

الجمهورية الجزائرية الديمقراطية الشعبية
République algérienne démocratique et populaire
وزارة التعليم العالي والبحث العلمي
Ministère de l'enseignement supérieur et de la recherche scientifique
جامعة عين تموشنت بلحاج بوشعيب
Université –Ain Temouchent- Belhadj Bouchaib
Faculté des Sciences et de Technologie
Département des Mathématiques et informatique



Projet de Fin d'Etudes
Pour l'obtention du diplôme de Master en : Probabilités et statistique
appliquées
Domaine : Mathématique et Informatique
Filière : Mathématiques
Spécialité : Probabilités et statistique appliquées
Thème

**Mélange fini basé sur la distribution gamma et la
distribution de Lindley généralisée**

Présenté Par :

1) Melle HADJ KADDOUR Ibtissem.

Devant le jury composé de :

Dr. KHIAR Hamid	M C B	UAT.B.B (Ain Temouchent)	Président
Dr. MESSABIHI Aicha	M C B	UAT.B.B (Ain Temouchent)	Examinatrice
Dr. MECENE Rahmouna	M C B	UAT.B.B (Ain Temouchent)	Encadrante

Année Universitaire 2023/2024

Remerciements

Avant tout, Il semble approprié d'entamer ce mémoire par des remerciements. D'abord je remercie dieu de m'avoir accordé la force et le courage de mener à terme ce modeste travail.

Toute ma reconnaissance et toute ma gratitude vont vers Dr. Rahmouna MECENE, qui m'a aidé et accompagné tout au long de cette expérience professionnelle avec beaucoup de patience et d'enthousiasme et de m'avoir guidé et orienté durant les différentes étapes de ce projet avec ça pédagogie et ça ferveur.

Je remercie également les membres du jury d'avoir accepter d'examiner et de juger mon travail.

Que tous ceux qui, de près ou de loin ont contribué, par leurs conseils, leurs encouragements ou leur amitié à l'aboutissement de ce travail, trouvent ici l'expression de mon profonde reconnaissance.

Pour leur encouragement, leur soutien moral et la patience qu'ils m'ont manifesté durant toute l'année, je remercie fortement tous les membres de ma familles.

Dédicace

Je dédie ce travail à :

A mon cher père

Pour son aide et soutien et son patience, cette aventure n'aurait certainement pas existé sans vous !

A ma chère mère

En témoignage de mon éternelle reconnaissance, que Dieu vous protège et vous prête bonne santé et longue vie.

A Dr. Rahmouna MECENE

Pour m'avoir encadrer et faire de leurs mieux afin de m'aider.

A Mes soeurs et Mes frères

Pour ses soutiens moral et leurs conseils précieux tout au long de mes études.

A mon fiancé

Pour l'encouragement et l'aide qu'il m'a toujours accordé.

A Mes neveux

Rayane, Islem et Sohaib.

Résumé

Dans ce mémoire, nous nous intéressons à l'introduction des nouveaux modèles probabilistes en se basant sur la stratégie de mélange de lois de probabilités. En premier lieu, nous donnons quelques rappels et notions de base en probabilités. Ensuite, nous présentons la définition formelle d'un modèle général de mélange fini et proposons deux méthodes d'estimation pour des paramètres : la méthode du maximum de vraisemblance et celle d'expectation-maximisation. Afin de se familiariser avec les modèles de mélange, nous montrons un exemple fréquemment utilisé. Celui est le mélange de lois gaussiennes. La variété des modèles existants dans la littérature rend le choix entre eux difficile. Pour cette raison, nous présentons deux critères classiques qui permettent de déterminer le meilleur modèle : le critère AIC et le critère BIC. Dans le dernier chapitre de ce mémoire, nous focalisons sur le mélange fini qui repose sur la distribution gamma et Lindley généralisée que ce soit unidimensionnel ou bidimensionnel. Dans ce contexte, nous étudions diverses propriétés de ces distributions telles que la représentation graphique, estimation des paramètres, etc.

Abstract

In this thesis, we focus on the introduction of new probabilistic models based on the strategy of mixing probability laws. First, we give a few reminders and basic notions of probability. Then, we present the formal definition of a general finite mixture model and propose two estimation methods for parameters : the maximum likelihood method and the expectation-maximization method. To familiarize ourselves with mixture models, we show a frequently used example. This is a mixture of Gaussian distributions. The variety of existing models in the literature makes it difficult to choose between them. For this reason, we present two classical criteria for determining the best model : the AIC criterion and the BIC criterion. In the final chapter of this dissertation, we focus on finite mixing based on the generalized gamma-Lindley distribution, whether one- or two-dimensional. In this context, we study various properties of these distributions, such as graphical representation, parameter estimation, and so on.

ملخص

في هذا البحث العلمي، نركز على إدخال النماذج الاحتمالية الجديدة القائمة على . استراتيجيات خلط قوانين للاحتمالات. أولاً، نقدم بعض التذكيرات والمفاهيم الأساسية والمفاهيم الأساسية للاحتمالات. بعد ذلك، نقدم التعريف الرسمي لنموذج الخليط المحدود العام ونقترح طريقتين لتقدير البارامترات: أقصى احتمال طريقة الحد الأقصى للاحتمالية القصوى وطريقة تعظيم التوقع. وللتعرف على نماذج المخلوط، نعرض مثالاً متكرر الاستخدام. مثال متكرر الاستخدام. هذا هو خليط القوانين الغاوسية. التوزيعات الغاوسية. إن تنوع النماذج الموجودة في الأدبيات يجعل من الصعب الاختيار بينهم. لهذا السبب، نقدم اثنين من أفضل النماذج: معيار انيبي ومعيار بنيبي. في الفصل الأخير من هذه الأطروحة نركز على الخليط المحدود، والذي يعتمد على توزيع ليندلي، سواء كان أحادي البعد أو ثنائي الأبعاد. المنطقي. في هذا السياق، ندرس الخصائص المختلفة مثل التمثيل البياني، والاعدادات إلخ

Table des matières

List of Figures	3
List of Tables	5
Introduction générale	6
1 Généralités	9
1.1 Tribu et Mesure	9
1.1.1 Tribu : Définition et Propriétés	9
1.1.2 Mesure et mesure de lebesgue sur l'ensemble des boréliennes	10
1.2 Loi de probabilité continue	12
1.2.1 Variable aléatoire continue	12
1.2.2 Fonction des moments	13
1.2.3 Exemple d'une loi continue fréquemment utilisée	14
1.3 Indépendance et conditionnement	14
2 Mélange fini de lois de probabilités	16
2.1 Présentation générale de modèle de mélange fini	16
2.2 Estimation de paramètres de la loi mélange	17
2.2.1 Estimation par la méthode du maximum de vrai- semblance	17
2.2.2 Algorithme d'expectation-maximisation (EM)	19
2.2.3 Test de Kolmogorov-Smirnov (K-S)	20
2.3 Modèle de mélange gaussien	21

2.3.1	Mélange de gaussiennes multinomiales	21
2.4	Choix entre modèle de mélange	23
2.4.1	Critère BIC (Bayesian Information Criterion) .	23
2.4.2	Critère AIC (Akaike Information Criterion) . .	23
3	Sur La distribution Gamma-Lindley généralisée	24
3.1	Distribution Gamma	25
3.2	Distribution de Lindley généralisée	26
3.2.1	Version bivariée de la distribution de Lindley généralisée	26
3.2.2	Version univariée de la distribution de Lindley généralisée	28
3.2.3	Forme de la distribution univariée de Lindley généralisée	29
3.2.4	Application réelle	31
3.3	Distribution bivariée Gamma-Lindley généralisée . . .	32
3.3.1	Forme de la distribution bivariée de gamma- lindley généralisée	33
3.3.2	Estimation des paramètres par L'algorithme EM	36
3.3.3	Distribution de la somme	39
3.3.4	Applications réelles	40
	Conclusion générale	44
	Bibliographie	45

Table des figures

1.1	Densité de la loi normale centrée réduite.	14
2.1	Un mélange gaussien dans \mathbb{R}^2 avec trois composantes.	22
3.1	Code MATLAB qui génère les graphes de la distribution BGLG	34
3.2	Graphe de la fonction de densité de la distribution BGLG pour quelques valeurs sélectionnées de paramètres $(\alpha, \beta, \gamma, \theta)$.	34
3.3	Code MATLAB qui donne les EMVs, AIC et BIC.	43

Liste des tableaux

3.1	Estimations et statistique de Kolmogorov-Smirnov . . .	32
3.2	Valeurs estimées de $\hat{\alpha}$, $\hat{\beta}$, $\hat{\gamma}$, $\hat{\theta}$ et $E(y)$	41
3.3	EMVs, AIC et BIC pour l'ensemble des données réelles de pluie et de neige pour la ville de Dijon.	42

Introduction générale

L'incertitude désigne l'absence de certitude ou de précision concernant la survenue ou l'issue d'un événement ou d'une situation. Autrement dit, c'est la possibilité qu'un événement ou situation se produise ou non. Elle peut être due à divers facteurs : le manque d'information, la complexité du système étudié, etc. En toute rigueur, l'incertitude peut être mesurée à l'aide d'une variable aléatoire ou un vecteur de variables aléatoires. Cependant, une variable aléatoire X est une application d'un espace probabilisé Ω (essentiellement défini comme un ensemble d'évènements possibles et muni d'une mesure P appelée probabilité), vers l'ensemble des réels \mathbb{R} . Ceci permet d'assigner un réel x au résultat $\omega \in \Omega$ d'un phénomène ou d'une expérience aléatoire. On écrit alors $X(\omega) = x$.

La notion d'incertitude liée aux variables aléatoires, est fondamentale dans de nombreux domaines, notamment la finance et la biologie. En finance, l'incertitude est abordée à travers des modèles probabilistes pour évaluer les risques et les rendements d'investissement. En effet, la théorie des options repose sur la modélisation de prix d'actifs financiers comme des variables aléatoires. La génétique des populations et l'écologie font appel à des modèles probabilistes pour étudier la variabilité génétique et les interactions entre espèces. Ces domaines utilisent des outils mathématiques telles que les distributions de probabilité et les statistiques pour quantifier et gérer l'incertitude associée aux phénomènes étudiés.

Comme les données résultantes de différents phénomènes sont souvent positives, alors les distributions à support positif sont importantes car elles permettent de représenter des variables qui ne peuvent pas prendre des valeurs négatives. Par exemple : le prix d'une action le nombre de clients, le taux de rendement d'investissement, etc. Elles sont aussi importantes dans la biologie pour modéliser la distribution des espèces. Par exemple : la concentration d'une substance dans un échantillon, le nombre de cellules dans un organe, le taux de croissance d'un organisme. La modélisation à l'aide des distributions à support positif devient très intéressante et ceci grâce à ses caractéristiques qui sont les suivantes :

1. La fonction de répartition est croissante. Cela signifie que la probabilité d'obtenir une valeur plus grande est plus élevée que la probabilité d'obtenir une valeur plus petite.
2. La fonction de densité est positive. Cela signifie que la probabilité d'obtenir une valeur donnée est non nulle.
3. L'espérance est finie. Cela signifie que la valeur moyenne de la variable est non nulle.

Un exemple classique de distribution à support positif est la distribution gamma représentée par sa densité :

$$f(x, \alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} \exp\left(-\frac{x}{\beta}\right), \quad x > 0 \quad (1)$$

Cette distribution est définie uniquement pour les valeurs positives. Elle est souvent utilisée dans l'analyse de survie, la fiabilité et les processus de file d'attente et d'autres domaines et cela en raison de ses propriétés flexibles et sa capacité à modéliser de nombreux types de données.

Ce mémoire est consacré à l'étude de la loi mélange, notamment la loi mélange de gamma et de Lindley généralisée. Il est composé de trois chapitres :

Dans le premier chapitre de ce mémoire, nous allons présenter les concepts théoriques nécessaires pour la définition de l'espace de probabilité. De plus, nous allons définir quelques notions de base relatives aux variables aléatoires et loi de probabilités.

Dans le deuxième chapitre nous donnons la présentation générale de modèle de mélange fini et nous élaborons deux méthodes d'estimation de paramètres pour la loi mélange. La première méthode est la méthode de maximum de vraisemblance et la deuxième est la méthode d'expectation-maximisation. De plus, nous exposons un exemple de loi mélange fréquemment utilisée qui est la loi mélange gaussien.

Dans le troisième chapitre, nous s'intéressons au mélange fini basé sur la distribution gamma et Lindley généralisée que ce soit unidimensionnel ou bidimensionnel. Après avoir présenter les propriétés de base de la distribution de Lindley généralisée, nous allons chercher les estimateurs de chaque extension de loi et entamer une application réelle qui montre la flexibilité de cette loi mélange.

1

Généralités

Dans ce premier chapitre, nous allons établir certaines notions dont nous aurons besoin tout au long de ce mémoire (voir [1] et [3]).

1.1) Tribu et Mesure

Les tribus et les mesures sont des outils mathématiques puissants qui permettent de quantifier la taille ou la probabilité d'ensembles. Elles ont de nombreuses applications dans divers domaines : mathématiques, scientifiques et de l'ingénierie.

1.1.1 Tribu : Définition et Propriétés

Dans cette partie, on note \mathbb{E} un ensemble quelconque et on introduit la notion de tribu, qui est une classe d'ensembles de parties de \mathbb{E} , c'est-à-dire une classe de sous ensembles de $\mathcal{P}(\mathbb{E})$.

Définition 1 Soit \mathbb{E} un ensemble, \mathcal{A} une famille de parties de \mathbb{E} . La famille \mathcal{A} est dite une tribu (on dit aussi une σ -algèbre sur \mathbb{E} si \mathcal{A} vérifie :

1. $\emptyset \in \mathcal{A}$ et $\mathbb{E} \in \mathcal{A}$.
2. \mathcal{A} est stable par union dénombrable, c'est-à-dire que pour toute famille dénombrable $(A_n)_{n \in \mathbb{N}} \subset \mathcal{A}$, on a $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{A}$.
3. \mathcal{A} est stable par passage au complémentaire, c'est-à-dire que pour tout $A \in \mathcal{A}$, on a $A^c \in \mathcal{A}$.

Exemples de tribus

- L'ensemble de parties $\mathcal{P}(A)$ est une tribu. C'est la tribu qui contient toutes les autres tribus sur A .
- $\mathcal{A} = \{\emptyset, A\}$ est une tribu sur A , elle s'appelle la tribu grossière.

Définition 2 On appelle espace topologique un couple (E, \mathcal{T}) où E est un ensemble et \mathcal{T} une famille de parties de E vérifiant :

1. $\emptyset \in \mathcal{T}$ et $E \in \mathcal{T}$.
2. Une intersection finie d'éléments de \mathcal{T} appartient à \mathcal{T} .
3. Une réunion quelconque d'éléments de \mathcal{T} appartient à \mathcal{T} .

Définition 3 (Tribu borélienne) Soit \mathbb{E} un espace topologique et O la famille des ouverts de \mathbb{E} . La tribu borelienn est également appelée tribu de Borel et on la note $B(\mathbb{E})$ la tribu engendrée par la famille O . En d'autres termes : $B(\mathbb{E}) = \sigma(O)$.

Propriétés 1 La tribu de Borel vérifie les propriétés suivantes :

- Elle contient tous les intervalles ouverts, fermés, demi-ouverts / fermés et infinis.
- Elle contient les union dénombrables de fermées.
- Elle contient les inersctions dénomrable d'ouverts.

1.1.2 Mesure et mesure de lebesgue sur l'ensemble des boréliennes

La mesure d'une partie nous donne une idée sur sa taille. En effet, une mesure naturelle d'un intervalle représente sa longueur, d'un disque représente l'aire et d'une boule donne son volume.

Définition 4 Une mesure sur \mathcal{A} est une application $\mu : \mathcal{A} \rightarrow [0, +\infty[$ vérifiant les propriétés suivantes :

1. $\mu(\emptyset) = 0$.
2. μ est σ -additive, c'est-à-dire que pour toute famille $(A_i)_{i \in I} \subset \mathcal{A}$ de parties disjointes deux à deux, on a :

$$\mu(\cup_{i \in I} A_i) = \sum_{i \in I} \mu(A_i).$$

Remarque 1 Si $\mu(\mathbb{E}) < \infty$ alors, μ est une mesure finie. De plus le triplet $(\mathbb{E}, \mathcal{A}, \mu)$ est appelé un espace mesuré.

Définition 5 La mesure de Lebesgue est une mesure définie sur la tribu borélienne de la droite réelle \mathbb{R} . On la note généralement λ telle que :

$$\lambda([a, b]) = b - a, \forall a, b \in \mathbb{R}, \text{ avec } a < b.$$

Mesure de lebesgue sur \mathbb{R}^n :

On considère a présent l'espace mesurable $(\mathbb{R}^n, B(\mathbb{R}^n))$ c'est a dire l'espace \mathbb{R}^n muni de sa tribu borélienne. On va voir qu'on peut définir la mesure de Lebesgue sur cet espace.

Théorème 1 Il existe une unique mesure sur l'espace $(\mathbb{R}^n, B(\mathbb{R}^n))$, qu'on notera λ_n telle que :

1. $\forall a \in \mathbb{R}^n$ et $\forall A \in B(\mathbb{R}^n)$ on a : $\lambda_n(a + A) = \lambda_n(A)$.
2. $\lambda_n([a, b]^n) = (b - a)^n, \forall a$ et $b \in \mathbb{R}$.

Cette mesure est appelée mesure de lebesgue sur $(\mathbb{R}^n, B(\mathbb{R}^n))$.

Exemple 1 Prenons le cas de \mathbb{R}^2 . La mesure de lebesgue λ_2 sur cet espace coincide sur les rectangles de la forme $[a, b] \times [c, d]$ avec la notion d'aire de ceux-ci. En effet, on prouve qu'on a $\lambda_2([a, b] \times [c, d]) = (b - a)(d - c)$. Plus généralement, la mesure de lebesgue λ_2 d'un sous-ensemble borélien de \mathbb{R}^2 correspond à notre définition intuitive de l'aire où la mesure de lebesgue d'un disque de rayon a est égale à $\pi.a^2$.

Définition 6 On appelle une probabilité sur l'espace $(\mathbb{E}, \mathcal{A})$ toute application P qui associe à chaque évènement de \mathcal{A} un réel de l'intervalle $[0, 1]$. La probabilité $P : \mathcal{A} \rightarrow [0, 1]$ satisfaisant :

1. $0 \leq P(A) \leq 1, \forall A \in \mathcal{A}$ avec $P(A)$ est la probabilité que A se produise.

2. $P(\mathbb{E}) = 1$.
3. $P(\cup_{i \in I} A_i) = \sum_{i \in I} P(A_i), \forall (A_i)_{i \in I} \in \mathcal{A}$.

Le triplet $(\mathbb{E}, \mathcal{A}, P)$ s'appelle un espace probabilisé.

1.2) Loi de probabilité continue

Les lois de probabilité continues sont un outil puissant pour modéliser des phénomènes aléatoires qui prennent des valeurs dans un intervalle continu. Elles ont de nombreuses applications dans divers domaines et sont essentielles pour comprendre le monde qui nous entoure.

1.2.1 Variable aléatoire continue

Une variable aléatoire est dite continue si elle peut prendre toutes les valeurs dans un intervalle donné. Cet intervalle peut être borné (avoir un début et une fin) ou non borné (s'étendre vers l'infini positif ou négatif). Par exemple, la masse corporelle des individus pour une espèce animale donnée et le taux de glucose dans le sang.

Définition 7 *La fonction de répartition F_X , d'une variable aléatoire X continue (au sens probabiliste) est continue (au sens analytique) et dérivable par morceaux. Elle est définie par :*

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

Sa dérivée f_X est appelée densité de probabilité de X :

$$f_X(x) = F'_X(x).$$

Propriétés 2

- $\forall x \in \mathbb{R}, f(x) \geq 0$.
- $\int_{-\infty}^{+\infty} f(x) dx = 1$
- $\int_{-\infty}^a f(x) dx = \mathbb{P}(X \leq a) = F(a)$.

Définition 8 La fonction de densité jointe d'un couple (X, Y) de variable aléatoires notée $f_{X,Y}$ vérifie les conditions suivantes :

- $\forall x, y \in \mathbb{R}, f_{X,Y}(x, y) \geq 0.$
- $\forall x, y \in \mathbb{R}, \int_{\mathbb{R}} \int_{\mathbb{R}} f_{X,Y}(x, y) dx dy = 1.$

1.2.2 Fonction des moments

Si X est une variable aléatoire, on appelle moment d'ordre n , s'il existe, le réel $E(X^n)$. Par exemple, l'espérance d'une variable aléatoire est son moment d'ordre 1.

Définition 9 Si X est une variable aléatoire admettant une densité f_X son moment d'ordre n se définit par la formule :

$$E(X^n) = \int_{\mathbb{R}} x^n f_X(x) dx.$$

En particulier, on trouve les deux définitions suivantes :

Définition 10 Si X est une variable aléatoire absolument continue de densité f_X , l'espérance de X notée $E[X]$ est définie par :

$$E(X) = \int_{\mathbb{R}} x f_X(x) dx.$$

Définition 11 Si X est une variable aléatoire absolument continue, la variance de X notée $Var(X)$ ou σ^2 est l'espérance des carrés des écarts par rapport à l'espérance est définie par :

$$\begin{aligned} Var(X) &= E(X - E(X))^2 \\ &= E(X^2) - E^2(X). \end{aligned}$$

Définition 12 Considérons une paire de variable aléatoire X et Y . Les moments joints des variables X et Y sont donnés par :

$$E\{X^i Y^k\} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x^i y^k f_{X,Y}(x, y) dx dy$$

où i et k sont deux entiers positifs. Le moment le plus important correspond à $i = k = 1$.

1.2.3 Exemple d'une loi continue fréquemment utilisée

Il existe de nombreuses lois de probabilité, chacune avec ses propres caractéristiques et applications. Parmi les plus courantes, on a la loi normale. Une loi normale permet d'approcher d'autres lois et ainsi de modéliser de nombreuses études scientifiques.

Définition 13 *On dit qu'une variable aléatoire X suit la loi normale ou la loi gaussienne de paramètres $m \in \mathbb{R}$ et σ^2 , avec $\sigma > 0$, ce que l'on note $X \hookrightarrow \mathcal{N}(m, \sigma^2)$ si elle est continue et admet pour densité :*

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right).$$

Remarque 2 *Si X suit une loi $\mathcal{N}(m, \sigma^2)$, alors X admet une espérance $m = E(X)$ et une variance $\sigma^2 = \text{Var}(X)$. De plus la variable $Z = (X - m)/\sigma$ suit une loi normale dite centrée réduite $\mathcal{N}(0, 1)$. La courbe de la densité de loi normale centrée réduite est représentée par la figure (1.1) :*

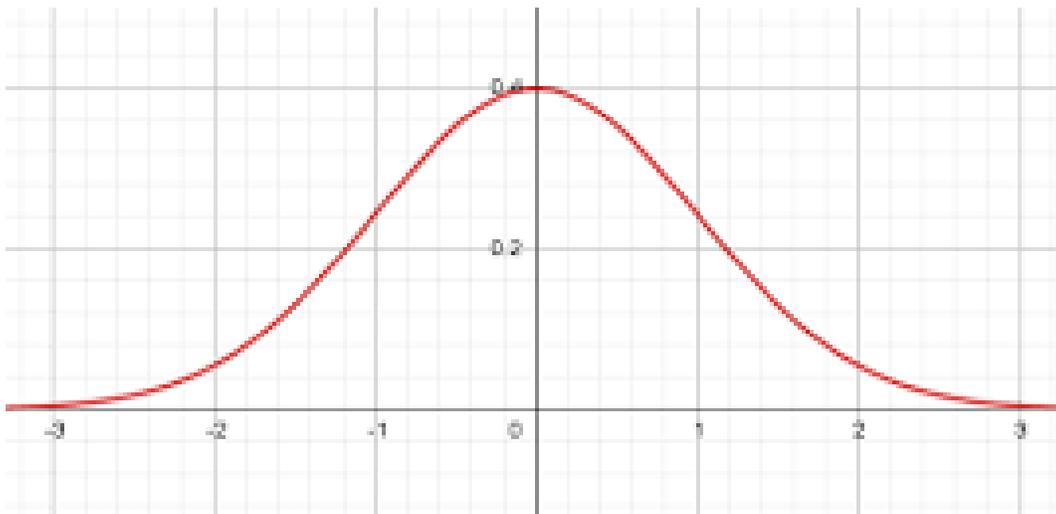


FIGURE 1.1 – Densité de la loi normale centrée réduite.

1.3) Indépendance et conditionnement

Nous présentons maintenant deux concepts essentiels de la théorie des probabilités. Le premier est le conditionnement qui permet

d'inclure une information supplémentaire dans le calcul d'une probabilité. Le second est l'indépendance qui explique que deux événements n'ont pas d'influence mutuelle, et donc que l'on peut évaluer la probabilité du premier sans tenir compte de l'existence ou non du second.

Définition 14 Soit un espace probabilisé (Ω, \mathcal{A}, P) et deux événements A et B appartenant à \mathcal{A} . On dit que deux événements A et B sont indépendants, et on note $A \perp B$, si :

$$P(A \cap B) = P(A)P(B).$$

Définition 15 On appelle probabilité conditionnelle de A sachant B , la probabilité que l'événement A se réalise sachant que l'événement B est réalisé. On la note : $P_B(A)$ ou $P(A \setminus B)$ et elle est définie par :

$$P(A \setminus B) = \frac{P(A \cap B)}{P(B)}, P(B) \neq 0.$$

Définition 16 On considère deux variable aléatoires X et Y , prenant respectivement les valeurs x_1, x_2, \dots, x_n et y_1, y_2, \dots, y_p . Dire que le couple (X, Y) est un couple de variables aléatoires indépendantes signifie que, pour tout entier i compris entre 1 et n et tout entier j compris entre 1 et p , les événements $(X = x_i)$ et $(Y = y_j)$ sont indépendants.

2

Mélange fini de lois de probabilités

Le mélange de lois de probabilités est une méthode efficace pour créer de nouvelles lois qui sont plus souples et adaptées à une variété de données. En particulier, le mélange de lois est un outil statistique puissant souvent utilisé pour modéliser des situations où les données proviennent de plusieurs populations distinctes, appelées composantes.

2.1) Présentation générale de modèle de mélange fini

Le mélange fini de loi joue un rôle essentiel dans la mise en place de nouvelles lois de probabilités, et ce pour diverses raisons. Dans cette section, nous donnons la définition formelle du mélange fini de lois de probabilités (voir [4]).

Définition 17 Soit $X = (X_1, \dots, X_n)$ un vecteur de n variables aléatoires et $f_{X_1}, f_{X_2}, \dots, f_{X_n}$ des fonctions de densité de probabilité associées à ses variables aléatoires. Un mélange fini de ces lois est une loi de probabilité f_X définie par :

$$f_X(x) = \sum_{k=1}^n \pi_k f_{X_k}(x). \quad (2.1)$$

où les f_{X_k} sont appelées les composantes du mélange et les π_k appelées proportions du mélange, avec $0 \leq \pi_k \leq 1$ et $\sum_1^n \pi_k = 1$.

Propriété 1 Le support de la loi de mélange est l'union des supports des lois de ses composantes.

2.2) Estimation de paramètres de la loi mélange

Cette section vise à présenter deux méthodes classiques utilisées dans l'estimation des paramètres du modèle mélange (voir [5]).

2.2.1 Estimation par la méthode du maximum de vraisemblance

Le principe d'estimation par maximum de vraisemblance consiste à affirmer que plus la probabilité d'obtenir les observations est élevée, plus le modèle est proche de la réalité. Donc, nous retenons le modèle pour lequel la probabilité de notre échantillon est la plus élevée :

$$\hat{\theta}^{MV} = \arg \max l(x_1, x_2, \dots, x_n; \theta),$$

avec $l(x_1, x_2, \dots, x_n; \theta)$ est la fonction de vraisemblance défini par :

$$l(x_1, x_2, \dots, x_n; \theta) = \prod_{i=1}^n f(x_i).$$

Dans la pratique, il est possible de résoudre directement un problème en raison de la présence du produit, mais il est nécessaire de prendre en compte le logarithme de la vraisemblance :

$$\hat{\theta}^{MV} = \arg \max L(x_1, x_2, \dots, x_n; \theta),$$

où

$$\begin{aligned} L(x_1, x_2, \dots, x_n; \theta) &= \ln l(x_1, x_2, \dots, x_n; \theta) \\ &= \ln \prod_{i=1}^n f(x_i) \\ &= \sum_{i=1}^n \ln f(x_i). \end{aligned}$$

Dans le but d'atteindre le maximum, on résoud l'équation du premier ordre :

$$\frac{\partial L(x_1, x_2, \dots, x_n; \theta)}{\partial \theta} \Big|_{\theta = \hat{\theta}^{MV}} = 0.$$

Dans certains cas, l'équation au premier ordre admet une solution explicite. Cependant, le cas le plus fréquent en pratique est celui où elle doit être résolue numériquement à l'aide de l'algorithme de Newton-Raphson.

Algorithme de Newton-Raphson

La méthode de Newton-Raphson est une méthode algorithmique qui permet de trouver la racine d'une fonction. C'est-à-dire trouver x tel que

$$f(x) = 0.$$

L'algorithme de Newton pour une fonction f consiste alors à générer une suite de paramètres $\{\theta^{(k)}\}$, où le nouvel itéré $\theta^{(k+1)}$ est calculé à partir de l'itéré courant $\theta^{(k)}$ par la récurrence suivante :

$$\theta^{(k+1)} = \theta^{(k)} - J^{-1}(\theta^{(k)})f(\theta^{(k)}),$$

où J est la matrice jacobienne définie par :

$$J = \begin{pmatrix} \frac{\partial f_1}{\partial \theta_1} & \cdots & \frac{\partial f_1}{\partial \theta_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial \theta_1} & \cdots & \frac{\partial f_n}{\partial \theta_n} \end{pmatrix}$$

$$\text{avec } f = \begin{pmatrix} f_1 \\ f_2 \\ \dots \\ f_n \end{pmatrix} \text{ et } \theta = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \dots \\ \theta_n \end{pmatrix}.$$

En particulier, une extension d'algorithme de Newton-Raphson est utilisée. Cette extension s'appelle l'algorithme de Fisher scoring et elle est souvent utilisée en statistique pour résoudre des équations du maximum de vraisemblance. L'algorithme est définie par

la relation suivante :

$$\theta^{(k+1)} = \theta^{(k)} + I^{-1}(\theta^{(k)})V(\theta^{(k)}),$$

où V est la fonction score et I est l'information de Fisher définies respectivement par :

$$V(\theta^{(k)}) = \begin{pmatrix} \frac{\partial L}{\partial \theta_1} \\ \vdots \\ \frac{\partial L}{\partial \theta_k} \end{pmatrix}$$

et

$$I = \begin{pmatrix} -E \left[\frac{\partial^2 \ln f(X; \theta)}{\partial \theta_1^2} \right] & \dots & -E \left[\frac{\partial^2 \ln f(X; \theta)}{\partial \theta_1 \theta_k} \right] \\ \vdots & \ddots & \vdots \\ -E \left[\frac{\partial^2 \ln f(X; \theta)}{\partial \theta_1 \theta_k} \right] & \dots & -E \left[\frac{\partial^2 f(X; \theta)}{\partial \theta_k^2} \right] \end{pmatrix}.$$

2.2.2 Algorithme d'expectation-maximisation (EM)

Algorithme expectation-maximisation, ou l'algorithme EM est une technique statistique utilisée pour estimer les paramètres d'un modèle probabiliste lorsque certaines variables du modèle sont cachées ou latentes, c'est-à-dire non directement observables dans les données. Cette méthode d'estimation est aussi importante car il permet d'analyser des données du monde réel souvent incomplètes. L'algorithme EM est un processus itératif qui se déroule en deux étapes à chaque itération :

- Etape E (Expectation) : À l'étape d'attente, on calcule l'espérance de la fonction de vraisemblance complète (tenant compte des variables cachées) conditionnellement aux données observées et aux

paramètres courants.

- Etape M (Maximisation) : À l'étape de maximisation, on maximise la fonction attendue calculée à l'étape E par rapport aux paramètres du modèle.

Ces deux étapes s'enchaînent jusqu'à ce que les paramètres du modèle convergent vers des valeurs stables.

2.2.3 Test de Kolmogorov-Smirnov (K-S)

Le test de Kolmogorov-Smirnov est un test qui compare la distribution observée d'un échantillon statistique à une distribution théorique. Autrement dit, il est utilisé pour tester l'adéquation d'un échantillon à une loi de probabilité spécifique. Le test K-S est aussi utilisé pour comparer la distribution de deux échantillons statistiques où il est basé sur la comparaison de leurs fonctions de répartition. On a alors :

- Données : p observations d'une variable aléatoire X et q observations d'une variable aléatoire Y .
- Hypothèse testée : "Les fonctions de répartition de X et de Y , notées respectivement F_X et F_Y sont égales" avec risque d'erreur α .
- Déroulement du test :
 1. On calcule $K = \sup |F_X(x) - F_Y(x)|$.
 2. On compare avec la valeur critique Δ de la loi de Kolmogorov-Smirnov, telle que : si $P(\Delta > b) = \alpha$ avec $\Delta = \sqrt{\frac{p+q}{p \times q}} b$ et $K < \Delta b$, alors on accepte l'hypothèse.

Dans la section suivante, nous allons proposer un exemple de modèle de mélange fréquemment utilisé [10].

2.3) Modèle de mélange gaussien

Un modèle de mélange gaussien (MMG) est une combinaison linéaire de plusieurs composantes gaussienne chacune caractérisée par ses propres paramètres : moyenne et covariance. Il est largement utilisé dans la classification de données, où chaque composante représente une classe différente ou bien pour modéliser des distributions de données complexes qui ne peuvent pas être représentées par une seule distribution gaussienne. La contribution relative de chaque distribution gaussienne au modèle global est pondérée par des coefficients appelés poids de mélange.

2.3.1 Mélange de gaussiennes multinomiales

Le mélange de gaussiennes multinomiales (MGM) est une technique statistique fondamentale en classification automatique. Il permet de modéliser la distribution des données dans un espace multidimensionnel en utilisant plusieurs distributions gaussiennes superposées.

Un mélange gaussien multinomial se caractérise par un ensemble de n distributions gaussiennes, chacune avec un vecteur de moyenne μ_k et une matrice de covariance Σ_k , où n est le nombre de composants du mélange et $k = 1, \dots, n$. On désigne par π_k la probabilité qu'un point de données x soit issu de la composante k , où $0 < \pi_k < 1$ et la somme des π_k est égale à 1.

Définition 18 *La densité de probabilité $f_X(x)$ d'un mélange gaus-*

sien multinomial est donnée par :

$$f_X(x) = \sum_{k=1}^n \pi_k N(x|\mu_k, \gamma_k),$$

où $N(x|\mu_k, \gamma_k)$ est la densité de probabilité de la k -ème distribution gaussienne.

Exemple 2 *La figure (2.1) illustre un mélange de trois gaussiens bidimensionnels ($n=3$) :*

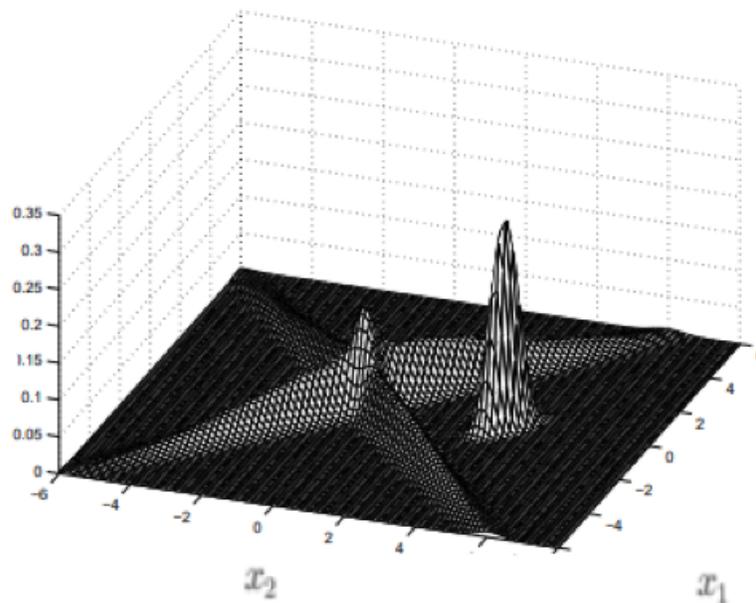


FIGURE 2.1 – Un mélange gaussien dans \mathbb{R}^2 avec trois composantes.

Remarque 3 *Dans la plupart des cas, on suppose également que chaque composante f_k fait partie d'une famille paramétrée $f(.;\theta_k)$ et on note $f(.;\Theta)$ la loi mélange associée à ce paramétrage, $\Theta = (\theta_1, \dots, \theta_k)$ désignant le paramètre de ce modèle. On note qu'il est possible de choisir seulement des familles $f(.;\theta_k)$ qui entraînent des lois mélanges identifiables, ou tout au moins qui le sont dans la plupart des situations d'intérêt (voir [9]).*

2.4) Choix entre modèle de mélange

Le BIC et l'AIC sont des critères de vraisemblance pénalisée couramment utilisés, ces critères sont essentiels pour choisir le nombre optimal de composants dans un modèle de mélange (voir [7] et [2]).

2.4.1 Critère BIC (Bayesian Information Criterion)

Comme mentionné précédemment, le BIC est largement utilisé pour comparer différents modèles de mélange. Il prend en compte à la fois la précision du modèle et sa complexité. Il est défini par

$$BIC = -2\log\hat{L} + k\log(n),$$

où n est la taille de l'échantillon, k est le nombre de paramètres à estimer et L la fonction de vraisemblance. Un modèle avec un BIC plus bas est généralement préféré.

2.4.2 Critère AIC (Akaike Information Criterion)

Similaire au BIC, l'AIC évalue la qualité du modèle en tenant compte à la fois de l'ajustement aux données et de la complexité du modèle. Il est défini par :

$$AIC = -2\log\hat{L} + 2k.$$

Comme pour le BIC, un modèle avec un AIC plus bas est préféré.

3

Sur La distribution Gamma-Lindley généralisée

Les distributions de Lindley peuvent prendre une variété de formes, allant des unimodales symétriques aux asymétriques avec des taux de défaillance croissants ou décroissants. Cette flexibilité les rendent idéales pour modéliser un large éventail de types de données relatives aux nombreux domaines scientifiques et pratiques, telles que les revenus, les tailles d'objets, les volumes d'eau et les durées de vie. La distribution de Lindley à un seul paramètre définie par la fonction de densité :

$$f(x) = \frac{\theta^2(1+x)e^{-\theta x}}{1+\theta}; x, \theta > 0, \quad (3.1)$$

a été introduite initialement par Lindley [10] lui même en 1958 pour montrer comment la distribution fiduciaire et la distribution a posteriori diffèrent l'une de l'autre. Sankaran [12] a introduit la distribution de Poisson-Lindley en la combinant avec un paramètre de Poisson. Gintany et al. [8] ont examiné plusieurs caractéristiques de (3.1). Lorsqu'il s'agit d'analyser divers types de données sur la durée de vie, l'unique paramètre de distribution Lindley limite sa flexibilité. Afin d'augmenter la flexibilité à des fins de modélisation, il sera utile de prendre en compte davantage de variations de cette distribution.

3.1) Distribution Gamma

Définition 19 On dit qu'une variable aléatoire X suit la distribution gamma de deux paramètres $\alpha, \beta > 0$ notée $\text{Gamma}(\alpha, \beta)$ si elle possède la densité de probabilité :

$$f_X(x) = \frac{x^{\alpha-1} e^{-\frac{x}{\beta}}}{\Gamma(\alpha) \beta^\alpha}, \quad (3.2)$$

avec $\Gamma(\cdot)$ est la fonction gamma d'Euler définie par :

$$\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} e^{-x} dx; \quad \alpha > 0,$$

où α est le paramètre de la forme et β est le paramètre d'échelle.

Remarque 4 Un α plus élevé donne une distribution plus aplatie, tandis qu'un α plus petit donne une distribution plus pointue. Un β plus élevé dilate la distribution, tandis qu'un β plus petit la contracte.

Alternativement, la distribution gamma peut être paramétrée à l'aide d'un paramètre d'intensité $\theta = 1/\beta$. Sa densité devient alors :

$$f_X(x) = \frac{\theta^\alpha x^{\alpha-1} e^{-\theta x}}{\Gamma(\alpha)}; \quad x, \theta, \alpha > 0. \quad (3.3)$$

Corollaire 1 Soient X_1 et X_2 deux variables aléatoires indépendantes et distribuées selon $\text{Gamma}(\alpha, \theta)$. La densité jointe de couple (X_1, X_2) est définie par :

$$f_{X_1, X_2}(x_1, x_2) = \frac{\theta^{2\alpha} (x_1 x_2)^{\alpha-1} e^{-\theta(x_1+x_2)}}{\Gamma^2(\alpha)}; \quad x, \theta, \alpha > 0. \quad (3.4)$$

Preuve 1 La fonction de densité de probabilité de X_1 est donnée par (3.3), et la fonction de densité de probabilité de X_2 est donnée par :

$$f_{X_2}(x_2) = \frac{\theta^\alpha (x_2)^{\alpha-1} e^{-\theta x_2}}{\Gamma(\alpha)}.$$

Comme X_1 et X_2 sont indépendantes, alors :

$$\begin{aligned} f_{X_1, X_2}(x_1, x_2) &= f_{X_1}(x_1)f_{X_2}(x_2) \\ &= \frac{\theta^{2\alpha}(x_1x_2)^{\alpha-1}e^{-\theta(x_1+x_2)}}{\Gamma^2(\alpha)}. \end{aligned}$$

3.2) Distribution de Lindley généralisée

Un grand nombre de distributions à deux variables ont été proposées dans la littérature. Dans cette section, nous présentons une famille de distributions bivariée dont les distributions marginales sont aussi des extensions de distributions de Lindley généralisées.

3.2.1 Version bivariée de la distribution de Lindley généralisée

Dans cette partie, on va utiliser la stratégie de mélange de lois pour construire une version bidimensionnelle de la distribution de Lindley generalisée.

Construction 1 Soit (V_1, V_2) et (W_1, W_2) deux vecteurs de variables aléatoires indépendantes, sachant que : (V_1, V_2) distribuées selon (3.4) et (W_1, W_2) distribuées selon le loi Gamma $(\alpha + 1, \theta)$. Pour $\gamma > 0$, On considère le couple aléatoires (X_1, X_2) tel que :

$$(X_1, X_2) = \begin{cases} (V_1, V_2) \text{ avec la probabilité } \theta/(\theta + \gamma); \\ (W_1, W_2) \text{ avec la probabilité } \gamma/(\theta + \gamma). \end{cases}$$

La fonction de densité jointe de couple (X_1, X_2) est donnée par :

$$f_{X_1, X_2}(x_1, x_2) = \frac{\theta}{\gamma + \theta} f_{V_1, V_2}(x_1, x_2) + \frac{\gamma}{\gamma + \theta} f_{W_1, W_2}(x_1, x_2), \quad (3.5)$$

avec :

$$f_{W_1, W_2}(x_1, x_2) = \frac{x_1^\alpha \theta^{\alpha+1} e^{-\theta x_1}}{\Gamma(\alpha + 1)} \frac{x_2^\alpha \theta^{\alpha+1} e^{-\theta x_2}}{\Gamma(\alpha + 1)}. \quad (3.6)$$

Donc, on trouve que :

$$f_{X_1, X_2}(x_1, x_2) = \frac{\theta^{\alpha+2} \theta x_1 x_2^{\alpha-1} (\alpha^2 + \gamma \theta x_1 x_2) e^{-\theta(x_1 x_2)}}{(\gamma + \theta) \Gamma^2(\alpha + 1)}; \alpha, \theta, x_1, x_2 > 0. \quad (3.7)$$

Remarque 5 Il convient de noter que la fonction de densité jointe (3.7) peut être écrite en termes de fonctions de densité gamma comme suit :

$$f_{X_1 X_2}(x_1, x_2) = \frac{\theta(\alpha^2 + \gamma \theta x_1 x_2)}{\alpha^2(\gamma + \theta)} f_g(x_1; \alpha, \theta) f_g(x_2; \alpha, \theta). \quad (3.8)$$

En particulier si $\gamma = 0$, les variables aléatoires X_1 et X_2 deviennent indépendantes et la fonction de densité bivariée (3.8) se réduit au produit de deux fonctions de densité gamma avec les mêmes paramètres.

Proposition 1 Pour chaque entier positif m et n , on peut obtenir l'expression suivante pour les moments :

$$E(X_1^n X_2^m) = \frac{\Gamma(\alpha + n) \Gamma(\alpha + m)}{\theta^{n+m} (\gamma + \theta) \Gamma^2(\alpha + 1)} (\alpha^2(\theta + \gamma) + \gamma(\alpha + m)(\alpha + n)).$$

Corollaire 2 Le coefficient de corrélation de X_1 et X_2 est donné par :

$$\text{Corr}(X_1, X_2) = \frac{\gamma \theta}{\alpha(\theta + \gamma)^2 + \gamma(2\theta + \gamma)}.$$

En particulier, lorsque $\gamma = 0$ on a $\text{Corr}(X_1, X_2) = 0$ et lorsque $\alpha \rightarrow 0$ et $\theta \rightarrow \infty$, on trouve que $\text{Corr}(X_1, X_2) \rightarrow \frac{1}{2}$.

Dans ce qui suit, nous définissons la propriété de dépendance de (3.7).

Définition 20 Une distribution bivariée est dite positivement dépendante du rapport de vraisemblance (PDRV) si la fonction de densité $f_{X,Y}(x, y)$ satisfait aux conditions suivantes :

$$f(x_1, y_1) f(x_2, y_2) > f(x_1, y_2) f(x_2, y_1). \quad (3.9)$$

Corollaire 3 Pour tous les $x_1 > x_2$ et $y_1 > y_2$. La fonction de densité bivariée (3.7) satisfait l'inégalité (3.9) qui est équivalente à :

$$(\alpha^2 + \theta\gamma x_1 y_1)(\alpha^2 + \theta\gamma x_2 y_2) \geq (\alpha^2 + \theta\gamma x_1 y_2)(\alpha^2 + \theta\gamma x_2 y_1),$$

où $(x_1 - x_2)(y_1 - y_2) > 0$. Donc, le couple (X, Y) est positivement dépendant de rapport de vraisemblance.

3.2.2 Version univariée de la distribution de Lindley généralisée

La version univariée de Lindley généralisée est une distribution de probabilité continue qui peut être utilisée pour modéliser une variété de phénomènes, tels que la durée de vie d'un produit, le temps de réponse d'un système ou la distance parcourue par une particule.

Définition 21 On dit qu'une variable aléatoire X suit la distribution univariée de Lindley généralisée, si elle possède la densité de probabilité suivante :

$$f_X(x) = \frac{\theta^2(\theta x)^{\alpha-1}(\alpha + \gamma x)e^{-\theta x}}{(\gamma + \theta x)\Gamma(\alpha + 1)}; \alpha, \beta, \theta > 0. \quad (3.10)$$

Cette distribution marginale peut être construite à partir d'un mélange de deux distributions gamma. En effet, on considère que :

$$X = \begin{cases} V_1 \text{ avec probabilité } \theta/(\theta + \gamma), \\ V_2 \text{ avec probabilité } \gamma/(\theta + \gamma). \end{cases}$$

En appliquant la relation du mélange, on trouve que :

$$f_X(x) = \frac{\theta}{\theta + \gamma} \left(\frac{\theta^\alpha x^{\alpha-1} e^{-\theta x}}{\Gamma(\alpha)} \right) + \frac{\gamma}{\theta + \gamma} \left(\frac{\theta^{\alpha+1} x^\alpha e^{-\theta x}}{\Gamma(\alpha + 1)} \right), \quad (3.11)$$

où

$$\frac{\theta}{\theta + \gamma} + \frac{\gamma}{\theta + \gamma} = 1.$$

Après simplification, on trouve que la densité marginale de la distribution univariée de Lindley généralisée est de la forme (3.10) .

3.2.3 Forme de la distribution univariée de Lindley généralisée

Les dérivées premières et secondes de la fonction $\log f_X(x)$ qui est définie par :

$$\log f_X(x) = \log\left(\frac{\theta^2(\theta x)^{\alpha-1}(\alpha + \gamma x)e^{-\theta x}}{(\gamma + \theta x)\Gamma(\alpha + 1)}\right), \quad (3.12)$$

sont les suivantes :

$$\begin{aligned} \frac{d}{dx} \log f_X(x) &= \frac{2\theta x^{\alpha-2}}{x} + (\alpha - 1)\frac{1}{x} + \frac{\alpha + \gamma x}{x} - \theta - \frac{(\gamma + \theta)\Gamma(\alpha + 1)}{\Gamma(\alpha + 1)x}, \\ &= \frac{(\alpha - 1)(\alpha + \gamma x) + (\gamma - \theta(\alpha + \gamma x))x}{x(\alpha + \gamma x)} \end{aligned}$$

et :

$$\frac{d^2}{dx^2} \log f_X(x) = \frac{(1 - \alpha)(\alpha + \gamma x)^2 - (\gamma x)^2}{x^2(\alpha + \gamma x)^2}.$$

Si $\alpha \geq 1$, alors $\frac{d^2}{dx^2} \log f_X(x) \leq 0$; c'est-à-dire que la fonction de densité $f_X(x)$ est log coconcave.

Distribution de la somme :

Il est bien connu que la distribution d'une somme de variables aléatoires gamma indépendantes ayant le même paramètre d'échelle est à nouveau une distribution gamma. Le résultat suivant montre que la distribution d'une somme de variables aléatoires indépendantes de la distribution LG peut être écrite comme un mélange des distributions gamma.

Proposition 2 Soit X_1, \dots, X_n des variables aléatoires indépendantes issues de la distribution LG avec les paramètres θ, γ et α_i pour $i = 1, \dots, n$. Alors la fonction de densité de $S = \sum_{i=1}^n X_i$ est donnée par :

$$f_S(x) = \sum_{k=0}^n \pi_k f_g(x; \alpha^* + k, \theta), \quad (3.13)$$

telles que f_g est la fonction de densité de la distribution gamma et $\alpha^* = \sum_{i=1}^n \alpha_i$.

Estimation par la méthode du maximum de vraisemblance

Dans cette partie, nous considérons l'estimation du maximum de vraisemblance (EMV) des paramètres. Si X_1, \dots, X_n est un échantillon aléatoire distribué selon la distribution LG, alors la fonction de log-vraisemblance L est :

$$L(\alpha, \theta, \gamma) = n(\alpha + 1)\log(\theta) - n\log(\theta + \gamma) - n\log\Gamma(\alpha + 1) \\ + (\alpha - 1)\sum_{i=1}^n \log(x_i) + \sum_{i=1}^n \log(\alpha + \gamma x_i) - \theta \sum_{i=1}^n (x_i).$$

Les dérivées de L par rapport aux paramètres α, θ et γ sont :

$$\frac{dL}{d\alpha} = n\log(\theta) - n\Psi(\alpha + 1) + \sum_{i=1}^n \log(x_i) + \sum_{i=1}^n \frac{1}{\alpha + \gamma x_i}, \quad (3.14)$$

$$\frac{dl}{d\theta} = \frac{n(\alpha + 1)}{\theta} - \frac{n}{\gamma + \theta} - \sum_{i=1}^n (x_i) \quad (3.15)$$

et

$$\frac{dl}{d\gamma} = \sum_{i=1}^n \frac{(x_i)}{\alpha + \gamma(x_i)} - \frac{n}{\gamma + \theta}, \quad (3.16)$$

où $\Psi(t) = \Gamma'(t)/\Gamma(t)$ représente la fonction digamma. Les équations (3.14), (3.15) et (3.16) peuvent être résolues simultanément pour trouver les estimateurs du maximum de vraisemblance de α, γ et θ .

La matrice de variance-covariance (I^{-1}), où I est la matrice d'information de Fisher est définie par les éléments suivants :

$$-E\left(\frac{d^2L}{d\alpha^2}\right) = n\Psi(1 + \alpha) + nj_0(\alpha, \theta, \gamma),$$

$$-E\left(\frac{d^2L}{d\alpha d\theta}\right) = -\frac{n}{\theta},$$

$$\begin{aligned}
-E\left(\frac{d^2l}{d\alpha d\gamma}\right) &= nj_1(\alpha, \theta, \gamma) \\
-E\left(\frac{d^2L}{d\theta^2}\right) &= \frac{n(\alpha + 1)}{\theta^2} - \frac{n^2}{\gamma + \theta}, \\
-E\left(\frac{d^2L}{d\theta d\gamma}\right) &= -\frac{n}{(\gamma + \theta)^2}
\end{aligned}$$

et

$$-E\left(\frac{d^2L}{d\gamma^2}\right) = nj_2(\alpha, \theta, \gamma) - \frac{n}{(\gamma + \theta)^2}.$$

Ici pour $i = 0, 1, 2$,

$$\begin{aligned}
j_i &= E\left(\frac{X^i}{(\alpha + \gamma X)^2}\right) \\
&= \frac{\theta^{1-i}}{\alpha\Gamma(\alpha + 1)(\gamma + \theta)} \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \left(\frac{\gamma}{\alpha\theta}\right)^k \Gamma(\alpha + k + i).
\end{aligned}$$

Dans ce qui suit, nous donnons des exemples numériques pour montrer la flexibilité de la distribution LG pour la modélisation des données.

3.2.4 Application réelle

Pour une application réelle de la distribution univariée de Lindley généralisée, nous considérons deux séries de données réelles tirées de Lawless (2003). Le premier ensemble de données représente les temps de défaillance (en minutes) d'un échantillon de 15 composants électroniques. Ces temps sont les suivants : 1.4, 5.1, 6.3, 10.8, 12.1, 18.5, 19.7, 22.2, 23, 30.6, 37.3, 46.3, 53.9, 59.8, 66.2. Le deuxième ensemble de données est le nombre de cycles jusqu'à la rupture pour 25 échantillons de 100 cm de fil, testés à un niveau de déformation particulier : 15, 20, 38, 42, 61, 76, 86, 98, 121, 146, 149, 157, 175,

176, 180, 180, 198, 220, 224, 251, 264, 282, 321, 325, 653.

En manipulant l'algorithme de Fisher Scoring sur le logiciel, on trouve les estimateurs de maximum de vraisemblance et la statistique de Kolmogorov-Smirnov comme présentés dans le tableau (3.1).

TABLE 3.1 – Estimations et statistique de Kolmogorov-Smirnov .

	distribution	$\hat{\alpha}$	$\hat{\theta}$	$\hat{\gamma}$	K-S
n=15	LG	1.203	0.064	0.083	0.095
	Gamma	1.442	0.052	/	0.100
n=25	LG	1.505	0.012	0.018	0.137
	Gamma	1.794	0.010	/	0.135

D'après le tableau (3.1), les distributions Gamma et Lindley généralisée ont des paramètres différents et des ajustements légèrement différents. Selon la statistique de Kolmogorov-Smirnov (K-S), pour $n = 5$ la distribution de Lindley généralisée semble mieux s'ajuster, tandis que pour $n = 25$ c'est la distribution Gamma qui donne un meilleur ajustement pour les données.

Remarque 6 Une petite valeur de la statistique K-S entraîne l'acceptation de l'hypothèse nulle H_0 .

3.3) Distribution bivariée Gamma-Lindley généralisée

Dans cette partie, On va étudier une classe de distributions bivariées appelée distribution Gamma-Lindley généralisée bivariée. La fonction densité de probabilité jointe de cette nouvelle distribution correspond à :

$$f_{X_1, X_2}(x_1, x_2) = \frac{\theta^{2\alpha+1}(x_1 x_2)^{\alpha-1} e^{-\theta(x_1+x_2)}}{\beta(\gamma + \theta)\Gamma^2(\alpha + 1)} (\alpha^2 + \theta(\beta\gamma + \beta\theta - \theta)x_1 + x_2). \quad (3.17)$$

Cette distribution bivariée est essentiellement basée sur un mélange

du produit de deux variables aléatoires indépendantes distribué selon la distribution gamma et de la distribution bivariée de Lindley généralisée (notée *BLG*, introduite par Zakerzadeh et Dolati (voir [14])). Les deux distributions sont définies respectivement par les fonctions de densité de probabilité jointe :

$$f_{gg}(x_1, x_2) = \frac{\theta^{2\alpha}(x_1x_2)^{\alpha-1}}{\Gamma^2(\alpha)} e^{-\theta(x_1+x_2)}; \quad x_1, x_2, \alpha, \theta > 0 \quad (3.18)$$

et :

$$f_{BLG}(x_1, x_2) = \frac{\theta^{2\alpha+1}(x_1x_2)^{\alpha-1}(\alpha^2 + \gamma\theta x_1x_2)}{(\gamma + \theta)\Gamma^2(\alpha + 1)} e^{-\theta(x_1+x_2)}; \quad x_1, x_2, \alpha, \theta > 0 \quad (3.19)$$

où $x_1, x_2, \alpha, \theta > 0, \gamma \geq 0$.

Construction 2 Soit (Y_1, Y_2) et (Z_1, Z_2) deux vecteurs aléatoires, tels que Y_1 et Y_2 sont des variables aléatoires indépendantes distribuées selon Gamma $(\alpha + 1, \theta)$ et que (Z_1, Z_2) est distribué selon *BLG* (α, θ, γ) . Pour $\beta \geq 1$, considérons la paire aléatoire (X_1, X_2) qui représente le mélange entre (Y_1, Y_2) avec la probabilité $(\beta - 1)/\beta$, et (Z_1, Z_2) avec la probabilité $1/\beta$. Cette loi de mélange est appelée distribution *BGLG* et la fonction de densité jointe de (X_1, X_2) est obtenue en utilisant la loi de probabilité totale :

$$\begin{aligned} f_{X_1, X_2}(x_1, x_2) &= \frac{(\beta - 1)}{\beta} f_{gg}(x_1, x_2; \alpha + 1, \theta) + \frac{1}{\beta} f_{BLG}(x_1, x_2; \alpha, \theta, \gamma) \\ &= \frac{\theta^{2\alpha+1}(x_1x_2)^{\alpha-1} e^{-\theta(x_1+x_2)}}{\beta(\gamma + \theta)\Gamma^2(\alpha + 1)} (\alpha^2 + \theta(\beta\gamma + \beta\theta - \theta)x_1x_2), \end{aligned} \quad (3.20)$$

où $x_1, x_2, \alpha, \theta > 0, \gamma \geq 0$ et $\Theta = (\alpha, \beta, \gamma, \theta)$.

3.3.1 Forme de la distribution bivariée de gamma-lindley généralisée

la distribution bivariée de gamma lindley généralisée *BGLG* s'écrit par la densité (3.20) présente de différentes formes qui dépendent

des paramètres α, β, γ et θ . Le code MATLAB (R2015a) ci-dessous généré quatre graphes de la distribution BGLG pour quelques valeurs α, β, γ et θ .

```

1 -   clc; clear all; figure;
2 -
3 -   x1=0:0.2:10;
4 -   x2=0:0.2:10;
5 -   [x11,x22]=meshgrid(x1,x2);
6 -
7 -   alpha=[ 2 2 7 7];
8 -   beta=[1.2 1.2 1.2 1.2];
9 -   gamma=[1.5 1.5 1.5 1.5];
10 -  theta=[0.3 2 0.3 3];
11 -
12 -  for i=1:length(a)
13 -
14 -      y=(t(i).^(2*a(i)+1).*(x11.*x22).^(a(i)-1).*exp(-t(i).*(x11+x22)))
15 -          .*((b(i).*g(i)+b(i).*t(i)-t(i)).*t(i).*x11.*x22+a(i).^2)./(b(i)
16 -          .*(g(i)+t(i)).*(gamma(a(i)+1).*gamma(a(i)+1)));
17 -
18 -      subplot(3,2,i)
19 -      surf(x11,x22,y)
20 -      |
21 -      xlabel('$x_{1}$','interpreter','latex')
22 -      ylabel('$x_{2}$','interpreter','latex')
23 -
24 -  end

```

FIGURE 3.1 – Code MATLAB qui génère les graphes de la distribution BGLG

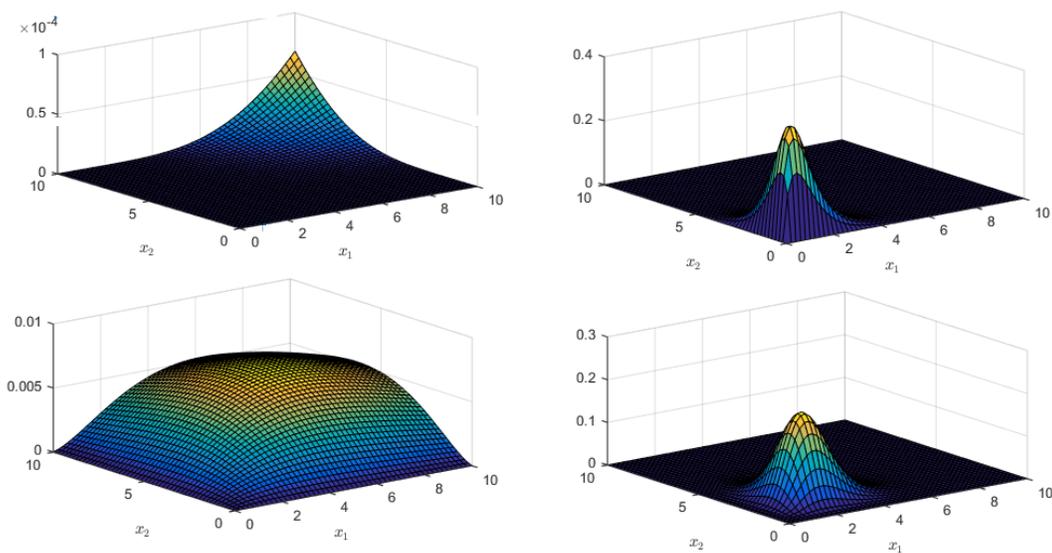


FIGURE 3.2 – Graphe de la fonction de densité de la distribution BGLG pour quelques valeurs sélectionnées de paramètres $(\alpha, \beta, \gamma, \theta)$.

Proposition 3 Pour chaque entier positif m et n , le moment joint des variables X_1 et X_2 ayant une distribution BGGL est exprimé comme suit :

$$E(X_1^m X_2^n) = \frac{\Gamma(\alpha + m)\Gamma(\alpha + n)}{\beta(\gamma + \theta)\Gamma^2(\alpha + 1)\theta^{m+n}}((\beta\theta + \beta\gamma - \theta)(\alpha + n)(\alpha + m) + \theta\alpha^2). \quad (3.21)$$

En particulier, pour $n = m = 1$, on trouve que :

$$E(X_1 X_2) = \frac{((\beta\theta + \beta\gamma - \theta)(\alpha + 1)^2 + \theta\alpha^2)}{\beta(\gamma + \theta)\theta^2}.$$

Ainsi le coefficient de corrélation entre X_1 et X_2 est :

$$\text{Corr}(X_1, X_2) = \frac{\theta(\beta\theta + \beta\gamma - \theta)}{(\beta^2 - 1)\theta^2 + \alpha\beta^2(\gamma + \theta)^2 + \beta^2\gamma(\gamma + 2\theta)}.$$

Remarque 7 Il est facile de remarquer que si $\alpha \rightarrow 0$, $\beta = 1$ et $\theta \rightarrow \infty$ ou si $\alpha \rightarrow 0$, $\gamma = 0$ et $\beta = 1$, on a $\text{Corr}(X_1, X_2) \rightarrow 1/2$.

Définition 22 Une distribution à deux variables est dite positivement dépendante du rapport de vraisemblance si la densité $f_{\Theta}(x_1, x_2)$ satisfait :

$$f_{\Theta}(x_1, y_1)f_{\Theta}(x_2, y_2) \geq f_{\Theta}(x_1, y_2)f_{\Theta}(x_2, y_1),$$

pour tout $x_1 > x_2$ et $y_1 > y_2$.

Proposition 4 Pour l'équation (3.17), l'inégalité ci-dessus se réduit à :

$$\begin{aligned} (\alpha^2 + \theta(\beta\gamma + \beta\theta - \theta)x_1y_1)(\alpha^2 + \theta(\beta\gamma + \beta\theta - \theta)x_2y_2) &\geq \\ (\alpha^2 + \theta(\beta\gamma + \beta\theta - \theta)x_1y_2)(\alpha^2 + \theta(\beta\gamma + \beta\theta - \theta)x_2y_1), \end{aligned}$$

ou $x_2y_2 + x_1y_1 = x_1y_2 + x_2y_1$, ce qui est clairement le cas. Par conséquent, la distribution BGLG dépend positivement du rapport de vraisemblance.

3.3.2 Estimation des paramètres par L'algorithme EM

L'algorithme espérance-maximisation (EM) (voir [6]) est un moyen important de trouver des estimations du maximum de vraisemblance (EMV) pour les paramètres d'un modèle lorsque les données sont incomplètes. Il s'agit d'une méthode itérative d'approximer la fonction de maximum de vraisemblance. Chaque itération de l'algorithme EM comporte deux étapes :

1. *Étape E* : nous calculons l'espérance conditionnelle notée par $Q(\Theta/\Theta^j)$, où $\Theta^j = (\alpha^{(j)}, \beta^{(j)}, \gamma^{(j)}, \theta^{(j)})$ est le vecteur des paramètres estimés de $\Theta = (\alpha, \beta, \gamma, \theta)$ à la $j^{\text{ème}}$ itération.
2. *Étape M* : nous maximisons $Q(\Theta/\Theta^j)$ en fonction du vecteur de paramètres Θ .

Dans cette section, nous examinons le problème du calcul des EMV des paramètres inconnus de la distribution BGLG pour un échantillon aléatoire donné de taille n , En utilisant la représentation de la densité de mélange de la distribution BGLG donnée par (3.20) la fonction de vraisemblance incomplète peut s'écrire comme suit

$$L(\Theta) = \prod_{i=1}^n \left(\frac{\beta-1}{\beta} f_{gg}(x_{i1}, x_{i2}, \Theta_1) + \frac{1}{\beta} f_{BL}(x_{i1}, x_{i2}, \Theta_2) \right),$$

dans lequel $\Theta_1 = (\alpha, \theta)$ et $\Theta_2 = (\alpha, \theta, \gamma)$. Pour compléter les données de cette fonction, on associe un vecteur aléatoire discret Δ , où Δ suit une distribution bivariée de Bernoulli avec des paramètres $p = \frac{\beta}{\beta-1}$ et $1-p = \frac{1}{\beta}$. Cela implique

$$P(\Delta = \delta_i) = P^{\delta_i} (1-p)^{1-\delta_i},$$

où $\delta_i \in \{0, 1\}^2$. Les échantillons complets sont les suivants :

$$(x_{11}, x_{12}, \delta_2), \dots, (x_{n1}, x_{n2}, \delta_n).$$

En se basant sur les observations complètes, la fonction de vraisemblance complète est la suivante :

$$l(\Theta) = \prod_{i=1}^n \left(\frac{\beta-1}{\beta} f_{gg}(x_{i1}, x_{i2}, \Theta_1) \right)^{\delta_i} \left(\frac{1}{\beta} f_{BL}(x_{i1}, x_{i2}, \Theta_2) \right)^{1-\delta_i}.$$

La fonction de log-vraisemblance devient donc

$$L(\Theta) = \log l(\Theta) = \sum_{i=1}^n \delta_i \left(\log \left(\frac{\beta-1}{\beta} \right) + \log(f_{gg}(x_{i1}, x_{i2}, \Theta_1)) \right) \\ + (1 - \delta_i) \left(\log(f_{BL}(x_{i1}, x_{i2}, \Theta_2)) + \log \left(\frac{1}{\beta} \right) \right).$$

Nous calculons maintenant l'espérance conditionnelle $\mathbb{Q}(\Theta, \Theta^j)$ de la fonction log vraisemblance complète, reposant sur les données observées et la valeur actuelle de la fonction de vraisemblance logarithmique $\Theta^j = (\alpha^{(j)}, \beta^{(j)}, \gamma^{(j)}, \theta^{(j)})$ à la $j^{\text{ème}}$ itération. Nous obtenons

$$\mathbb{Q}(\Theta, \Theta^j) = E(L(\Theta)/\Theta^j) \\ = \sum_{i=1}^n E(\delta_i/\Theta^j) \left(\log \left(\frac{\beta-1}{\beta} \right) + \log(f_{gg}(x_{i1}, x_{i2}, \Theta_1)) \right) \\ + \sum_{i=1}^n E((1-\delta_i)/\Theta^j) \left(\log \left(\frac{1}{\beta} \right) + \log(f_{BLG}(x_{i1}, x_{i2}, \Theta_2)) \right).$$

où

$$E(\delta_i/\Theta^j) = \tau_i^j = \frac{\frac{(\beta^{(j)}-1)}{\beta^{(j)}} f_{gg}(x_{i1}, x_{i2}, \Theta_1)}{\frac{(\beta^{(j)}-1)}{\beta^{(j)}} f_{gg}(x_{i1}, x_{i2}, \Theta_1) + \frac{1}{\beta^{(j)}} f_{BLG}(x_{i1}, x_{i2}, \Theta_2)},$$

et

$$E((1-\delta_i)/\Theta^j) = \zeta_i^j = 1 - \tau_i^j.$$

On note que τ_i^j est la probabilité postérieure calculée à la $j^{\text{ème}}$ itération.

Dans l'étape de maximisation de l'algorithme EM, nous essayons de trouver le maximum de $\Theta^{(j+1)}$ de la fonction $Q(\Theta, \Theta^j)$ par rapport à $\Theta = (\alpha, \beta, \gamma, \theta)$.

$$\Theta^{(j+1)} = \underset{\Theta}{\operatorname{argmax}} Q(\Theta/\Theta^j),$$

où

$$\begin{aligned} Q(\Theta/\Theta^j) = & \sum_{i=1}^n \tau_i^j \left(\log\left(\frac{\beta-1}{\beta}\right) + (2\alpha+2)\log(\theta) + \alpha(\log(x_{i1}) + \log(x_{i2})) \right. \\ & - \theta(x_{i1} + x_{i2}) - 2\log(\gamma(\alpha+1))) + \zeta_i^{(j)} (-\log(\beta) + (2\alpha+1)\log(\theta) \\ & + (\alpha-1)(\log(x_{i1}) + \log(x_{i2})) + \log(\alpha^2 + \gamma\theta(x_{i1})(x_{i2}) - \theta(x_{i1}) + (x_{i2}) \\ & \left. - \log(\gamma + \theta) - 2\log(\gamma(\alpha+1))). \quad (3.22) \end{aligned}$$

En dérivant la fonction (3.22) par rapport à chaque paramètre, nous obtenons le système suivant (voir [6]) :

$$\frac{\partial Q}{\partial \alpha} = 2n \log(\theta) - 2n \Psi(\alpha+1) + \sum_{i=1}^n (\log(x_{i1}) + \log(x_{i2})) + 2\alpha \sum_{i=1}^n \frac{\zeta_i^j}{\alpha^2 + \gamma\theta x_{i1} x_{i2}},$$

$$\frac{\partial Q}{\partial \beta} = \frac{1}{\beta(\beta-1)} + \sum_{i=1}^n \tau_i^j - \frac{1}{\beta} \sum_{i=1}^n \zeta_i^j,$$

$$\frac{\partial Q}{\partial \theta} = \frac{2n(\alpha+1)}{\theta} - \sum_{i=1}^n (x_{i1} + x_{i2}) + \sum_{i=1}^n \zeta_i^j \left(\frac{\gamma x_{i1} x_{i2}}{\alpha^2 + \gamma\theta x_{i1} x_{i2}} - \frac{1}{\gamma + \theta} - \frac{1}{\theta} \right),$$

et

$$\frac{\partial Q}{\partial \gamma} = \sum_{i=1}^n \zeta_i^j \left(\frac{\theta x_{i1} x_{i2}}{\alpha^2 + \gamma\theta x_{i1} x_{i2}} - \frac{1}{\gamma + \theta} \right).$$

Dans lequel $\Psi(\cdot)$ est la fonction digamma définie par $\Psi(\cdot) = \Gamma'(\cdot)/\Gamma(\cdot)$. Ensuite, nous mettons les équations différentielles précédentes égales à zéro et nous utilisons le logiciel mathématique MATLAB pour résoudre numériquement ce système d'équations à l'aide de l'algorithme de Newton-Raphson. Nous répétons cette étape jusqu'à ce que

$$\| \Theta^{(j+1)} - \Theta^j \| < \varepsilon,$$

où ε est un seuil de convergence fixe (voir [13]).

Remarque 8 L'algorithme EM converge en un nombre fini d'itérations et donne les estimations du maximum de vraisemblance des paramètres. Ces solutions donnent les estimateurs EM, notés $\hat{\alpha}, \hat{\beta}, \hat{\gamma}$ et $\hat{\theta}$.

L'algorithme : En outre, nous élaborons l'algorithme suivant pour calculer les EMV des paramètres inconnus.

— Étape 1 : Choisir quelques valeurs initiales de Θ , disent $\theta^{(0)} = (\alpha^{(0)}, \beta^{(0)}, \gamma^{(0)}, \theta^{(0)})$.

— Étape 2 : Obtenir

$$\theta^{(1)} = \arg \max_{\Theta} \mathbb{Q}(\Theta / \Theta^{(0)})$$

— Étape 3 : poursuivre le processus jusqu'à ce qu'il y ait convergence.

3.3.3 Distribution de la somme

Dans cette partie, nous dérivons des expressions explicites pour fonctions de densité de probabilité et les moments de $Y = X_1 + X_2$, lorsque le vecteur aléatoire (X_1, X_2) est défini selon la distribution BGLG.

Proposition 5 Si X_1 et X_2 sont distribués selon (3.17), alors la fonction de densité de probabilité de Y est exprimée par :

$$f_Y(y) = \frac{\theta^{2\alpha+1} y^{2\alpha-1} e^{-\theta y}}{\beta(\gamma + \theta)\Gamma(2\alpha)} + \frac{(\beta\gamma + \beta\theta - \theta)\theta y^2}{2\alpha(2\alpha + 1)}; y > 0. \quad (3.23)$$

Preuve 2 En substituant $x_1 = yz$ et $x_2 = y(1 - z)$ avec le jacobien $J(x_1, x_2 \rightarrow y, z) = y$, dans la fonction densité jointe de X_1 et X_2 , nous obtenons la fdp jointe de Y et Z comme suit :

$$f_{Y,Z}(y, z) = \frac{\theta^{2\alpha+1} z^{\alpha-1} (1-z)^{\alpha-1} y^{2\alpha-1} e^{-\theta y}}{\beta(\gamma + \theta)\Gamma^2(\alpha + 1)} [(\beta\gamma + \beta\theta - \theta)\theta y^2 z(1-z) + \alpha^2], \quad (3.24)$$

où $z \in (0, 1)$ et $y > 0$. En intégrant cette équation de manière appropriée, nous obtenons les densités marginales de Y et Z .

Proposition 6 Si X_1 et X_2 sont distribués selon (3.20) alors la fonction de moment de la variable Y est exprimée par :

$$E(Y^n) = \sum_{k \geq 0} \binom{n}{k} \frac{(n-k+\alpha)\Gamma(\alpha+k)}{\beta(\gamma+\theta)\Gamma^2(\alpha+1)\theta^n} ((\beta\theta + \beta\gamma - \theta)(\alpha+n-k)(\alpha+k) + \theta\alpha^2).$$

3.3.4 Applications réelles

Nous considérons dans un premier temps les données de pluie et de neige collectées dans trois villes françaises et présentons une application du modèle donné par (3.7).

Les précipitations sont toutes les formes d'eau qui se forment dans l'atmosphère terrestre et qui tombent à la surface de la Terre. Les nuages finissent par être trop chargés en vapeur d'eau et les précipitations se transforment en liquide (pluie) ou en solide (neige). La pluie désigne les précipitations de différentes tailles, des grosses gouttes lourdes aux gouttes légères, tandis que la neige est l'un des types de précipitations solides. Elle est constituée d'eau qui a été gelée.

Nous considérons les données disponibles sur le site <https://en.tutiempo.net/climate> recueillies auprès des trois villes françaises suivantes (situées dans différentes parties du pays) concernant la pluie et la neige : Dijon, Paris et Toulouse. L'ensemble des données comprend le nombre de jours de chaque année où la pluie et la neige sont apparues au cours de la période allant de 1990 à 2018. Nous considérons les variables suivantes :

- X_1 : proportion de jours de pluie ;
- X_2 : proportion de jours avec de la neige ;
- $Y = X_1 + X_2$: proportion de jours avec précipitations (pluie ou

neige).

Le tableau (3.2) présente les estimations de $\hat{\alpha}$, $\hat{\beta}$, $\hat{\gamma}$ et $\hat{\theta}$ qui ont été obtenues en utilisant l'algorithme EM considéré dans la section précédente.

TABLE 3.2 – Valeurs estimées de $\hat{\alpha}$, $\hat{\beta}$, $\hat{\gamma}$, $\hat{\theta}$ et $E(y)$

City	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\theta}$	$E(y)$
Dijon	0.7191	1.1963	5.6880	4.9055	0.5431
Paris	0.4957	1.0001	7.7070	4.1536	0.5516
Toulouse	0.3318	1.1444	8.6421	4.6344	0.4431

On constate que la ville de Toulouse présente la proportion de jours de précipitations (pluie ou neige) la plus faible.

Enfin, nous examinons les résultats expérimentaux en utilisant l'ensemble de données réelles de la pluie et de la neige pour la ville de Dijon dans la fiabilité de différents modèles statistiques tels que les distributions BGL et BGLG. Afin de comparer ces modèles, nous utilisons les deux critères suivants : Critère d'information d'Akaike (AIC) et critère d'information bayésien (BIC), qui sont définis comme suit :

$$AIC = -2\log\hat{L} + 2k$$

$$BIC = -2\log\hat{L} + k \log n$$

où k est le nombre de paramètres libres dans le modèle, n est la taille de l'échantillon et $\log\hat{L}$ est la log-vraisemblance. Pour l'ajustement d'un ensemble de données, le meilleur modèle est celui qui présente la plus petite valeur d'AIC et de BIC.

Le tableau suivant présente les estimations du maximum de vraisemblance (EMV) des paramètres des modèles ajustés et les valeurs des statistiques suivantes : AIC, BIC. Sur la base des valeurs de ces

statistiques, nous concluons que la distribution BGLG peut fournir de bons ajustements pour les données réelles. Notez que des résultats et interprétations similaires peuvent être traités pour les autres ensembles de données.

TABLE 3.3 – EMVs, AIC et BIC pour l'ensemble des données réelles de pluie et de neige pour la ville de Dijon.

Distr	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\theta}$	AIC	BIC
BGLG	0.733	1.255	5.248	4.999	-59.526	-54.197
BGL	0.719	—	7.767	4.905	-10.355	-6.358

D'après le tableau [3.3](#) la distribution bivariée de Gamma-Lindley généralisée (BGLG) présente les valeurs les plus petites des critère AIC et BIC. Ce résultat montre que la distribution BGLG donne la meilleure ajustement de données simulée. Les résultats du ce tableau ont été trouvés en exécutant le programme suivant.

```

1      clear all , clc
2      t=readtable('dodij.xlsx'); % Data of Dijon
3      X=[t.RDJ t.SDj];
4      err=0.00001 ;
5      p=[0.7;1.2;4.9;4.6]; %p=[alpha;beta;gamma;theta]
6      pl=[0.2;0.5;0.7];
7      [n,k]=size(X);
8      nbt_itr_n=100 ;
9      v_BGLG=4;
10
11     for i=1:nbt_itr_n
12         %~~~~~ Gradient de la BGLG ~~~~~
13         g1=(2*n*log(p(4)))-(2*n*psi(p(1)+1))+sum(log(X(:,1)))+sum(log(X(:,2)))_
14         2*p(1)*sum(1/((p(1)^2)+(p(2)*p(4)+p(2)*p(3)-p(4))*p(4)*X(:,1).*X(:,2)));
15         g2=(-n/p(2))+((p(4)*p(3)+(p(4)^2)*sum((X(:,1).*X(:,2))./(p(1)^2)
16         +(p(2)*p(3)+p(2)*p(4)-p(4))*p(4)*X(:,1).*X(:,2))));
17         g3=(-n/(p(3)+p(4)))+(p(4)*p(2))*sum((X(:,1).*X(:,2))./(p(1)^2)+(p(2)*p(3)
18         +p(2)*p(4)-p(4))*p(4)*X(:,1).*X(:,2)));
19         g4=(-n/(p(3)+p(4)))+(n*((2*p(1)+1)/p(4))-sum(X(:,1)+X(:,2)))+(p(2)*p(3)
20         +2*p(2)*p(4)-2*p(4))*sum((X(:,1).*X(:,2))./(p(1)^2)+(p(4)*(p(2)*p(3)
21         +p(2)*p(4)-p(4))*X(:,1).*X(:,2))));
22         grad_bgg1=[g1;g2;g3;g4];
23         %~~~~~ Matrice Hessienne de BGLG ~~~~~
24         H(1,1)=-2*n*psi(1,p(1)+1)+2*sum(((p(2)*p(4)+p(2)*p(3)-p(4))*p(4)*X(:,1)
25         .*X(:,2)-(p(1)^2))./(p(1)^2+(p(2)*p(4)+p(2)*p(3)-p(4))*p(4)*X(:,1).*X(:,2)
26         H(2,2)=(n/(p(2)^2))-((p(4)*p(3)+(p(4)^2)^2)*sum(((X(:,1).*X(:,2)).^2)
27         ./((p(1)^2)+(p(2)*p(4)+p(2)*p(3)-p(4))*p(4)*X(:,1).*X(:,2)).^2));
28         H(3,3)=(n/((p(3)+p(4))^2))-((p(2)*p(4))^2)*sum(((X(:,1).*X(:,2)).^2)
29         ./((p(1)^2)+(p(2)*p(4)+p(2)*p(3)-p(4))*p(4)*X(:,1).*X(:,2)).^2));
30         H(4,4)=(n/((p(3)+p(4))^2))-((n*((2*p(1)+1)/p(4)^2)+sum(((2*p(2))-2)
31         .*X(:,1).*X(:,2))./(p(1)^2+(p(2)*p(4)+p(2)*p(3)-p(4))*p(4)*X(:,1).*X(:,2)
32         -sum(((p(2)*p(3)+2*p(2)*p(4)-2*p(4))*X(:,1).*X(:,2)).^2)./(p(1)^2+(p(2)
33         *p(4)+p(2)*p(3)-p(4))*p(4).*X(:,1).*X(:,2)).^2));
34         H(2,1)=-sum((2*p(1)*(p(3)*p(4)+(p(4)^2))*X(:,1).*X(:,2))./(p(1)^2+(p(2)
35         *p(4)+p(2)*p(3)-p(4))*p(4).*X(:,1).*X(:,2)).^2));
36         H(1,2)=H(2,1);
37         H(3,1)=-2*p(1)*p(2)*p(4)*sum((X(:,1).*X(:,2))./((p(1)^2+(p(2)*p(4)+p(2)
38         *p(3)-p(4))*p(4)*X(:,1).*X(:,2)).^2));
39         H(1,3)=H(3,1);
40         H(4,1)=(2*n)/p(4)-2*p(1)*(p(2)*p(3)+(2*p(2)*p(4)-(2*p(4))) *sum(((p(1)
41         +(p(2)*p(4)+p(2)*p(3)-p(4))*p(4).*X(:,1).*X(:,2)).^2));
42         H(1,4)=H(4,1);
43         H(3,2)=p(4)*sum((X(:,1).*X(:,2))./(p(1)^2+(p(2)*p(4)+p(2)*p(3)-p(4))*p(4)
44         .*X(:,1).*X(:,2)))-p(4)*p(2)*(p(4)*p(3)+(p(4)^2))*sum(((X(:,1).*X(:,2)).^2)
45         ./((p(1)^2)+(p(2)*p(4)+p(2)*p(3)-p(4))*p(4)*X(:,1).*X(:,2)).^2));
46         H(2,3)=H(3,2);
47         H(2,4)=(p(3)+2*p(4))*sum((X(:,1).*X(:,2))./((p(1)^2+(p(2)*p(4)+p(2)*p(3)
48         -p(4))*p(4)*X(:,1).*X(:,2)).^2))-((p(4)*p(3)+(p(4)^2))*p(2)*p(3)+2*p(2)
49         *p(4)-2*p(4))*sum(((X(:,1).*X(:,2)).^2)./(p(1)^2+(p(2)*p(4)+p(2)*p(3)
50         -p(4))*p(4).*X(:,1).*X(:,2)).^2));
51         H(4,2)=H(2,4);
52         H(3,4)=(n/((p(3)+p(4))^2))+p(2)*sum((X(:,1).*X(:,2))./(p(1)^2+(p(2)*p(4)
53         +p(2)*p(3)-p(4))*p(4)*X(:,1).*X(:,2))-p(4)*p(2)*(p(2)*p(3)+(2*p(2)*p(4)
54         -(2*p(4)))*sum(((X(:,1).*X(:,2)).^2)./(p(1)^2+(p(2)*p(4)+p(2)*p(3)-p(4)
55         *p(4)*X(:,1).*X(:,2)).^2));
56         H(4,3)=H(3,4);
57         %~~~~~ Itération de l'algorithme ~~~~~
58         pr=p;
59         p=p-inv(H)*grad_bglg;
60
61         if abs(pr(1)-p(1))<err & abs(pr(2)-p(2))<err & abs(pr(3)-p(3))<err
62             & abs(pr(4)-p(4))<err
63
64             break
65         end
66         %~~~~~ Afficher p (le vecteur d'estimations) ~~~~~
67         display(p)
68         %~~~~~ Calcul de L^ ~~~~~
69         L_BGLG=n*((2*p(1)+1)*log(p(4)))-log(p(2))-log(p(3)+p(4))-
70         (2*log(gamma(p(1)+1)))+(p(1)-1)*sum(log(X(:,1)))+(p(1)-1)
71         *sum(log(X(:,2)))-p(4)*sum(X(:,1)+X(:,2))+sum(log((p(1)^2)
72         +(p(4)*p(2)+(p(3)*p(4)-p(4))*p(4)*X(:,1).*X(:,2))))
73         %~~~~~ Calcul de AIC pour ~~~~~
74         Aic_BGLG=(-2*log(L_BGLG))+(2*k)
75         Bic_BGLG=(-2*log(L_BGLG))+(k*log(n))

```

FIGURE 3.3 – Code MATLAB qui donne les EMVs, AIC et BIC.

Conclusion générale

Ce mémoire a présenté une introduction aux lois de probabilité continues et à leurs applications en modélisation statistique.

Le premier chapitre a posé les bases théoriques en introduisant les concepts de tribu, mesure et loi de probabilité d'une variable aléatoire continue. La loi gaussienne multidimensionnelle a été présentée comme un exemple important de loi continue.

Le deuxième chapitre s'est concentré sur les modèles de mélange de lois finies. Deux méthodes d'estimation des paramètres de ces modèles ont été détaillées : la méthode du maximum de vraisemblance et la méthode d'expectation-maximisation.

Enfin, le troisième chapitre a exploré les différentes versions de la distribution de Lindley généralisée, en versions univariée et bivariée. L'accent a été mis sur la flexibilité de cette famille de distributions et son application dans des contextes concrets.

Bibliographie

- [1] Aliu, F., Pavelkova, D. and Dehning, B. (2017). *Portfolio risk-return analysis : The case of the automotive industry in the Czech Republic. J. Inter. Studies, 10 (4) : 72-83.*
- [2] Akaike, H. (1974). *A new look at statistical model identification. IEEE Trans. Autom. Control, AC-19 : 716–723.*
- [3] Barrera, J., Huillet, T. and Paroissin, C. (2005). *Size-biased permutation of Dirichlet partitions and search-cost distribution. Probab. Eng. Inform. Sc., 19 (1) : 83-97.*
- [4] Christophe, B. *Pourquoi les mélanges de modèles pour la classification , CNRS et Université de Lille 1, Villeneuve d’Ascq, France.*
- [5] Christophe, C., Caen. (2018). *cour sur l’estimateur du Maximun de Vraisemblance.*
- [6] Dempster, A. P. Laird, N. M. Rubin, D. B. (1977). *Maximum likelihood from incomplete data via the EM algorithm, J. R. Stat. Soc. Ser. B. Stat. Methodol. 39(1), 1–38.*
- [7] Gideon, S. (1978). *Estimating the number of components in a finite mixture model. Ann. Stat, 6 : 461–464.*
- [8] Ghitany, B. Atieh, and S. Nadarajah.(2008). *Lindley distribution and its applications. IMACS, 78, 493– 506.*
- [9] Gradshteyn, I. S. ryzhik, I. M. (2000). *Table of Integrals, Series, and Products, 6th Ed. AP, NY.*

- [10] Lindley, D. V.(1958). *Fiducial distributions and Bayes theorem*, *J. Roy. Statist. Soc. Ser. B* 20, 102–107.
- [11] Samuel, K, N. Balakrishnan, Norman L. Johnson. (2000). *Continuous multivariate distributions. Vol. 1, second ed.*, Wiley, New York.
- [12] Sankaran. (1970). *The discrete Poisson.Lindley distribution*. *Biom.*, 26, 145-149.
- [13] WU, C. J. (1983). *On the convergence properties of the EM algorithm*, *Ann. Stat.* 11(1) , 95–103.
- [14] Zakerzadeh, H. Dolati, A. (2009). *Generalized Lindley distribution*, *J. Math. Ext.* 3(2), 13–25.