

MA-510-52-1

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche
Scientifique

Université de Blida 1
Faculté des sciences
Département de Mathématiques

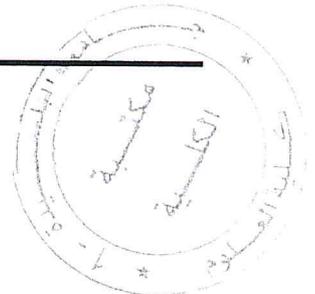


Mémoire de Master en Statistiques

Thème:

Méthode des L-Moments et Application
pour régionaliser la Distribution des
Précipitations

Présenté par: **KOUDIL Sihem**
NEZAYAI Somia



Soutenue le 06 Juillet 2017: devant le jury composé de :

TAMI Omar	Université de Blida 1	M.A.A.	Président
FRIHI Redouane	Université de Blida 1	M.A.A.	Examinateur
RASSOUL Abdelaziz	ENSH, Blida	M.C.A.	Promoteur

Année universitaire: 2016/2017

MA-510-52-1

Remerciements

Je tiens tout d'abord à exprimer ma reconnaissance à tous les professeurs de l'université de Blida qui ont assuré notre formation.

*J'exprime aussi toute ma gratitude et remerciements :
À mon promoteur Monsieur RASSOUL Pour leurs conseils enrichissants et leurs orientations pertinentes d'une façon ou d'une autre a contribué au succès de ce mémoire.*

Mon respect aux membres de jury qui me feront l'honneur de juger ce travail, je les remercie vivement.

Dédicace

A ma très chère mère Yamina

*Affable, honorable, aimable : Tu représentes pour moi le
Symbole de la bonté par excellence, la source de tendresse et
L'exemple du dévouement qui n'a pas cessé de m'encourager et
De prier pour moi.*

A la mémoire de mon Père el marhoum Mhamed

*Aucune dédicace ne saurait exprimer l'amour,
L'estime, le dévouement et le respect que j'ai toujours eu
Pour vous.*

En hommage a mon frère el marhoum Ben Youcef

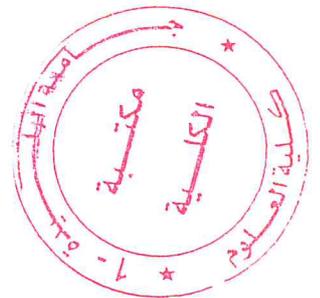
*Rien au monde ne vaut les efforts fournis jour et
Nuit pour mon éducation et mon bien être.
Ce travail est le fruit de tes sacrifices que tu as
Consentis pour mon éducation et ma formation.*

A ma très chère sœur Chahira

A mon très cher mari Amine

A mes frères et mes belle sœur

A ma belle famille



sihem

Dédicace

A ma très chère mère Mira

*Affable, honorable, aimable : Tu représentes pour moi le
Symbole de la bonté par excellence, la source de tendresse et
L'exemple du dévouement qui n'a pas cessé de m'encourager et
De prier pour moi.*

A mon Père Mohamed

*Aucune dédicace ne saurait exprimer l'amour,
L'estime, le dévouement et le respect que j'ai toujours eu
Pour vous.*

A mes très chères sœurs et beaux frères

A mes frères et mes belles sœurs

somia

ملخص

عادة ما تستخدم طريقة العزوم لتوصيف توزيع الاحتمالات أو مجموعة البيانات الملحوظة. استخدام طريقة العزوم لتقدير المعلمة هو أيضا شائع في بناء التوزيع الوسيط المناسب لمجموعة معينة من البيانات. طريقة العزوم لا تنتج دائما نتائج مرضية. ومن الصعب تحديد بالضبط ما هي المعلومات المتعلقة بشكل التوزيع التي يعبر عنها لعزوم من الترتيب الثالث والأعلى.

وفي حالة العينات الصغيرة على وجه الخصوص، يمكن أن تكون القيم العددية لعزوم العينة مختلفة جدا عن القيم المقابلة للعزوم النظرية لتوزيع الاحتمالات ذات الصلة التي تأتي منها العينة العشوائية. وغالبا ما تكون تقديرات المعلمات لتوزيع الاحتمالات التي يتم إجراؤها بواسطة طريقة العزوم أقل دقة بكثير من تلك التي يتم الحصول عليها باستخدام أساليب أخرى، لا سيما في حالة العينات الصغيرة. وتتناول هذه الورقة نهجا بديلا لبناء توزيع وسيطي مناسب لمجموعة البيانات التي تم النظر فيها باستخدام إحصاءات النظام.

:

Résumé :

Les moments classique sont couramment utilisés pour caractériser la distribution de probabilité ou l'ajustement des données observé. Malheureusement cette méthode du moment ne produit pas toujours de résultats satisfaisants. Il est difficile de déterminer exactement quelles informations concernant la forme de la distribution s'expriment par ses moments de troisième ordre et d'ordre supérieur. Dans le cas de petits échantillons en particulier, les valeurs numériques des moments d'échantillonnage peuvent être très différentes des valeurs correspondantes des moments théoriques de la distribution de probabilité pertinente à partir de laquelle l'échantillon aléatoire vient. Les estimations de paramètres de la distribution de probabilité effectuées par la méthode du moment sont souvent nettement moins précises que celles obtenues en utilisant d'autres méthodes, en particulier dans le cas de petits échantillons. Le présent document traite d'une approche alternative pour la construction d'une répartition paramétrique appropriée pour l'ensemble de données considéré en utilisant les statistiques d'ordre.

Abstract :

Moments are commonly used to characterize the probability distribution or observed data set. The use of the moment method of parameter estimation is also common in the construction of an appropriate parametric distribution for a certain data set. The moment method does not always produce satisfactory results. It is difficult to determine exactly what information concerning the shape of the distribution is expressed by its moments of the third and higher order. In the case of small samples in particular, numerical values of sample moments can be very different from the corresponding values of theoretical moments of the relevant probability distribution from which the random sample comes. Parameter estimations of the probability distribution made by the moment method are often considerably less accurate than those obtained using other methods, particularly in the case of small samples. The present paper deals with an alternative approach to the construction of an appropriate parametric distribution for the considered data set using order statistics.

ABBREVIATIONS

$P(A)$: La probabilité de A .
$E(X)$: L'espérance mathématique de X .
$x(F)$: Quantile fonctionnent.
$Var(X)$: Variance de X .
σ	: Ecart-type.
τ	: Coefficient de variation.
s	: Skewness.
k	: Kurtosis.
1_A	: Fonction indicatrice de l'ensemble A .
$\mathcal{N}(0,1)$: Loi normal standard.
$\mathcal{N}(\mu, \sigma^2)$: Loi normale de moyenne μ et de variance σ^2 .
IID	: indépendantes et identiquement distribuées.
GEV	: Distribution généralisée des valeurs extrêmes.
GLO	: Distribution logistique.
GPA	: Distributions uniformes, exponentielle et Pareto.
P_3	: pearson type 3.

TABLE DES MATIÈRES

1	Inférence statistique	9
1.1	Probabilité	9
1.2	Notion de variable aléatoire	9
1.2.1	Variabes aléatoires réelles discrètes	10
1.2.1.1	Variabes aléatoires réelles continues	10
1.3	Caractéristiques des variables aléatoires	11
1.3.1	Espérance	11
1.3.2	Variance	11
1.3.3	Ecart-type	12
1.3.4	Coefficient d'asymétrie ou Skewness	12
1.3.5	Coefficient d'aplatissement ou Kurtosis	12
1.4	Lois de probabilité discrètes	13
1.4.1	Loi de Bernouli	13
1.4.2	Loi binomiale	13
1.4.3	Loi hypergéométrique	13
1.4.4	Loi de Poisson	13
1.4.5	Loi de Pascal	13
1.5	Quelques lois de probabilité continues	14
1.5.1	Loi normale ou de Laplace Gauss	14
1.5.2	Loi du KHI deux	14
1.5.3	Loi de Student	14
1.5.4	Loi de Fisher	14
1.6	Lois à deux dimensions	14
1.7	Estimations	15

TABLE DES MATIÈRES

1.7.1	Généralités	15
1.7.2	Estimateur convergent	15
1.7.3	Estimateur sans biais	16
1.7.4	Estimateur de variance minimum	17
1.7.5	Une méthode générale d'estimation : Le maximum de vraisemblance	18
1.7.6	Une bricole sur le théorème central limit	19
1.8	Tests d'hypothèses	19
1.8.1	Généralités	19
1.8.2	Hypothèse	19
1.8.3	Définition des risques	19
1.9	Tests paramétriques et non paramétriques	20
1.9.1	Tests classiques	20
1.9.1.1	Comparaisons portant sur les variances	20
1.9.1.2	Comparaisons portant sur les moyennes	21
1.9.2	Comparaisons portant sur les proportions	21
1.9.2.1	Test bilatéral	22
1.9.2.2	Test unilatéral	22
1.9.3	Comparaison de deux proportions	22
1.9.4	Test exact	23
1.9.5	Test de conformité a une loi de proba	23
1.9.5.1	Test de Kolmogorov-Smirnov (KS)	24
1.9.5.2	Test d'ajustement du Khi-deux	25
2	L-MOMENTS	27
2.1	Introduction	27
2.2	Limites des moments classiques	28
2.3	L-Moments des Distributions de probabilités	28
2.4	L-moments et statistiques d'ordre	36
2.5	Propriétés	38
2.5.1	Les L-moments possèdent des propriétés générales intéressantes . .	38
2.5.1.1	Exemples de calage pour des lois usuelles	39
2.5.2	Lois de probabilités généralisées	39
2.5.3	L-moments pour certaines distributions spécifiques	40
2.5.3.1	Distribution logistique généralisée	40
2.5.3.2	Distribution de valeur extrême généralisée	41
2.5.3.3	Distribution Kappa	42

2.6	Estimation empirique du L-moment	43
2.6.1	L-moments empirique	43
3	Exemples de simulation sous R	47
3.1	Simulations des distributions de trois paramètres	47
3.1.1	Generalized Extreme Value distribution (GEV)	47
3.1.2	Generalized Logistic (GLO) distribution :	47
3.1.3	Gamma distribution	48
3.1.4	Generalized Pareto distribution (GPD)	48
3.1.5	Pearson type III distribution (PE3)	48
3.2	Application sur la régionalisation des précipitations	48
3.2.1	Test d'homogénéité	48
3.2.2	Identification de la distribution régionale	50
3.2.3	Estimation des paramètres et quantiles de la distribution régionale .	51
3.2.4	Region d'étude et données	51

INTRODUCTION GÉNÉRALE

Les L-moments constituent la base d'une théorie générale qui inclut la synthèse et la description des distributions théoriques de probabilité et les ensembles d'échantillons obtenus, l'estimation des paramètres des distributions de probabilité théoriques et le test d'hypothèses des valeurs des paramètres pour les distributions théoriques de probabilité.

La théorie des moments L comprend les méthodes établies telles que l'utilisation des statistiques d'ordre et la différence moyenne de Gini Cela conduit à des innovations prometteuses dans le domaine de la mesure de l'asymétrie et de la kurtosis de la distribution et fournit des méthodes relativement nouvelles d'estimation des paramètres pour une Distribution individuelle. Les L-moments peuvent être définis pour toute variable aléatoire dont la valeur attendue existe. l'avantage principal des L-moments par rapport aux moments conventionnels est qu'ils peuvent être estimés par des fonctions linéaires de valeurs d'échantillon et sont plus résistants à l'influence de la variabilité de l'échantillon. Les L moments sont plus robustes que les moments classiques à l'existence de valeurs aberrantes dans les données, ce qui facilite de meilleures conclusions sur la base de petits échantillons de la distribution de probabilité de base. Les L-moments parfois apportent des estimations de paramètres encors plus efficaces de la distribution paramétrique que celles estimées par la méthode du maximum de vraisemblance pour les petits échantillons en particulier, voir Hosking (1990).

Les L-moments ont certains avantages théoriques par rapport aux moments conventionnels consistant en la capacité de caractériser une gamme plus large de la distribution (c'est-à-dire la gamme de valeurs que la variable aléatoire peut prendre incluant les valeurs extrêmes). Ils sont également plus résistants et moins propices au biais d'estimation, l'approximation par la distribution normale asymptotique étant plus précise dans les échantillons finis.

Les L- moments sont analogues aux moments conventionnels. Ils peuvent être estimés

sur la base de combinaisons linéaires de statistiques d'ordre d'échantillon, c'est-à-dire de L-statistiques. L-moments sont un système alternatif décrivant la forme de la distribution de probabilité.

Notre recherche est basé essentiellement sur trois chapitres, le premier chapitre présente des notions et des définitions sur les probabilités et les statistiques en se mentionnant les lois des probabilités discrètes et continues avec quelques méthodes d'estimation, et enfin des tests classiques. Le deuxième chapitre est centré sur L-moments dont on a commencé par les limites des moments classiques, ainsi que L-moments des distributions de probabilités et aussi les statistiques d'ordre, en complétant ce chapitre par des estimations empiriques de L-moments. Ensuite on a estimé cette recherche par un troisième chapitre sur des exemples de simulation sous R basé sur l'application sur la régionalisation des fréquences.

Enfin, on a conclu par les résultats trouvés en appliquant la méthode de L-moment.

CHAPITRE 1

INFÉRENCE STATISTIQUE

Introduction

La théorie des probabilités constitue un cadre mathématique pour la description du hasard et de la variabilité, ainsi que pour le raisonnement en univers incertain. Elle forme un tout cohérent dont les concepts, les méthodes et les résultats interviennent dans de très nombreux domaines des sciences et des technologies, parfois de manière Fondamentale.

1.1 Probabilité

Définition 1.1 Soit Ω un univers associé à une expérience aléatoire et soit A l'ensemble des parties de Ω .

Une probabilité P sur l'espace (Ω, A) est une application de A dans $[0, 1]$ telle que :

- $P(\emptyset) = 0$,
- $P(\Omega) = 1$,
- Si $(A_n)_{n \geq 1}$ une famille d'événements de A , 2 à 2 incompatibles, alors :

$$P(\cup_{n=1}^{+\infty} A_n) = \sum_{n=1}^{\infty} P(A_n).$$

Le triplet (Ω, A, P) est appelé espace de probabilité.

1.2 Notion de variable aléatoire

Définition 1.2 Etant donné un univers Ω , une variable aléatoire réelle (v.a.r.) est une application de Ω dans \mathbb{R} :

$$X : \omega \in \Omega : X(\omega) \in \mathbb{R}.$$

Définition 1.3 Soit Ω un univers muni d'une probabilité P , et soit X une v.a.r. On appelle loi de probabilité de X , notée P_X , l'application qui à toute partie A de \mathbb{R} associe :

$$P_x = P(\{\omega \in \Omega : X(\omega) \in A\}).$$

1.2.1 Variables aléatoires réelles discrètes

Une v.a.r. X à valeurs dans un ensemble χ fini ou dénombrable est appelée v.a.r. discrète. Dans ce cas, la loi de X est déterminée par l'ensemble des probabilités :

$$P_X(x) = P(X = x), x \in \chi.$$

Ainsi, pour toute partie A de X , on a alors :

$$P_X(A) = P(X \in A) = \sum_{x \in A} P(X = x) \text{ et } P_X(\chi) = \sum_{x \in \chi} P(X = x) = 1.$$

1.2.1.1 Variables aléatoires réelles continues

Définition 1.4 Soit X une v.a.r. qui prend un nombre infini non dénombrable de valeurs. Si F_X est une fonction continue, on dit que X est une v.a.r. continue. Dans ce cas, la loi de X est déterminée par l'ensemble des probabilités $P(a < X < b)$, pour tout $a < b$.

Si l'on peut écrire la fonction de répartition d'une variable continue sous la forme

$$F_X(t) = \int_{-\infty}^t f_X(x) dx$$

où f_X est une fonction de \mathbb{R} dans \mathbb{R} , alors on dit que f_X est la densité de probabilité de la v.a.r. X .

Ceci implique que l'on a pour tout $a < b$:

$$P(a < X < b) = F_X(b) - F_X(a) = \int_a^b f_X(x) dx.$$

Ce intégrale étant positive pour tout $a < b$, il en résulte que $f_X \geq 0$. De plus, puisque

$$\lim_{t \rightarrow +\infty} F_X(t) = 1, \text{ ou } \int_{-\infty}^{+\infty} f_X(x) dx = 1.$$

Une densité de probabilité est donc une fonction positive ou nulle, d'intégrale 1, et qui caractérise la loi d'une v.a.r. continue. De plus, en tout point $x_0 \in \mathbb{R}$ où F_X est dérivable, on a $f_X(x_0) = F_X'(x_0)$.

1.3 Caractéristiques des variables aléatoires

1.3.1 Espérance

Définition 1.5 Soit X une v.a.r. et h une application de \mathbb{R} dans \mathbb{R} . Donc $h(X)$ est elle aussi une v.a.r.

Soit X une v.a.r. et h une application de \mathbb{R} dans \mathbb{R} . Donc $h(X)$ est elle aussi une v.a.r., Si X est discrète à valeurs dans un ensemble X , l'espérance de $h(X)$ est la quantité

$$E(h(X)) = \sum_{x \in X} h(x)P(X = x)$$

pourvu que cette série converge (dans le cas où X est infini). Si X est continue et admettant une densité, l'espérance de $h(X)$ est la quantité

$$E(h(X)) = \int_{-\infty}^{+\infty} h(x)f_X(x)dx$$

à condition que cette intégrale soit convergente. Notons que si $h(x) = x$, on obtient $E(X)$ appelée espérance mathématique (ou moyenne) de la v.a.r. X . Par ailleurs, si l'on définit la v.a.r. suivante :

$$1_{\{X \in A\}} = \begin{cases} 1 & \text{si } X \in A \subset \mathbb{R} \\ 0 & \text{sinon} \end{cases}$$

qui est appelée fonction caractéristique de l'événement $\{X \in A\}$, alors l'espérance de cette v.a.r. est :

$$E(1_{\{X \in A\}}) = P(X \in A) = P_X(A)$$

d'où le lien étroit entre probabilité et espérance.

1.3.2 Variance

La variance mesure la distance des réalisations de la variable par rapport à la moyenne.

La variance est définie comme un moment d'ordre 2.

$$Var(X) = E[(X - E(X))^2]$$

En réécrivant la variable $(X - E(X))^2$ sous la forme $(X - E(X))^2 = X^2 - 2E(X)X + (E(X))^2$, on réécrit sa moyenne

comme

$$E[(X - E(X))^2] = E(X^2) - 2E(X) * E(X) + (E(X))^2 = E(X^2) - E(X)^2,$$

d'où une autre formule de la variance

$$\text{Var}(X) = E(X^2) - E(X)^2$$

1.3.3 Ecart-type

L' écart type est la racine de la variance. On s'intéresse a la racine du moment d'ordre deux, a n d'avoir une mesure qui est comparable a la variable et en particulier aux paramètres de position.

L' écart-type est la racine de la variance

$$\sigma = \sqrt{\text{Var}(X)}$$

1.3.4 Coefficient d'asymétrie ou Skewness

C'est un moment d'ordre 3. Le coefficient d'asymétrie ou Skewness est le moment d'ordre 3 centré

$$\mu_3 = E[(X - E(X))^3]$$

Le coefficient d'asymétrie ou Skewness de Fisher est relatif

$$S = \frac{\mu_3}{\sigma^3}$$

1.3.5 Coefficient d'aplatissement ou Kurtosis

C'est un moment d'ordre 4. Le coefficient d'aplatissement ou Kurtosis est le moment centré d'ordre 4

$$\mu_4 = E[(X - E(X))^4]$$

Pearson a défini le coefficient d'aplatissement (Kurtosis) qui permet d' étudier la forme plus ou moins pointue ou aplatie :

$$K = \frac{\mu_4}{\sigma^4}$$

Fisher propose d' étudier $K' = K - 3$ ce qui permet de faire référence a une distribution particulière qui est la loi normale pour laquelle K vaut 3. Les logiciels statistiques vous donnent la valeur de K' .

1.4 Lois de probabilité discrètes

1.4.1 Loi de Bernouli

La loi de probabilité d'une variable de Bernoulli est définie par :

$$P(X = 1) = p$$

$$P(X = 0) = 1 - p = q$$

1.4.2 Loi binomiale

X suit une loi binomiale de paramètre N et p on note :

$$X \rightsquigarrow B(N, p)$$

$$P(X = k) = C_N^k p^k (1 - p)^{N-k} \text{ et ceci pour } k = 0, 1 \dots N.$$

1.4.3 Loi hypergéométrique

X suit une loi hypergéométrique de paramètre N, n, N_1 on note :

$$X \rightsquigarrow \mathcal{H}\left(N, n, \frac{N_1}{N}\right)$$

1.4.4 Loi de Poisson

On dit que la va X suit une loi de poisson de paramètre λ , et on note :

$$X \rightsquigarrow \mathcal{P}(\lambda)$$

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!} \quad k = 0, 1 \dots$$

1.4.5 Loi de Pascal

On dit que la va X suit une loi de Pascal de paramètre p si :

$$P(X = k) = p(1 - p)^{k-1}; \quad k = 1, 2 \dots$$

1.5 Quelques lois de probabilité continues

1.5.1 Loi normale ou de Laplace Gauss

La loi normale est caractérisée par sa densité de probabilité. Pour une loi normale de moyenne m et de variance σ^2 elle est donnée par :

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

1.5.2 Loi du KHI deux

Soient X_1, X_2, \dots, X_n n va indépendantes de même loi $\mathcal{N}(0, 1)$, et soit :

$$K = X_1^2 + X_2^2 + \dots + X_n^2 = \sum_{i=1}^n X_i^2$$

alors, K suit une loi du Khi 2 à n degrés de liberté :

$$K \rightsquigarrow X_n^2$$

1.5.3 Loi de Student

Soit X une v.a de loi $\mathcal{N}(0, 1)$ et K une v.a qui suit la loi du X_n^2 (khi 2 à n degrés de liberté), on suppose de plus que K et X sont insépendantes.

$$T_n = \frac{X}{\sqrt{\frac{K}{n}}}$$

1.5.4 Loi de Fisher

Soient K_1 et K_2 deux variables aléatoires indépendantes de loi respectives X_n^2 et X_p^2 alors la quantité

$$F_{n,p} = \frac{K_1(n)}{K_2(p)}$$

suit une loi de Fisher à n et p degrés de liberté.

1.6 Lois à deux dimensions

Loi normale a deux dimensions : Dans la suite, nous supposons que le couple $(X; Y)$ suit une loi normale à deux dimensions. La loi normale deux à dimensions est définies par

5 paramètres : sa moyenne (m_x, m_y) et sa matrice de variance covariance :

$$V = \begin{pmatrix} \sigma_x^2 & cov(X, Y) \\ cov(X, Y) & \sigma_y^2 \end{pmatrix}$$

Si on note ρ le coefficient de corrélation entre X et Y , la densité de la loi normale à deux dimensions s'exprime par la formule :

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(x-m_x)^2}{\sigma_x^2} - 2\rho\frac{(x-m_x)(y-m_y)}{\sigma_x\sigma_y} + \frac{(y-m_y)^2}{\sigma_y^2}\right]\right)$$

1.7 Estimations

1.7.1 Généralités

L'estimation consiste à rechercher la valeur numérique d'un ou plusieurs paramètres inconnus d'une loi de probabilité à partir d'observations (valeurs prises par la v.a. qui suit cette loi de probabilité). On utilise pour cela un estimateur fonction de la v.a. étudiée : quand la v.a. prend comme valeur l'observation, la valeur de l'estimateur est appelée estimation. Le mot estimation se réfère à la valeur numérique utilisée pour approximer. En général un estimateur est une variable aléatoire, en d'autres termes l'estimation du paramètre dépend des individus présents dans l'échantillon.

Si un autre échantillon avait été considéré, une autre estimation du paramètre aurait été obtenue. Le choix de l'estimateur se fait selon des critères qui mesurent sa proximité au paramètre inconnu. Nous allons dans ce qui suit présenter la liste des critères les plus souvent utilisés pour définir les « qualités » d'un estimateur.

1.7.2 Estimateur convergent

Une des propriétés élémentaires que doit remplir un estimateur est d'être convergent. En d'autres termes, lorsque la taille de l'échantillon tend vers l'infini, il faut que l'estimateur se rapproche du paramètre qu'il estime. Il existe plusieurs façons de mesurer cette proximité qui donnent lieu à la définition de plusieurs types de convergence.

Définition 1.6 *L'estimateur T_n est convergent en moyenne quadratique si :*

$$\text{var}(T_n) \rightarrow 0$$

$$\text{Quand } n \rightarrow \infty$$

Rappelons que la variance d'une variable aléatoire est définie par :

$$\text{var}(T_n) = E(T_n - E(T_n))^2 = E(T_n - \theta)^2.$$

Dire que T_n converge en moyenne quadratique signifie en fait que lorsque n tend vers l'infini, la distance moyenne qui sépare T_n de θ tend vers 0. Il est facile d'établir que $\text{var}(\bar{X}_n) = \frac{\sigma^2}{n}$. par conséquent lorsque $n \rightarrow \infty$, $\text{var}(\bar{X}_n) \rightarrow 0$. De même $\text{var}(\hat{p}_n) = \frac{p(1-p)}{n}$ tend vers 0 quand n tend vers ∞ .

Définition 1.7 L'estimateur T_n est convergent en probabilité si : pour tout $\varepsilon > 0$ fixé la quantité

$$P(\|T_n - \theta\| > \varepsilon)$$

tend vers 0 quand n tend vers ∞ .

Définition 1.8 L'estimateur T_n est presque sûrement convergent si :

$$P(\lim_{n \rightarrow \infty} T_n \neq \theta) = 0$$

On voit à travers cette définition que la convergence presque sûre est une convergence beaucoup plus "forte" que la convergence en probabilité : elle implique la convergence en probabilité. Pour obtenir une convergence presque sûre, il est nécessaire que la convergence en proba soit suffisamment rapide pour que n assez grand un très faible pourcentage de réalisations de T_n ne tombent en dehors de l'intervalle que nous avons défini précédemment. En réfléchissant un peu, on peut voir que si T_n converge en probabilité alors, il est possible de trouver une sous suite de $(T_n)_n$ qui converge presque sûrement.

1.7.3 Estimateur sans biais

Un estimateur peut être sans biais. Un estimateur est sans biais si, à taille d'échantillon finie et fixée, les différentes estimations d'un même paramètre sur différents échantillons admettent le paramètre estimé comme barycentre, ou plus simplement, si T est un estimateur de θ , $E(T) = \theta$. l'opérateur $E(.)$ est utilisé pour symboliser la moyenne de population de la variable aléatoire sur laquelle il opère.

Un estimateur classiquement utilisé pour la variance est :

$$\widehat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2.$$

C'est un estimateur biaisé de la variance, il sous-estime en moyenne la variance de population, en effet

$$E(\widehat{\sigma}_n^2) = \left(1 - \frac{1}{n}\right)\sigma^2.$$

On voit à partir de la formule précédente qu'un estimateur sans biais de la variance est donné par :

$$\widehat{\sigma}_{n-1}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Si la moyenne de population m est connue, il est facile de montrer qu'un estimateur sans biais de la variance est donné par :

$$\widehat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - m)^2.$$

Plus généralement, si g est une fonction non linéaire, et si T est un estimateur sans biais de θ , alors

$$E(g(t)) \neq g(\theta)$$

Ainsi, en prenant $g(x) = \sqrt{x}$ on obtient :

$$E(\sqrt{\widehat{\sigma}_{n-1}^2}) \neq \sigma$$

la quantité $\sqrt{\widehat{\sigma}_{n-1}^2}$ n'est donc pas un estimateur sans biais de l'écart type σ .

1.7.4 Estimateur de variance minimum

Définition 1.9 Soit $x = (x_1, \dots, x_n)$ une observation d'un échantillon. (X_1, \dots, X_n) de taille n dont la densité $f_\theta(x)$ dépend d'un paramètre θ (à estimer). On définit la vraisemblance de l'échantillon par :

$$L(x_1, \dots, x_n, \theta) = f(x_1, \theta) \dots f(x_n, \theta).$$

Les n observations étant indépendantes, la vraisemblance apparaît comme la probabilité d'obtention de l'échantillon dans le cas discret et comme la densité de cette probabilité dans le cas continu.

Sous certaines conditions de régularité de la vraisemblance, on a l'inégalité suivante (Cramer-Rao) : Soit T un estimateur d'une fonction $g(\theta)$ alors :

$$\text{var}(T) \geq \frac{[g'(\theta)]^2}{E\left[\left(\frac{\partial}{\partial \theta} \ln L(x_1, \dots, x_n, \theta)\right)^2\right]}$$

avec

$$\ln L(x_1, \dots, x_n, \theta) = \sum_{i=1}^n \ln f(x_i, \theta)$$

On voit donc que si T est un estimateur sans biais de θ alors $g(\theta) = \theta$ et $g'(\theta) = 1$. De plus, si f vérifié certaines conditions de régularité alors :

$$\text{var}(T) \geq \frac{1}{E\left(\frac{\partial^2 \ln f_{\theta}}{\partial \theta^2}\right)}$$

Cette inégalité montre qu'à taille d'échantillon finie, la variance d'un estimateur sans biais ne peut être inférieure à une certaine limite. Il est donc illusoire de penser qu'il est possible d'accéder aux paramètres de population sur un échantillon de taille finie). Un estimateur est efficace si sa variance atteint la borne inférieure de Cramer-Rao en d'autres termes si :

$$\text{var}(T) = \frac{1}{E\left(\frac{\partial^2 \ln f_{\theta}}{\partial \theta^2}\right)} = \text{borne inf de Cramer - Rao}$$

1.7.5 Une méthode générale d'estimation : Le maximum de vraisemblance

Fisher a proposé une méthode basée sur la remarque suivante : les meilleures valeurs du paramètre inconnu θ sont celles qui donnent à l'événement observé (x_1, \dots, x_n) la plus grande probabilité.

On a vu que cette probabilité peut être "représentée" par la vraisemblance

$$L(x, \theta) = f(x_1, \theta) \dots f(x_n, \theta).$$

L'estimation "maximum de vraisemblance" de θ sera une fonction des observations qui rend $L(x, \theta)$ maximum.

Remarque 1.1 *Il est équivalent de rendre maximum*

$$\ln L(x, \theta) = \sum_{i=1}^n \ln f(x_i, \theta).$$

1.7.6 Une bricole sur le théorème central limit

Un théorème important sera souvent évoqué dans ce cours, le théorème “central limit”. En voici un énoncé un peu formel : Soient X_1, \dots, X_n n variables aléatoires indépendantes identiquement distribuées de moyenne m et de variance σ^2

$$\lim_{n \rightarrow +\infty} L\left(\sqrt{n} \frac{\bar{X} - m}{\sigma}\right) = \mathcal{N}(0, 1).$$

ou encors : $\forall a, b \in \mathbb{R}$ ou Φ est la fonction de répartition d’une loi normale $\mathcal{N}(0, 1)$ Ce théorème signifie, que si un grand nombre de quantités aléatoires indépendantes, de même variance sont ajoutées, alors la distribution de la somme est une loi normale. C’est une des raisons qui justifie l’utilisation de la loi normale pour les opérations sur les moyennes, même quand la population n’est pas normalement distribuée.

1.8 Tests d’hypothèses

1.8.1 Généralités

Un test d’hypothèses sert à répondre à une question. Répondre à une question suppose que soient déjà définis : la question (des hypothèses) et, une façon d’y répondre (une règle de décision). L’objet de ce chapitre est d’examiner plus précisément les questions (les hypothèses) et les règles de décision, en d’autres termes les tests d’hypothèses.

1.8.2 Hypothèse

Une hypothèse est un ensemble de valeurs des paramètres inconnus. Une hypothèse est dite simple si elle contient une unique valeur des paramètres inconnus, elle est composée dans le cas contraire.

1.8.3 Définition des risques

Le risque de première espèce est le risque de rejeter (avec la règle de décision) l’hypothèse H_0 alors qu’en réalité cette hypothèse est vraie.

Le risque de seconde espèce est le risque d’accepter (avec la règle de décision) l’hypothèse H_0 alors qu’en réalité cette hypothèse est fautive.

En général on présente ces risques dans le tableau suivant La quantité $1 - \beta$

ou encore

	<i>Décision</i>	
<i>Réalité</i>	<i>H₀vraie</i>	<i>H₁vraie</i>
<i>H₀vraie</i>	1 - α	A
<i>H₁vraie</i>	B	1 - β

est une probabilité de bonne décision appelée puissance du test.

les risques α et β sont liés et varient en sens inverse. A taille d'échantillon donnée, se fixer un risque α équivaut à se fixer un risque β . Un test statistique est par nature négatif.

Accepter H_0 ne signifie pas que cette hypothèse est vraie mais seulement que les observations disponibles ne sont pas incompatibles avec cette hypothèse et que l'on n'a pas de raison suffisante de lui préférer l'hypothèse H_1 compte tenu des résultats expérimentaux.

1.9 Tests paramétriques et non paramétriques

Un test paramétrique est un test pour lequel des hypothèses sur la distribution des populations sont requises. La plupart des tests paramétriques qui seront abordés dans ce cours sont construits en faisant l'hypothèse de normalité des distributions. On qualifie de non paramétriques les méthodes statistiques qui sont applicables dans les conditions générales quant aux distributions des populations. Les anglo-saxons utilisent l'expression "distribution free", qui bien mieux que "non paramétriques", décrit ce dont il s'agit.

1.9.1 Tests classiques

1.9.1.1 Comparaisons portant sur les variances

La comparaison de variances est un outil essentiel des statistiques, nous l'utiliserons intensivement en régression multiple et en analyse de la variance. Supposons que nous disposons de p échantillons gaussiens indépendants de tailles respectives n_1, \dots, n_p , on peut pour chaque échantillon, calculer un estimateur sans biais de la variance de la population. Par exemple, pour le $k^{\text{ième}}$ échantillon, un estimateur sans biais de la variance de population σ_k^2 est donné par :

$$\hat{\sigma}_k^2 = \frac{1}{n_k - 1} \sum_{i=1}^{n_k} (X_i^k - X^k)^2$$

où X_i^k est la $i^{\text{ième}}$ donnée de l'échantillon k , et X^k est la moyenne de l'échantillon k .

1.9.1.2 Comparaisons portant sur les moyennes

La plupart des techniques permettant de comparer deux moyennes ne peuvent être utilisées que si un certain nombre d'hypothèses sont vérifiées.

Dans un premier temps, donnons-nous des notations et précisons ces hypothèses. Supposons que nous disposons de deux échantillons de taille respective n et p que nous noterons X_1, X_2, \dots, X_n et Y_1, Y_2, \dots, Y_p les $(X_i)_{i=1 \dots n}$ suivent une loi $\mathcal{N}(m_X, \sigma_x^2)$ et sont indépendantes, les $(Y_i)_{i=1 \dots p}$ suivent une loi $\mathcal{N}(m_Y, \sigma_Y^2)$ elles sont indépendantes et elles sont indépendantes des $(X_i)_{i=1 \dots n}$. Le fait de supposer que toutes les variables aléatoires $(X_i)_{i=1 \dots n}$ par exemple suivent une même loi de probabilité, signifie simplement que toutes les observations dont nous pouvons disposer doivent provenir d'une même population et que, pour cette population, la variable étudiée (X par exemple) ait une moyenne m_x et une variance σ_x^2 .

L'indépendance, signifie que la valeur que va prendre X_2 par exemple ne doit pas être "influencée" par les autres valeurs (pas de phénomène de contagion).

Comme nous disposons d'échantillons, nous ne pouvons avoir accès aux valeurs de populations de la moyenne et de la variance, les seules informations dont nous disposons sont des estimations de ces valeurs. Donnons donc un nom à ces estimations. Nous noterons \bar{X} et \bar{Y} les moyennes respectives des (x_i) et des (y_i) soit :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \text{ et } \bar{y} = \frac{1}{p} \sum_{i=1}^p y_i$$

Les variances de population sont estimées sans biais par :

$$\widehat{\sigma}_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \text{ et } \widehat{\sigma}_y^2 = \frac{1}{p-1} \sum_{i=1}^p (y_i - \bar{y})^2.$$

Rappelons enfin que la moyenne \bar{X} est aléatoire (la valeur qu'elle prend dépend d'échantillon), elle a une variance $\frac{\sigma_x^2}{n}$ de même \bar{Y} a une variance égale à $\frac{\sigma_Y^2}{p}$.

1.9.2 Comparaisons portant sur les proportions

Comparaison d'une proportion à une valeur donnée considérons une population infinie d'individus possédant l'un ou l'autre de deux caractères opposés de laquelle on prélève un échantillon aléatoire d'effectif n : On note X le nombre d'individus qui possèdent le premier caractère, $\widehat{P} = \frac{X}{n}$ est alors un estimateur sans biais de la proportion p d'individus de la population qui possèdent ce caractère.

On peut se poser un certain nombre de questions sur p : par exemple savoir si Cette

proportion est égale à une proportion donnée p_0 fixée a priori). Pour répondre à cette question, deux tests d'hypothèses peuvent être réalisés selon que l'hypothèse alternative est unilatérale ou bilatérale.

Ces deux tests ne sont à utiliser que si x et $n - x$ sont assez grands (la valeur 5 est généralement la valeur minimale tolérée par les biologistes).

1.9.2.1 Test bilatéral

$$H_0 : p = p_0 \text{ contre } H_1 : p \neq p_0$$

Deux règles de décision sont usuellement utilisées :

1) on rejette H_0 si :

$$U_{obs} = \frac{|X - np_0|}{\sqrt{np_0(1 - p_0)}} > U_{1-\alpha/2}$$

2) on rejette H_0 si :

$$U_{obs} = 2\sqrt{n} \left| \arcsin \sqrt{\frac{x}{n}} - \arcsin \sqrt{p_0} \right| > U_{1-\alpha/2}$$

$U_{1-\alpha/2}$ est la valeur limite au seuil $1-\alpha/2$ d'une loi $N(0;1)$ et arcsin est la fonction réciproque de la fonction sinus.

ATTENTION :

Si vous utilisez la seconde règle de décision, il faut qu'au moment du calcul de arcsin, les angles soient exprimés en radians, pas en degrés.

1.9.2.2 Test unilatéral

$$H_0 : P = p_0 \text{ contre } H_1 : P > p_0.$$

Deux règles de décision sont usuellement utilisées :

1. on rejette H_0 si :

$$U_{obs} = \frac{|X - np_0|}{\sqrt{np_0(1 - p_0)}} > U_{1-\alpha}$$

2. on rejette H_0 si :

$$U_{obs} = 2\sqrt{n} \left| \arcsin \sqrt{\frac{x}{n}} - \arcsin \sqrt{p_0} \right| > U_{1-\alpha}$$

1.9.3 Comparaison de deux proportions

Souvent, on veut comparer la proportion d'individus d'une population à une autre proportion d'individus, ou encore comparer P_1 ET P_2 .

Les données dont nous disposons sont, d'une part les effectifs n_1 et n_2 des deux échantillons, d'autre part la répartition de ces n_1 et n_2 individus en fonction du caractère étudié.

Les données peuvent être présentées dans une table de contingence qui a la forme suivante :

	échantillon 1	échantillon 2	Totaux
caractère 1	a	b	$a + b$
caractère 2	c	d	$c + d$
Totaux	$a + c$	$b + d$	$a + b + c + d$
	ou n_1	n_2	$n_1 + n_2$

Les symboles a , b , c , d représentent les effectifs observés correspondants aux quatre cellules de ce tableau.

Test des hypothèses : $H_0 : p_1 = p_2$ contre $H_1 : p_1 \neq p_2$.

1.9.4 Test exact

Les tests usuellement utilisés, sont des tests asymptotiques tout à fait acceptables pour des effectifs assez élevés. Dans certains cas, les effectifs sont trop faibles pour faire raisonnablement confiance au risque annoncé par les logiciels, il reste alors une solution : utiliser un test exact. La loi Hypergéométrique permet de déterminer la probabilité de rencontrer, lorsque H_0 est vraie, une hypothèse aussi anormale que celle réellement observée.

On obtient :

$$P(a) = \frac{C_{a+c}^a C_{b+d}^b}{C_{a+b+c+d}^{a+b}} = \frac{(a+c)!(b+d)!(a+b)!(c+d)!}{a!b!c!d!(a+b+c+d)!}$$

Si la probabilité d'observer un effectif égal à a ou un effectif plus anormal (quand l'hypothèse H_0 est vraie) est faible, c'est à dire si la répartition observée n'est pas compatible avec l'hypothèse H_0 alors, on rejette cette hypothèse.

1.9.5 Test de conformité a une loi de proba

Une loi de probabilité est définie par "la probabilité" qu'elle donne à chaque point.

Pour les variables continues (poids, tailles) une fonction appelée densité caractérise complètement la loi de probabilité.

La densité n'est, en fait, que l'histogramme des fréquences construit sur la totalité de la population quand les classes sont réduites à un point.

A partir de la densité, on peut construire d'autres fonctions, comme par exemple, la fonction de répartition F . Cette dernière peut s'interpréter comme la fonction des fréquences cumulées. Comme la densité, cette fonction définit complètement la loi de probabilité.

Un histogramme est un estimateur de la densité, la fonction des fréquences cumulées F est un estimateur de la fonction de répartition. La plupart des tests de conformité à une loi de probabilité, sont construits en comparant soit la fonction de répartition empirique à la fonction de répartition, soit, l'histogramme à la densité.

1.9.5.1 Test de Kolmogorov-Smirnov (KS)

Il permet de comparer la fonction de répartition empirique (construite à partir de l'échantillon) à la fonction de répartition théorique F d'une loi normale. De façon plus précise, pour un échantillon z_1, z_2, \dots, z_n de taille n , F est définie comme le pourcentage d'observations inférieures ou égale à z , ou encore

$$\hat{F}(z) = \frac{1}{n} \sum_{i=1}^n 1_{[z_i \leq z]}$$

avec

$$\begin{aligned} 1_{[z_i \leq z]} &= 1 \text{ si } z_i \leq z \\ 1_{[z_i \leq z]} &= 0 \text{ sinon} \end{aligned}$$

Le test de KS permet de tester les hypothèses :

H_0 : La distribution de la population dont est issu l'échantillon est normale, contre.

H_1 : La distribution de la population dont est issu l'échantillon n'est pas normale.

Ce test est basé sur la statistique :

$$K = \sqrt{n} \left[\max_i \left| F(z_i) - \frac{i - 0,5}{n} \right| + \frac{1}{2n} \right]$$

qui mesure l'éloignement de la fonction de répartition empirique et de la fonction de répartition théorique.

La règle de décision est la suivante :

pour $\alpha = 0,05$, on rejette H_0 si $K \geq 1,36$

pour $\alpha = 0,01$, on rejette H_0 si $K \geq 1,63$.

Note : Pour un même jeu de données, il est courant que plusieurs distributions ne puissent être rejetées par ce test.

1.9.5.2 Test d'ajustement du Khi-deux

Il permet de comparer la densité d'une loi normale à l'histogramme construit à partir des observations. Le problème avec l'utilisation de l'histogramme, est le choix toujours arbitraire des classes, supposons néanmoins que p classes sont choisies. Le principe du test du χ^2 est de comparer le pourcentage d'observations observé dans la classe numéro i , que nous noterons \hat{p}_i , au pourcentage d'observation que contiendrait cette même classe, que nous noterons P_i , si la distribution de la population était normale. Le test du χ^2 repose donc sur le calcul d'une distance entre P_i et \hat{p}_i , et ceci pour chaque classe, ou, pour être plus précis,

$$\chi_{obs}^2 = n \sum_{i=1}^n \frac{(\hat{p}_i - P_i)^2}{P_i}$$

ce qui peut aussi s'exprimer avec les effectifs de chaque classe n_i :

$$\chi_{obs}^2 = \sum_{i=1}^n \frac{(n_i - nP_i)^2}{nP_i}$$

Pour tester les hypothèses :

H_0 : La distribution de la population dont est issu l'échantillon est normale, contre

H_1 : La distribution de la population dont est issu l'échantillon n'est pas normale.

pour un risque de première espèce α , la règle de décision est la suivante :

on rejette H_0 si : $\chi_{obs}^2 \geq \chi_{1-\alpha}^2$ où $\chi_{1-\alpha}^2$ est la valeur limite au seuil $1 - \alpha$ d'une loi du χ^2 à $p - 3$ degrés de liberté. Ce test peut être utilisé si pour tout i les quantités nP_i sont assez grandes (en général on impose à ces quantités d'être au moins supérieures à 5). Dans le cas contraire, il faut faire des regroupements des classes jusqu'à ce que cette condition soit vérifiée.

2.1 Introduction

Les moments classiques (moyenne, variance, skewness, kurtosis, ...) sont le plus souvent utilisés pour caractériser une distribution de probabilité. Cependant, dans le cas de distributions à queues épaisses, il peut arriver que ces moments n'existent pas. Il est donc difficile d'utiliser uniquement ce type de moments. Plus récent, le concept de L-moments, dispose de propriétés intéressantes et générales pour le calage de loi notamment. Ces moments sont définis à partir de combinaisons linéaires de la fonction quantile et existent sous des conditions plus générales.

L'idée de L-moments, est apparue progressivement dans différents travaux sur les combinaisons linéaires de statistiques d'ordre (Silitto, 1969 [3] ; David, 1968 [5] ; Chernoff et al., 1967 [6] ; Greenwood et al., 1979 [9]), puis fut conceptualisée dans un cadre théorique général par Hosking (1990) [10].

Alors que les premières applications ont concerné l'hydrologie, la météorologie, le contrôle qualité (Dewar et Wallis, 1990 [11] ; Smithers et Schulze, 2001 [12] ; Chen and Tung, 2003 [13]), on trouve depuis de nombreuses applications aux marchés actions (Karvanen 2006 [14] ; Brown et Konstantinos 2006 [8] ; Chu et Salmon 2008 [15]), estimation de VaR (Martins-Filho, Yao 2006 [16]), l'allocation d'actifs (Jurcenzo, Maillet 2006 [17]) et plus récemment aux hedge funds (Darolles, Gourieroux, Jasiak 2008 [18])

2.2 Limites des moments classiques

Traditionnellement une variable aléatoire X sur un espace de probabilité $(\Omega; \mathcal{F}; P)$ est caractérisée par ses moments ordinaires

$$\mu = E(X), \quad \mu_n = E[(X - \mu)^n], n > 1.$$

Le moment d'ordre n et tous les moments précédents existent pour les variables aléatoires X appartenant à $L^n(\Omega; \mathcal{F}; P)$.

Si X possède une distribution de probabilité continue, les moments permettent de caractériser sa forme. En effet

- le moment d'ordre 1 (μ) ou moyenne mesure la position du centre de la distribution
- l'écart-type σ représente la dispersion de la distribution autour de sa moyenne
- le skewness s permet de mesurer l'asymétrie de la distribution :

$$s = \frac{\mu_3}{\sigma^3},$$

- le kurtosis k , permet de mesurer l'aplatissement de la distribution :

$$k = \frac{\mu_4}{\sigma^4}.$$

Si X appartient à $L^4(\Omega; \mathcal{F}; P)$, les quantités $\mu; \sigma; s$ et k existent et on a

$$\mu \in \mathbb{R}, \quad \sigma \geq 0, \quad s \in \mathbb{R}, \quad k \geq 0$$

De plus l'inégalité $k \geq 1 + s^2$ fournit une borne supérieure de l'asymétrie pour un niveau d'aplatissement donné par le kurtosis.

Si la variable aléatoire admet des moments de tous ordres, et sous certaines conditions, la donnée de l'ensemble des moments permet de caractériser la loi de distribution.

2.3 L-Moments des Distributions de probabilités

La sortie des L-moments est discutée, par exemple, dans Adamowski (2000) [3] ou Ulrych et al. (2000).

Soit X une variable aléatoire continue et distribuée avec la fonction $F(x)$ de distribution et quantile fonctionnent $x(F)$.

Soit $X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n:n}$ Les statistiques d'ordre d'un échantillon aléatoire de la taille de l'échantillon n qui provient de la distribution de la variable aléatoire X . L-moment d'ordre r de la variable aléatoire X est défini comme suit :

$$\lambda_r = \frac{1}{r} \sum_{j=0}^{r-1} (-1)^j \binom{r-1}{j} \mathbb{E}(X_{r-j:r}), r = 1, 2, \dots \quad (2.1)$$

Une valeur attendue de la statistique de la $r^{ième}$ ordre de l'échantillon aléatoire de la taille de l'échantillon n à la forme :

$$\mathbb{E}(X_{r:n}) = \frac{n!}{(r-1)!(n-r)!} \int_0^1 x [F(x)]^{r-1} \cdot [1-F(x)]^{n-r} dF(x). \quad (2.2)$$

Si nous substituons l'équation (2.2) à l'équation (2.1), les ajustements en arrière nous obtenons :

$$\lambda_r = \int_0^1 x(F) \cdot P_{r-1}^* [F(x)] dF(x), r = 1, 2, \dots, \quad (2.3)$$

où :

$$P_r^* [F(x)] = \sum_{j=0}^r P_{r,j}^* \cdot [F(x)]^j \quad \text{a} \quad P_{r,j}^* = (-1)^{r-j} \cdot \binom{r}{j} \cdot \binom{r+j}{j}, \quad (2.4)$$

$P_r^* [F(x)]$ étant le polynôme de Legendre modifié d'ordre r . Ayant substitué l'expression (2.4) à l'expression (2.3), nous avons également obtenu :

$$\lambda_r = \frac{1}{r} \cdot \sum_{j=0}^{r-1} (-1)^j \cdot \binom{r-1}{j} \cdot \frac{r!}{(r-j-1)! \cdot j!} \cdot \int_0^1 x \cdot [F(x)]^{r-j-1} \cdot [1-F(x)]^j dF(x), r = 1, 2, \dots \quad (2.5)$$

La lettre "L" dans "L-moments" indique que le $r^{ième}$ L- moment λ_r est une fonction linéaire de la valeur d'attente d'une certaine combinaison linéaire de statistiques d'ordre. L'estimation du $r^{ième}$ L-moment λ_r , basée sur l'échantillon, c'est donc la combinaison linéaire de valeurs de données d'ordre, c'est-à-dire des L- statistiques.

Les quatre premiers L-moments de la distribution de probabilité sont maintenant définis comme suit :

$$\lambda_1 = \mathbb{E}(X_{1:1}) = \int_0^1 x(F) dF(x), \quad (2.6)$$

$$\lambda_2 = \frac{1}{2} \mathbb{E}(X_{2:2} - X_{1:2}) = \int_0^1 x(F) \cdot [2F(x) - 1] dF(x); \quad (2.7)$$

$$\lambda_3 = \frac{1}{3} \mathbb{E}(X_{3:3} - 2X_{2:3} + X_{1:3}) = \int_0^1 x(F) \cdot \{6[F(x)]^2 - 6F(x) + 1\} dF(x), \quad (2.8)$$

$$\lambda_4 = \frac{1}{4} \mathbb{E}(X_{4:4} - 3X_{3:4} + 3X_{2:4} - X_{1:4}) = \int_0^1 x(F(x)) \cdot \{20[F(x)]^3 - 30[F(x)]^2 + 12[F(x)] - 1\} dF(x). \quad (2.9)$$

La distribution de probabilité peut être spécifiée par ses L- moments, même si certains de ses moments conventionnels n'existe pas, le contraire, cependant, n'est pas vrai.

On peut prouver que le premier L-moment λ_1 est un Caractéristique de localisation, le second L- moment λ_2 étant une caractéristique de variabilité. Il est souvent souhaitable de normaliser les L-moments supérieurs $\lambda_r, r \geq 3$, afin qu'ils puissent être indépendants des unités spécifiques du hasard Variable X .

Théorème 2.1 (i) *Les L-moments $\lambda_r, r = 1, 2, \dots$, d'une variable aléatoire réelle X existe si et seulement si X a une valeur définie*

(ii) *Une distribution dont la moyenne existe est caractérisée par ses L-moments $\{\lambda_r, r = 1, 2, \dots\}$.*

Preuve. Une moyenne définitive implique des attentes infinies de toutes les statistiques d'ordre (David 1981, p.33 [5]), d'où partie (i) suit immédiatement.

Pour la partie (ii) Nous montrons tout d'abord qu'une répartition se caractérise par l'ensemble $\{\mathbb{E}(X_{r:r}), r = 1, 2, \dots\}$. Cela a été prouvé par Chan et Konheim (1971) : Soit X et Y des variables aléatoires avec des fonctions de distribution cumulative F et G et des fonctions quantile $x(u)$ et $y(u)$ respectivement soit :

$$\xi_r^{(X)} = \mathbb{E}X_{r:r} = r \int x\{F(x)\}^{r-1} dF(x),$$

$$\xi_r^{(Y)} = \mathbb{E}X_{r:r} = r \int x\{G(x)\}^{r-1} dG(x)$$

puis

$$\begin{aligned} \xi_{r+2}^{(X)} - \xi_{r+1}^{(X)} &= \int_0^1 \{(r+2)u^{r+1} - (r+1)u^r\} x(u) du \\ &= \int_0^1 u^r \cdot u(1-u) dx(u) \text{ par parties} \\ &= \int_0^1 u^r \cdot dz_X(u) \end{aligned}$$

où $z_X(u)$, définie par $dz_X(u) = u(1-u)dx(u)$, est une fonction croissante sur $(0, 1)$.

Si $\xi_r^{(X)} = \xi_r^{(Y)}$, $r = 1, 2, \dots$, puis

$$\int_0^1 u^r dz_X(u) = \int_0^1 u^r dz_Y(u), \quad r = 0, 1, \dots$$

Ainsi, z_X et z_Y sont des distributions qui ont les mêmes moments sur l'intervalle fini $(0, 1)$, par conséquent (Feller, 1970, p.222-224), $z_X = z_Y$. Cela implique que $x(u) = y(u)$.

Nous avons montré qu'une distribution avec une moyenne finie se caractérise par l'ensemble $\{\xi_r : r = 1, 2, \dots\}$. En utilisant 2.3 nous avons

$$\lambda_r = \sum_{k=1}^r p_{r-1, k-1}^* k^{-1} \xi_k,$$

$$\xi_r = \sum_{k=1}^r \frac{(2k-1)r!(r-1)!}{(r-k)!(r-1+k)!} \lambda_k.$$

Ainsi, un ensemble donné λ_r détermine un ensemble unique de ξ_r . De sorte que la caractérisation d'une distribution en termes de ces dernières quantités s'étend à l'ancienne. Ainsi, une distribution peut être spécifiée par ses L-moments, même si certains de ses moments conventionnels n'existent pas. En outre, une telle spécification est toujours unique, cela n'est évidemment pas vrai des moments conventionnels. En effet, la preuve du Théorème(1) montre (dans un sens) pourquoi L-moments caractérise une Distribution alors que les moments conventionnels en général ne caractérisent pas par L-moments se réduit au problème du moment classique sur un intervalle fini, le (problème du moment de Hausdorff)

"donnée $s_r = \int_0^1 u^r dz(u)$, $r = 0, 1, \dots$, trouver $z(u)$ ".

Tandis que la caractérisation par moments conventionnels est le problème du moment classique sur un intervalle infini (le problème du moment de la Hamburger)

"donnée $s_r = \int_{-\infty}^{\infty} u^r dz(u)$, $r = 0, 1, \dots$, trouver $z(u)$ ".

Seul le problème de Hausdorff a une solution unique

Comme le montre Hosking (1990), λ_2 est une mesure de l'échelle ou de la dispersion de la variable aléatoire X . Il est souvent préférable de normaliser les moments supérieurs $\lambda_r, r \geq 3$. Afin qu'ils soient indépendants des unités de mesure de X Définir donc Les rapports L-moment de X pour être la quantité

$$\tau_r = \lambda_r / \lambda_2, \quad r = 3, 4, \dots$$

■

Il est également possible de définir une fonction des L-moments qui est analogue au coefficient de variation c'est le L-CV, $\tau = \lambda_2 / \lambda_1$. Des limites sur les valeurs numériques des rapports L-Moment et L-CV sont données par le théorème suivant :

Théorème 2.2 Soit X une variable aléatoire non dégénérée avec une moyenne finie. Alors

les rapports L-moment de X satisfont $|\tau_r| < 1, r \geq 3$, si en plus $X \geq 0$ presque sûrement, puis τ , le L-CV de X satisfait $0 < \tau < 1$.

Preuve. On définit $Q_r(t)$ par

$$t(1-t)Q_r(t) = \frac{(-1)^r}{r!} \frac{d^r}{dt^r} \{(1-t)^{r+t} :$$

$Q_r(t)$ Est-ce que le polynôme Jacobi est $P_r^{(1,1)}(2t-1)$ De Szego (1959 Chapitre 4) Il s'ensuit que

$$\frac{d}{dt} \{t(1-t)Q_r(t)\} = -(r+1)P_{r+1}^*(t),$$

alors en intégrant 2.12 par parties donne

$$\begin{aligned} \lambda_r &= [-xF(x)\{1-F(x)\}(r-1)^{-1}Q_{r-2}(F(x))] \\ &+ \int F(x)\{1-F(x)\}(r-1)^{-1}Q_{r-2}(F(x))dx. \end{aligned}$$

Le terme intégré disparaît pour la finalité de la moyenne assure que $xF(x)\{1-F(x)\} \rightarrow 0$ Alors que x s'approche des points finaux de la distribution : Ainsi

$$\lambda_r = \int F(x)\{1-F(x)\}(r-1)^{-1}Q_{r-2}(F(x))dx.$$

Depuis $Q_0(t) = 1$ Le cas $r = 2$ Donne

$$\lambda_2 = \int F(x)\{1-F(x)\}dx.$$

Maintenant $0 \leq F(x) \leq 1$ Pour tout x , Et parce que X n'est pas dégénéré, il existe un ensemble de mesure non nulle sur laquelle $0 < F(x) < 1$ Ainsi : $\lambda_2 > 0$ Puisque $F(x)\{1-F(x)\} \geq 0$ Pour tout x , il s'ensuit que

$$\begin{aligned} |\lambda_r| &\leq (r-1)^{-1} \sup |Q_{r-2}(t)| \int F(x)\{1-F(x)\}dx \\ &= (r-1)^{-1} \sup |Q_{r-2}(t)| \lambda_2. \end{aligned}$$

de szego(1959,p.166) Il s'ensuit que

$$\sup_{0 \leq t \leq 1} |Q_r(t)| = r+1.$$

Avec le suprême étant atteint seulement à $t = 0$ Ou $t = 1$ Ainsi $|\lambda_r| \leq \lambda_2$ Avec égalité Seulement si $F(x)$ ne peut prendre que les valeurs 0 et 1, i.e. seulement si X est dégénéré.

Ainsi, une distribution non dégénérée $|\lambda_r| \leq \lambda_2$ Qui, avec $\lambda_2 > 0$ Implique $|\tau| < 1$.

Si $X \geq 0$ Presque certainement alors $\lambda_1 = EX > 0$ et $\lambda_2 > 0$ alors $\tau = \lambda_2/\lambda_1 > 0$ En outre, $EX_{1:2} > 0$, alors

$$\tau - 1 = (\lambda_2 - \lambda_1)/\lambda_1 = -EX_{1:2}/\lambda_1 < 0.$$

Nous considérons la limite des rapports L-moment comme un avantage. Intuitivement, Il semble plus facile d'interpréter une mesure telle que τ_3 , Qui est contraint de se situer dans l'intervalle $(-1, 1)$ Que l'asymétrie conventionnelle skewness, Qui peut prendre des valeurs arbitrairement importantes.

Des limites plus strictes sur le τ_r peuvent être trouvées. La preuve du théorème 1 Implique une séquence ξ_1, ξ_2, \dots , Peut-être le " $EX_{r,r}$ " D'une variable aléatoire X si et seulement s'il existe une fonction croissante z telle que :

$$\xi_{r+2} - \xi_{r+1} = \int u^r dz(u), \quad r = 0, 1, \dots$$

L'établissement de conditions pour l'existence de la fonction z fait partie du problème classique, Et a été résolu par Hausdorff (1923) Akhiezer (1965, p.74)

Montre que z existe si et seulement si les formes quadratiques suivantes ne sont pas négatives :

$$\sum_{i,j=0}^m (\xi_{i+j+3} - \xi_{i+j+2})x_i x_j,$$

$$\sum_{i,j=0}^m (-\xi_{i+j+3} + 2\xi_{i+j+2} - \xi_{i+j+1})x_i x_j,$$

pour $r = 2m - 1$, et

$$\sum_{i,j=0}^m (\xi_{i+j+2} - \xi_{i+j+1})x_i x_j,$$

$$\sum_{i,j=0}^m (-\xi_{i+j+4} + 2\xi_{i+j+3} - \xi_{i+j+2})x_i x_j,$$

pour $r = 2m - 1$: ces conditions peuvent s'exprimer en fonction de la non-négativité des déterminants de certaines matrices dont les éléments sont des combinaisons linéaires du ξ_r Mallows (1973) Donne un énoncé exact du résultat. ■

Théorème 2.3 (Mallows (1973), Théorème 2(ii)). La séquence ξ_1, \dots, ξ_n peuvent être considérés comme les attentes des statistiques de commandes les plus importantes d'échantillons de taille $1, \dots, n$ d'une vraie variable aléatoire réelle si et seulement si soit (a), $A_k > 0, k =$

$2, \dots, n$ Et $B_k > 0, k = 3, \dots, n$ où (b) pour certains même $m, 2 \leq m < n$. Nous avons $A_k > 0$ et $B_k > 0$ pour $k = 2, \dots, m, A_m = 0, B_m > 0$ et $A_k = B_k = 0$ pour $k = m + 1, \dots, n$ ici A_k et B_k sont des déterminants des matrices, comme suit :

$$\begin{aligned} A_{2k} &= \det[\xi_{i+j} - \xi_{i+j-1}]_{i,j=1,\dots,k}, \\ A_{2k+1} &= \det[\xi_{i+j+1} - \xi_{i+j}]_{i,j=1,\dots,k}, \\ B_{2k} &= \det[-\xi_{i+j} + 2\xi_{i+j-1} - \xi_{i+j-2}]_{i,j=2,\dots,k}, \\ B_{2k+1} &= \det[-\xi_{i+j+1} + 2\xi_{i+j} - \xi_{i+j-1}]_{i,j=1,\dots,k}, \end{aligned}$$

Du théorème 3 et l'équation (3), on peut obtenir les valeurs possibles des L-moments d'une distribution. En particulier, pour une distribution non dégénérée, les contraintes sur ξ_1, ξ_2, ξ_3 et ξ_4 sont :

$$\begin{aligned} \xi_2 - \xi_1 &> 0, \quad \xi_3 - \xi_2 > 0, \quad -\xi_3 + 2\xi_2 - \xi_1 > 0, \\ (\xi_4 - \xi_3)(\xi_2 - \xi_1) - (\xi_3 - \xi_2)^2 &\geq 0, \quad -\xi_4 + 2\xi_3 - \xi_2 > 0, \end{aligned}$$

Et donc les contraintes sur $\lambda_1, \lambda_2, \tau_3$ et τ_4 sont :

$$0 < \lambda_2, \quad -1 < \tau_3 < 1, \quad \frac{1}{4}(5\tau_3^2 - 1) \leq \tau_4 < 1.$$

Ici, pour une bonne mesure, les contraintes τ_3 et τ_6 sont :

$$\frac{1}{5}\tau_3(7\tau_4 - 2) - \frac{7(1 - \tau_4)(1 + 4\tau_4 - 5\tau_3^2)}{5(1 + \tau_3)} \leq \tau_5 \leq \frac{1}{5}\tau_3(7\tau_4 - 2) - \frac{7(1 - \tau_4)(1 + 4\tau_4 - 5\tau_3^2)}{5(1 + \tau_3)}$$

$$\frac{1}{25}(42\tau_4^2 - 14\tau_4 - 3) + \frac{6(2\tau_3 - 7\tau_3\tau_4 + 5\tau_5^2)}{35(1 + 4\tau_4 - 5\tau_3^2)} \leq \tau_6 \leq \frac{1}{10}(3 + 7\tau_4) - \frac{15(\tau_3 - 5\tau_3^2)}{14(1 - \tau_4)}.$$

L'égalité dans ces limites ne peut être atteinte que par une distribution qui peut prendre un nombre fini, (m dire) des valeurs distinctes. Une telle distribution satisfait la limite inférieure de τ_{2m} et les limites inférieures et supérieures sur $\tau_r, r > 2m$.

Le rapport des L-moments du $r^{ième}$ ordre de la variable aléatoire X est défini comme suit :

$$\tau_r = \frac{\lambda_r}{\lambda_2}, r = 3, 4, \dots \quad (2.10)$$

On peut aussi définir la fonction des moments L qui est analogue au coefficient classique de variation,

Le L-coefficient de variation :

$$\tau = \frac{\lambda_2}{\lambda_1} \tag{2.11}$$

Le rapport des L- moments τ_3 est une caractéristique d'asymétrie, le rapport des L-moments τ_4 étant une caractéristique de kurtosis de la distribution de probabilité correspondante. Les principales propriétés de la distribution de probabilité sont très bien résumée par les quatre caractéristiques suivantes :

- L-location λ_1 ,
- L-variabilité λ_2 ,
- L-skewness τ_3
- L-kurtosis τ_4 .
- Les L-moments λ_1 et λ_2 , le L-coefficient de variation τ et les rapports des L-moments τ_3 et τ_4 sont les caractéristiques les plus utiles pour la synthèse de la distribution de probabilité.

Leur principale propriété est l'existence (si la valeur attendue de la distribution est définie, alors il existe tous ses L-moments) et l'unicité (si la valeur attendue de la distribution est définie, alors les L-moments définissent la seule distribution, C'est-à-dire qu'aucune distribution distincte n'a les mêmes L-moments).

En utilisant les équations 2.6, 2.9 et 2.10, on obtient à la fois les expressions pour les rapports L-moments et L-moments pour les distributions probabilistes logarithmiques et génériques pareto, voir Tableau 1.

Tableau 1 Formules pour les fonctions de distributions ou de quantiles, les L-moments et leurs ratios pour les distributions de probabilité lognormal et paréto généralisé ou quantile $x(F)$

Distribution	CDF $F(x)$ ou quantile $x(F)$	L-moments et ratios de L-moments
lognormal	$F(x) = \Phi \left\{ \frac{\ln[x(F)-\xi]-\mu}{\sigma} \right\}$	$\lambda_1 = \xi + \exp\left(\mu + \frac{\sigma^2}{2}\right)$ $\lambda_2 = \exp\left(\mu + \frac{\sigma^2}{2}\right) \cdot \operatorname{erf}\left(\frac{\sigma}{2}\right)$ $\tau_3 = 6\pi^{-\frac{1}{2}} \frac{\int_0^{\frac{\sigma}{2}} \operatorname{erf}\left(\frac{x}{\sqrt{3}}\right) \cdot \exp(-x^2) dx}{\operatorname{erf}\left(\frac{\sigma}{2}\right)}$
Pareto généralisé	$x(F) = \xi + \alpha \cdot \frac{1-[1-F(x)]^k}{k}$	$\lambda_1 = \xi + \frac{\alpha}{1+k}$ $\lambda_2 = \frac{\alpha}{(1+k) \cdot (2+k)}$ $\tau_3 = \frac{1-k}{3+k}$ $\tau_4 = \frac{(1-k) \cdot (2-k)}{(3+k) \cdot (4+k)}$

2.4 L-moments et statistiques d'ordre

L'introduction des L-moments permet de caractériser la loi d'une classe de variables aléatoires plus grande que celles admettant des moments de tous ordres. Les L-moments existent si et seulement si la distribution a une moyenne finie. Plus encore, une distribution dont la moyenne est finie est caractérisée par ses L-moments (Hosking, 1990)[10]. Cette propriété est essentielle quand on s'intéresse à des distributions à queues épaisses. En effet, sans les hypothèses d'intégrabilité, nous pouvons considérer des distributions dont les moments classiques n'existent pas.

Nous avons également vu dans la section précédente que les valeurs du skewness et du kurtosis sont contraintes par une relation d'inégalité. Ce n'est pas le cas pour les Lmoments d'ordre supérieur. Cette propriété illustre bien l'intérêt des L-moments.

Quand on se situe au « voisinage » de la distribution gaussienne, skewness et kurtosis permettent bien d'évaluer la distance entre une distribution donnée et l'hypothèse gaussienne. Quand la distribution est trop « non gaussienne », fonder l'étude de ses caractéristiques sur les moments d'ordre supérieur classiques est une mauvaise idée car leur valeur est contrainte (quand elle existe). Le fait de changer de base et d'utiliser les L-moments donne plus de souplesse. Donnons d'abord une construction heuristique nous conduisant vers la définition des L-moments.

Prenons un échantillon de n tirages d'une variable aléatoire X de la loi étudiée, arrangé en ordre croissant $x_{1,n} \leq x_{2,n} \leq \dots \leq x_{n,n}$.

- ◇ L'échantillon de taille 1, $x_{1,1}$, est une seule observation, il contient l'information sur la localisation de la distribution, i.e. sa moyenne.
- ◇ Un échantillon de taille 2, constitué de 2 observations $x_{1,2}$ et $x_{2,2}$ permettent d'en dire plus. Si la distribution n'a que peu de dispersion, les deux valeurs auront tendance à être proches, dans le cas contraire, elles seront plutôt éloignées. Il est donc possible de poser $x_{2,2} - x_{1,2}$ comme mesure de la dispersion de la distribution.
- ◇ Ensuite, l'échantillon de taille 3 permet de caractériser l'asymétrie de la distribution : si la distribution est symétrique, on s'attend à ce que $x_{3,3} - x_{2,3} \simeq x_{2,3} - x_{1,3} \simeq x_{3,3} - 2x_{2,3} + x_{1,3} \simeq 0$.

Si la distribution est asymétrique à droite (respectivement à gauche), la dernière expression sera typiquement positive (respectivement négative). On peut donc retenir $x_{3,3} - 2x_{2,3} + x_{1,3}$ (différence finie d'ordre 2 sur l'échantillon ordonné de 3 observations) comme mesure de l'asymétrie d'une distribution.

- ◇ Enfin, avec le même raisonnement, on peut montrer que la quantité $x_{4,4} - 3x_{3,4} + 3x_{2,4} - x_{1,4}$ (différence finie d'ordre 3 sur l'échantillon ordonné de 4 observations)

caractérise l'aplatissement d'une distribution comme le fait le kurtosis.

Une généralisation du raisonnement à des échantillons de toutes tailles conduit à la définition des 4 premiers L-moments. On considère k exemplaires i.i.d. d'une variable aléatoire X et les statistiques d'ordre associées à cet échantillon conceptuel $(X_{i:k})_{i=1}^k$ et l'on définit les L-moments d'ordres 1 à 4 comme :

$$\lambda_1 = E[X_{1,1}]$$

$$\lambda_2 = \frac{1}{2}E[X_{2,2} - X_{1,2}]$$

$$\lambda_3 = \frac{1}{3}E[X_{3,3} - 2X_{2,3} + X_{1,3}]$$

$$\lambda_4 = \frac{1}{4}E[X_{4,4} - 3X_{3,4} + 3X_{2,4} - X_{1,4}]$$

Il est important de remarquer que k n'est pas le nombre de tirages observés (que l'on notera n par la suite), mais la taille du vecteur aléatoire de variables aléatoires i.i.d. que l'on utilise pour définir le L-moment d'ordre k (et qui correspond à une taille d'information).

Plus généralement, en suivant la même logique on peut définir les L-moments de tout ordre k . Soit $k \geq 1$ et X_1, X_2, \dots, X_k un échantillon i.i.d de variables aléatoires de même loi que X . on note $X_{1,k} \leq X_{2,k} \leq \dots \leq X_{k,k}$ les statistiques d'ordres associées à cet échantillon. Le L-moment d'ordre k est (Hosking, 1990) [10]

$$\lambda_k = \frac{1}{k} \sum_{i=0}^{k-1} (-1)^i C_{k-1}^i E[X_{k-i,k}] \quad (2.12)$$

λ_k est une combinaison linéaire des moyennes des statistiques d'ordre. Considérons un échantillon de taille n , ordonné, $x_{1,n} \leq x_{2,n} \leq \dots \leq x_{n,n}$

On a la formule suivante pour le calcul des espérances dans (2.13) :

$$E[X_{k-1,i}] = k C_{k-1}^{k-i-1} Q(u) u^{i-1} (1-u)^{k-i} du$$

où $Q(u)$ est la fonction quantile de la distribution. En substituant dans (2.13) il vient :

$$\lambda_k = \int_0^1 Q(u) P_{k-1}^*(u) du = E[X P_{k-1}^*(F(X))] \quad (2.13)$$

où F est la fonction de répartition de X , $P_k^*(u) = P_k(2u-1)$ et $P_k(u)$ sont les polynômes de Legendre.

Remarquons que cette formulation des L-moments, faisant apparaître explicitement la fonction quantile, présente deux avantages notables dans les applications :

- ★ le quantile permet de construire un lien formel avec la notion de valeurs extrêmes
- ★ la construction d'estimateurs de L-moments bénéficie des travaux précédents sur l'estimation des fonctions quantiles
- ★ Les polynômes P_k^* ont les propriétés suivantes :
 1. P_k^* est un polynôme de degré k
 2. $P_k^*(1) = 1$
 3. si $k \neq s$, $\int_0^1 P_k^*(u)P_s^*(u)du = 0$
 4. les polynômes P_k^* forment une base orthonormée sur $[0,1]$.
- ★ Ainsi, pour X variable aléatoire intégrable, les L-moments s'interprètent comme les projections de sa fonction quantile sur la base constituée par ces polynômes.
- ★ L'ensemble des L-moments de X permet de reconstruire la fonction quantile de X et permet de caractériser la loi de toute variable aléatoire intégrable.
- ★ Les quantités :

$$\tau_3 = \frac{\lambda_3}{\lambda_2} \quad \tau_4 = \frac{\lambda_4}{\lambda_2}$$

appelées L-skewness et L-kurtosis, sont similaires aux skewness et kurtosis issus des moments ordinaires, et sont définies pour des variables aléatoires admettant un moment d'ordre 1.

2.5 Propriétés

2.5.1 Les L-moments possèdent des propriétés générales intéressantes

- ♣ Les conditions d'existence des L-moments sont plus générales que celles des moments classiques. En effet, les L-moments sont définis pour toute distribution de premier moment fini.
- ♣ Les L-moments, sont moins sensibles que les moments classiques à la présence de données atypiques (« outliers»). Ceci résulte du fait que les skewness et kurtosis classiques contiennent des puissances 3 et 4, et donc surpondèrent les données extrêmes ce qui n'est pas le cas des L-moments qui sont des combinaisons linéaires de statistiques d'ordre.
- ♣ Les L-moments ont des propriétés agréables pour le calage de lois

- ◇ l'ensemble des λ_k , pour $k \geq 1$ caractérise entièrement la distribution puisque sa fonction quantile peut être reconstruite à partir de sa décomposition sur la base des polynômes orthogonaux (pour une preuve complète cf. Hosking J.R.M., 1989) [19].
- ◇ Hosking (1990) [10] montre que pour des petits échantillons le calage par les L-moments donne de meilleurs résultats que la méthode du maximum de vraisemblance (repris dans Karvanen 2006 [14]; et Brown et Tolikas 2006 [7]); ce qui est une propriété intéressante lorsque l'on s'intéresse à la distribution de valeurs extrêmes qui par définition correspondent à des événements plus rares.
- ◇ les 4 premiers L-moments permettent de caractériser les lois usuelles (uniforme; exponentielle; normale) et les lois dites « classiques généralisées » (Generalized Logistic distribution, Generalized Extreme Value, Generalized Normal distribution, Generalized Pareto distribution, Pearson type III distribution).

2.5.1.1 Exemples de calage pour des lois usuelles

- ♣ **Loi uniforme** : $\tau_3 = 0$ et $\tau_4 = 0$. Inversement, si pour une distribution on a : $\tau_3 = 0$ et $\tau_4 = 0$, elle est nécessairement uniforme, les deux premiers L-moments donnent l'intervalle de définition de la distribution.
- ♣ **Loi exponentielle** : $\tau_3 = \frac{1}{3}$ et $\tau_4 = \frac{1}{6}$. Inversement, si pour une distribution on a : $\tau_3 = \frac{1}{3}, \tau_4 = \frac{1}{6}$ elle est nécessairement exponentielle, les deux premiers L-moments permettent de spécifier entièrement la distribution.
- ♣ **Loi normale** : $\tau_3 = 0$ et $\tau_4 \simeq 0.123$. Toute distribution ayant ces valeurs de τ_3 et τ_4 est nécessairement normale et les deux premiers L-moments permettent de la spécifier complètement.

2.5.2 Lois de probabilités généralisées

Les 4 premiers L-moments permettent de caractériser la distribution de lois dites classiques généralisées.

Ces distributions constituent une généralisation des lois classiques

- ◆ : **Generalized Logistic distribution (GLO)** : Cas particulier : distribution logistique.
- ◆ : **Generalized Extreme Value distribution (GEV)** : Cas particulier : distribution de Gumbel.
- ◆ : **Generalized Pareto distribution (GPD)** : Cas particulier : distributions uniformes, exponentielle et Pareto.

◆ **Pearson type III distribution (PE3)** : Cas particulier : distribution exponentielle, normale et Gamma.

◆ : **Distribution Kappa.**

Les trois paramètres des distributions généralisées $(\beta; \alpha; k)$ sont systématiquement le paramètre de localisation de la distribution β , celui de sa dispersion α et celui de forme de la distribution k .

Les deux premiers ne coïncident pas généralement avec la moyenne et la variance de la distribution.

Les L-moments de ces distributions généralisées peuvent être calculés soit explicitement, soit numériquement en fonction des paramètres $(\beta; \alpha; k)$. Pour chaque loi généralisée il existe une relation entre les L-ratios τ_3 et τ_4 qui peut être représentée par une courbe dans le plan $(\tau_3; \tau_4)$. Le calage de loi peut être effectué à l'aide d'un diagramme où l'on repère en abscisse et en ordonnée les ratios τ_3 et τ_4 : les lois courantes y sont décrites par des courbes, et les estimateurs de L-moments par des points, le calage consiste à identifier les lois dont les courbes sont les plus proches des points.

2.5.3 L-moments pour certaines distributions spécifiques

2.5.3.1 Distribution logistique généralisée

Définition 2.1 Paramètres : β (location), α (dispersion), k (forme)

Gamme de x : $-\infty < x \leq \beta + \alpha/k$ si $k > 0$; $-\infty < x < \infty$ si $k = 0$; $\beta + \alpha/k \leq x < \infty$ si $k < 0$.

$$f(x) = \frac{\alpha^{-1} e^{-(1-k)y}}{(1 + e^{-y})^2}, \quad y = \begin{cases} -k^{-1} \log\{1 - k(x - \beta)/\alpha\}, & k \neq 0 \\ (x - \beta)/\alpha, & k = 0 \end{cases}$$

$$F(x) = 1/(1 + e^{-y})$$

$$x(F) = \begin{cases} \beta + \alpha[1 - \{(1 - F)/F\}^k]/k, & k \neq 0 \\ \beta - \alpha \log\{(1 - F)/F\}, & k = 0 \end{cases}$$

Cas particuliers : $k = 0$ est la distribution logistique. Cette généralisation de la distribution logistique diffère des autres qui ont été définis dans la littérature. C'est une version reparamétrisée de la distribution log-logistique d'Ahmad et al. (1988). Le nom est choisi pour refléter la similitude de la distribution avec le Pareto généralisé et les distributions de valeur extrême généralisées.

Les L-moments sont définis pour $-1 < k < 1$.

$$\begin{aligned}\lambda_1 &= \beta + \alpha(1/k - \pi/\sin k\pi) \\ \lambda_2 &= \alpha k\pi/\sin k\pi \\ \tau_3 &= -k \\ \tau_4 &= (1 + 5k^2)/6\end{aligned}$$

les paramètres

$$k = -\tau_3, \quad \alpha = \frac{\lambda_2 \sin k\pi}{k\pi}, \quad \beta = \lambda_1 - \alpha\left(\frac{1}{k} - \frac{\pi}{\sin k\pi}\right).$$

2.5.3.2 Distribution de valeur extrême généralisée

Définition 2.2 *gamme de x : $-\infty < x \leq \beta + \alpha/k$ si $k > 0$; $-\infty < x < \infty$ si $\beta + \alpha/k \leq x < \infty$ si $k < 0$.*

$$f(x) = \alpha^{-1} e^{-(1-k)y - e^{-y}}, \quad y = \begin{cases} -k^{-1} \log\{1 - k(x - \beta)/\alpha\}, & k \neq 0 \\ (x - \beta)/\alpha, & k = 0 \end{cases}$$

$$F(x) = \begin{cases} \beta + \alpha\{1 - (-\log F)^k\}/k, & k \neq 0 \\ \beta - \alpha \log(-\log F), & k = 0 \end{cases}$$

Cas particuliers : $k = 0$ est la distribution de Gumbel ; $K = 1$ est une distribution exponentielle inverse ; C'est-à-dire que $1 - F(-x)$ est la fonction de distribution cumulative d'une distribution exponentielle. Les distributions de valeur extrême sont souvent classées en trois types avec des fonctions de distribution cumulée comme suit :

$$\begin{aligned}\text{type 1} : F(x) &= \exp(e^{-x}), & -\infty < x < \infty, \\ \text{type 2} : F(x) &= \exp(-e^{-\delta}), & 0 \leq x < \infty, \\ \text{type 3} : F(x) &= \exp(-|x|^\delta), & -\infty < x \leq 0.\end{aligned}$$

La répartition généralisée de la valeur extrême englobe chacun de ces types, types 1, 2 et 3 correspondant à $k = 0, k < 0$ et $k > 0$, respectivement. La distribution de Weibull définie par

$$F(x) = 1 - \exp[-\{(x - \zeta)/\xi\}^\delta], \quad \zeta \leq x < \infty,$$

Est une distribution de valeur extrême générique inversée avec des paramètres :

$$k = 1/\delta, \quad \alpha = \xi/\delta, \quad \beta = \zeta - \xi.$$

Les L-moments sont définis pour $k > -1$.

$$\lambda_1 = \beta + \alpha\{1 - \Gamma(1 + k)\}/k$$

Ici $\Gamma(\cdot)$ Désigne la fonction gamma

$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt.$$

2.5.3.3 Distribution Kappa

Définition 2.3 Paramètres (4) : ξ (emplacement), α (échelle), k, h . Gamme de x : limite supérieure est $\xi + \alpha/k$ si $k > 0$, ∞ si $k \leq 0$, La limite inférieure est $\xi + \alpha(1 - h^{-k})/k$ si $h > 0$, $\xi + \alpha/k$ si $h \leq 0$ et $k < 0$, et $-\infty$ si $h \leq 0$ et $k \geq 0$.

$$f(x) = \alpha^{-1} \{1 - k(x - \xi)/\alpha\}^{1/k-1} \{F(x)\}^{1-h} \quad (2.14)$$

$$F(x) = [1 - h\{1 - k(x - \xi)/\alpha\}^{1/k}]^{1/h} \quad (2.15)$$

$$x(F) = \xi + \frac{\alpha}{k} \left\{1 - \left(\frac{1 - F^h}{h}\right)^k\right\} \quad (2.16)$$

Les cas $h = 0$ et $k = 0$ sont inclus implicitement en tant que limites continues de (2.14)-(2.16).

Cas particuliers :

- ⊙ $h = -1$ est la distribution logistique généralisée,
- ⊙ $h = 0$ est la distribution de valeur extrême généralisée,
- ⊙ $h = 1$ est la distribution généralisée de Pareto.

La distribution kappa à trois paramètres de Mielke et Johnson (1973) est une spéciale cas de la distribution kappa définie ici. Sa fonction de distribution cumulative

$$F(x) = (x/b)^\theta \{a + (x/b)^{a\theta}\}^{-1/a}, x \geq 0, a, b, \theta > 0 \quad (2.17)$$

Est obtenu à partir de l'équation (2.14) en prenant $\xi = b$, $a = b/(a\theta)$, $k = -1/(a\theta)$, et $h = -a$.

La distribution kappa est une distribution à quatre paramètres qui inclut comme spécial cas, la logistique généralisée, la valeur extrême généralisée et le Pareto généralisé Distributions. La gamme de valeurs de paramètres la plus utile est $h \geq -1$. Sujet à Cette restriction, les L-moments de la distribution couvrent une large zone de (τ_3, τ_4) , Avion pour ces raisons, il est utile en tant que distribution générale avec qui consiste à comparer l'ajustement des distributions à deux et à trois paramètres et à utiliser dans simulation de données artificielles afin d'évaluer l'exactitude des méthodes statistiques.

$$l_3 = \frac{1}{3} \binom{n}{3}^{-1} \cdot \sum_{i>j>k} \sum \sum (x_{i:n} - 2x_{j:n} + x_{k:n}), \quad (2.22)$$

$$l_4 = \frac{1}{4} \binom{n}{4}^{-1} \cdot \sum_{i>j>k>l} \sum \sum \sum (x_{i:n} - 3x_{j:n} + 3x_{k:n} - x_{l:n}). \quad (2.23)$$

Les statistiques U sont largement utilisées, en particulier dans les statistiques non paramétriques. Leurs propriétés positives sont l'absence de biais, la normalité asymptotique et une légère résistance en raison de l'influx des valeurs aberrantes, voir Hosking (1990).

Lors du calcul du moment du $r^{\text{ième}}$ exemple, il n'est pas nécessaire de répéter le processus sur tous les sous-ensembles de la taille d'échantillon r , car cette statistique peut être exprimée directement comme une combinaison linéaire de statistiques d'ordre d'un échantillon aléatoire de la taille de l'échantillon n .

Si nous supposons une estimation de $E(X_{r,r})$ obtenue avec l'utilisation de statistiques U , elle peut être écrite comme $r.b_{r-1}$, où :

$$b_r = \frac{1}{n} \binom{n-1}{r}^{-1} \cdot \sum_{j=r+1}^n \binom{j-1}{r} \cdot x_{j:n}, \quad (2.24)$$

à savoir :

$$b_0 = \frac{1}{n} \cdot \sum_{j=1}^n x_{j:n}, \quad (2.25)$$

$$b_1 = \frac{1}{n} \cdot \sum_{j=2}^n \frac{(j-1)}{(n-1)} \cdot x_{j:n}, \quad (2.26)$$

$$b_2 = \frac{1}{n} \cdot \sum_{j=3}^n \frac{(j-1) \cdot (j-2)}{(n-1) \cdot (n-2)} \cdot x_{j:n}, \quad (2.27)$$

Et de manière générale :

$$b_r = \frac{1}{n} \cdot \sum_{j=r+1}^n \frac{(j-1) \cdot (j-2) \dots (j-r)}{(n-1) \cdot (n-2) \dots (n-r)} \cdot x_{j:n}, \quad (2.28)$$

Ainsi, le premier échantillon L-moments peut être écrit comme suit :

$$l_1 = b_0 \quad (2.29)$$

$$l_2 = 2b_1 - b_0 \quad (2.30)$$

$$l_3 = 6b_2 - 6b_1 + b_0 \quad (2.31)$$

$$l_4 = 20b_3 - 30b_2 + 12b_1 - b_0 \quad (2.32)$$

Nous pouvons donc écrire en général :

$$l_{r+1} = \sum_{k=0}^r P_{r,k}^* b_k, \quad r = 0, 1, 2, \dots, n-1 \quad (2.33)$$

où :

$$P_{r,k}^* = (-1)^{r-k} \binom{r}{k} \binom{r+k}{k} = \frac{(-1)^{r-k} (r+k)!}{(k!)^2 (r-k)!} \quad (2.34)$$

l'échantillon de L-moments sont utilisés de la même manière que l'exemple des L-moments conventionnels, résumant les propriétés de base de la distribution de l'échantillon, qui sont l'emplacement (niveau), la variabilité, l'asymétrie et la kurtosis.

Ainsi, l'échantillon de L-moments permettent d'estimer les propriétés correspondantes de la distribution de probabilité à partir desquelles l'échantillon est originaire et peuvent être utilisés pour estimer les paramètres de la distribution de probabilité pertinente. Nous préférons souvent les L-moments aux moments classiques dans de telles applications, car l'échantillon de L-moments en tant que fonctions linéaires des valeurs d'échantillon - sont moins sensibles à la variabilité de l'échantillon ou aux erreurs de mesure dans les observations extrêmes que les moments conventionnels. Les L-moments conduisent donc à des estimations plus précises et robustes des caractéristiques ou des paramètres de la distribution de probabilité de base. l'échantillon de L-moments ont été utilisés précédemment dans les statistiques, mais pas dans le cadre d'une théorie unifiée.

Le premier échantillon L-moment l_1 est un échantillon de L-localisation (moyenne de l'échantillon), le deuxième instant d'échantillonnage l_2 étant un échantillon de L-variabilité . L'estimation naturelle du rapport L-moments (10) est le rapport échantillon des L-moments :

$$t_r = \frac{l_r}{l_2}, \quad r = 3, 4, \dots \quad (2.35)$$

Par conséquent, t_3 est un échantillon de L-skewness et t_4 est un échantillon de L-kurtosis. Les rapports d'échantillons des moments-L t_3 et t_4 peuvent être utilisés comme caractéristiques d'asymétrie et de kurtosis d'un ensemble de données d'échantillons.

Tableau 2 Formules pour les estimations de paramètres réalisées par la méthode de L-moments de lognormal et des répartitions de probabilité Pareto généralisées

Distribution	Estimation des paramètres
lognormal	$z = \sqrt{\frac{8}{3}} \cdot \Phi^{-1} \left(\frac{1+t_3}{2} \right)^2$ $\widehat{\sigma} = 0.999281z - 0.006118z^3 + 0.000127z^5$ $\widehat{\mu} = \ln \frac{l_2}{\operatorname{erf}(\frac{z}{2})} - \frac{\widehat{\sigma}^2}{2}$ $\widehat{\xi} = l_1 - \exp \left(\widehat{\mu} + \frac{\widehat{\sigma}^2}{2} \right)$
Pareto généralisées	(ξ) $\widehat{k} = \frac{l_1}{l_2} - 2$ $\widehat{\alpha} = (1 + \widehat{k}) \cdot l_1$

La différence significative de Gini se rapporte à la fois aux échantillon de L-moments, ayant la forme de :

$$G = \binom{n}{2}^{-1} \cdot \sum_{i>j} \sum (x_{i:n} - x_{j:n}),$$

Et le coefficient de Gini qui dépend uniquement d'un seul paramètre $U3c3$ dans le cas de la distribution lognormal à deux paramètres, en fonction toutefois des valeurs des trois paramètres dans le cas de la distribution lognormal trois paramétrique. Le tableau 2 présente les expressions pour les estimations de paramètres des distributions de probabilité normale et partielle de Pareto obtenues en utilisant la méthode des L-moments .

Pour plus de détails, voir, par exemple, Bílková (2010), Bílková (2011), Bílková (2012), Bílková, Malá (2012), Hosking (1990) ou Kyselý, Pícek (2007).

CHAPITRE 3

EXEMPLES DE SIMULATION SOUS R

Pour illustrer l'utilisation de la méthode L-moment par des résultats de simulation, on va générer une variable aléatoire X qui suit une loi de probabilité à trois paramètres, les trois paramètres des distributions généralisées $(\beta; \alpha; k)$ sont systématiquement le paramètre de localisation de la distribution β , celui de sa dispersion α et celui de forme de la distribution k . Les deux premiers ne coïncident pas généralement avec la moyenne et la variance de la distribution.

On va générer la loi selon les 3 paramètres et appliquer le package `lmom` pour calculer les quatre L-moments.

3.1 Simulations des distributions de trois paramètres

3.1.1 Generalized Extreme Value distribution (GEV)

GEV	λ_1	λ_2	τ_3	τ_4
$(\beta = 0; \alpha = 2; k = -0.25)$	1.8033336	1.8548605	0.3410373	0.2408491
$(\beta = 1; \alpha = 2; k = 0)$	2.154431	1.386294	0.169925	0.150375
$(\beta = 0.25; \alpha = 1.5; k = -0.4)$	2.0844709	1.7842826	0.4543469	0.3258053

3.1.2 Generalized Logistic (GLO) distribution :

On va générer la loi GLO selon les 3 paramètres et appliquer le package `lmom` pour calculer les 4 L-moments

GLO	λ_1	λ_2	τ_3	τ_4
$(\beta = 0; \alpha = 1; k = -0.5)$	1.141593	1.570796	0.500000	0.375000
$(\beta = 4; \alpha = 1; k = 0)$	4.0000000	1.0000000	0.0000000	0.1666667
$(\beta = 3; \alpha = 5; k = -0.2)$	4.723983	5.344797	0.200000	0.200000

3.1.3 Gamma distribution

On va généré la loi GPA selon les 3 paramètres et appliqué le package lmom pour calculer les 4 L-moments

Gamma	λ_1	λ_2	τ_3	τ_4
$(\beta = 2; \alpha = 2)$	4.0000000	1.5000000	0.2345679	0.1415895
$(\beta = 1; \alpha = 5)$	5.0000000	2.5000000	0.3333333	0.1666667
$(\beta = 2.5; \alpha = 1)$	2.5000000	0.8488264	0.2092834	0.1372565

3.1.4 Generalized Pareto distribution (GPD)

On va généré la loi GPD selon les 3 paramètres et appliqué le package lmom pour calculer les 4 L-moments.

GPD	λ_1	λ_2	τ_3	τ_4
$(\beta = -0.71; \alpha = 4; k = -0.35)$	5.4438462	3.7296037	0.5094340	0.3279917
$(\beta = 0; \alpha = 6.4; k = 0.62)$	3.95061728	1.50786919	0.10497238	0.03135538
$(\beta = 0.78; \alpha = 1; k = 0)$	1.7800000	0.5000000	0.3333333	0.1666667

3.1.5 Pearson type III distribution (PE3)

PE3	λ_1	λ_2	τ_3	τ_4
$(\beta = 0; \alpha = 1.9; k = 5)$	0.0000000	0.6300146	0.7056335	0.4507001
$(\beta = 1; \alpha = 8.3; k = 0)$	1.0000000	4.6827735	0.0000000	0.1226017
$(\beta = 2; \alpha = 3; k = 0.4)$	2.00000000	1.68412759	0.06527693	0.12387312

3.2 Application sur la régionalisation des précipitations

Les paramètres de la distribution régionale sont estimés à partir de l'ensemble des statistiques locales de la région homogène. L'estimation des statistiques locales est obtenue par la méthode des L-moments.

3.2.1 Test d'homogénéité

Pour tester l'homogénéité d'une région (groupe de stations) en termes de rapports des L-moments le test d'homogénéité statistique proposé par Hosking et Wallis (1993)

[21] est utilisé, dans ce test, les paramètres représentatifs d'une région sont les moyennes pondérées des statistiques des L-moments.

Ainsi, pour une région de N stations chacune de taille n_i , les rapports des L-moments et les L-moments régionaux sont calculés de manière suivante :

$$\bar{t}_r = \frac{\sum_{i=1}^N n_i t_r^{(i)}}{\sum_{i=1}^N n_i}, \quad r \leq 3, \quad (3.1)$$

où $t_r^{(i)}$ sont les rapports des L-moments d'ordre r à la station i .

Le test de simulation de Monte Carlo est utilisé afin de tester l'homogénéité d'une région. Un grand nombre des données de régions sont générées à partir de la distribution Kappa à quatre paramètres. La distribution Kappa est ajustée en utilisant les rapports des L-moments moyens pondérés.

Chacune des simulations doit refléter la configuration de la base de données de la région considérée. Plus précisément, au cours d'une simulation le nombre de sites et le nombre d'observations à chaque site doivent être reproduits. Pour chaque région générée, les trois mesures suivantes de la variabilité intersites des rapports des L-moments sont calculées comme suit (Hosking et Wallis (1993) [21]) :

La variance pondérée du L-Cv :

$$V_1 = \frac{\sum_{i=1}^N n_i (t^{(i)} - \bar{t})^2}{\sum_{i=1}^N n_i}$$

L'écart-type pondéré pour les rapports des L-moments t et t_3 :

$$V_2 = \frac{\sum_{i=1}^N n_i \sqrt{(t^{(i)} - \bar{t})^2 + (t_3^{(i)} - \bar{t}_3)^2}}{\sum_{i=1}^N n_i}$$

L'écart-type pondéré pour les rapports des L-moments t_3 et t_4 :

$$V_3 = \frac{\sum_{i=1}^N n_i \sqrt{t_3^{(i)} - \bar{t}_3)^2 + (t_4^{(i)} - \bar{t}_4)^2}}{\sum_{i=1}^N n_i}$$

où $t^{(i)}, t_3^{(i)}$ et $t_4^{(i)}$ désignent respectivement les L-Cv, L-Cs et L-Ck du site i ; \bar{t} , \bar{t}_3 et \bar{t}_4 désignent respectivement les L-Cv, L-Cs et L-Ck régionaux calculées selon l'équation (3.1) et N est le nombre de sites. Si on désigne par V l'une de ces trois valeurs, v_1 , v_2 et v_3 le critère d'homogénéité d'une région est calculé comme suit :

$$H_V = \frac{V_{obs} - \mu_V}{\sigma_V}$$

où V_{obs} est la valeur observée de V et μ_V et σ_V sont respectivement la moyenne et l'écart-type de V obtenu par simulations. La variable H mesure la dispersion des observations à celle des simulations. Selon Hosking et Wallis (1993) [21], une région peut être considérée comme étant acceptablement homogène si $H < 1$, probablement hétérogène si $1 \leq H \leq 2$ et définitivement hétérogène si $H \geq 2$.

3.2.2 Identification de la distribution régionale

Parmi les différentes distributions, la distribution de GEV, la distribution Gumbel et la distribution logistique (GLO) sont les plus souvent utilisées en Algérie, dans l'analyse fréquentielle des précipitations extrêmes dans un site donné. Ces lois ont trois paramètres est aussi largement utilisées dans différents contextes climatiques. Leurs comportement asymptotique des distributions sont remis en question par Koutsoyiannis (2004).

Alila (1999) émet certaines réserves quant à l'application de la loi Gumbel dans un contexte régional. Dans son étude de la régionalisation des précipitations de courte durée au Canada, les différentes distributions ont été testées. Parmi ces distributions la distribution GEV a été identifiée comme étant la distribution régionale la plus appropriée adéquate. La distribution GEV est la plus utilisée tant pour l'analyse fréquentielle régionale des précipitations que pour celle des crues. Par exemple, Overeem et al. (2007) l'ont utilisée pour la régionalisation des précipitations de courte durée pour l'ensemble de la Hollande. Djerboua (2001), Versiani et al. (1999), Cannarozzo et al. (1995) ont choisi comme modèle statistique régional des précipitations journalières maximales annuelles la distribution à deux composantes TCEV (Two Component Extreme Value). Pour déterminer les courbes de croissance régionales des précipitations de courte durée Sveinsson et al. (2002) ont utilisé une approche régionale fréquentielle basée sur la méthode d'indice de crue en prenant en considération les différentes distributions : log-normale à trois paramètres (LN3), GEV, log-normale (LN) et Pearson type 3 (P3). Dans cette étude, nous faisons l'hypothèse que les distributions GEV, LN3, P3 et GLO (Generalized Logistic) s'ajustent aux séries des valeurs réduites des précipitations maximales annuelles journalières de la région d'étude. La pertinence de l'ajustement de chacune de ces distributions à trois paramètres est évaluée en termes de différence entre le L-applatissement théorique de la distribution ajustée et le L-applatissement régional. La signification de cette différence est estimée par la statistique Z (Hosking et Wallis, 1993 [21]) :

$$Z^{DIST} = \frac{\tau_4^{DIST} - \bar{\tau}_4 + \beta_4}{\sigma_{\bar{\tau}_4}}$$

où $\bar{\tau}_4$ est le L-applatissement régional observé, τ_4^{DIST} est le L-applatissement théorique

de la distribution (DIST) estimé à partir du L-asymétrie régional observé β_4 et $\sigma_{\bar{t}_4}$ sont respectivement le biais et l'écart-type de \bar{t}_4 obtenus par simulations d'une région homogène avec la distribution Kappa. La statistique Z^{DIST} est basée sur la normalité asymptotique et l'ajustement, elle est satisfait au niveau de 90% si $Z^{DIST} \leq 1.64$.

3.2.3 Estimation des paramètres et quantiles de la distribution régionale

Pour estimer les paramètres de la distribution régionale les trois premiers L-moments régionaux sont utilisés. La courbe de croissance régionale sera établie sur la base des paramètres de la distribution régionale en appliquant la moyenne comme facteur d'échelle. Dans cette approche les L-coefficients d'asymétrie et de variation régionaux sont supposés constants. Ainsi, pour estimer les précipitations associées aux différentes périodes de retour en un site donné d'une région homogène, il faut multiplier les valeurs du quantile régional correspondant aux mêmes périodes de retour par la moyenne des précipitations journalières maximales.

3.2.4 Region d'étude et données

Pour illustrer l'application de la distribution GLO en utilisant l'approche L-moments, un ensemble de maximum Série de pluies quotidiennes sélectionnées chaque année est présenté ici. L'ensemble de données utilisé comprend 35 pluies maximales annuelles de la période 1979–2014 sur 3 station du bassin de la Hodna les caractéristiques des 3 stations sont présenté dans le tableau suivant :

Nom de la station	code	Longitude(m)	Latitude	Altitude	période d'observation
Baniou	05 18 01	4,380	35.440	404	1979-2014
Ben srou	05 21 01	4,687	35.438	369	1979-2014
chellal	05 07 01	5	35.438	371	1979-2014

les maximum annuelles des précipitations sont présenté dans le tableau suivants :

année	max	année	max	année	max
1979	19,15399	1979	24,34506	1979	29,85
1980	32,84912	1980	33,11004	1980	31,18
1981	60,75096	1981	62,368	1981	70,21
1982	34,82667	1982	25,77152	1982	26,55
1983	10,04219	1983	11,75881	1983	15,81
1984	11,93562	1984	18,86387	1984	37,15
1985	17,94376	1985	15,78768	1985	23,46
1986	19,39087	1986	23,03009	1986	28,27
1987	21,3564	1987	33,20105	1987	38,28
1988	15,47013	1988	23,10219	1988	28,69
1989	30,748	1989	34,51165	1989	41,12
1990	17,42364	1990	16,92066	1990	16,28
1991	78,32569	1991	60,71913	1991	51,38
1992	46,63873	1992	43,05654	1992	39,53
1993	15,75679	1993	16,05034	1993	17,28
1994	71,84716	1994	15,09247	1994	21,76
1995	71,847166	1995	75,99109	1995	75,67
1996	55,48093	1996	60,66424	1996	41,97
1997	59,13735	1997	52,75153	1997	37,25
1998	27,39373	1998	39,96966	1998	60,66
1999	33,7998	1999	32,20706	1999	36,85
2000	5,645941	2000	7,815745	2000	9,13
2001	7,453534	2001	8,201982	2001	8,54
2002	13,35526	2002	13,36212	2002	15,98
2003	12,17422	2003	10,60524	2003	11,83
2004	12,42657	2004	12,43515	2004	15,21
2005	22,67991	2005	21,71173	2005	18,97
2006	57,03624	2006	44,33497	2006	21,82
2007	11,64895	2007	9,793282	2007	14,46
2008	19,74621	2008	22,8447	2008	25,04
2009	27,15683	2009	31,08788	2009	38,95
2010	21,12812	2010	34,09023	2010	29,15
2011	29,24767	2011	21,75122	2011	18,23
2012	20,9736	2012	28,08723	2012	25,87
2013	22,56318	2013	13,08017	2013	14,73
2014	19,44923	2014	16,5413	2014	16,44

les caractéristiques descriptives de la série pluviométriques :

min	1Q	med	moyenne	3Q	max
10,04219	10,60524	11,64895	11,75881	11,83	11,93562

l'estimation des 3 paramètres de chaque lois

loi	γ	α	k	h
GEV	3.4894458	-0.1962443	1.6361435	/
GLO	0.2825000	6.5098814	-0.7397933	/
KAPPA	0	0.05737772	0.93291262	1.10526418

l'application du package lmom pour estimer les 4 L-moments pour les différentes distribution :

distribution	λ_1	λ_2	τ_3	τ_4	$L - cv$
kappa	0.03115113	0.009328264	0.0421214	-5.751347e-05	0.0004227403
GEV	3.545748	-0.1195448	-0.5400145	0.308649	0.9285187
GLO	22.06618	20.48886	0.7439045	0.6301605	0.9285187

Résultat : le coefficients de variation de la distribution Kappa est petit par rapport les autres

l'application du package lmom pour estimer les 4 L-moments pour les différentes distribution :sur la station Baniou :

distribution	λ_1	λ_2	τ_3	τ_4	$L - cv$
kappa	0	0	0.1265904	0.1642923	0
GEV	0.8945717	1.277738	0.05642036	0.1162469	1.428324
GLO	26.29646	24.73387	0.7441074	0.6304167	0.9405779

l'application du package lmom pour estimer les 4 L-moments pour les différentes distribution :sur la station srouur :

distribution	λ_1	λ_2	τ_3	τ_4	$L - cv$
kappa	0	0	-0.3683781	0.07833581	0
GEV	0.259181	1.455682	-0.2500052	0.1330197	5.616469
GLO	-0.3618315	0.6466452	-0.579649	0.3460263	-1.787145

l'application du package lmom pour estimer les 4 L-moments pour les différentes distribution :sur la station Chellal :

distribution	λ_1	λ_2	τ_3	τ_4	$L - cv$
kappa	0	0	0.04028067	0.05050975	0
GEV	0.3037721	0.3270389	0.2779056	0.2021027	1.076593
GLO	-5.009671	-4.712269	0.74473	0.6312032	0.9406344

la Z^{DIST} statistique a été calculée pour chaque distribution. Les valeurs de cette dernière obtenues par la réalisation de 108 simulations en utilisant la distribution Kappa

et les valeurs de Laplatissement théorique de chaque distribution ajustées sont données dans le tableau 9 :

Dist	Région1		Région2		Région3	
	τ_4	Z	τ_4	Z	τ_4	Z
Kappa	0.1642923	-0.005579	0.07833581	-0.01873	0.05050975	-0.0262
gev	0.1162469	-0.01387	0.1330197	0.005375	0.2021027	-0.011303
glo	0.6304167	-0.017315	0.3460263	0.015674	0.6312032	-0.021515

CONCLUSION GÉNÉRALE

L'information sur les valeurs des précipitations extrêmes et leurs fréquences est nécessaire pour les différents projets liés à l'eau. Dans cette étude, l'analyse fréquentielle régionale basée sur les L-moments a été utilisée pour améliorer la qualité d'estimation des quantiles des précipitations journalières maximales annuelles du bassin l'Hodna situé au coeur de la steppe Algérienne, il est parmi les 5 premiers bassins versant de l'Algérie située au sud des hauts plateaux ayant pour capitale M'sila ,située au Sud-Est à 248 km d'Alger. La méthode d'analyse régionale fréquentielle basée sur les L-moments introduite par Hosking et Wallis (1997) [21] a été utilisée. L'application de test d'homogénéité a permis de diviser la zone d'étude en trois régions statiquement homogènes Le groupement des stations présente une cohérence géographique et climatique.L'étape importante de l'analyse régionale a été l'indentification de la distribution régionale de chaque région homogène. Les distributions à trois paramètres GEV, GLO et Kappa ont été testées. Parmi ces dernières, selon le test de la Z-statistique la distribution Kappa a été identifiée comme la plus adéquate distribution régionale pour les 3 régions.

BIBLIOGRAPHIE

- [1] Introduction à la statistique inférentielle Didier Concordet Unité de Biométrie Ecole Vétérinaire de Toulouse.
- [2] Cours d'introduction à l'analyse statistique 4, L3 LISS - Université Paris-Dauphine, Arnold Chassagnon, LEDa-SDFi, Octobre 2010
- [3] Adamowski, K. (2000). Regional analysis of annual maximum and partial duration flood data by nonparametric and L-moment methods. *Journal of Hydrology*, 229(3), 219-231.
- [4] Sillitto G., 1969, "Derivations of approximate to the inverse distribution function of a continuous univariate population from the order statistics of a sample", *Biometrika* 56(3), 641-650.
- [5] David H.A. , 1968, "Gini's mean difference rediscovered", *Biometrika* 55, 573- 575.
- [6] Chernoff H., Gastwirth J.L., Johns M.V., 1967, "Asymptotic distribution of linear combinations of functions of order statistics with applications to estimation", *The annals of Mathematical Statistics* 38(1), 52-72.
- [7] Brown, R. A., Tolikas, K. (2006). The Distribution of the Extreme Daily Share Returns in the Athens Stock Exchange.
- [8] Brown, Richard A., and Konstantinos Tolikas. "The Distribution of the Extreme Daily Share Returns in the Athens Stock Exchange." (2006).
- [9] Greenwood J.A., Landwehr J.M., Matalas N.C., Wallis J.R., 1979, "Probab Weighted moments : definition in relation to parameters of several distributions expressable in inverse form", *Water Ressources Research* 15, 1049-1054.
- [10] Hosking J., 1990, "L-moments : Analysis and estimation of distributions using linear combinations of order statistics", *Journal of Royal Statistical Society B* 52 (1),105-124.

- [11] Dewar R.E., Wallis J.R., 1990, "Geographical patterning of interannual rainfall variability in the tropics and near tropics : an L-moment approach", *Journal of Climate* 12 (12), 3457-3466.
- [12] Smithers J.C., Schulze R.E., 2001, "A methodology for the estimation of short duration design storms in South Africa using a regional approach based on L-moments, *Journal of Hydrology* 241 (1-2), 42-52.
- [13] Chen X., Tung Y.-K., 2003, "Investigation of polynomial normal transform", *Structural Safety* 25(4), 423-445.
- [14] Karvanen J., 2006, « Estimation of Quantile Mixtures via L-moments and Trimmed L-moments », working paper.
- [15] Chu B., Salmon M., 2008, « Testing Distributional Assumptions : A L-moment Approach », working paper.
- [16] Martins-Filho C., Yao F., 2006, « Estimation of Value-at-Risk and Expected Shortfall based on Nonlinear Models of Return Dynamic and Extreme Value Theory », working paper.
- [17] Jurcenzo E., Maillet B., Yanou G., 2007, « Portfolio Selection with Higher-order L-moments : A Robust Non-parametric Multi-moment Efficient Frontier », working paper.
- [18] Darolles S., Gouriéroux C., Jasiak J., 2008, « L-Performance with an Application to Hedge Funds », *Journal of Empirical Finance*, 4, 671-685.
- [19] Hosking J.R.M., 1989, « Some theoretical results concerning L-moments » Research Report RC14492, IBM Research Division, Yorktown Heights, N.Y.
- [20] Brooks C., Kat H., 2001, « The Statistical Properties of Hedge Funds Index Returns and Their Implications for Investors », Cass Business School Research Paper.
- [21] Hosking J.R.M., Wallis J.R., 1997, "Regional Frequency Analysis", Cambridge University Press.

