loi Weibull alors ln(1/X) suit une loi Gumbel.

7 Conclusion

Nous avons donc dans ce mémoire présenté le contexte général de la théorie des valeurs extrêmes et ses principales lois, les lois de Gumbel, Fréchet et Weibull. Ces lois vont permettre de représenter les minimums et maximums d'une variable aléatoire, ce qui est extrêmement utile dans de nombreux domaines tels que la finance, la météorologie ou en hydrologie pour prévoir des événements extrêmes (catastrophes naturelles, grandes crues, grandes sinistres, …). Nous pouvons nous demander si il ne serait pas possible de regrouper ces 3 lois en une seule grande loi qui permettrait de regrouper toutes les propriétés des lois dont elle est issue.

8 Résumé

Dans ce mémoire, nous avons présenté la théorie des valeurs extrêmes, nous avons commencé par présenter les généralités sur les lois de probabilités et nous avons montré comment elles étaient définies, nous avons ensuite présenté rapidement l'histoire de la loi normale et comment elle avait été découverte, notamment par l'intuition de Abraham de Moivre et son rapprochement avec la loi binomiale. Nous avons ensuite présenté le théorème central limite, théorème donnant la loi permettant de représenter la moyenne d'une somme de variable aléatoire. Nous nous sommes penchés sur le sujet de la théorie des valeurs extrêmes et ses 3 principales lois, Gumbell, Fréchet et Weibull, qui permettent quant-à elle de représenter le minimum et le maximum d'une somme de variable aléatoire si la loi de probabilité étudiée respecte le théorème de la valeur extrême .

9 Abstract

In this mémoire, we presented the theory of extreme values, we began by presenting generalities on the probability distributions and showed how they were defined, we then briefly presented the history of the normal distribution and how it had been discovered, in particular by the intuition of Abraham de Moivre and its approximation to binomial distribution. We then presented the central limit theorem, a theorem giving the distribution allowing to represent the average of a random variable sum. Finally, we looked at the subject of extreme value theory and its 3 main distributions, Gumbell, Fréchet and Weibull, which allow it to represent the minimum and maximum of a random variable sum.

10 Sources

1. The Doctrine of Chance – Abraham de Moivre

2. Fondement de la théorie des valeurs extrêmes, ses principales applications et son apport à la gestion des risques du marché pétrolier – Bechir Raggad

3. La théorie des valeurs extrêmes : Présentation et premières applications en finance – François Longin

4. Théorie des valeurs extrêmes – Thierry Roncalli

5. De la loi de Bernoulli à la loi Normale en suivant le programme de statistique de terminal – IREM Marseille

6. Introduction à la théorie des valeurs extrêmes – Anne Sabourin

7. Théorie des valeurs extrêmes – Julie Carreau

8. Introduction à la théorie des valeurs extrêmes – Armelle Guillou & Alexandre You

9. Wikipédia.org

10. jybaudot.fr