

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 de Mathématiques et Informatique



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

PROPRIÉTÉS DES DISTRIBUTIONS DE DIRICHLET

Présenté Par :

M^{me}. BELABBES BENGRAA Imène

Devant le jury composé de :

Dr. HELLAL Meryem	M C A	UAT.B.B (Ain Témouchent)	Présidente
Dr. BENNAFLA Djamila	M C B	UAT.B.B (Ain Témouchent)	Examinatrice
Dr. MECENE Rahmouna	M C B	UAT.B.B (Ain Témouchent)	Encadrante

Année Universitaire 2023/2024

Dédicaces

À mes parents,

dont le sacrifice et le dévouement ont tracé la voie vers la réussite, je vous suis reconnaissante pour votre soutien indéfectible.

À mon mari,

dont le soutien inébranlable a été ma force motrice tout au long de cette aventure académique.

À mes enfants, Lina et Youcef,

dont les sourires ont illuminé les jours les plus sombres et ont donné un sens profond à chaque effort entrepris.

À mes frères, Saïd et Mohamed, et à ma sœur Fériel,

pour leur présence réconfortante et leurs conseils précieux. En particulier à mon frère Saïd, qui a toujours été là pour moi, je te remercie pour tout.

Enfin, à mes amies,

pour leur amitié sincère et leurs encouragements constants, vous avez rendu ce voyage non seulement possible, mais aussi enrichissant et mémorable.

Remerciements

Je tiens tout d'abord à exprimer ma profonde gratitude envers **Dr. MECENE Rahmouna** pour son soutien inestimable, son expertise et sa guidance tout au long de ce travail de recherche. Ses conseils éclairés et son dévouement ont grandement contribué à la réussite de ce projet de fin d'études.

Je tiens également à remercier chaleureusement le **Dr. HELLAL Meryem** et le **Dr. BENNAFLA Djamilia** pour avoir consacré leur temps précieux à évaluer ce travail et pour leurs commentaires constructifs qui ont enrichi cette étude.

Enfin, je souhaite exprimer ma reconnaissance envers l'ensemble des enseignants qui m'ont accompagné tout au long de mon parcours académique. Leurs enseignements, leur encouragement et leur inspiration ont joué un rôle essentiel dans ma formation et dans le développement de mes compétences.

Merci infiniment à tous ceux qui ont contribué, de près ou de loin, à la réalisation de ce travail.

Résumé

Ce mémoire propose une analyse approfondie des distributions de Dirichlet, mettant en lumière les propriétés fondamentales de ces distributions complexes. En se basant sur les concepts clés de la théorie des probabilités, l'étude explore les relations entre les distributions de Liouville multivariées et la distribution de Dirichlet, offrant ainsi une perspective enrichissante sur la modélisation des données compositionnelles. L'objectif de ce mémoire se concentre principalement sur la propriété de scaling des distributions de Dirichlet de type 3, démontrant leur capacité à être transformées tout en préservant leur distribution caractéristique. Des méthodes efficaces pour calculer les fonctionnelles de ces distributions sont présentées, avec une application concrète dans l'analyse des proportions de sérum sanguin chez des canetons de Pékin blancs. Cette recherche ouvre de nouvelles voies pour l'application pratique de ces modèles en statistique et en modélisation des données, offrant ainsi des perspectives prometteuses pour l'avancement des connaissances et la résolution de problèmes réels.

Abstract

This master's thesis provides an in-depth analysis of Dirichlet distributions, highlighting the fundamental properties of these complex distributions. Based on key concepts of probability theory, the study explores the relationships between multivariate Liouville distributions and Dirichlet distribution, offering an enriching perspective on compositional data modeling. The primary focus of this work is on the scaling property of type 3 Dirichlet distributions, demonstrating their ability to be transformed while preserving their characteristic distribution. Efficient methods for calculating the functionals of these distributions are presented, with a concrete application in the analysis of blood serum proportions in white Pekin ducklings. This research opens new avenues for the practical application of these models in statistics and data modeling, thus offering promising prospects for advancing knowledge and solving real-world problems.

ملخص

نقترح من خلال هذه المذكرة تحليلاً معمقاً لتوزيعات ديريشليه ، مسلطين الضوء على الخصائص الأساسية لهذه التوزيعات المعقدة بالاعتماد على المفاهيم الرئيسية لنظرية الاحتمالات، تستكشف هذه الدراسة العلاقات بين توزيعات ليوفيل متعددة المتغيرات وتوزيع ديريشليه، مما يقدم منظوراً غنياً حول نمذجة البيانات التركيبية. يركز جوهر هذه المذكرة على خاصية القياس لتوزيعات ديريشليه من النوع الثالث، مما يثبت قدرتها على التحول مع الحفاظ على توزيعها المميز. يتم تقديم طرق فعالة لحساب الدوال لهذه التوزيعات، مع تطبيق عملي في تحليل نسب مصل الدم لدى البط الأبيض البكيني. تفتح هذه المذكرة آفاقاً جديدة للتطبيق العملي لهذه النماذج في الإحصاء ونمذجة البيانات، مما يقدم آفاقاً واعدة لتقدم المعرفة وحل المشكلات الواقعية.

Table des matières

Introduction générale	6
1 Préliminaire	8
1.1 Tribu et mesure	8
1.1.1 Tribu et tribu borélienne	8
1.1.2 Mesure	9
1.2 Loi de Probabilité Multivariée	10
1.2.1 Fonctions de base	11
1.2.2 Espérance mathématique	12
1.2.3 Changement de variable	13
1.2.4 Indépendance	14
1.3 Distributions de base	14
1.4 Maximum de vraisemblance	16
1.4.1 Estimateur du maximum de vraisemblance	16
1.4.2 Algorithme de Fisher scoring	18
1.4.3 La matrice de covariance	19
2 Distribution de Liouville multivariée	20
2.1 Distribution de Liouville multivariée	21
2.1.1 Intégrale de Dirichlet-Liouville	22
2.2 Distribution de Dirichlet de type 1	22
2.2.1 Construction à partir de la distribution Gamma	23
2.2.2 Structure de RAM (Residual Allocation Model) de la partition de Dirichlet	23
2.2.3 Propriété de Scaling	25
2.2.4 Estimation des paramètres	25

2.3	Distribution de Dirichlet inverse	27
2.3.1	Construction à partir de la distribution Gamma	27
2.3.2	Calcul de la fonctionnelle avec la distribution de Dirichlet inverse	28
2.4	Conclusion	28
3	Distribution de Dirichlet de type 3	30
3.1	Propriétés de la distribution de Dirichlet de type 3	30
3.1.1	Distribution de Dirichlet de type 3	31
3.1.2	Structure de RAM de la distribution de Dirichlet de type 3	33
3.1.3	Propriété de Scaling	34
3.1.4	Calcul de fonctionnelle	36
3.2	Application de la distribution de Dirichlet de type 3	37
3.3	Conclusion	40
	Conclusion et perspectives	41
	Bibliographie	42

Introduction générale

La modélisation des phénomènes possédants des données positives représente un défi significatif qui transcende divers domaines telles que l'analyse médicale, les études atmosphériques, la microbiologie, la science environnementale, la gestion financière, etc. Malgré l'utilisation répandue de distributions paramétriques traditionnelles, l'adaptation de ces modèles aux particularités des données positives devient souvent insatisfaisante, mettant en lumière la nécessité d'introduction des nouvelles distributions probabilistes qui sont plus novatrices et adaptatives pour mieux appréhender la variabilité inhérente à ces types de données.

Particulièrement, la mécanique statistique offre une approche fascinante pour caractériser les systèmes complexes, où un système défini par des paramètres macroscopiques tels que la pression, l'énergie, la densité, le volume, ou la fréquence, peut être représenté par une multitude d'états microscopiques dans un espace nommé espace de phase. Autrement dit, cette méthode repose sur l'idée selon laquelle une précision absolue de l'état microscopique n'est pas indispensable pour décrire de manière raisonnable le comportement observable du système. En effet, elle propose d'introduire d'une variable (aléatoire) d'état, fonction des variables microscopiques, dotée d'une nature intrinsèquement probabiliste. Cela ouvre ainsi la voie à des approches probabilistes plus flexibles et adaptées à la complexité des données positives. Ce genre de modélisation considère la distribution de Liouville définie proportionnellement par :

$$f \left(\sum_{i=1}^n x_i \right) \prod_{i=1}^n x_i^{a_i-1}.$$

comme un outil puissant pour explorer la variabilité et la nature intrinsèquement probabiliste des données positives résultantes de ce système. La problématique centrale de cette recherche se cristallise autour de la manière dont la distribution de Liouville, en lien avec la mécanique statistique, peut être exploitée de manière novatrice pour modéliser efficacement des données positives dans des contextes divers. Cette interrogation

guide notre exploration vers des approches mathématiques plus avancées, cherchant à dépasser les limitations des modèles traditionnels.

Ce mémoire est consacré à l'étude des distributions de Liouville, avec une attention particulière portée aux distributions de Dirichlet avec ses trois types 1, 3 et inverse. Le mémoire est composé de trois chapitres :

Le premier chapitre présente les bases essentielles pour la compréhension du mémoire. Il aborde les concepts clés de tribu et de mesure, ainsi que la mesure de Lebesgue, cruciale en analyse probabiliste. Ensuite, il explore les lois de probabilité multivariées, l'indépendance entre variables aléatoires et le maximum de vraisemblance pour l'estimation des paramètres. Ce chapitre a pour objectif de présenter les fondements théoriques indispensables pour les développements ultérieurs sur les distributions de probabilité et les modèles statistiques.

Le deuxième chapitre introduit les distributions de Liouville multivariées, essentielles pour modéliser les données compositionnelles. Nous définissons la distribution et examinons sa relation avec d'autres distributions, comme la distribution de Dirichlet. L'intégrale de Dirichlet-Liouville est introduite pour évaluer le volume sur le support. Nous abordons également des aspects avancés comme la structure de RAM (Residual Allocation Model). Ce chapitre offre une base solide pour l'application pratique de ces concepts en statistiques et en modélisation des données.

Le chapitre 3 explore en profondeur les distributions de Dirichlet de type 3. La première partie de ce chapitre définit la propriété de scaling, démontrant comment les variables aléatoires sont transformées tout en conservant leur distribution. La deuxième partie développe des théorèmes et des corollaires permettant de calculer efficacement les fonctionnelles de ces distributions complexes en fonction de fonctionnelles plus simples de gamma. En utilisant ces résultats théoriques, le chapitre illustre ensuite une application concrète de la distribution de Dirichlet de type 3 dans l'analyse des proportions de sérum sanguin chez des canetons de Pékin blancs, démontrant ainsi l'importance pratique de ces modèles dans le domaine de la biologie.

Préliminaire

Dans ce chapitre, nous rappelons quelques définitions et propriétés nécessaires pour la bonne compréhension du reste du mémoire. Pour plus de détails voir [25], [27] et [29].

1.1 Tribu et mesure

Dans le cadre de la théorie de la mesure en mathématiques, deux concepts fondamentaux émergent : la "tribu" et la "mesure". Ces deux notions sont intimement liées et jouent un rôle crucial dans la définition et la compréhension des ensembles mesurables ainsi que dans l'attribution de valeurs numériques à ces ensembles.

1.1.1 Tribu et tribu borélienne

Dans cette partie, on présente les concepts de tribu et tribu borélienne :

Définition 1.1. Soit Ω un ensemble et \mathcal{A} une famille de parties de Ω . On dit que \mathcal{A} est une tribu si \mathcal{A} satisfait les conditions suivantes :

1. $\emptyset \in \mathcal{A}$ et $\Omega \in \mathcal{A}$.
2. Si $A \in \mathcal{A}$ alors $A^c \in \mathcal{A}$.
3. Si $(A_i)_{i \in \mathbb{N}}$ est une suite d'éléments de \mathcal{A} alors $\bigcup_{i \in \mathbb{N}} A_i \in \mathcal{A}$.

On dit alors que (Ω, \mathcal{A}) est un espace mesurable.

Définition 1.2. Un espace topologique est un couple (Ω, T) , où Ω est un ensemble et T une topologie sur Ω . Un ensemble de parties de Ω que l'on appelle les ouverts de (Ω, T) vérifiant les propriétés suivantes :

1. L'ensemble vide et Ω appartiennent à T .
2. Toute réunion quelconque d'ouverts est un ouvert, c'est-à-dire que si $(O_i)_{i \in I}$ est une famille d'éléments de T indexée par un ensemble I quelconque (pas nécessairement fini, ni même dénombrable) alors :

$$\bigcup_{i \in I} O_i \in T.$$

3. Toute intersection finie d'ouverts est un ouvert, c'est-à-dire que si O_1, \dots, O_n sont des éléments de T alors :

$$O_1 \cap \dots \cap O_n \in T.$$

Définition 1.3. La tribu borélienne (également appelée tribu de Borel ou tribu des boréliens) sur un espace topologique Ω est la plus petite tribu sur Ω contenant tous les ensembles ouverts. Les éléments de la tribu borélienne sont appelés des boréliens.

1.1.2 Mesure

Dans cette partie, on définit la notion de mesure :

Définition 1.4. Soit (Ω, \mathcal{A}) un espace mesurable. Une application μ définie sur \mathcal{A} à valeurs dans $[0, +\infty[$ est appelée une mesure lorsque les deux propriétés suivantes sont satisfaites :

1. L'ensemble vide a une mesure nulle, c'est à dire : $\mu(\emptyset) = 0$.
2. L'application μ est σ -additive, c'est-à-dire : si $(A_i)_{i \in \mathbb{N}}$ est une famille dénombrable de parties de Ω appartenant à \mathcal{A} et si ces parties sont deux à deux disjointes alors :

$$\mu \left(\bigcup_{i=1}^{\infty} A_i \right) = \sum_{i=1}^{\infty} \mu(A_i).$$

Proposition 1.1. Il existe une plus petite mesure définie sur une tribu de \mathbb{R}^n qui soit complète et coïncide sur les pavés avec leur volume (c'est-à-dire le produit des longueurs de leurs côtés). Cette mesure est appelée la mesure de Lebesgue.

Remarque 1. *La mesure de Lebesgue est la mesure complétée de sa restriction à la tribu des boréliens. Cette restriction aux boréliens de la mesure de Lebesgue est parfois dénommée mesure de Borel-Lebesgue.*

Définition 1.5. *Soit (Ω, \mathcal{A}) un espace mesuré. On appelle probabilité une application*

$$\mathbb{P} : \mathcal{A} \rightarrow [0, 1]$$

vérifiant les axiomes suivants :

1. $\mathbb{P}(\Omega) = 1$.
2. *Pour toute suite d'événements disjoints $(A_i)_{i \in \mathbb{N}}$ de \mathcal{A} , on a*

$$\mathbb{P} \left(\bigcup_{i=1}^{\infty} A_i \right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

Le triplet $(\Omega, \mathcal{A}, \mathbb{P})$ s'appelle alors un espace de probabilité.

Définition 1.6. *Une famille finie ou dénombrable d'ensemble (non vide) $(A_i)_{i \in \mathbb{I}}$ est dite partition de Ω si on a :*

1. *La réunion de tous les ensembles A_i couvre l'ensemble Ω , c'est-à-dire :*

$$\Omega = \bigcup_{i \in \mathbb{I}} A_i.$$

2. *Les ensembles A_i sont deux à deux disjoints, c'est-à-dire que pour tous indices i et j distincts, on a :*

$$A_i \cap A_j = \emptyset \quad \text{pour } i \neq j.$$

1.2 Loi de Probabilité Multivariée

Dans certains problèmes interviennent simultanément plusieurs variables aléatoires. Mis à part les cas particuliers de variables indépendantes et de variables liées fonctionnellement, cela introduit la notion de loi de probabilité à plusieurs variables autrement appelée loi jointe. Pour éviter la complexité des formules, la suite est limité à deux variables.

1.2.1 Fonctions de base

Dans cette partie, nous donnons les définitions de quelques fonctions de base qui caractérisent la loi d'une variable ou vecteur aléatoire.

- La probabilité pour que la variable aléatoire X prenne une valeur numérique inférieure à x alors que Y prend une valeur inférieure à y définit la fonction de répartition :

$$\begin{aligned} F_{X,Y}(x, y) &= \mathbb{P}(X \leq x, Y \leq y), \\ &= \int_{-\infty}^y \int_{-\infty}^x f_{x,y}(t, s) dt ds. \end{aligned}$$

Celle-ci est non décroissante en x et en y entre la valeur 0 lorsque les deux variables tendent vers $-\infty$ et la valeur 1 lorsqu'elles tendent toutes deux vers $+\infty$.

- La densité de probabilité jointe ou loi jointe s'obtient par une double dérivation :

$$f_{X,Y}(x, y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y).$$

- Une intégration par rapport à y (resp. x) donne la densité de probabilité marginale ou loi marginale de X (resp. Y) :

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy.$$

- Le rapport de la densité de probabilité jointe à la densité marginale de Y (concernant toutes les valeurs x) représente la densité de probabilité conditionnelle de X sous la condition $Y = y$:

$$f_{X|Y}(x) = \frac{f_{X,Y}(x, y)}{f_Y(y)}.$$

- La fonction génératrice des moments d'une variable aléatoire X est la fonction M_X définie par

$$\begin{aligned} M_X(t) &= \mathbb{E}(e^{tX}), \\ &= \int_{\mathbb{R}} e^{tx} f_X(x) dx, \end{aligned}$$

pour tout réel t tel que cette espérance existe. Cette fonction, comme son nom l'indique, est utilisée afin d'engendrer les moments associés à la distribution de probabilités de la variable aléatoire X .

1.2.2 Espérance mathématique

L'espérance mathématique, souvent simplement appelée "espérance", est un concept fondamental en probabilités et en statistiques. Elle représente la valeur moyenne attendue d'une variable aléatoire.

Définition 1.7. *L'espérance mathématique d'une fonction g de deux variables généralise la formule donnée pour une seule variable :*

$$\mathbb{E}[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f_{XY}(x, y) dx dy.$$

Proposition 1.2. *L'opérateur espérance est linéaire ; en particulier, l'espérance (la moyenne) d'une somme de deux variables aléatoires est la somme des moyennes :*

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y].$$

Définition 1.8. *Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité et $X : \Omega \rightarrow \mathbb{R}$ une variable aléatoire. Lorsque X^2 est d'espérance finie, on appelle variance de X le réel $V(X)$ défini par :*

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

et écart-type de X le réel $\sigma(X)$ défini par :

$$\sigma_X = \sqrt{V(X)}.$$

Définition 1.9. *Considérons une paire de variables aléatoires X et Y . Les moments joints des variables X et Y sont donnés par*

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

où n et m sont deux entiers positifs. On note que le moment qui correspond à $n = m = 1$ est nécessaire pour le calcul de la covariance.

Définition 1.10. *la fonction caractéristique d'un couple (X, Y) de variables aléatoires est définie par :*

$$\varphi_{XY}(\theta, \psi) = \mathbb{E}[e^{i(\theta X + \psi Y)}].$$

Proposition 1.3. *La fonction caractéristique de la somme $Z = X + Y$ de variable X et Y est alors égale au produit des fonctions caractéristiques individuelles :*

$$\varphi_Z(\theta) = \varphi_X(\theta) \varphi_Y(\theta).$$

1.2.3 Changement de variable

Dans cette partie nous allons déterminer la loi de probabilité d'une variable aléatoire écrite en fonction d'une autre variable aléatoire de loi de probabilité connue (voir[11]).

Théorème 1.4. Soient Y_1 et Y_2 deux variables aléatoires de densité jointe f_{Y_1, Y_2} . Soit :

$$Y = \{(y_1, y_2), f_{Y_1, Y_2}(y_1, y_2) > 0\}.$$

Soit la fonction h définie de E vers F par :

$$\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = h \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

tels que :

- les dérivées partielles de h sont continues sur E ,
- le jacobien :

$$J = \det \begin{pmatrix} \frac{\partial y_1}{\partial u_1} & \frac{\partial y_1}{\partial u_2} \\ \frac{\partial y_2}{\partial u_1} & \frac{\partial y_2}{\partial u_2} \end{pmatrix}$$

est non nul pour $(u_1, u_2) \in F$. Alors, la densité jointe des variables aléatoires (U_1, U_2) est donnée par :

$$f_{U_1, U_2}(u_1, u_2) = \begin{cases} |J| f_{Y_1, Y_2}(h^{-1}(u_1, u_2)) & \text{pour } (u_1, u_2) \in F, \\ 0 & \text{sinon.} \end{cases}$$

Exemple 1. Soit la densité suivante :

$$f_{Y_1, Y_2}(y_1, y_2) = \begin{cases} 1 & \text{si } 0 \leq y_1 \leq 1 \text{ et } 0 \leq y_2 \leq 1 \\ 0 & \text{sinon.} \end{cases}$$

On demande la densité jointe de :

$$\begin{pmatrix} U_1 \\ U_2 \end{pmatrix} = \begin{pmatrix} Y_1 + Y_2 \\ Y_2 - Y_1 \end{pmatrix}.$$

On peut écrire :

$$\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \implies \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$$

Le jacobien J est donné par :

$$J = \det \begin{pmatrix} \frac{\partial y_1}{\partial u_1} & \frac{\partial y_1}{\partial u_2} \\ \frac{\partial y_2}{\partial u_1} & \frac{\partial y_2}{\partial u_2} \end{pmatrix} = \det \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} = \frac{1}{2}$$

Ainsi,

$$|J| = \frac{1}{2}$$

Donc, la densité jointe des variables aléatoires (U_1, U_2) est :

$$f_{U_1, U_2}(u_1, u_2) = \begin{cases} \frac{1}{2} & \text{pour } (u_1, u_2) \in U \\ 0 & \text{sinon} \end{cases}$$

1.2.4 Indépendance

L'indépendance en probabilité se réfère à une relation entre deux variables aléatoires où la réalisation ou le comportement de l'une n'affecte pas celui de l'autre.

Définition 1.11. Si la densité de probabilité conditionnelle de X par rapport à Y est identique à la densité marginale :

$$f_X(x) = \frac{f_{XY}(x, y)}{f_Y(y)},$$

on dit que les deux variables sont indépendantes. L'égalité se réécrit :

$$f_{XY}(x, y) = f_X(x) f_Y(y).$$

Proposition 1.5. Si les deux variables X et Y sont indépendantes, alors elles sont décorrélées, l'inverse n'est pas vrai, car l'indépendance implique tous les moments au lieu d'un seul. Donc la notion de variables décorrélées est plus faible que celle d'indépendance et est loin d'avoir la même utilité.

1.3 Distributions de base

Dans cette partie, nous allons définir quelques distributions de base qui sont fréquemment utilisées dans ce mémoire.

Définition 1.12. On dit qu'une variable aléatoire distribuée selon la distribution gamma, avec un paramètre de forme $\theta > 0$ et un paramètre d'échelle 1 si sa fonction de densité est donnée par :

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

et sa fonction de moment est donnée par ($q > -\theta$) :

$$\mathbf{E}[X^q] = \frac{\Gamma(\theta + q)}{\Gamma(\theta)} =: (\theta)_q,$$

où $\Gamma(\theta)$ désigne la fonction gamma d'Euler définie par :

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

Définition 1.13. On dit qu'une variable aléatoire X suit une distribution beta de type 1 avec des paramètres $\alpha, \beta > 0$ (disons $X \stackrel{d}{\sim} \beta(\alpha, \beta)$), si sa fonction de densité est donnée par

$$f_X(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}; \quad 0 < x < 1,$$

de sorte que sa fonction de moment est

$$\mathbf{E}[X^q] = \frac{(\alpha)_q}{(\alpha + \beta)_q}.$$

Définition 1.14. Soit ${}_2F_1$ la fonction hypergéométrique de Gauss donnée par :

$${}_2F_1(a, b; c; x) = \sum_{j \geq 0} \frac{(a)_j (b)_j}{(c)_j j!} x^j.$$

L'expression intégrale de la fonction hypergéométrique de Gauss est donnée par :

$${}_2F_1(a, b; c; x) = \frac{\Gamma(c)}{\Gamma(a)\Gamma(c-a)} \int_0^1 t^{a-1} (1-t)^{c-a-1} (1-xt)^{-b} dt, \quad (1.1)$$

où $\text{Re}(c) > \text{Re}(a) > 0$.

Définition 1.15. Si une variable aléatoire X suit une distribution beta de type 3 avec des paramètres $\alpha, \beta > 0$ (disons $X \stackrel{d}{\sim} \beta^3(\alpha, \beta)$), alors sa fonction de densité est donnée par

$$f_X(x) = 2^\alpha \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} (1+x)^{-(\alpha+\beta)}; \quad 0 < x < 1,$$

de sorte que sa fonction de moment est

$$\mathbf{E}[X^q] = \frac{2^{-\beta} (\alpha)_q}{(\alpha + \beta)_q} {}_2F_1\left(\beta, \alpha + \beta; \alpha + \beta + q; \frac{1}{2}\right).$$

Définition 1.16. La variable aléatoire continue X suit une loi normale de paramètres μ et σ^2 , notée $X \sim \mathcal{N}(\mu, \sigma^2)$, si sa densité de probabilité est donnée par :

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

pour tout $x \in \mathbb{R}$, où :

- $\mu \in \mathbb{R}$ est la moyenne de la distribution.
- $\sigma > 0$ est l'écart-type (la racine carrée de la variance σ^2).

Définition 1.17. Soit $\mathbf{s}_n := (s_1, \dots, s_n) \in \mathbb{R}^n$ et $f : \mathbb{R}^n \rightarrow \mathbb{R}$. Si $f(x\mathbf{s}_n) = x^d f(\mathbf{s}_n)$ pour $x > 0$, alors f est dite homogène de degré d .

1.4 Maximum de vraisemblance

Le maximum de vraisemblance est une méthode théorique d'estimation des paramètres. On considère n valeurs observées (x_1, x_2, \dots, x_n) d'un échantillon de n variables aléatoires indépendantes et de même loi. Dans son principe, la méthode du maximum de vraisemblance consiste à choisir comme estimation de θ la valeur qui maximise la vraisemblance $L(\theta; x_1, x_2, \dots, x_n)$, c'est-à-dire la valeur pour laquelle les observations effectuées avaient « le plus de chance » de se produire. Cette justification empirique est confirmée par des théorèmes de mathématiques et cette méthode générale fournit des estimateurs qui sont très souvent les meilleurs.

1.4.1 Estimateur du maximum de vraisemblance

Soit X une variable aléatoire continue, de loi \mathcal{D}_θ de paramètre θ inconnu et de densité $f(x; \theta)$. On appelle vraisemblance de θ au vu des observations (x_1, \dots, x_n) d'un n -échantillon indépendamment et identiquement distribué selon la loi \mathcal{D}_θ , le nombre :

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

À (x_1, \dots, x_n) fixé, on cherche à trouver le maximum de cette vraisemblance pour que les probabilités des réalisations observées soient aussi maximum. Ceci est un problème

d'optimisation. Ainsi, un estimateur du maximum de vraisemblance est tout estimateur $\hat{\theta}$ de θ vérifiant :

$$\sup_{\theta} \ln L(x_1, \dots, x_n; \theta) = L(x_1, \dots, x_n; \hat{\theta}).$$

À priori, il n'y a ni existence, ni unicité d'un estimateur du maximum de vraisemblance. Cependant, dans la plupart des cas, il existe et il est unique.

Exemple 2. : Soit X une variable aléatoire continue qui suit la loi normale. La fonction de vraisemblance doit être définie à partir de la densité par :

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

Donc :

$$\begin{aligned} L(x_1, \dots, x_n; m) &= \prod_{i=1}^n \left[\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x_i-m}{\sigma}\right)^2} \right] \\ &= \left(\frac{1}{\sigma\sqrt{2\pi}} \right)^n e^{-\frac{1}{2} \sum_{i=1}^n \left(\frac{x_i-m}{\sigma}\right)^2}. \end{aligned}$$

Puis, on a la fonction log-vraisemblance :

$$\ln L(x; m) = -n \ln(\sigma\sqrt{2\pi}) - \frac{1}{2} \sum_{i=1}^n \left(\frac{x_i - m}{\sigma}\right)^2,$$

où sa dérivée par rapport à m est donnée par :

$$\frac{\partial}{\partial m} \ln L(x, m) = \sum_{i=1}^n \frac{x_i - m}{\sigma^2}.$$

Ainsi, \hat{m}^{MV} est déterminé par l'équation suivante :

$$\begin{aligned} \sum_{i=1}^n \left(\frac{x_i - m}{\sigma^2}\right) = 0 &\Rightarrow \sum_{i=1}^n (x_i - m) = 0 \\ &\Rightarrow \hat{m}^{MV} = \sum_{i=1}^n \frac{x_i}{n} = \bar{X}. \end{aligned}$$

Considérons maintenant l'EMV de σ^2 . La dérivée de la fonction log-vraisemblance par rapport à σ^2 est trouvée comme :

$$\frac{\partial}{\partial \sigma^2} \ln(L(x, m, \sigma^2)) = \frac{-n}{2\sigma^2} + \frac{1}{2} \sum_{i=1}^n \frac{(x_i - m)^2}{\sigma^4};$$

L'EMV $\hat{\sigma}^{2MV}$ est déterminé, ainsi par l'équation suivante :

$$\begin{aligned} \frac{-n}{2\sigma^2} + \sum_{i=1}^n \frac{(x_i - m)^2}{2\sigma^4} &= 0 \\ \Rightarrow \frac{-n\sigma^2 + \sum_{i=1}^n (x_i - m)^2}{2\sigma^4} &= 0 \\ \Rightarrow -n\sigma^2 + \sum_{i=1}^n (x_i - m)^2 &= 0 \end{aligned}$$

Alors

$$\hat{\sigma}^{2MV} = \frac{1}{n} \sum_{i=1}^n (x_i - m)^2.$$

Donc on remplace par sa valeur devient

$$\hat{\sigma}^{2MV} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{X})^2 = S^2.$$

1.4.2 Algorithme de Fisher scoring

L'algorithme de scoring, également connu sous le nom de méthode de Fisher scoring, est un algorithme de recherche de racines (ou zéros) d'une fonction réelle. Il est utilisé en statistique pour résoudre numériquement des équations de maximum de vraisemblance.

Soient X_1, \dots, X_n des variables aléatoires, indépendantes et identiquement distribuées avec une densité de probabilité à deux dérivées $f(x; \theta)$, et nous souhaitons calculer l'estimateur du maximum de vraisemblance (EMV) de θ . Tout d'abord, supposons que nous avons un point de départ pour notre algorithme θ_0 . Alors, l'algorithme de Fisher scoring est défini par la relation suivante :

$$\theta^{(m+1)} = \theta^{(m)} + \mathcal{I}^{-1}(\theta^{(m)})S(\theta^{(m)}).$$

avec $\mathcal{I}(\theta)$ est la matrice d'information de Fisher qui est généralement définie par :

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

et S est la fonction score définie par :

$$S(\theta) = \nabla_{\theta} \log \mathcal{L}(\theta) = \left(\frac{\partial \log \mathcal{L}(\theta)}{\partial \theta_i} \right)_{i=1 \dots m}.$$

On note que $\mathcal{I}(\theta)$ est une matrice symétrique définie positive.

1.4.3 La matrice de covariance

La matrice de covariance d'un vecteur de n variables aléatoires $X = (X_1, \dots, X_n)$ dont chacune possède une variance, est la matrice carrée dont le terme générique est donné par :

$$a_{i,j} = \text{Cov}(X_i, X_j).$$

La matrice de covariance, notée parfois Σ , est définie par

$$\Sigma \equiv \text{E} [(X - \text{E}(X))(X - \text{E}(X))^T].$$

En développant les termes :

$$\begin{aligned} \Sigma &= \begin{pmatrix} \text{Var}(X_1) & \text{Cov}(X_1, X_2) & \cdots & \text{Cov}(X_1, X_p) \\ \text{Cov}(X_2, X_1) & \ddots & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(X_p, X_1) & \cdots & \cdots & \text{Var}(X_p) \end{pmatrix} \\ &= \begin{pmatrix} \sigma_{x_1}^2 & \sigma_{x_1x_2} & \cdots & \sigma_{x_1x_p} \\ \sigma_{x_2x_1} & \ddots & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{x_px_1} & \cdots & \cdots & \sigma_{x_p}^2 \end{pmatrix}. \end{aligned}$$

Remarque 2. On note que la matrice de covariance Σ représente l'inverse de la matrice d'information de Fisher \mathcal{I} .

Distribution de Liouville multivariée

Les ensembles de données compositionnelles, caractérisant la répartition des différentes parties ou constituants d'un tout, sont incontournables dans une multitude de domaines scientifiques allant de la chimie à la médecine. Que ce soit pour comprendre la composition des sols en géologie, la distribution des espèces dans un écosystème en biologie, ou encore la répartition des éléments chimiques dans un composé, ces données fournissent des informations cruciales pour la compréhension des phénomènes naturels et des processus biologiques. L'analyse statistique précise de ces ensembles de données revêt alors une importance capitale afin d'extraire des informations significatives et fiables. Une approche adéquate pour modéliser la variabilité de ces données est nécessaire. Dans cette perspective, la classe de distribution probabiliste de Liouville, qui englobe notamment les distributions de Dirichlet, émerge comme un cadre théorique approprié. Cette classe offre une représentation mathématique robuste pour caractériser la structure complexe des données compositionnelles, comme discuté de manière approfondie dans [1].

La famille de distributions de Liouville est une famille de distributions probabilistes qui puise ses origines dans les contributions majeures du mathématicien Joseph Liouville. Les premières investigations sur cette distribution remontent aux travaux novateurs de Marshall et Olkin [18], jetant ainsi les bases de son exploration. Par la suite, cette distribution a été soumise à un examen minutieux par une pléiade de chercheurs éminents, parmi lesquels figurent Gupta et Richards [14], ainsi que Gupta et Kabe [15].

Leurs études ont porté sur une analyse approfondie de certaines propriétés essentielles de cette distribution, éclairant ainsi sa nature et son comportement probabiliste.

2.1 Distribution de Liouville multivariée

Dans cette section, nous donnons une définition formelle détaillée de la distribution de Liouville multivariée (voir [14]).

Définition 2.1. Soit $\mathbf{X}_n := (X_1, \dots, X_n)$ un vecteur de n variables aléatoires positives (absolument continues). On dit que \mathbf{X}_n suit la distribution de Liouville multivariée si sa densité de probabilité est proportionnelle à :

$$g \left(\sum_{i=1}^n x_i \right) \prod_{i=1}^n x_i^{\theta_i - 1}. \quad (2.1)$$

Où les variables s'étendent sur l'octant généralisé $\Delta_n^+ := \{(x_1, \dots, x_n) : x_i > 0, i = 1, \dots, n\}$, les paramètres $\theta_1, \dots, \theta_n$ sont positifs et la fonction g est positive et continue.

Plus rigoureusement, si le support Δ_n^+ de $g(\cdot)$ est non compact, on dit que X_n suit une distribution de Liouville de première espèce. Alors que, si le support de $g(\cdot)$ est compact, nous disons que \mathbf{X}_n suit une distribution de Liouville de deuxième espèce.

La distribution de Liouville englobe les différents type de distributions de Dirichlet classiques qui se présentent comme des cas particuliers. En effet, si on considère la fonction de densité définie par $g(t) = (1 - t)^{\theta_{n+1} - 1}$, avec $0 < t < 1$ et $\theta_{n+1} > 0$ où $t = \sum_{i=1}^n x_i$, alors le vecteur aléatoire \mathbf{X}_n possède la fonction de densité jointe suivante :

$$A \left(1 - \sum_{i=1}^n x_i \right)^{\theta_{n+1} - 1} \prod_{i=1}^n x_i^{\theta_i - 1}, \quad (2.2)$$

avec A est la constante de normalisation donnée par $A = \Gamma \left(\sum_{i=1}^n \theta_i \right) / \prod_{i=1}^n \Gamma(\theta_i)$. Cette dernière, correspond à la fonction de densité de la distribution de Dirichlet de type 1.

De plus, si la fonction g est définie par $g(t) = (1 + t)^{-(\theta_1 + \dots + \theta_{n+1})}$, avec $t > 0$ et $\theta_{n+1} > 0$, alors le vecteur aléatoire \mathbf{X}_n possède la fonction de densité définie comme suit :

$$B \left(1 + \sum_{i=1}^n x_i \right)^{-(\theta_1 + \dots + \theta_{n+1})} \prod_{i=1}^n x_i^{\theta_i - 1}. \quad (2.3)$$

avec $B = \Gamma \left(\sum_{i=1}^{n+1} \theta_i \right) / \prod_{i=1}^{n+1} \Gamma(\theta_i)$. Cette dernière représente la fonction de densité de la distribution de Dirichlet inverse.

On peut voir clairement que la distribution de Dirichlet inverse représente une distribution de Liouville de première espèce, tandis que la distribution de Dirichlet de type 1 manifeste une distribution de Liouville de deuxième espèce.

2.1.1 Intégrale de Dirichlet-Liouville

Dans cette partie, nous définissons un type d'intégrale approprié aux distributions de Liouville. Cette intégrale est appelée l'intégrale de Dirichlet-Liouville. Pour plus de détails voir [24].

Définition 2.2. On appelle une intégrale de Dirichlet-Liouville toute intégrale définie sur le support de dimension n .

$$\Delta_n^+ = \left\{ x_1, \dots, x_n; x_i \geq 0 \text{ et } i = 1, \dots, n, \text{ avec } s \leq \sum_{i=1}^n x_i \leq t \right\}$$

Par la formule suivante :

$$\int_{\Delta_n} g \left(\sum_{i=1}^n x_i \right) \prod_{i=1}^n x_i^{a_i-1} dx_i = \frac{\prod_{i=1}^n \Gamma(a_i)}{\Gamma \left(\sum_{i=1}^n a_i \right)} \int_s^t h(z) z^{\sum_{i=1}^n a_i - 1} dz \quad (2.4)$$

où $h(x)$ est une fonction continue et $a_i > 0, i = 1, \dots, n$. Le cas $g(x) = 1$ correspond à l'intégrale de Dirichlet.

Dans les deux sections suivantes, nous montrons que ces deux distributions peuvent être introduites à partir de la distribution Gamma.

2.2 Distribution de Dirichlet de type 1

La distribution de Dirichlet de type 1 joue un rôle crucial dans diverses applications statistiques, notamment les statistiques d'ordre, l'analyse de fiabilité et de survie, et l'analyse bayésienne. Sa polyvalence et sa flexibilité en font un outil précieux pour modéliser des proportions et des probabilités dans une grande variété de contextes [19].

2.2.1 Construction à partir de la distribution Gamma

En manipulant des variables aléatoires distribuées selon la loi gamma, on expose le résultat suivant (voir [16]) :

Définition 2.3. Soient X_1, \dots, X_n des variables aléatoires indépendantes et identiquement distribuées qui suivent la loi Gamma (θ_m) , $m = 1, \dots, n$. $\tilde{X} = \sum_{m=1}^n X_m \stackrel{d}{\sim}$ Gamma $(\sum_{m=1}^n \theta_m)$. Alors, $\mathbf{S}_n = (S_1, \dots, S_n) := (X_1/\tilde{X}, \dots, X_n/\tilde{X})$ possède la distribution de Dirichlet notée $D_n(\boldsymbol{\theta}_n)$ avec $\boldsymbol{\theta}_n = (\theta_1, \dots, \theta_n)$ et est indépendante de \tilde{X} . La densité de (S_1, \dots, S_n) est définie sur le support :

$$\Xi_n = \left\{ (s_1, \dots, s_n) \in [0, 1]^n, \sum_{m=1}^n s_m = 1 \right\}$$

par

$$f_{S_1, \dots, S_n}(s_1, \dots, s_n) = \frac{\Gamma(\sum_{m=1}^n \theta_m)}{\prod_{m=1}^n \Gamma(\theta_m)} \prod_{m=1}^n s_m^{\theta_m - 1}. \quad (2.5)$$

Nous définissons $\mathbf{S}_n \stackrel{d}{\sim} D_n(\boldsymbol{\theta}_n)$ sur Ξ_n et $\mathbf{S}_{n-1} \stackrel{d}{\sim} D_{n-1}(\boldsymbol{\theta}_{n-1}; \theta_n)$ sur Λ_{n-1} où :

$$\Lambda_{n-1} = \left\{ (s_1, \dots, s_{n-1}) \in [0, 1]^{n-1}, \sum_{m=1}^{n-1} s_m < 1 \right\}.$$

Il convient de noter que si $(S_1, \dots, S_{n-1}) \stackrel{d}{\sim} D_{n-1}(\boldsymbol{\theta}_{n-1}; \theta_n)$ et $S_n = 1 - \sum_{m=1}^{n-1} S_m$, alors $(S_1, \dots, S_n) \stackrel{d}{\sim} D_n(\boldsymbol{\theta}_n)$.

De plus, la fonction des moments joints de \mathbf{S}_n est donnée par : ($q_m > -\theta_m$, $m = 1, \dots, n$) :

$$E \left(\prod_{m=1}^n S_m^{q_m} \right) = \frac{\Gamma(\sum_{m=1}^n \theta_m)}{\Gamma(\sum_{m=1}^n (\theta_m + q_m))} \prod_{m=1}^n \frac{\Gamma(\theta_m + q_m)}{\Gamma(\theta_m)}. \quad (2.6)$$

2.2.2 Structure de RAM (Residual Allocation Model) de la partition de Dirichlet

Dans cette partie, on s'intéresse à présenter une propriété fondamentale de la distribution de Dirichlet.

Proposition 2.1. Soit (Z_1, \dots, Z_{n-1}) un vecteur de $(n-1)$ variables aléatoires indépendantes telles que : $Z_m \stackrel{d}{\sim} \text{beta}(\theta_m, \sum_{l=1}^n \theta_l - \sum_{l=1}^m \theta_l)$. Avec $\bar{Z}_m = 1 - Z_m \stackrel{d}{\sim} \text{beta}(\sum_{l=1}^n \theta_l - \sum_{l=1}^m \theta_l, \theta_m)$ et $\prod_{m=1}^0 \bar{Z}_i = 1$, on définit

$$S_m = \left(\prod_{i=1}^{m-1} \bar{Z}_i \right) Z_m, \quad m = 1, \dots, n-1. \quad (2.7)$$

$$S_n = 1 - \sum_{m=1}^{n-1} S_m = \prod_{m=1}^{n-1} \bar{Z}_m. \quad (2.8)$$

Alors $S_n \stackrel{d}{\sim} D_n(\theta_n)$.

Preuve : En utilisant l'indépendance des Z_m , on trouve que la densité jointe du vecteur aléatoire (Z_1, \dots, Z_{n-1}) est donnée par

$$\frac{\Gamma(\sum_{l=1}^n \theta_l)}{\prod_{m=1}^n \Gamma(\theta_m)} \prod_{m=1}^{n-1} z_m^{\theta_m-1} (1-z_m)^{\sum_{l=1}^n \theta_l - \sum_{l=1}^m \theta_l - 1}. \quad (2.9)$$

Prenons la transformation $s_m = \prod_{i=1}^{n-1} (1-z_i) z_m$, $m = 1, \dots, n-1$ avec le jacobien

$$J(z_1, \dots, z_{n-1} \rightarrow s_1, \dots, s_{n-1}) = \prod_{m=1}^{n-1} \left(1 - \sum_{l=1}^{m-1} s_l \right)^{-1} \quad (2.10)$$

dans l'équation (2.9), on trouve que la fonction de densité jointe de S_1, \dots, S_n avec $S_n = 1 - \sum_{m=1}^{n-1} S_m$ est donnée par

$$\frac{\Gamma(\sum_{l=1}^n \theta_l)}{\prod_{m=1}^n \Gamma(\theta_m)} \prod_{m=1}^{n-1} s_m^{\theta_m-1} \left(1 - \sum_{l=1}^{m-1} s_l \right)^{-\theta_m} \times \left(1 - \left(1 - \sum_{l=1}^{m-1} s_l \right) s_m \right)^{\sum_{l=1}^n \theta_l - \sum_{l=1}^m \theta_l - 1}. \quad (2.11)$$

Or,

$$1 - \left(1 - \sum_{l=1}^{m-1} s_l \right)^{-1} s_m = \frac{1 - \sum_{l=1}^m s_l}{1 - \sum_{l=1}^{m-1} s_l}. \quad (2.12)$$

En simplifiant cette expression, on aura

$$\frac{\Gamma(\sum_{m=1}^n \theta_m)}{\prod_{m=1}^n \Gamma(\theta_m)} \prod_{m=1}^{n-1} s_m^{\theta_m-1} \left(1 - \sum_{m=1}^{n-1} s_m \right)^{\theta_n-1}. \quad (2.13)$$

Par conséquent, $(S_1, \dots, S_n) \stackrel{d}{\sim} D_n(\theta_n)$.

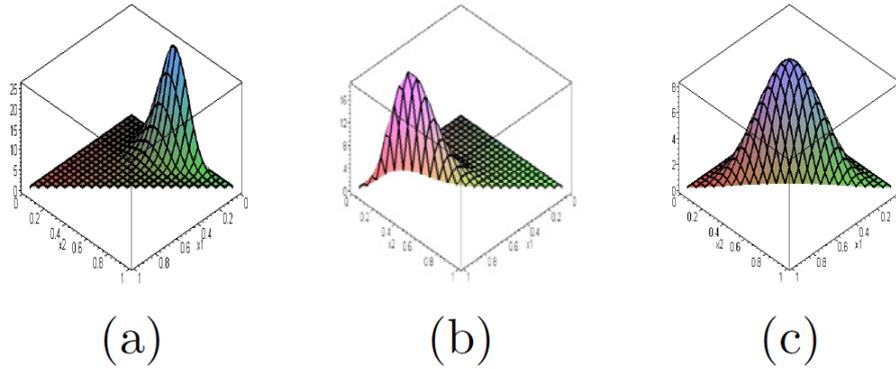


FIGURE 2.1 – Graphe de la distribution de Dirichlet de type 1 pour quelques paramètres choisis : (a) $\theta_1 = 8.5, \theta_2 = 7.5, \theta_3 = 1.5$. (b) $\theta_1 = 10.5, \theta_2 = 3.5, \theta_3 = 3.5$. (c) $\theta_1 = 3.5, \theta_2 = 3.5, \theta_3 = 3.5$.

2.2.3 Propriété de Scaling

Il est connu que la distribution de Dirichlet peut être générée par $S_m = X_m / \sum_{m=1}^n X_m$, avec $\sum_{m=1}^n X_m$ la somme de n variables aléatoires indépendantes distribuées selon une loi gamma(θ_m). Ainsi, sa fonction de densité jointe est écrite comme suit :

$$f_{S_1, \dots, S_n}(s_1, \dots, s_n) = \frac{\Gamma(\sum_{m=1}^n \theta_m)}{\prod_{m=1}^n \Gamma(\theta_m)} \prod_{m=1}^n s_m^{\theta_m - 1}.$$

Définition 2.4. Soit $\mathbf{S}_n(x) := (X_1(x) / \sum_{m=1}^n X_m, \dots, X_n(x) / \sum_{m=1}^n X_m)$ où $X_m(x) = xX_m$; $m = 1, \dots, n$ et $x > 0$. Alors, la distribution de $\mathbf{S}_n(x) := (S_m(x); m = 1, \dots, n)$ est donnée par

$$f_{S_1(x), \dots, S_n(x)}(s_1, \dots, s_n) = \frac{\Gamma(\sum_{m=1}^n \theta_m)}{\prod_{m=1}^n \Gamma(\theta_m)} \frac{\prod_{m=1}^n s_m^{\theta_m - 1}}{x^{\sum_{m=1}^n \theta_m - 1}}. \quad (2.14)$$

Sur le support

$$\Omega_n = \left\{ (s_1, \dots, s_n) \in [0, 1]^n, \sum_{m=1}^n s_m = x \right\}.$$

2.2.4 Estimation des paramètres

L'idée des estimations par maximum de vraisemblance est de trouver les valeurs des paramètres qui correspondent le mieux aux données observées, c'est-à-dire qui maximisent la fonction de vraisemblance (ou la fonction de log-vraisemblance). Étant

donné un échantillon aléatoire de k observations provenant d'une distribution de Dirichlet de type 1, la fonction de log-vraisemblance peut être écrite comme suit [22] :

$$L = \prod_{i=1}^k \left[\frac{\Gamma(\sum_{m=1}^n \theta_m)}{\prod_{m=1}^n \Gamma(\theta_m)} \prod_{m=1}^{n-1} s_m^{\theta_m - 1} \right]. \quad (2.15)$$

$$\log L = k \left\{ \log \Gamma \left(\sum_{m=1}^n \theta_m \right) - \sum_{m=1}^n \log \Gamma(\theta_m) \right\} + k \sum_{m=1}^n (\theta_m - 1) \sum_{i=1}^k \log G_i, \quad (2.16)$$

où :

$$G_i = \left[\prod_{m=1}^n S_{nm} \right]^{1/k}; \quad i = 1, 2, \dots, n-1 \text{ et } G_n = \left[\prod_{i=1}^k \left(\sum_{m=1}^{n-1} X_{im} \right) \right]^{1/k}.$$

En prenant les dérivées de la fonction de log-vraisemblance, les équations de vraisemblance peuvent être écrites comme suit :

$$\frac{\partial \log L}{\partial \theta_m} = k \psi \left(\sum_{m=1}^n \theta_m \right) - k \psi(\theta_j) + k \log G_i = 0 \text{ avec } i = 1, \dots, n, \quad (2.17)$$

où $\psi(\cdot)$ est la fonction digamma définie par $\Gamma'(\cdot)/\Gamma(\cdot)$. Les dérivées partielles secondes et les dérivées partielles mixtes sont :

$$\frac{\partial^2 \log L}{\partial \theta_m^2} = k \psi' \left(\sum_{m=1}^n \theta_m \right) - k \psi'(\theta_m); \quad m = 1, \dots, n.$$

$$\frac{\partial^2 \log L}{\partial \theta_i \partial \theta_j} = k \psi' \left(\sum_{m=1}^n \theta_m \right). \quad (2.18)$$

La matrice d'information (I) est :

$$I = \{I_{i,m}\} = -E \left[\frac{\partial^2 \log L}{\partial \theta_i \partial \theta_j} \right],$$

$$I_{i,m} = -k \psi' \left(\sum_{m=1}^n \theta_m \right), \quad i \neq j, \quad (2.19)$$

$$I_{i,m} = k \psi'(\theta_m) - k \psi' \left(\sum_{m=1}^n \theta_m \right).$$

2.3 Distribution de Dirichlet inverse

Bdiri et al ont développé plusieurs modèles qui utilisent la distribution de Dirichlet inversée pour représenter et modéliser des données non gaussiennes. Ils ont introduit des modèles de mélanges finis (voir[3] et [4]) et infinis (voir[5]) de distributions de Dirichlet inversées en utilisant la technique de Newton-Raphson pour estimer les paramètres et le processus de Dirichlet pour modéliser les mélanges infinis. Bdiri et al ont également utilisé la distribution de Dirichlet inversée pour proposer une approche de génération de noyaux de machine à vecteurs de support basée sur l'inférence bayésienne (voir[6]) et une autre approche pour établir une classification hiérarchique (voir [7] et [8]).

2.3.1 Construction à partir de la distribution Gamma

Définition 2.5. Soit X_1, \dots, X_{n+1} des variables aléatoires indépendantes distribuées respectivement selon Gamma (θ_m). On définit

$$(Y_1, \dots, Y_n) = \left(\frac{X_1}{X_{n+1}}, \dots, \frac{X_n}{X_{n+1}} \right). \quad (2.20)$$

Alors Y_n suit une distribution de Dirichlet inverse notée $ID_n(\theta_n, \theta_{n+1})$ avec les paramètres $\theta_n = (\theta_1, \dots, \theta_n)$ et θ_{n+1} sur $\Lambda_n = \{0 < y_1 < \infty, \dots, 0 < y_n < \infty\}$. On note que :

$$Y_n \stackrel{d}{\sim} \mu(dy_n) = f_{Y_1, \dots, Y_n}(y_1, \dots, y_n) \prod_{m=1}^n dy_m, \quad (2.21)$$

où f_{Y_1, \dots, Y_n} est la fonction de densité jointe donnée par :

$$f_{Y_1, \dots, Y_n}(y_1, \dots, y_n) = \frac{\Gamma(\sum_{m=1}^{n+1} \theta_m)}{\prod_{m=1}^{n+1} \Gamma(\theta_m)} \prod_{m=1}^n y_m^{\theta_m-1} \left(1 + \sum_{m=1}^n y_m \right)^{\sum_{m=1}^{n+1} \theta_m} \quad (2.22)$$

Alternativement, la loi de $Y_n := (Y_1, \dots, Y_n)$ peut être caractérisée par sa fonction de moment jointe ($q_m > -\theta_m, m = 1, \dots, n$ et $\theta_{n+1} > \sum_{m=1}^n q_m$) :

$$\mathbb{E} \left(\prod_{m=1}^n Y_m^{q_m} \right) = \frac{\Gamma(\theta_{n+1} - \sum_{m=1}^n q_m)}{\Gamma(\theta_{n+1})} \prod_{m=1}^n (\theta_m)_{q_m} \quad (2.23)$$

Remarque 3. Il découle de l'Équation (2.23) que la loi marginale Y_m avec ($m = 1, \dots, n$) suit la distribution bêta inversée avec les paramètres θ_m et θ_{n+1} .

2.3.2 Calcul de la fonctionnelle avec la distribution de Dirichlet inverse

Cette section vise à présenter un résultat fondamental qui permet de calculer le fonctionnel de Dirichlet en terme de fonctionnel plus simple de gamma.

Théorème 2 Soit $\mathbf{Y}_n = (Y_1, \dots, Y_n)$ un vecteur aléatoire qui suit la distribution de Dirichlet inverse.

- i) Soit $\mathbf{X}_n(t) := (X_m(t); m = 1, \dots, n)$, n variables aléatoires indépendantes définies par $\mathbf{X}_m(t) = X_m \setminus t$ avec $t > 0$ et X_m suit la distribution de Gauss θ_n . Alors, pour une fonction mesurable de Borel $f : \mathbb{R}^n \rightarrow \mathbb{R}$ satisfaisant

$$\int_0^\infty E(|f(\mathbf{X}_n(t))|) t^{\theta_{n+1}-1} e^{-tp} dt < \infty \quad (2.24)$$

on a :

$$E(f(\mathbf{Y}_n(p))) = \frac{p^{\theta_{n+1}}}{\Gamma(\theta_{n+1})} \int_0^\infty E(f(\mathbf{X}_n(t))) t^{\theta_{n+1}-1} e^{-tp} dt. \quad (2.25)$$

- ii) Si f est homogène de degré d , et si $E(|f(\mathbf{X}_n)|) < \infty$, alors, avec $\mathbf{X}_n := (X_1, \dots, X_n)$

$$E(f(\mathbf{Y}_n)) = \frac{\Gamma(\theta_{n+1} - d)}{\Gamma(\theta_{n+1})} E(f(\mathbf{X}_n)); \quad \text{pour } \theta_{n+1} > d. \quad (2.26)$$

L'importance de l'énoncé du Théorème peut être vue lorsqu'il est utilisé pour calculer la fonction caractéristique de la distribution de Dirichlet inverse donnée dans [13]. Si nous considérons la fonction

$$f(Y_1, \dots, Y_n) = \left(\sum_{l=1}^k Y_{m_l} \right)^q,$$

nous observons que f est homogène de degré $d = q$. En appliquant la deuxième partie du Théorème 2 à ce cas particulier, nous obtenons

$$E \left[\left(\sum_{l=1}^k Y_l \right)^q \right] = \frac{\Gamma(\theta_{n+1} - q)}{\Gamma(\theta_{n+1})} \left(\sum_{l=1}^k \theta_l \right)^q. \quad (2.27)$$

Ce qui montre que $\sum_{l=1}^k Y_l \stackrel{d}{\sim} \beta \left(\sum_{l=1}^k \theta_l, \theta_{n+1} \right)$.

2.4 Conclusion

En conclusion, nous avons exploré en détail la distribution de Liouville multivariée et ses implications dans la modélisation des ensembles de données compositionnelles.

Nous avons commencé par définir formellement la distribution de Liouville et avons discuté de ses propriétés fondamentales, notamment ses conditions d'applicabilité dans les premier et deuxième espaces. Ensuite, nous avons examiné les relations entre la distribution de Liouville et d'autres distributions statistiques couramment utilisées, telles que la distribution de Dirichlet de type 1 et la distribution de Dirichlet inverse.

Nous avons également abordé l'intégrale de Dirichlet-Liouville, qui fournit une formule pour le volume sur le support de dimension n , et avons établi des liens entre la distribution de Dirichlet de type 1 et la distribution Gamma. Enfin, nous avons exploré la structure de RAM (Residual Allocation Model) de la partition de Dirichlet et discuté des propriétés de scaling de la distribution de Dirichlet de type 1.

Distribution de Dirichlet de type 3

Parmi les distributions simples les plus connues, nous mentionnons la distribution de Dirichlet. Malgré ses nombreuses propriétés mathématiques et statistiques, elle est inadaptée pour modéliser la plupart des données compositionnelles en raison de la faible structure de dépendance qu'elle implique. En effet, à bien des égards, elle peut être considérée comme la référence standard pour modéliser les relations d'indépendance les plus fortes compatibles avec les données compositionnelles (voir par exemple [2], [23] et [28]). Comme la distribution de Dirichlet de type 3 admet le même simplexe que la distribution de Dirichlet, elle peut présenter le même intérêt d'application et sera utilisée pour modéliser les données compositionnelles relatives à des problèmes plus complexes. De plus, la distribution de Dirichlet de type 3 est couramment utilisée comme distribution a priori en statistique bayésienne, et est en fait une distribution a priori conjuguée de la distribution multinomiale.

3.1 Propriétés de la distribution de Dirichlet de type 3

Dans cette Section, nous présentons quelques propriétés principales de la distribution de Dirichlet de type 3. En particulier, nous trouvons le modèle d'allocation résiduelle (RAM) similaire à celui donné par Devroye [12] pour la distribution de Dirichlet. Ensuite, nous prouvons la propriété de stabilité sous mise à l'échelle de la distribution de Dirichlet de type 3. De plus, nous dérivons une formule qui permet de

calculer de nombreux fonctionnels, ce qui rend la distribution de Dirichlet de type 3 une distribution plus flexible [12].

3.1.1 Distribution de Dirichlet de type 3

Définition 3.1. Soient X_1, \dots, X_{n+1} des variables aléatoires indépendantes suivant des lois Gamma respectives Gamma (θ_1) , ..., Gamma (θ_{n+1}) , et définissons

$$\mathbf{S}_n = \left(\frac{X_1}{\sum_{m=1}^n X_m + 2X_{n+1}}, \dots, \frac{X_n}{\sum_{m=1}^n X_m + 2X_{n+1}} \right).$$

Alors, on dit que $\mathbf{S}_n := (S_1, \dots, S_n)$ est distribué selon la distribution de Dirichlet de type 3 notée $D_n^3(\boldsymbol{\theta}_n; \theta_{n+1})$ de paramètres $(\theta_1, \dots, \theta_{n+1})$ sur le support :

$$\Lambda_n = \left\{ 0 < s_m < 1, m = 1, \dots, n, \sum_{m=1}^n s_m < 1 \right\},$$

par la fonction de densité jointe

$$\begin{aligned} f_{S_1, \dots, S_n}(s_1, \dots, s_n) &= \frac{2^{\sum_{m=1}^n \theta_m} \Gamma(\sum_{m=1}^{n+1} \theta_m)}{\prod_{m=1}^{n+1} \Gamma(\theta_m)} \\ &\times \left(\prod_{m=1}^n s_m^{\theta_m - 1} \right) \left(1 - \sum_{m=1}^n s_m \right)^{\theta_{n+1} - 1} \left(1 + \sum_{m=1}^n s_m \right)^{-\sum_{m=1}^{n+1} \theta_m}. \end{aligned} \quad (3.1)$$

Définition 3.2. Soit $\mathbf{S}_n := (S_1, \dots, S_n)$ un vecteur aléatoire qui suit la distribution de Dirichlet de type 3. Alors, la fonction de moments jointe est définie par :

$$\begin{aligned} \mathbf{E} \left[\prod_{m=1}^n S_m^{q_m} \right] &= \frac{2^{-\theta_{n+1}} \prod_{m=1}^n (\theta_m)_{q_m}}{(\sum_{m=1}^{n+1} \theta_m)_{\sum_{m=1}^n q_m}} \\ &\times {}_2F_1 \left(\theta_{n+1}, \sum_{m=1}^{n+1} \theta_m; \sum_{m=1}^n (\theta_m + q_m) + \theta_{n+1}; \frac{1}{2} \right). \end{aligned} \quad (3.2)$$

Remarque 4. Il découle de l'équation (3.1) que la distribution de S_m possède la fonction de densité suivante (avec $0 < s_m < 1$)

$$\begin{aligned} f_{S_m}(s_m) &= \frac{2^{-\theta_{n+1}} \Gamma(\sum_{m=1}^{n+1} \theta_m)}{\Gamma(\theta_m) \Gamma(\sum_{m=1}^{n+1} \theta_m - \theta_m)} s_m^{\theta_m - 1} (1 - s_m)^{\sum_{m=1}^{n+1} \theta_m - \theta_m - 1} \\ &\times {}_2F_1 \left(\theta_{n+1}, \sum_{m=1}^{n+1} \theta_m; \sum_{l=1}^{n+1} \theta_l - \theta_m; \frac{1 - s_m}{2} \right). \end{aligned}$$

En théorie des probabilités bayésienne, si la distribution a posteriori est de la même famille de distributions de probabilité que la distribution a priori, la distribution a priori et la distribution a posteriori sont alors appelées distributions conjuguées, et la distribution à priori est appelée à priori conjuguée pour la fonction de vraisemblance.

Proposition 3.1. *La distribution de Dirichlet de type 3 est une distribution à priori conjuguée pour la distribution multinomiale.*

Preuve : Soit $\mathbf{X}_n := (X_1, \dots, X_n)$ et $\mathbf{S}_n := (S_1, \dots, S_n)$ deux vecteurs aléatoires distribués respectivement selon la distribution multinomiale et la distribution de Dirichlet de type 3, telles que :

$$p(x_1, \dots, x_n, k | s_1, \dots, s_n) = \binom{x_1 + \dots + x_n + k}{x_1, \dots, x_n, k} s_1^{x_1} \dots s_n^{x_n} (1 - \sum_{m=1}^n s_m)^k,$$

et

$$\begin{aligned} p(s_1, \dots, s_n) &= C(\theta_1, \dots, \theta_n, \theta_{n+1}, \sum_{m=1}^{n+1} \theta_m) \left(\prod_{m=1}^n s_m^{\theta_m - 1} \right) \\ &\quad \times \left(1 - \sum_{m=1}^n s_m \right)^{\theta_{n+1} - 1} \left(1 + \sum_{m=1}^n s_m \right)^{-\sum_{m=1}^{n+1} \theta_m} \\ &= \frac{2^{\sum_{m=1}^n \theta_m} \Gamma(\sum_{m=1}^{n+1} \theta_m)}{\prod_{m=1}^{n+1} \Gamma(\theta_m)} \left(\prod_{m=1}^n s_m^{\theta_m - 1} \right) \\ &\quad \times \left(1 - \sum_{m=1}^n s_m \right)^{\theta_{n+1} - 1} \left(1 + \sum_{m=1}^n s_m \right)^{-\sum_{m=1}^{n+1} \theta_m}, \end{aligned}$$

où $0 < s_m < 1$, $m = 1, \dots, n$, $\sum_{m=1}^n s_m < 1$, et

$$[C(\theta_1, \dots, \theta_n, \theta_{n+1}, \alpha)]^{-1} = \frac{\prod_{m=1}^{n+1} \Gamma(\theta_m)}{\Gamma(\sum_{m=1}^{n+1} \theta_m)} {}_2F_1 \left(\sum_{m=1}^n \theta_m, \alpha; \sum_{m=1}^{n+1} \theta_m; -1 \right).$$

Alors,

$$\begin{aligned} p(s_1, \dots, s_n | x_1, \dots, x_n, k) &= C(\theta_1 + x_1, \dots, \theta_n + x_n, \theta_{n+1} + k, \sum_{m=1}^{n+1} \theta_m) \\ &\quad \times \left(\prod_{m=1}^n s_m^{\theta_m + x_m - 1} \right) \left(1 - \sum_{m=1}^n s_m \right)^{\theta_{n+1} + k - 1} \left(1 + \sum_{m=1}^n s_m \right)^{-\sum_{m=1}^{n+1} \theta_m}. \end{aligned}$$

Ainsi, la distribution de Dirichlet de type 3 est une distribution à priori conjuguée pour la distribution multinomiale (voir [21]).

3.1.2 Structure de RAM de la distribution de Dirichlet de type 3

L'objectif de cette section est de présenter une propriété fondamentale de la distribution de Dirichlet de type 3 (voir [26]).

Définition 3.3. *Un vecteur aléatoire (W_1, \dots, W_n) avec des valeurs dans $[0, 1]^n$ tel que $\sum_{m=1}^n W_m = \tilde{W}$ a une représentation RAM si et seulement si il existe $(n-1)$ variables aléatoires indépendantes à valeurs dans $[0, 1]$, B_1, \dots, B_{n-1} , mutuellement indépendantes avec \tilde{W} , telles que $W_1 = \tilde{W} \cdot B_1$ et*

$$\begin{aligned} W_m &= \tilde{W} \cdot B_m \prod_{k=1}^{m-1} (1 - B_k), \quad m = 2, \dots, n-1, \\ W_n &= \tilde{W} \cdot \prod_{m=1}^{n-1} (1 - B_m). \end{aligned}$$

Cette représentation s'appelle la structure de RAM et est utilisé en statistiques bayésiennes non paramétriques (voir par exemple [17]). À cet égard, nous avons le résultat suivant :

Théorème 3.2. *Soient A_1, \dots, A_{n-1} $(n-1)$ variables aléatoires indépendantes avec $A_m \stackrel{d}{\sim} \beta^3(\theta_m, \sum_{l=1}^n \theta_l - \sum_{l=1}^m \theta_l)$, $m = 1, \dots, n-1$. Soient $B_m := 2A_m/(1 + A_m)$; $m = 1, \dots, n-1$, et $\tilde{S} \stackrel{d}{\sim} \beta^3(\sum_{m=1}^n \theta_m, \theta_{n+1})$ une variable aléatoire indépendante de (B_1, \dots, B_{n-1}) . Avec $\prod_{k=1}^0 (1 - B_k) := 1$, définissons*

$$S_m := \tilde{S} \cdot B_m \prod_{k=1}^{m-1} (1 - B_k), \quad m = 1, \dots, n-1, \quad (3.3)$$

$$S_n = \tilde{S} - \sum_{m=1}^{n-1} S_m = \tilde{S} \cdot \prod_{m=1}^{n-1} (1 - B_m).$$

Alors, $\mathbf{S}_n \stackrel{d}{\sim} D_n^3(\boldsymbol{\theta}_n; \theta_{n+1})$.

Preuve. Tout d'abord, en utilisant la propriété d'indépendance

$$\mathbf{E} \left[\prod_{m=1}^n S_m^{q_m} \right] = \left[\prod_{m=1}^{n-1} \mathbf{E} \left(B_m^{q_m} (1 - B_m)^{\sum_{l=m+1}^n q_l} \right) \right] \mathbf{E} \left[\tilde{S}^{\sum_{m=1}^n q_m} \right].$$

De plus, il est facile de déduire que si $A_m \stackrel{d}{\sim} \beta^3(\theta_m, \sum_{l=1}^n \theta_l - \sum_{l=1}^m \theta_l)$, $m = 1, \dots, n$, alors $B_m \stackrel{d}{\sim} \beta^1(\theta_m, \sum_{l=1}^n \theta_l - \sum_{l=1}^m \theta_l)$, $m = 1, \dots, n$ (voir par exemple [10]). De là, nous

pouvons vérifier que

$$\mathbf{E} \left(B_m^{q_m} (1 - B_m)^{\sum_{l=m+1}^n q_l} \right) = \frac{(\theta_m)_{q_m} \Gamma \left(\sum_{l=1}^n \theta_l - \sum_{l=1}^{m-1} \theta_l \right)}{\Gamma \left(\sum_{l=1}^n \theta_l - \sum_{l=1}^m \theta_l \right)} \times \frac{\Gamma \left(\sum_{l=1}^n \theta_l - \sum_{l=1}^m \theta_l + \sum_{l=m+1}^n q_l \right)}{\Gamma \left(\sum_{l=1}^n \theta_l - \sum_{l=1}^{m-1} \theta_l + \sum_{l=m}^n q_l \right)}.$$

D'autre part, la fonction de moment de \tilde{S} est donnée par

$$\mathbf{E} \left[\tilde{S}^{\sum_{m=1}^n q_m} \right] = \frac{2^{-\theta_{n+1}} \Gamma \left(\sum_{m=1}^{n+1} \theta_m \right) \Gamma \left(\sum_{m=1}^n (\theta_m + q_m) \right)}{\Gamma \left(\sum_{m=1}^n \theta_m \right) \Gamma \left(\sum_{m=1}^n (\theta_m + q_m) + \theta_{n+1} \right)} \times {}_2F_1 \left(\theta_{n+1}, \sum_{m=1}^{n+1} \theta_m; \sum_{m=1}^n (\theta_m + q_m) + \theta_{n+1}; \frac{1}{2} \right).$$

Par conséquent, nous obtenons

$$\mathbf{E} \left[\prod_{m=1}^n S_m^{q_m} \right] = \frac{\Gamma \left(\sum_{m=1}^n \theta_m \right) \prod_{m=1}^n (\theta_m)_{q_m}}{\Gamma \left(\sum_{m=1}^n (\theta_m + q_m) \right)} \times \frac{2^{-\theta_{n+1}} \Gamma \left(\sum_{m=1}^{n+1} \theta_m \right) \Gamma \left(\sum_{m=1}^n (\theta_m + q_m) \right)}{\Gamma \left(\sum_{m=1}^n \theta_m \right) \Gamma \left(\sum_{m=1}^n (\theta_m + q_m) + \theta_{n+1} \right)} \times {}_2F_1 \left(\theta_{n+1}, \sum_{m=1}^{n+1} \theta_m; \sum_{m=1}^n (\theta_m + q_m) + \theta_{n+1}; \frac{1}{2} \right).$$

Cela confirme que $\mathbf{S}_n \stackrel{d}{\sim} D_n^3(\boldsymbol{\theta}_n; \theta_{n+1})$.

3.1.3 Propriété de Scaling

Dans cette section, notre attention principale est portée sur la propriété de mise à l'échelle de la distribution de Dirichlet de type 3 qui porte le nom de propriété de scaling.

Définition 3.4. Soient X_1, \dots, X_{n+1} ($n+1$) variables aléatoires indépendantes distribuées respectivement selon $\text{Gamma}(\theta_1), \dots, \text{Gamma}(\theta_{n+1})$ nous définissons

$$\mathbf{S}_n(x) := \left(\frac{X_1(x)}{\sum_{m=1}^n X_m + 2X_{n+1}}, \dots, \frac{X_n(x)}{\sum_{m=1}^n X_m + 2X_{n+1}} \right),$$

où $X_m(x) = x X_m$; $m = 1, \dots, n$. Dans ce cas, la distribution de $\mathbf{S}_n(x) := (S_m(x); m = 1, \dots, n)$ est exprimée par (avec $0 < \sum_{m=1}^n s_m < x$)

$$f_{S_1(x), \dots, S_n(x)}(s_1, \dots, s_n) = \frac{2^{\sum_{m=1}^n \theta_m} \Gamma \left(\sum_{m=1}^{n+1} \theta_m \right)}{x^{-1} \prod_{m=1}^{n+1} \Gamma(\theta_m)} \left(\prod_{m=1}^n s_m^{\theta_m - 1} \right)$$

$$\times \left(x - \sum_{m=1}^n s_m \right)^{\theta_{n+1}-1} \left(x + \sum_{m=1}^n s_m \right)^{-\sum_{m=1}^{n+1} \theta_m}. \quad (3.4)$$

Proposition 3.3. (i) Soit $\mathbf{S}_n \stackrel{d}{\sim} D_n^3(\boldsymbol{\theta}_n; \theta_{n+1})$ la distribution de type 3 de Dirichlet. Alors, elle peut également être définie de manière conditionnelle par

$$\mathbf{S}_n \stackrel{d}{=} \left(\frac{X_1}{\sum_{m=1}^n X_m + 2X_{n+1}}, \dots, \frac{X_n}{\sum_{m=1}^n X_m + 2X_{n+1}} \mid \tilde{X} = 1 \right).$$

(ii) La propriété de mise à l'échelle appelée propriété de scaling est donnée par $xS_m \stackrel{d}{=} S_m(x)$, $m = 1, \dots, n$.

Preuve. (i) En utilisant l'indépendance de X_m , la fonction de densité jointe de X_1, \dots, X_{n+1} est exprimée par

$$\frac{e^{-\sum_{m=1}^{n+1} x_m}}{\prod_{m=1}^{n+1} \Gamma(\theta_m)} \prod_{m=1}^{n+1} x_m^{\theta_m-1}. \quad (3.5)$$

En effectuant la transformation $s_m = x_m/(x_{n+1} + \tilde{x})$; $m = 1, \dots, n$, où $\tilde{x} = \sum_{m=1}^{n+1} x_m$ et le Jacobien

$$J(x_1, \dots, x_{n+1} \rightarrow s_1, \dots, s_n, \tilde{x}) = \frac{(2\tilde{x})^n}{(1 + \sum_{m=1}^n s_m)^{n+1}},$$

dans l'Eq. (3.5), la fonction de densité jointe de $S_1, \dots, S_n, \tilde{X}$ est donnée par

$$e^{-\tilde{x}\tilde{x}^{\sum_{m=1}^{n+1} \theta_m-1}} \cdot \frac{2^{\sum_{m=1}^n \theta_m}}{\prod_{m=1}^{n+1} \Gamma(\theta_m)} \left(\prod_{m=1}^n s_m^{\theta_m-1} \right) \\ \times \left(1 - \sum_{m=1}^n s_m \right)^{\theta_{n+1}-1} \left(1 + \sum_{m=1}^n s_m \right)^{-\sum_{m=1}^{n+1} \theta_m}.$$

Ainsi, nous concluons que $\mathbf{S}_n = \left(\frac{X_1}{\sum_{m=1}^n X_m + 2X_{n+1}}, \dots, \frac{X_n}{\sum_{m=1}^n X_m + 2X_{n+1}} \right)$ et \tilde{X} sont indépendants. Par conséquent, la fonction de densité jointe de

$$\left(\frac{X_1}{\sum_{m=1}^n X_m + 2X_{n+1}}, \dots, \frac{X_n}{\sum_{m=1}^n X_m + 2X_{n+1}} \right)$$

conditionnée à $\tilde{X} = 1$ est trouvée être la même que celle de $\mathbf{S}_n \stackrel{d}{\sim} D_n^3(\boldsymbol{\theta}_n; \theta_{n+1})$. La première déclaration suit.

(ii) De même, on peut vérifier que

$$\left(x \frac{X_1}{\sum_{m=1}^n X_m + 2X_{n+1}}, \dots, x \frac{X_n}{\sum_{m=1}^n X_m + 2X_{n+1}} \mid x\tilde{X} = x \right) \stackrel{d}{=} \mathbf{S}_n(x).$$

Ainsi, nous obtenons la propriété de mise à l'échelle $xS_m \stackrel{d}{=} S_m(x)$, $m = 1, \dots, n$.

3.1.4 Calcul de fonctionnelle

Dans cette partie, nous donnons une formule qui permet de calculer le fonctionnelle de Dirichlet de type 3 en fonction des fonctionnelle de plus simple de gamma.

Théorème 3.4. *Considérons la distribution de Dirichlet $\mathbf{Y}_n \stackrel{d}{\sim} D_n(\boldsymbol{\theta}_n)$. Soit $f : \mathbb{R} \rightarrow \mathbb{R}$ une fonction mesurable de Borel pour laquelle*

$$\int_0^\infty \mathbb{E}(|f(\mathbf{Y}_n(x))|) x^{\sum_{m=1}^n \theta_m - 1} e^{-yx} dx < \infty.$$

Alors, avec $\mathbf{X}_n(y) := (X_m(y); m = 1, \dots, n)$, n variables aléatoires indépendantes définies par $X_m(y) = (1/y) X_m$, $y > 0$, $m = 1, \dots, n$ où $X_m \stackrel{d}{\sim} \gamma_{\theta_m}$, nous avons

$$\int_0^\infty \mathbb{E}(f(\mathbf{Y}_n(x))) x^{\sum_{m=1}^n \theta_m - 1} e^{-yx} dx = \Gamma\left(\sum_{m=1}^n \theta_m\right) y^{-\sum_{m=1}^n \theta_m} \mathbb{E}(f(\mathbf{X}_n(y))). \quad (3.6)$$

En conséquence, on a le théorème suivant :

Théorème 3.5. *Considérons la distribution de type 3 de Dirichlet $\mathbf{S}_n \stackrel{d}{\sim} D_n^3(\boldsymbol{\theta}_n; \theta_{n+1})$.*

(i) Avec $0 < s < 1$, soit $\mathbf{X}_n\left(\frac{y}{1-s}\right) = (X_m\left(\frac{y}{1-s}\right); m = 1, \dots, n)$, n variables aléatoires indépendantes définies comme précédemment dans le Théorème 3.4. Alors, pour $p > 0$ et une fonction mesurable de Borel $f : \mathbb{R}^n \rightarrow \mathbb{R}$ satisfaisant

$$\int_0^1 \mathbf{E}\left(\left|f\left(\mathbf{X}_n\left(\frac{y}{1-s}\right)\right)\right|\right) s^{\theta_{n+1}-1} (1-s)^{\sum_{m=1}^n \theta_m - 1} \left(1 - \frac{s}{2}\right)^{-\sum_{m=1}^{n+1} \theta_m} ds < \infty,$$

on a

$$\begin{aligned} \mathbf{E}(f(\mathbf{S}_n(p))) &= \frac{2^{-\theta_{n+1}} \Gamma(\sum_{m=1}^{n+1} \theta_m)}{p^{\sum_{m=1}^n \theta_m - 1} \Gamma(\theta_{n+1})} \\ &\times \frac{1}{2\pi i} \int_L e^{py} y^{-\sum_{m=1}^n \theta_m} \int_0^1 \mathbf{E}\left(f\left(\mathbf{X}_n\left(\frac{y}{1-s}\right)\right)\right) \\ &\times s^{\theta_{n+1}-1} (1-s)^{\sum_{m=1}^n \theta_m - 1} \left(1 - \frac{s}{2}\right)^{-\sum_{m=1}^{n+1} \theta_m} ds dy, \end{aligned} \quad (3.7)$$

(ii) Si f est homogène de degré d , et si $\mathbf{E}(|f(\mathbf{X}_n)|) < \infty$, alors, avec $\mathbf{X}_n := (X_1, \dots, X_n)$,

$$\mathbf{E}(f(\mathbf{S}_n)) = \frac{2^{-\theta_{n+1}}}{(\sum_{m=1}^{n+1} \theta_m)_d} {}_2F_1\left(\theta_{n+1}, \sum_{m=1}^{n+1} \theta_m; \sum_{m=1}^{n+1} \theta_m + d; \frac{1}{2}\right) \mathbf{E}(f(\mathbf{X}_n)). \quad (3.8)$$

3.2 Application de la distribution de Dirichlet de type 3

Dans de nombreux cas, les biologistes s'intéressent aux proportions en raison des difficultés pratiques à mesurer les nombres réels. Nous prenons en compte les données de proportions de sérum sanguin (pré-albumine, albumine et globuline) chez des canetons de Pékin blancs âgés de 3 semaines. Nous utilisons 23 ensembles de données, chaque ensemble correspondant à un régime alimentaire différent. Les mesures des proportions de sérum sanguin sont affichées dans [20].

Pour confirmer les performances de la distribution de Dirichlet de type 3, nous présentons un exemple concret. Soit X_m une variable aléatoire dont la valeur est proportionnelle au type particulier de niveau de sérum sanguin. Alors, $S_m = X_m / (\sum_{l=1}^2 X_l + 2X_3)$ correspond aux proportions. De plus, nous supposons que X_m peut être considéré comme étant indépendant les uns des autres et de $\sum_{l=1}^2 X_l + 2X_3$. Par conséquent, le vecteur $\mathbf{S}_2 = (S_1, S_2)$ suivrait la distribution de Dirichlet de type 3 avec les paramètres θ_1, θ_2 et θ_3 .

Ces paramètres inconnus sont estimés en utilisant la méthode du maximum de vraisemblance et en mettant en œuvre la méthode de Fisher scoring (voir [22] et [30]).

$$\frac{\partial \log L}{\partial \theta_1} = n \log(2) + \psi \sum_{m=1}^3 \theta_m - \psi(\theta_1) + \sum_{i=1}^n \log s_{i1} - \sum_{i=1}^{k-1} \log(1 + s_{i1} + s_{i2}) \quad (3.9)$$

$$\frac{\partial \log L}{\partial \theta_2} = n \log(2) + \psi \sum_{m=1}^3 \theta_m - \psi(\theta_2) + \sum_{i=1}^n \log s_{i2} - \sum_{i=1}^{k-1} \log(1 + s_{i1} + s_{i2}) \quad (3.10)$$

$$\frac{\partial \log L}{\partial \theta_3} = n \log(2) + \psi \sum_{m=1}^3 \theta_m - \psi(\theta_3) + \sum_{i=1}^n \log s_{i3} - \sum_{i=1}^{k-1} \log(1 + s_{i1} + s_{i2}) \quad (3.11)$$

La matrice d'information relative à la distribution est trouvée comme suit :

$$I_{i,j} = n\psi'(\theta_i) - n\psi' \left(\sum_{m=1}^3 \theta_m \right); \quad i = j. \quad (3.12)$$

$$I_{i,j} = -n\psi' \left(\sum_{m=1}^3 \theta_m \right); \quad i \neq j. \quad (3.13)$$

En implémentant l'algorithme de Fisher sur le logiciel MATLAB (R2015a), on trouve que les valeurs estimées des paramètres θ_1 , θ_2 et θ_3 sont comme donnés dans le tableau (3.1).

TABLE 3.1 – Valeurs estimées de θ_1 , θ_2 et θ_3 .

$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$
3.6112	23.2517	12.6186

La matrice de variance-covariance correspondante est donnée par :

$$\begin{pmatrix} 0.5796 & 3.2148 & 1.7131 \\ 3.2148 & 24.3069 & 12.4250 \\ 1.7131 & 12.4250 & 7.1480 \end{pmatrix}$$

La figure 3.1 représente le programme exécuté sur Matlab version R2015a, qui a permis de trouver les estimations des paramètres et la matrice de variance-covariance.

```

1      clc; clear all;
2      %~~~~~ Importtion de données ~~~~~
3      T=readtable('fish_alg.xlsx');
4      S=[T.S1 T.S2 T.S3]
5      %~~~~~ Initialisation ~~~~~
6      [n,k]=size(S)
7      nbr_iteration=100;
8      err=0.000001;
9      it=0;
10     grad=zeros(1,k);
11     s_min=min( min(S) )
12     par_in=[s_min;s_min;s_min];
13
14     for r=1:nbr_iteration
15         it=it+1;
16         %~~~~~ Gradient ~~~~~
17
18         for i=1:k-1
19             V(i)=1+S(i,1)+S(i,2);
20         end
21         s1=sum(log(V));
22         for t=1:k-1
23             grad(t)=n*(log(2)+ psi(sum(par_in))-psi(par_in(t))+sum(log(S(:,t)))-s1;
24         end
25         grad(3)=n*(psi(sum(par_in))-psi(par_in(3))+sum(log(S(:,3)))-s1;
26
27         %~~~~~ Matrice d'information ~~~~~
28         for i=1:k
29             for j=1:k
30                 if i==j
31                     inf(i,j)=n*psi(1,par_in(i))-n*psi(1,sum(par_in));
32                 else
33                     inf(i,j)=-n*psi(1,sum(par_in));
34                 end
35             end
36         end
37         %~~~~~ Matrice de variance-covariance ~~~~~
38         COV=inv(inf);
39         %~~~~~ Algorithme de Fisher ~~~~~
40         par=par_in;
41         par_in=par_in+COV*grad';
42         %~~~~~ Test de convergence ~~~~~
43         if abs(par_in(1)-par(1))<err & abs(par_in(2)-par(2))<err
44             & abs(par_in(2)-par(2))<err
45             break
46         end
47     end
48     display(it)
49     display(par_in)
50     %~~~~~ Matrice d'information calculée aux estimations ~~~~~
51     for i=1:k
52         for j=1:k
53             if i==j
54                 inf(i,j)=n*psi(1,par_in(i))-n*psi(1,sum(par_in));
55             else
56                 inf(i,j)=-n*psi(1,sum(par_in));
57             end
58         end
59     end
60     %~~~~~ Matrice estimée de variance-covariance ~~~~~
61     Cov_mat=inv(inf)

```

FIGURE 3.1 – Programme d'estimation des paramètres et la matrice de var-cov

3.3 Conclusion

Le chapitre 3 explore la distribution de Dirichlet de type 3, mettant en évidence ses propriétés et son utilité pour modéliser des données compositionnelles complexes, contrairement à la distribution de Dirichlet classique qui présente une faible structure de dépendance. Cette distribution est particulièrement avantageuse en statistique bayésienne, où elle sert de distribution a priori conjuguée pour la distribution multinomiale. Les propriétés clés étudiées incluent le modèle d'allocation résiduelle (RAM), la stabilité sous mise à l'échelle, et une formule pour calculer divers fonctionnels, démontrant ainsi sa flexibilité et son applicabilité étendue.

L'application de la distribution de Dirichlet de type 3 aux proportions de sérum sanguin chez des canetons de Pékin blancs âgés de trois semaines offre des résultats prometteurs quant à sa capacité à modéliser efficacement ces données. Les estimations des paramètres $\hat{\theta}_1$, $\hat{\theta}_2$ et $\hat{\theta}_3$, obtenues par la méthode du maximum de vraisemblance et la méthode de Fisher scoring, sont respectivement de 3.6112, 23.2517 et 12.6186, confirmant ainsi la fiabilité du modèle dans ce contexte spécifique. De plus, la matrice de variance-covariance des paramètres estimés fournit des indications sur la précision de ces estimations, renforçant la validité des résultats obtenus. Ces résultats soulignent la pertinence de la distribution de Dirichlet de type 3 dans la modélisation des données positives, ouvrant ainsi la voie à son utilisation dans d'autres domaines de la biologie ainsi que dans d'autres disciplines où la modélisation précise des proportions est cruciale.

Conclusion et perspectives

La recherche présentée dans ce mémoire explore en profondeur les fondements théoriques et les applications pratiques des distributions de probabilité dans le contexte de la modélisation des données positives. En commençant par les bases essentielles de la théorie des probabilités dans le premier chapitre, nous avons posé les fondements nécessaires à la compréhension des développements ultérieurs. La discussion sur les tribus, les mesures, la mesure de Lebesgue, ainsi que les lois de probabilité multivariées et l'estimation des paramètres par maximum de vraisemblance a jeté les bases théoriques indispensables.

Le deuxième chapitre a approfondi notre compréhension en examinant les distributions de Liouville multivariées, qui jouent un rôle crucial dans la modélisation des données compositionnelles. En mettant en lumière la relation entre la distribution de Liouville et d'autres distributions telles que la distribution de Dirichlet, ce chapitre a facilité la transition vers des applications pratiques en statistique et en modélisation des données.

Enfin, le chapitre 3 a constitué le cœur de cette étude en explorant les distributions de Dirichlet de type 3, en mettant en évidence leur propriété de scaling et en proposant des méthodes efficaces pour calculer leurs fonctionnelles. En illustrant une application concrète dans le domaine de la biologie, notamment dans l'analyse des proportions de sérum sanguin chez des canetons de Pékin blancs, ce chapitre a démontré l'importance pratique de ces modèles dans la résolution de problèmes réels.

Cette recherche apporte une contribution importante à la modélisation des données positives en proposant des méthodes théoriques et en montrant leur utilité pratique. Cette étude crée de nouvelles opportunités de recherche et d'application dans divers domaines, de la biologie à la finance, offrant ainsi des perspectives prometteuses pour l'avancement des connaissances et la résolution de problèmes complexes.

Bibliographie

- [1] Aitchison, J. (1981). A new approach to null correlations of proportions. *J. Int. Assoc. Math. Geol.*, Springer Nature, 13 :175–189.
- [2] Arashi, M., Bekker, A., de Waal, D., and Makgai, S. (2020). Constructing multivariate distributions via the Dirichlet generator. *Comput. Method. Statist. Biostatist.*, Springer, pages 159–186.
- [3] Bdiri, T., and Bouguila, N. (2012). Positive vectors clustering using inverted Dirichlet finite mixture models. *Expert Systems with Applications*, 39(2) :1869–1882. DOI : [10.1016/j.eswa.2011.08.063](https://doi.org/10.1016/j.eswa.2011.08.063).
- [4] Bdiri, T., and Bouguila, N. (2011). Learning Inverted Dirichlet Mixtures for Positive Data Clustering. In *Rough Sets, Fuzzy Sets, Data Mining and Granular Computing*, volume 6743 of *Lecture Notes in Computer Science*, pages 265–272. DOI : [10.1007/978-3-642-21881-1_42](https://doi.org/10.1007/978-3-642-21881-1_42).
- [5] Bdiri, T., and Bouguila, N. (2011). An Infinite Mixture of Inverted Dirichlet Distributions. In *Neural Information Processing*, volume 7063 of *Lecture Notes in Computer Science*, pages 71–78. DOI : [10.1007/978-3-642-24958-7_9](https://doi.org/10.1007/978-3-642-24958-7_9).
- [6] Bdiri, T., and Bouguila, N. (2013). Bayesian learning of inverted Dirichlet mixtures for SVM kernels generation. *Neural Computing and Applications*, 23(5) :1443–1458. DOI : [10.1007/s00521-012-1094-z](https://doi.org/10.1007/s00521-012-1094-z).
- [7] Bdiri, T., Bouguila, N., and Ziou, D. (2013). Visual Scenes Categorization Using a Flexible Hierarchical Mixture Model Supporting Users Ontology. In *2013 IEEE 25th International Conference on Tools with Artificial Intelligence*, pages 262–267. DOI : [10.1109/ICTAI.2013.48](https://doi.org/10.1109/ICTAI.2013.48).
- [8] Bdiri, T., Bouguila, N., and Ziou, D. (2014). Object clustering and recognition using multi-finite mixtures for semantic classes and hierarchy modeling. *Expert Systems with Applications*, 41(4) :1218–1235. DOI : [10.1016/j.eswa.2013.08.005](https://doi.org/10.1016/j.eswa.2013.08.005).

- [9] Bouguila, N., Ziou, D., and Vaillancourt, J. (2003). Novel mixtures based on the Dirichlet distribution : application to data and image classification. In *International Workshop on Machine Learning and Data Mining in Pattern Recognition*, Springer Berlin Heidelberg, pages 172–181.
- [10] Cardeno, L., Nagar, D. K., and Sanchez, L. E. (2005). Beta type 3 distribution and its multivariate generalization. *Tamsui Oxf. j. math. sci.*, **21**(2) : pages 225–242.
- [11] Deschamps, P. (2004). Cours d'économétrie. Université de Fribourg.
- [12] Devroye, L. (1986). Non-Uniform Random Variable Generation. *Springer-Verlag*, New York.
- [13] Ghorbel, M. (2009). On the inverted Dirichlet distribution. *Commun. Stat. Theory and Methods*, Taylor & Francis, 39(1) :21–37. DOI : <https://doi:10.1080/03610920802627062>.
- [14] Gupta, R.D., and Richards, D.St P. (1987). Multivariate Liouville distributions. *J. Multivar. Anal.*, Elsevier, 23(2) :233–256.
- [15] Gupta, R.D., and Kabe, D.G. (1999). On multivariate Liouville distribution. *Metron - International Journal of Statistics*, University of Rome, vol. 0(1-2), pages 173-179.
- [16] Huillet, T. (2007). Ewens sampling formulae with and without selection. *J. Comput. Appl. Math.*, Elsevier, 206(2) :755–773.
- [17] Ishwaran, R., James, L.F. (2001). Gibbs sampling methods for stick-breaking priors. *J. Amer. Statist. Assoc.*, **96**(453) : 161–173.
- [18] Marshall, A.W., Olkin, I., and Arnold, B.C. (1979). Inequalities : theory of majorization and its applications. Springer.
- [19] Mayamol, E.V. (2010). Some properties of a generalized type-1 Dirichlet distribution. *Statistica*, 70(1) :95–111. DOI : <https://doi.org/10.6092/issn.1973-2201/3571>.
- [20] Mosimann, J.E. (1962). On the compound multinomial distribution, the multivariate β -distribution, and correlations among proportions. *Biometrika*, **49** : pages 65–82.
- [21] Nagar, D.K., Bedoya-Valencia, D., and Nadarajah, S. (2015). Multivariate generalization of the Gauss hypergeometric distribution. *Hacet. J. Math. Stat.*, **44**(4) : 933–948.

- [22] Narayanan, A. (1992). A note on parameter estimation in the multivariate beta distribution. *Computers Math. Applic.*, **24** : 11–17.
- [23] Ongaro, A., and Migliorati, S. (2013). A generalization of the Dirichlet distribution. *J. Multivar. Anal.*, **114** : 412–426.
- [24] Olkin, I. (1979). Matrix extensions of Liouville-Dirichlet-type integrals. *Linear Algebra and its Applications*, Elsevier, pages 28 :155–160.
- [25] Park, K.I. (2018). *Fundamentals of Probability and Stochastic Processes with Applications to Communications*. Springer. ISBN 978-3-319-68074-3.
- [26] Patil, G.P., and Taillie, C. (1977). Diversity as a concept and its implications for random communities. *Bull. Int. Statist. Inst.*, **47** : 497–515.
- [27] Purba, S.A. (2021). The Application of Fisher Scoring Algorithm on Parameter Estimation of Normal Distributed Data. *J. Res. Math. Trends and Technology*, 3(1) :20–24. DOI : [https://doi.org/\[DOI_number\]](https://doi.org/[DOI_number]).
- [28] Rayens, W.S., and Srinivasan, C. (1994). Dependence properties of generalized Liouville distributions on the simplex. *J. Am. Stat. Assoc.*, **89** : 1465–1470.
- [29] Revuz, D. (1997). *Mesure et intégration*. Paris, Hermann, 212 p. (ISBN 2-7056-6350-9).
- [30] Ronning, G. (1989). Maximum likelihood estimation of Dirichlet distributions. *J. Stat. Comput. Simul.*, **32**(4) : 215–221.