Caractérisation de la diversité d'une population à partir de mesures quantifiées d'un modèle non-linéaire. Application à la plongée hyperbare

Youssef Bennani

Université de Nice - Sophia Antipolis

ÉCOLE DOCTORALE STIC

Sciences et Technologies de l'Information et de la Communication

THÈSE

pour obtenir le titre de

Docteur en Sciences

de l'Université de Nice - Sophia Antipolis

Mention : Automatique, Traitement du Signal et des Images

présentée et soutenue par

Youssef BENNANI

CARACTÉRISATION DE LA DIVERSITÉ D'UNE POPULATION À PARTIR DE MESURES QUANTIFIÉES D'UN MODÈLE NON-LINÉAIRE. APPLICATION À LA PLONGÉE HYPERBARE.

Thèse dirigée par Luc PRONZATO et Maria João RENDAS

Soutenue publiquement le 10 décembre 2015 devant le jury composé de

Ali MOHAMMAD-DJAFARI	Directeur de recherche CNRS (Supélec - Paris Sud)	Rapporteur
Julien HUGON	Docteur - Ingénieur de recherche	Examinateur
Sylvie ICART	Maître de Conférence (Nice - Sophia Antipolis)	Examinateur
Adeline LECLERCQ SAMSON	Professeur des Universités (Joseph Fourier - Grenoble)	Rapporteur
Luc PRONZATO	Directeur de recherche CNRS (Nice - Sophia Antipolis)	Directeur de thèse
Maria João RENDAS	Chargée de recherche CNRS (Nice - Sophia Antipolis)	Co-directrice de thèse

TABLE DES MATIÈRES

notation	signification	page
Θ	domaine paramétrique contenant le paramètre d'intérêt θ.	8
π_θ	densité de probabilité du paramètre $\theta \in \Theta$.	8
$\mathscr{P}$	ensemble des distribution de probabilité sur Θ.	8
$\{\mathcal{R}^j\}_{j=1}^J$	liste de J partitions de Θ.	9
$\left\{\mathcal{R}_\ell^j\right\}_{\ell=1}^{L^j}$	liste des L^j éléments de la partition $\mathcal{R}^j$.	9
$\mathbf{X}_n$	jeu de données censurées de taille n.	10
$p_{\pi,\mathcal{R}}$	loi de probabilité induite par la distribution π_θ.	10
$\mathcal{S}^{L+}$	simplexe probabiliste de taille L.	10
$\{\mathcal{Q}_m^J\}_{m=1}^M$	liste des régions élémentaires.	10
$\mathscr{L}(\pi, \mathbf{X}_n)$	vraisemblance d'une distribution π quand on observe $\mathbf{X}_n$.	12
$\mathcal{L}(\pi, \mathbf{X}_n)$	log-vraisemblance d'une distribution π quand on observe $\mathbf{X}_n$.	12
$\hat{\pi}^{\mathcal{L}}$	ENPMV.	12
$\hat{\mathbf{w}}^{\mathcal{L}}$	ENPMV considéré dans le simplexe $\mathcal{S}^{M+}$.	12
n_ℓ^j	nombre de fois où la région $\mathcal{R}_\ell^j$ a été observée.	13
$\mathbf{B}$	matrice binaire indiquant les intersections entre régions de censure et régions élémentaires.	13
$\mathbf{n}$	vecteur résultant de la concaténation de n_ℓ^j.	13
$\mathbf{q}_j(\mathbf{n})$	vecteur des fréquences d'observation des éléments de $\mathcal{R}^j$.	15
$\mathbf{z}_j(\mathbf{w})$	vecteur des fréquences estimées par $\mathbf{w}$.	15
$H_1(\cdot)$	entropie de Shannon.	15
$D_{KL}(\cdot\|\cdot)$	divergence de Kullback-Leibler.	15
$\mathbf{q}(\mathbf{n})$	vecteur résultant de la cncaténation des vecteurs $\mathbf{q}_j(\mathbf{n})$.	15
$H_\alpha(\cdot)$	entropie de Rényi.	19
$\hat{\pi}_\star^{\mathcal{L}}$	ENPMV maximisant l'entropie de Rényi d'ordre 2.	20
$h^2(\cdot,\cdot)$	distance de Hellinger.	22
$\mathcal{G}(V,E)$	graphe dont V est l'ensemble des sommets et E est l'ensemble des arêtes.	25
$\mathcal{H}(V,E)$	hyper-graphe dont V est l'ensemble des sommets et E. est l'ensemble des hyper-arêtes.	25
$\mathcal{H}_{\max}(\mathcal{G})$	hyper-graphe dont les hyper-arrêtes sont les cliques maximales de $\mathcal{G}$.	26
$S_{\mathcal{L}}^\star$	support de l'ENPMV.	28
$\mathrm{P}(H,\mathscr{C})$	problème de maximisation de la fonction H sous les contraintes $\mathscr{C}$.	60
$\hat{\pi}_{\mathscr{C}}^H$	solution du problème $\mathrm{P}(H,\mathscr{C})$.	60
$Q(\mathbf{g})$	famille des densités exponentielles relatives à l'ensemble des fonctions $\mathbf{g}$.	61
$\epsilon_\star$	le plus petit niveau de relaxation assurant l'existance de l'estimateur MaxEnt	69
ϵ^{MV}	le niveau de relaxation correspondant à l'estimateur MaxEnt-MV.	69
$\hat{\pi}_\epsilon^H$	estimateur MaxEnt corrspondant au niveau de relaxation ϵ.	69
$\hat{\pi}_{\epsilon^{MV}}^H$	estimateur MaxEnt-MV.	69

- CHAPITRE 1 -

INTRODUCTION

1.1 POSITION DU PROBLÈME

Cette thèse propose un nouveau critère pour l'estimation de densités de probabilité à partir d'observations censurées.

Il s'agit d'un problème classique en analyse statistique. Il survient, par exemple, quand on souhaite déterminer la distribution de l'instant l'occurrence $\tau \geqslant 0$ d'un événement d'intérêt (la guérison d'un patient dans des essais thérapeutiques, la défaillance d'une pièce dans des études de fiabilité, ...) à partir d'observations réalisées à une série d'instants où l'on vérifie si l'événement s'est déjà produit ou pas. Pour chaque entité i observée (patient, pièce,...) l'ensemble des observations réalisées à des instants $t_j^i, j = 1, 2, \ldots$, permet uniquement de déduire que l'événement d'intérêt s'est produit entre les deux instants de mesure t_ℓ^i et $t_{\ell+1}^i$ où l'observation a changé. Dans les cas évoqués ci-dessus on parlera de censure par des intervalles réels, contenus dans $\mathbb{R}^+$. Ce type d'observations peut d'une façon équivalente être modélisé par des opérateurs de quantification,

$$Q^i(\tau) = \ell \Leftrightarrow \tau \in I_\ell^i \stackrel{\triangle}{=} \left]t_\ell^i, t_{\ell+1}^i\right] \quad ,$$

qui appliqueraient tous les instants d'occurrence entre deux instants d'observation consécutifs t_ℓ^i et $t_{\ell+1}^i$ sur l'intervalle I_ℓ^i. Nous remarquons que les collections d'intervalles $\mathcal{I}^i \stackrel{\triangle}{=} \{I_\ell^i, \ell = 0, 1, \cdots\}$ sont des partitions de $\mathbb{R}^+$. En général ces partitions varient d'une entité à une autre.
Quand on s'intéresse à la distribution conjointe de $d \geqslant 1$ variables aléatoires, et les observations de chacune de ces variables sont indépendamment censurées (ou quantifiées) par des intervalles, il est alors question de censure par des intervalles de $\mathbb{R}^d$, produits de d intervalles réels.

Le mécanisme d'observation considéré dans notre étude est plus complexe que celui qui vient d'être décrit, et est schématisé dans la Figure 1.1. Nos observations correspondent à une quantifica-

tion de fonctions scalaires $\{F_j(\cdot)\}_{j=1}^J$ appliquées à une variable aléatoire $\theta = (\theta_1, \ldots, \theta_d)^\mathsf{T} \in \Theta \subset \mathbb{R}^d$, $d \geqslant 1$, dont on souhaite estimer la distribution conjointe.

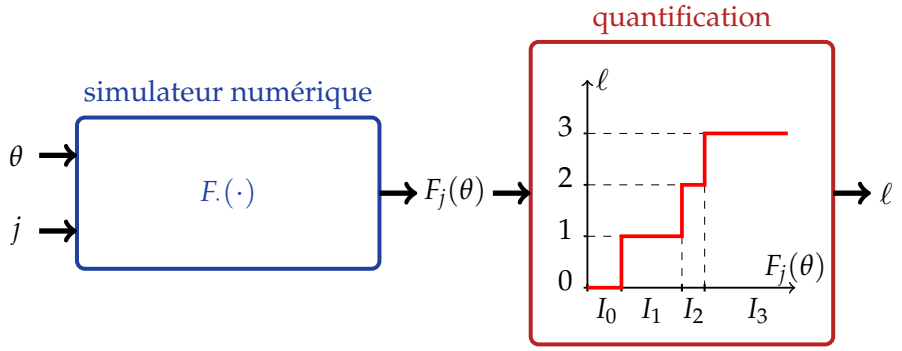


FIGURE 1.1 – *Mécanisme de censure des observations dans le cadre de notre étude.*

Maintenant, l'observation du niveau de quantification ℓ indique que $F_j(\theta) \in I_\ell$, ou encore que $\theta \in F_j^{-1}(I_\ell)$. Nous sommes donc toujours en présence d'observations censurées de θ, mais la censure est faite dans notre cas par l'ensemble de régions $\mathcal{R}_\ell^j \triangleq \{\theta;\ F_j(\theta) \in I_\ell\}$.

Les fonctions $F_j(\cdot)$ pouvant être des fonctions non-linéaires arbitraires, la géométrie des régions $\mathcal{R}_\ell^j$, appelées par la suite *régions de censure*, est également arbitraire. Comme on le verra, l'abandon de l'hypothèse d'une géométrie simple (intervalles) des régions de censure requiert plusieurs modifications des techniques existantes pour l'estimation à partir d'observations censurées, qui seront traitées dans ce manuscrit.

Notre étude s'éloigne encore du cadre habituel d'estimation à partir d'observations censurées par un second aspect. La majorité des travaux présentés dans la littérature n'imposent pas de structure particulière sur les partitions $\mathcal{I}^i$, qui peuvent varier d'une réalisation à une autre. Le nombre d'intervalles de censure agissant sur les réalisations de la variable d'intérêt est potentiellement infini, chacun d'entre eux étant, en général, observé une seule fois (cela correspond bien au cadre classique d'études cliniques, où chaque patient peut être observé à des instants différents des autres patients). A contrario, nos régions de censure $\mathcal{R}_\ell^j$ sont toutes issues d'un même quantificateur $Q(\cdot)$ et l'ensemble de fonctions scalaires $\{F_j(\cdot)\}_{j=1}^J$ que l'on peut observer est fini. Le nombre de régions de censure est donc nécessairement fini, chacune d'entre elles étant observée en général plusieurs fois. Cette modification du modèle classique d'observations censurées permettra une formulation du problème d'estimation de densité qui met en jeu un ensemble de lois empiriques associées à chacune des fonctions scalaires $F_j(\cdot)$. Cette formulation du problème est à la base du nouveau critère d'estimation proposé dans la thèse. Elle serait impossible dans le cadre classique non-contraint, où une infinité de quantificateurs est directement appliquée aux réalisations de θ.

Nous venons de préciser le modèle d'observations considéré dans notre étude, et comment il

s'éloigne de certaines hypothèses classiquement admises. Pour finaliser la formulation du problème d'estimation, nous devons indiquer le cadre de modélisation admis (que sait-on sur la densité ?) et un critère d'estimation (laquelle de deux densités estimées préfère-t-on ?).

Comme nous l'indiquons dans la prochaine section, le cadre applicatif de cette dissertation ne fournit pas de connaissances permettant de formuler l'hypothèse d'appartenance de la densité recherchée à une quelconque famille paramétrique de distributions. Ceci nous a naturellement conduit à considérer des estimateurs non-paramétriques de la densité de θ. Au delà de cette motivation d'ordre pratique, ce choix nous a conduit à devoir étudier les limites d'identifiabilité de la densité de θ sous les conditions admises d'observation censurée. Bien sur, la restriction à une famille paramétrique donnée devrait permettre de dépasser certaines des limitations observées dans le cas non-paramétrique, mais cela se ferait au prix d'une éventuelle dégradation dans le cas où cette restriction serait fausse.

Sous certaines conditions de régularité, les estimateurs par maximum de vraisemblance sont convergents et asymptotiquement efficaces. Ces propriétés expliquent l'intérêt qu'ont porté de nombreux auteurs au problème d'estimation non-paramétrique de densités de probabilité par maximum de vraisemblance [ABE+55, KM58, Efr67, Pet73, Tur76, Mal86, GV01].

Ces travaux, dont une brève revue sera présentée, ont conduit à plusieurs algorithmes d'estimation, ainsi qu'à la caractérisation de certaines de ses propriétés et la mise en évidence de quelques limitations dans le contexte de données censurées. Globalement, ces conclusions restent valables dans le cadre précis de notre étude. En particulier, des problèmes tels que la concentration du support de l'estimateur sur un ensemble réduit de régions isolées ainsi que sa non-unicité persistent, indiquant que le maximum de vraisemblance ne semble pas être le critère le plus pertinent pour construire l'estimateur de la densité d'une population naturelle à partir d'observations censurées.

Le principe du maximum d'entropie fournit un cadre alternatif au maximum de vraisemblance pour l'estimation de densités de probabilité : il choisit la densité contenant le moins d'information qui est compatible avec l'ensemble d'observations disponibles. Originellement utilisé par Boltzmann (1871) et Gibbs (1902) dans leurs travaux en mécanique statistique, l'utilisation de ce principe a connu un large succès en estimation de densités, depuis son introduction en théorie d'information par Jaynes [Jay57]. Dans ce cadre formel, les observations déterminent un ensemble de contraintes sur la densité à estimer, imposant le plus habituellement des égalités entre des moyennes empiriques et leurs espérances statistiques sous la densité estimée. Il arrive, surtout pour un faible nombre d'observations, qu'il n'existe pas de densité satisfaisant l'ensemble de toutes les contraintes. Alors, l'estimateur du maximum d'entropie n'existe que si les contraintes sont relaxées. Plusieurs auteurs [SJ00, KT03, DPS07] ont proposé de transformer les contraintes d'égalité en des inégalités bornant les déviations entre moyennes empiriques et espérances statistiques. Le degré d'éloignement consenti par ces inégalités devient alors un paramètre de réglage, contrôlant le compromis entre la régularité de l'estimateur (son entropie) et sa fidélité aux données observées.

Cette thèse présente un critère alternatif pour l'estimation non-paramétrique de densités de probabilité, choisissant le paramètre de réglage du maximum d'entropie par un critère de vraisemblance. Exploitant la dualité de ces deux critères, il conduit à un estimateur non-paramétrique qui présente une bonne capacité de généralisation, tout en gardant un fort attachement aux données.

1.2 Cadre applicatif de l'étude

Le problème d'estimation de densité étudié dans cette thèse a été motivé par le souhait de prédire le risque d'accidents de décompression en plongée hyperbare pour une population de plongeurs. Plus précisément, il s'agissait de caractériser la distribution statistique des paramètres biophysiques d'un modèle mathématique décrivant le phénomène de production de bulles circulant dans l'organisme du plongeur en phase de décompression, phénomène largement accepté comme étant à l'origine des accidents de plongée. Associé à un modèle reliant la quantité de bulles au risque d'accidents, la connaissance de cette distribution doit permettre l'estimation du risque de n'importe quelle plongée, ouvrant la porte à une amélioration des tables de plongée existantes. En effet, en se basant sur la distribution estimée des paramètres biophysiques, on sera en mesure de définir pour tout couple profondeur/durée de plongée, un protocole de remontée à la surface garantissant que le volume de bulles de gaz libéré n'excède pas un certain seuil pour une très large partie de la population étudiée. Pour garder une bonne performance de prédiction du risque pour des nouveaux profils de plongée la distribution estimée doit décrire exhaustivement la variabilité biophysique de la population, ce qui évitera de négliger des individus potentiellement à risque.

Les données qui ont été mises à notre disposition pour estimer la densité des paramètres biophysiques sont pauvres. La technologie actuelle ne permet pas de mesurer continuellement le volume instantané des bulles de gaz libéré par les tissus dans le sang des plongeurs. En effet, la majorité des bases de données concernant les accidents de plongée ne consignent que des mesures quantifiées – habituellement désignées par "grades de plongée" – de la quantité de bulles observées à un faible nombre (souvent égal à un) d'instants de mesure, plaçant le problème de l'estimation de densité dans le cadre d'observations censurées exposé dans la section précédente.

La mise en œuvre de notre estimateur dans ce cadre précis a donné lieu à des résultats additionnels, spécifiques au domaine de la plongée hyperbare, notamment au niveau du modèle formel de production de bulles. Selon la loi de Henry, la quantité de gaz dissout dans un liquide est proportionnelle à la pression partielle qu'exerce ce gaz sur le liquide. Ainsi, en plongée sous-marine, la diminution de la pression ambiante (phase de remontée) peut induire l'apparition de bulles gazeuses au sein des tissus du corps du plongeur. Nous avons mis en place un simulateur numérique efficace qui décrit les mécanismes subis par un plongeur avec des paramètres biophysiques θ, pendant la décompression le long d'un profil de plongée P_j, prédisant le volume instantané des bulles formées

dans les tissus et transférées dans le sang veineux, $F_j(\cdot, \theta)$. Ce simulateur est construit à partir du modèle développé dans [Hug10] qui repose sur un système d'équations différentielles non linéaires décrivant la cinétique des échanges gazeux et les dynamiques de recrutement et de croissance des bulles. Dans le souci de minimiser son temps d'exécution tout en gardant une fidélité au modèle initial, nous avons apporté plusieurs modifications aux équations originelles proposées dans [Hug10]. La stabilité du code produit a été testée sur un grand nombre de plongées et de larges domaines paramétriques.

La base de données que nous avons pu traiter contient une seule mesure de grade pour chaque plongée, et ne donne pas d'indication concernant l'instant des mesures. Nous avons ainsi supposé que le grade enregistré est la quantification du volume maximal observé. Avec cette supposition, les mesures de grade sont une quantification du scalaire $F_j(\theta) = \max_t F_j(t, \theta)$. Idéalement, nous aurions dû estimer aussi les caractéristiques du quantificateur $Q(\cdot)$, c'est à dire l'ensemble de seuils sur le volume maximum de bulles, $F_j(\theta)$, qui définissent les 5 grades observés (allant de 0 à 4). Nous avons néanmoins déterminé ces seuils avec une procédure ad-hoc, de manière à respecter les rapports entre deux seuils consécutifs du codage correspondant aux signaux issus de l'échographie et rapportés en [Eft07] ; l'ensemble des seuils ainsi déterminés sous des contraintes d'ordre a été validé par Julien Hugon, expert en plongée sous-marine. Nous avons pu de la sorte découpler l'étude de l'identifiabilité de la densité des paramètres du modèle de décompression du problème de caractérisation du mécanisme d'observation. De nouveaux instruments de mesure développés dans le cadre du projet dont fait partie notre étude doit permettre la résolution de ce problème directement à partir de couples de mesure (volumes de gaz/grades).

1.3 Contributions

La contribution principale de cette thèse est la proposition d'un nouveau critère d'estimation de densité pour des données censurés, s'appuyant à la fois sur les approches d'estimation par maximum de vraisemblance et par maximum d'entropie. En se basant sur une étude détaillée des estimateurs classiques, la solution proposée allie les deux critères déjà mentionnés et conduit à des densités estimées qui offrent une bonne capacité de généralisation, tout en gardant une forte attache aux données. Une attention spéciale a été portée à l'optimisation de l'exploitation de l'information disponible, par la prise en compte des dépendances statistiques entre l'ensemble des contraintes sur les moments empiriques que doit satisfaire l'estimateur du maximum d'entropie.

Une deuxième contribution est liée à la nature arbitraire des régions de censure, qui ne sont plus de simples intervalles comme dans les travaux classiques sur données censurées, permettant l'estimation de densités multivariées à partir d'observations fortement quantifiées d'un ensemble de fonctions scalaires des variables d'intérêt.

A notre connaissance, c'est la première fois qu'un modèle numérique d'un phénomène naturel de complexité élevée, dont on ne dispose que d'observations quantifiées, est effectivement utilisé

pour caractériser la diversité d'une population. Ce cadre général est beaucoup plus large que celui de la censure par intervalles, et ouvre la porte à de nombreux domaines d'intérêt pratique dans lesquels des modèles de grande complexité ont été développés et pour lesquels on ne dispose que d'observations ordinales, correspondant par exemple à des classifications effectuées par des experts.

Cette mise en œuvre nous a demandé, en plus du développement d'outils efficaces pour la mise en jeu de l'estimateur, de fournir des outils appropriés pour la représentation et la manipulation des régions de censure induites par les fonctions $F_j(\theta)$. En effet, et comme nous l'avons exposé dans la section précédante, du point de vue de l'observation de θ les fonctions scalaires $F_j(\theta)$ n'interviennent dans le problème d'estimation de la densité qu'à travers la définition des régions $\mathcal{R}_\ell^j$. La détermination de ces régions pour un nombre d de paramètres élevé est en soi-même un problème complexe qui doit être résolu pour la mise en œuvre de l'estimateur. Dans notre cas, $d = 2$, et nous avons pu fait appel à des techniques d'interpolation non-paramétriques pour construire des méta-modèles des fonctions $F_j(\theta)$, permettant un calcul rapide de leurs valeurs pour tout $\theta \in \Theta$. Ainsi, il a été possible d'approximer les valeurs de $F_j(\cdot)$ aux points d'une grille assez fine du domaine paramétrique, construisant une représentation approximée des différentes régions de censure $\mathcal{R}_\ell^j$ de notre problème. Cependant cette approche n'est pas transposable aux cas où d est grand, et on devra alors faire appel à d'autres outils, tels que la géométrie algorithmique ou la théorie de l'apprentissage.

Finalement, on savait que les cliques maximales du graphe d'intersection des régions de censure jouent un rôle fondamental dans la détermination de la complexité du problème d'estimation pour des données censurées par des intervalles. Nous montrons que pour des régions de forme arbitraire comme nous le considérons ici, les cliques importantes ne sont pas nécessairement les cliques maximales du graphe d'intersection, et nous proposons un algorithme pour leur détermination.

1.4 Organisation du manuscrit

Le Chapitre 2 est consacré au maximum de vraisemblance. Nous y présentons une analyse de la performance de plusieurs algorithmes existants dans la littérature qui permettent de calculer cet estimateur. Une analyse des résultats de l'estimation non-paramétrique par maximum de vraisemblance confirme que ce critère n'est pas adapté à notre problème. En plus d'être potentiellement affecté par plusieurs formes de non-unicité, cet estimateur a la singularité de concentrer la masse de probabilité dans un ensemble de petites zones isolées du domaine paramétrique. Dans ce cas les densités résultantes ne sont pas plausibles en tant que modèles de populations naturelles. Au delà de ce manque de réalisme, le rétrécissement du support de la densité est dangereux dans le cadre de la prédiction du risque, pouvant conduire à des prévisions optimistes si la possibilité d'existence d'individus à risque est sous-estimée.

Ayant mis en évidence les limitations du maximum de vraisemblance, nous présentons dans le Chapitre 3 la nouvelle méthode d'estimation qui est proposée dans cette thèse. Nous montrons dans un premier temps que le problème d'estimation peut être formulé en termes d'un ensemble de lois empiriques associées aux différentes fonctions scalaires $F_j(\theta)$, et qu'il est possible de déduire de ces lois un ensemble de contraintes sur la densité estimée. Nous présentons ensuite la définition du nouvel l'estimateur, qui fait appel aux critères du maximum d'entropie et de maximum de vraisemblance, et nous abordons la résolution du problème d'optimisation numérique associé. Finalement, nous étudions par simulation la perfomance de l'estimateur introduit, en particulier son comportement quand le nombre J de fonctions $F_j(\cdot)$ augmente, confirmant sa supériorité par rapport à l'estimateur classique.

Le Chapitre 4 traite du modèle biophysique de décompression, en précisant son exploitation dans le cadre de l'estimation, c'est à dire, la réalisation de simulateurs numériques des fonctions $F_j(\theta)$. Dans le cadre concret de la plongée ces différentes fonctions correspondent à l'exécution de différents profils de plongée. Nous présentons brièvement les mécanismes biophysiques intervenant dans la formation de bulles gazeuses et leur modélisation mathématique, et nous identifions les variables sur lesquelles portera l'estimation de densité. Nous abordons ensuite la caractérisation du quantificateur $Q(\cdot)$ et nous présentons la procédure utilisée pour fixer les seuils qui définissent les 5 grades de plongée. Une étude du modèle de plongée a été réalisée et est présentée en annexe de ce Chapitre. En particulier, nous avons étudié la sensibilité du modèle à plusieurs paramètres, et nous avons fait un travail de réduction du temps de calcul des simulations qui a conduit à l'introduction d'approximations de certaines équations du modèle originel.

Le Chapitre 5 s'intéresse à la mise en œuvre des estimateurs étudiés dans le cadre du problème de plongée hyperbare, pour le jeu de données disponibles. Nous abordons dans un premier temps la détermination des régions $\mathcal{R}_j^\ell$ nécessaires au calcul des estimateurs. Dans un deuxième temps, nous comparons la performance de la solution proposée à celle de l'estimateur classique. La qualité de l'estimation de la densité, ainsi que son pouvoir prédictif, ont été l'objet d'études portant à la fois sur des données simulées et sur le jeu de données réelles disponibles. Les résultats obtenus montrent la primauté de la solution proposée, qui fournit une description plausible d'un point de vue biophysique de la population étudiée (en évitant les singularités déjà rencontrées) tout en présentant une plus grande capacité à prédire les probabilités de grades pour des profils de plongée non-explorés.

Les études numériques présentées dans le Chapitre 5 confirment que l'alliance des natures duales du maximum de vraisemblance et du maximum d'entropie fournissent un cadre approprié au problème traité dans la thèse, et ouvrent de nombreuses perspectives de travail futur : Comment mieux caractériser la solution proposée ? Comment mettre cette méthode en place pour des problèmes de dimension supérieure à 2 ? Le Chapitre 6 liste quelques perspectives pour la continuation du travail présenté ici, ainsi qu'un résumé des contributions apportées. Les résultats obtenus dans cette thèse ont déjà fait l'objet des publications [BPR14a, BPR14b, BPR15a, BPR15b].

- CHAPITRE 2 -

ESTIMATION NON-PARAMÉTRIQUE PAR MAXIMUM DE VRAISEMBLANCE

Dans ce chapitre nous étudions l'estimateur non-paramétrique du maximum de vraisemblance (ENPMV) avec observations censurées sur un ensemble de régions.

Après la formulation du problème de l'ENPMV dans la section 2.1, nous nous intéressons aux propriétés d'unicité et de convergence de cet estimateur dans la section 2.2. Ensuite nous mettons en évidence le rôle joué par la théorie des graphes dans la détermination de son support dans la section 2.3.1. Nous abordons la question de la maximisation de la fonction de la vraisemblance en comparant la performance de plusieurs méthodes d'optimisation dans la section 2.3.2. Nous présentons enfin dans la section 2.4 des résultats d'application de cet estimateur sur des jeux de données simulées qui mettent en évidence les limitations de l'ENPMV.

2.1 FORMULATION DU PROBLÈME

Soient Θ un sous-ensemble compact de $\mathbb{R}^d$ et π_θ une distribution de probabilité sur Θ. Soit Y une variable aléatoire à valeurs dans Θ qui suit la distribution π_θ :

$$\pi_\theta \in \mathscr{P} \stackrel{\Delta}{=} \{\text{toutes les distributions de probabilité sur } \Theta\}.$$

La singularité de notre étude réside dans la façon dont nous observons les réalisations y_i de Y. En effet, nous avons uniquement indication d'une région $\mathcal{A}_i \subset \Theta$ à laquelle y_i appartient, selon un mécanisme illustré dans la Figure 2.1. Nous désignerons ce type de mécanisme d'observation par *censure par régions*. Il est la généralisation de la notion de censure par intervalles, fréquente par exemple en pharmacocinétique [LBW+00].

Dans le problème qui motive cette thèse nous nous intéressons à l'estimation de π_θ à partir d'observations censurées par des régions de forme quelconque. Nous soulignons la différence entre ce cas et la notion d'observations avec erreurs bornées. Alors que dans l'estimation à erreurs bornées, les

régions modélisent l'incertitude liée à la mesure, dans notre cas les régions de censure [1] sont liées à des limitations physiques du mécanisme d'observation et peuvent varier d'une réalisation à l'autre, comme indiqué dans la Figure 2.1.

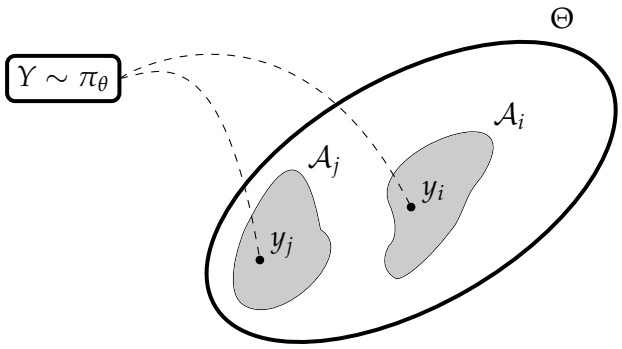


FIGURE 2.1 – *Censure par régions des réalisations de Y.*

Avant de s'attaquer à l'estimation de π_θ nous introduisons des notations qui seront utiles par la suite.

2.1.1 Notations

Dans le problème d'estimation de la densité π_θ à partir de données censurées, nous allons admettre que chaque *région de censure* $\mathcal{A}_i$ est élément d'une partition $\mathcal{R}^j$ choisie dans un ensemble fini $\{\mathcal{R}^1, \mathcal{R}^2, \ldots, \mathcal{R}^J\}$ de partitions de Θ :

$$\mathcal{A}_i \in \mathcal{R}^j \triangleq \{\mathcal{R}^j_\ell\}_{\ell=1}^{L^j},$$

où nous avons implicitement défini L^j comme la taille de la partition $\mathcal{R}^j$, et par définition, pour tout j,

$$\cup_{\ell=1}^{L^j} \mathcal{R}^j_\ell = \Theta \text{ et } \ell_1 \neq \ell_2 \Rightarrow \mathcal{R}^j_{\ell_1} \cap \mathcal{R}^j_{\ell_2} = \emptyset.$$

Exemple 1. La Figure 2.2 représente les quatre régions de censure $\left\{\mathcal{R}^j_\ell\right\}_{j=1,\ell=1}^{2,2}$, correspondant à deux partitions $\mathcal{R}^1 = \{\mathcal{R}^1_1, \mathcal{R}^1_2\}$ et $\mathcal{R}^2 = \{\mathcal{R}^2_1, \mathcal{R}^2_2\}$. Dans ce cas, $J = 2$, et $L^1 = L^2 = 2$.

Nos observations sont donc des événements du type $\left\{y_i \in \mathcal{R}^j_\ell\right\}$. Ainsi, une observation est déterminée par la réalisation de deux variables discrètes :

– la première, associée à la variable S_p à valeurs dans $\{1, \ldots, J\}$, correspond à la sélection d'une partition $\mathcal{R}^{s_p}$ parmi les J partitions possibles $\left\{\mathcal{R}^j\right\}_{j=1}^J$,

1. Comme expliqué dans la section 5.1, page 106, dans le cadre d'application à la caractérisation de la densité des paramètres biophysiques (N^{max}, A), les régions de censure sont imposées par le modèle biophysique de décompression.

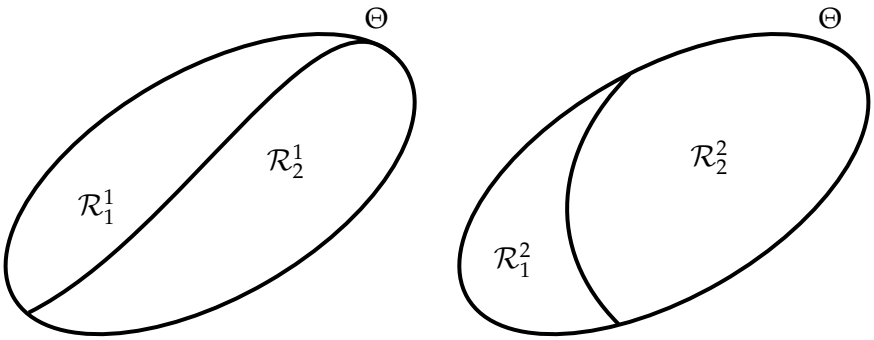


FIGURE 2.2 – *Régions de censure $\mathcal{R}_\ell^j$ issues de deux partitions $\mathcal{R}^1 = \{\mathcal{R}_1^1, \mathcal{R}_2^1\}$ et $\mathcal{R}^2 = \{\mathcal{R}_1^2, \mathcal{R}_2^2\}$.* □

– la seconde correspond à la réalisation s_r de la variable aléatoire S_r (sélection de la région de censure) à valeurs dans $\{1, \ldots, L^{s_p}\}$ indiquant la région $\mathcal{R}_{s_r}^{s_p}$ sélectionnée dans la partition $\mathcal{R}^{s_p}$.

Pour un ensemble d'observations de taille n, et en notant $s_p(i)$ et $s_r(i)$ les i-ièmes réalisations respectives de S_p et S_r, le jeu de données est complètement déterminé par les réalisations de ces deux variables

$$\mathbf{X}_n \stackrel{\Delta}{=} \{x_i\}_{i=1}^n = \left\{(s_p(i), s_r(i))\right\}_{i=1}^n .$$

Définition 1. *($p_{\pi,\mathcal{R}}$) : Nous notons $p_{\pi,\mathcal{R}}$ la loi de probabilité induite par la distribution $\pi \in \mathscr{P}$ sur les éléments de la partition $\mathcal{R} = \{\mathcal{R}_\ell\}_{\ell=1}^L$ de Θ :*

$$p_{\pi,\mathcal{R}}(\ell) \stackrel{\Delta}{=} p_{\pi,\mathcal{R}_\ell} = \int_{\mathcal{R}_\ell} d\pi, \qquad \ell = 1, \ldots, L. \tag{2.1}$$

□

Bien sûr, $p_{\pi,\mathcal{R}}$ appartient à $\mathcal{S}^{L+}$, le simplexe probabiliste de taille L

$$\mathcal{S}^{L+} \stackrel{\Delta}{=} \{\mathbf{w} \in \mathbb{R}^L : w_\ell \geq 0 \text{ et } \sum_{\ell=1}^L w_\ell = 1\}.$$

Définition 2. ***(régions élémentaires)*** *: Soit $\left\{\mathcal{R}^{s_p(i)}\right\}_{i=1}^n$ l'ensemble des partitions de Θ correspondant aux observations $\mathbf{X}_n$ où $s_p(i) \in \{1, \ldots, J\}$ pour tout $i = 1, \ldots, n$. Nous notons $\mathcal{Q}^J = \{\mathcal{Q}_m^J\}_{m=1}^M$ la partition la plus fine de Θ telle que $\sigma(\mathcal{Q}^J)$, la tribu engendrée par $\mathcal{Q}^J$, contient tous les éléments $\left\{\mathcal{R}_{s_r(i)}^{s_p(i)}\right\}_{i=1}^n$, avec $s_r(i) \in \{1, \ldots, L^{s_p(i)}\}$. Nous appelons régions élémentaires les éléments de $\mathcal{Q}^J$.* □

Nous remarquons que $\mathcal{Q}^J$ dépend des observations $\mathbf{X}^n$ à travers la géométrie des J partitions $\left\{\mathcal{R}^j\right\}_{j=1}^J$. La Figure 2.3 illustre la définition de la partition $\mathcal{Q}^J$ dans le cas où tous les éléments des deux partitions $\mathcal{R}^1$ et $\mathcal{R}^2$ dans l'exemple de la Figure 2.2 sont observés. On remarque que les régions élémentaires forment toujours une partition de Θ et sont donc disjointes 2 à 2.

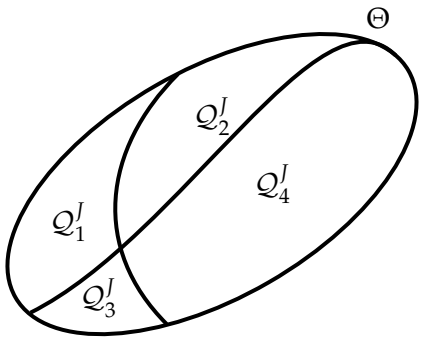


FIGURE 2.3 – *Exemple des régions élémentaires* $\{\mathcal{Q}_m^J\}_{m=1}^4$ *induites par les partitions* $\mathcal{R}^1$ *et* $\mathcal{R}^2$ *de la Figure 2.2.*

Définition 3. *(*ξ_ℓ^j*) : Pour tout élément* $\mathcal{R}_\ell^j$*, nous définissons l'ensemble* ξ_ℓ^j *des régions élémentaires contenues dans* $\mathcal{R}_\ell^j$ *:* $\xi_\ell^j \triangleq \{m \in \{1,\dots,M\}\,;\, \mathcal{Q}_m^J \subset \mathcal{R}_\ell^j\}$. □

Notons qu'avec cette définition, nous avons pour $j = 1,\dots,J$ et pour $\ell = 1,\dots,L^j$

$$\mathcal{R}_\ell^j = \bigcup_{m\in\xi_\ell^j} \mathcal{Q}_m^J \quad \text{et} \quad \mathcal{Q}_m^J = \bigcap_{m\in\xi_\ell^j} \mathcal{R}_\ell^j \ . \tag{2.2}$$

Nous venons d'introduire les notations nécessaires afin d'écrire la fonction de vraisemblance pour l'estimation de π_θ.

2.1.2 La fonction de vraisemblance

Comme nous l'avons déjà mentionné, nous admettons que le manque d'information sur π_θ ne nous permet pas de formuler une hypothèse d'appartenance à une quelconque famille paramétrique de distributions. Nous portons donc notre intérêt à l'estimation non-paramétrique de π_θ. Il est connu que sous certaines conditions de régularité, l'estimateur par maximum de vraisemblance est convergent (*"consistent"*) et asymptotiquement efficace, ce qui explique son utilisation fréquente. Nous nous intéressons dans ce chapitre à l'estimation non-paramétrique par maximum de vraisemblance de π_θ.

La vraisemblance d'une distribution de probabilité π quand nous observons $\mathbf{X}_n$ est donnée par :

$$\mathscr{L}(\pi, \mathbf{X}_n) \triangleq \prod_{i=1}^n p_{\pi,\mathcal{R}_{sr(i)}^{sp(i)}} = \prod_{i=1}^n \prod_{\ell=1}^{L^{sp(i)}} \left(p_{\pi,\mathcal{R}_\ell^{sp(i)}}\right)^{\delta_{\ell,sr(i)}}, \quad \pi \in \mathscr{P},$$

avec $\delta_{\ell,s_r(i)} = 1$ si $\ell = s_r(i)$ et $\delta_{\ell,s_r(i)} = 0$ sinon. La log-vraisemblance (normalisée) de π est donc :

$$\mathcal{L}(\pi, \mathbf{X}_n) \triangleq \frac{1}{n} \sum_{i=1}^{n} \sum_{\ell=1}^{L^{sp(i)}} \delta_{\ell,s_r(i)} \log \left(p_{\pi,\mathcal{R}_\ell^{sp(i)}} \right). \tag{2.3}$$

Nous notons $\hat{\pi}^{\mathcal{L}}$ l'ENPMV [2] :

$$\hat{\pi}^{\mathcal{L}} \triangleq \underset{\pi \in \mathscr{P}}{\operatorname{argmax}}\, \mathcal{L}(\pi, \mathbf{X}_n). \tag{2.4}$$

Lemme 1. *Chaque loi de probabilité* $\mathbf{w}$ *dans* $\mathcal{S}^{M+}$ *définit dans* $\mathscr{P}$ *une classe d'équivalence* $\mathscr{P}^{\mathbf{w}}$

$$\mathscr{P}^{\mathbf{w}} \triangleq \left\{ \pi \in \mathscr{P};\ p_{\pi,\mathcal{Q}^J} = \mathbf{w} \right\}.$$

L'ensemble $\mathscr{P}^{\mathbf{w}}$ *est caractérisé par une même valeur de la log-vraisemblance égale à*

$$\mathcal{L}(\mathbf{w}, \mathbf{X}_n) \triangleq \frac{1}{n} \sum_{i=1}^{n} \sum_{\ell=1}^{L^{sp(i)}} \delta_{\ell,s_r(i),\ell} \log \left(\sum_{m \in \xi_\ell^{sp(i)}} w_m \right).$$

Nous notons

$$\hat{\mathbf{w}}^{\mathcal{L}} \triangleq \underset{\mathbf{w} \in \mathcal{S}^{M+}}{\operatorname{argmax}}\, \mathcal{L}(\mathbf{w}, \mathbf{X}_n).$$

$\hat{\mathbf{w}}^{\mathcal{L}}$ *existe toujours et toutes les distributions* $\hat{\pi}^{\mathcal{L}}$ *appartenant à la classe d'équivalence définie par* $\hat{\mathbf{w}}^{\mathcal{L}}$ *sont des estimateurs de maximum de vraisemblance. En utilisant la notation de la Définition 1, nous avons* $\hat{\mathbf{w}}^{\mathcal{L}} = p_{\hat{\pi}^{\mathcal{L}},\mathcal{Q}^J}$. □

Le Lemme 1 affirme que nous pouvons restreindre la recherche de l'ENPMV à la famille $\left\{ p_{\pi_{\mathbf{w}},\mathcal{Q}^J}, \mathbf{w} \in \mathcal{S}^{M+} \right\}$ des classes d'équivalence paramétrées par les éléments de $\mathcal{S}^{M+}$, où

$$\pi_{\mathbf{w}} = \sum_{m=1}^{M} w_m \delta_{\mathcal{Q}_m^J},$$

et $\delta_{\mathcal{Q}_m^J}$ est une mesure dont le support est inclus dans la région élémentaire $\mathcal{Q}_m^J$. Nous en déduisons que, compte tenu de la nature censurée des observations, nous ne pouvons connaître π_θ au delà des masses attribuées aux régions élémentaires $\{\mathcal{Q}_m^J\}_{m=1}^M$.

Preuve : La démonstration de ce Lemme découle immédiatement de l'écriture de $\mathcal{R}_\ell^j$ dans l'expression (2.2). Soient $\pi \in \mathscr{P}$ une distribution de probabilité et $\mathbf{w} \in \mathcal{S}^{M+}$ tel que $\mathbf{w} = p_{\pi,\mathcal{Q}^J}$. Pour toute

2. Une étude de l'unicité de cet estimateur suivra.

région de censure $\mathcal{R}_\ell^j$, nous avons

$$\pi(\mathcal{R}_\ell^j) = \sum_{m \in \xi_\ell^j} \pi(\mathcal{Q}_m^J) = \sum_{m \in \xi_\ell^j} w_m,$$

et donc

$$\mathcal{L}(\pi, \mathbf{X}_n) = \mathcal{L}(\mathbf{w}, \mathbf{X}_n).$$

Puisque $\mathcal{L}(\mathbf{w}, \mathbf{X}_n)$ est une fonction continue en $\mathbf{w}$ sur l'ensemble $\mathcal{S}^{M+}$, un borné fermé dans $\mathbb{R}^M$, donc compact, $\hat{\mathbf{w}}^{\mathcal{L}}$ existe toujours. □

Introduisons quelques notations qui nous aideront à mettre en évidence certaines propriétés de l'ENPMV.

Soient pour tout $j \in \{1, \ldots, J\}$ et tout $\ell \in \{1, \ldots, L^j\}$, n_ℓ^j le nombre de fois où $\mathcal{R}_\ell^j$ est "observée"

$$n_\ell^j \triangleq \#\left\{s_p(i) = j \text{ et } s_r(i) = \ell;\ \ i \in \{1, \ldots, n\}\right\},$$

et $\mathbf{B}^{(j)}$ la matrice binaire de taille $L^j \times M$ définie par

$$\mathbf{B}_{\ell m}^{(j)} \triangleq \begin{cases} 1 & \text{si } \mathcal{Q}_m^J \subset \mathcal{R}_\ell^j \\ 0 & \text{sinon} \end{cases} \qquad \text{pour } m = 1, \ldots, M. \tag{2.5}$$

La matrice $\mathbf{B}^{(j)}$ indique les régions élémentaires qui composent chaque élément de $\mathcal{R}^j$. Nous notons que

$$\mathbf{1}_{L^j}^\mathsf{T} \mathbf{B}^{(j)} = \mathbf{1}_M^\mathsf{T},$$

avec $\mathbf{1}_d$ le vecteur colonne de taille d dont tous les éléments sont égaux à 1. Chaque matrice $\mathbf{B}^{(j)}$ applique $\mathcal{S}^{M+}$ dans $\mathcal{S}^{L^j+}$. On peut remarquer que cette application est nécessairement surjective mais elle n'est pas injective sauf dans le cas trivial où $\mathcal{Q}^J = \mathcal{R}^j$.

Nous posons $K = \sum_{j=1}^J L^j$ le nombre total de régions de censure possibles.

Lemme 2. *Nous avons*[3]

$$\mathcal{L}(\mathbf{w}, \mathbf{X}_n) = \mathcal{L}(\mathbf{w}, \mathbf{n}, \mathbf{B}) \triangleq \frac{1}{n} \sum_{k=1}^K n_k \log \mathbf{B}_{k.}\mathbf{w}. \tag{2.6}$$

3. Dans l'équation 2.6 nous remplaçons la notation $\mathcal{L}(\mathbf{w}, \mathbf{X}_n)$ par $\mathcal{L}(\mathbf{w}, \mathbf{n}, \mathbf{B})$ car la log-vraisemblance ne dépend des données $\mathbf{X}_n$ qu'à travers la matrice $\mathbf{B}$ et le vecteur $\mathbf{n}$.

avec $\mathbf{B}$ *la matrice de taille* $K \times M$ *(généralement* $M > K$*) formée par les* J *sous-matrices* $\mathbf{B}^{(j)}$ *:*

$$\mathbf{B} \triangleq \begin{bmatrix} \mathbf{B}^{(1)} \\ \vdots \\ \mathbf{B}^{(J)} \end{bmatrix}, \tag{2.7}$$

$\mathbf{B}_{k.}$ *la k-ième ligne de* $\mathbf{B}$ *et* $\mathbf{n}$ *le vecteur de taille* K

$$\mathbf{n} \triangleq (n_1, \ldots, n_K)^\mathsf{T} = \left(n_1^1, \ldots, n_{L^1}^1, \cdots, n_1^J, \ldots, n_{L^J}^J\right)^\mathsf{T} .$$

□

Le Lemme 2 montre que $\mathcal{L}(\mathbf{w}, \mathbf{X}_n)$ dépend de $\mathbf{X}_n$ uniquement à travers du vecteur $\mathbf{n}$ et de la matrice $\mathbf{B}$. La matrice $\mathbf{B}$ décrit la géométrie du mécanisme de censure en précisant les intersections entre les régions de censure et les régions élémentaires, tandis que le vecteur $\mathbf{n}$ spécifie le nombre de fois où les différentes régions de censure ont été observées.

Nous remarquons que l'application linéaire définie par $\mathbf{B}$

$$\mathbf{B} \ : \ \mathcal{S}^{M+} \to \mathcal{S}^{L^1+} \times \cdots \times \mathcal{S}^{L^J+},$$

n'est généralement pas surjective. Nous remarquons aussi que toutes les colonnes de $\mathbf{B}$ sont différentes. En effet, si deux colonnes de $\mathbf{B}$ coïncident, les régions élémentaires correspondant à ces deux colonnes coïncident d'après l'équation (2.2), et cela contredit la définition de $\mathcal{Q}^J$ (Définition 2).

Exemple 2 (exemple 1 (cont.))**.** Revenons à l'exemple 1 (Figures 2.2 et 2.3, page 10) afin de donner une illustration simple des notations que nous venons de définir. Dans cet exemple, $J = 2$, $L^1 = L^2 = 2$ et $K = M = 4$. La matrice $\mathbf{B}$ décrivant la décomposition des éléments de chacune des deux partitions dans les quatre régions élémentaires (les éléments de $\mathcal{Q}^J$) est donnée par

$$\mathbf{B} = \begin{bmatrix} \mathbf{B}^{(1)} \\ \\ \mathbf{B}^{(2)} \end{bmatrix} = \left[\begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \right].$$

Remarquons que la matrice $\mathbf{B}$ n'est pas de rang plein : sa première colonne est égale à la somme des colonnes 2 et 3 moins la colonne 4. La log-vraisemblance dans cet exemple est

$$\mathcal{L}(\mathbf{w}, \mathbf{n}, \mathbf{B}) = \frac{n_1}{n} \log(w_1 + w_2) + \frac{n_2}{n} \log(w_3 + w_4) + \frac{n_3}{n} \log(w_1 + w_3) + \frac{n_4}{n} \log(w_2 + w_4), \tag{2.8}$$

et elle est fonction uniquement des masses de probabilité w_1, w_2, w_3, w_4 attribuées aux quatre régions élémentaires de la Figure 2.3. □

Le Lemme 3 ci-dessous présente une expression alternative de la fonction de log-vraisemblance qui permet d'établir le lien entre l'ENPMV et des outils de la théorie d'information. Commençons d'abord par introduire les notations nécessaires.

Notons $\mathbf{q}_j(\mathbf{n}) \in \mathcal{S}^{L^j+}$ le vecteur qui regroupe les fréquences relatives d'observation des éléments de $\mathcal{R}^j$

$$\{\mathbf{q}_j(\mathbf{n})\}_\ell \triangleq \frac{n_\ell^j}{n^j},$$

avec $n^j \triangleq \sum_{\ell=1}^{L^j} n_\ell^j$ le nombre de fois où des éléments de $\mathcal{R}^j$ ont été observés. De façon similaire, nous notons $\mathbf{z}_j(\mathbf{w}) \in \mathcal{S}^{L^j+}$ le vecteur de taille L^j de composantes $\{\mathbf{z}_j(\mathbf{w})\}_\ell \triangleq \mathbf{B}_{\ell.}^{(j)}\mathbf{w}$.

Soient $\mathbf{w} = (w_1, \ldots, w_M)^\mathsf{T}$ et $\mathbf{v} = (v_1, \ldots, v_M)^\mathsf{T}$ deux lois dans $\mathcal{S}^{M+}$. L'entropie de Shannon H_1 [Sha01] de $\mathbf{w}$ et la divergence de Kullback-Leibler [Kul68] D_{KL} de $\mathbf{v}$ par rapport à $\mathbf{w}$ sont définies par

$$H_1(\mathbf{w}) \triangleq -\sum_{m=1}^{M} w_m \log w_m, \quad \text{et} \quad D_{KL}(\mathbf{w}\|\mathbf{v}) \triangleq \sum_{m=1}^{M} w_m \log \frac{w_m}{v_m}. \tag{2.9}$$

Nous avons $H_1(\mathbf{w}) \geqslant 0$ et $D_{KL}(\mathbf{w}\|\mathbf{v}) \geqslant 0$ pour tous $\mathbf{w}$ et $\mathbf{v}$ dans $\mathcal{S}^{M+}$. De plus, $D_{KL}(\mathbf{w}\|\mathbf{v}) = 0$ si et seulement si $\mathbf{w} = \mathbf{v}$ presque partout.

Lemme 3.

$$\mathcal{L}(\mathbf{w}, \mathbf{n}, \mathbf{B}) = -\sum_{j=1}^{J} \frac{n^j}{n} D_{KL}(\mathbf{q}_j(\mathbf{n})\|\mathbf{z}_j(\mathbf{w})) - \sum_{j=1}^{J} \frac{n^j}{n} H_1(\mathbf{q}_j(\mathbf{n})). \tag{2.10}$$

□

La démonstration est immédiate d'après les définitions de H_1 et D_{KL}. En effet, nous avons

$$\mathcal{L}(\mathbf{w}, \mathbf{n}, \mathbf{B}) = \sum_{j=1}^{J} \frac{n^j}{n} \sum_{\ell=1}^{L^j} \frac{n_\ell^j}{n^j} \log \mathbf{B}_{\ell.}^{(j)}\mathbf{w} = -\sum_{j=1}^{J} \frac{n^j}{n} \left[D_{KL}(\mathbf{q}_j(\mathbf{n})\|\mathbf{z}_j(\mathbf{w})) + H_1(\mathbf{q}_j(\mathbf{n})) \right].$$

Le terme $\sum_{j=1}^{J} \frac{n^j}{n} H_1(\mathbf{q}_j(\mathbf{n}))$ ne dépend pas de $\mathbf{w}$, et donc la maximisation de la vraisemblance est équivalente à la minimisation du premier terme $\sum_{j=1}^{J} \frac{n^j}{n} D_{KL}(\mathbf{q}_j(\mathbf{n})\|\mathbf{z}_j(\mathbf{w}))$, par rapport à $\mathbf{w} \in S^{M+}$. Ceci montre que l'ENPMV minimise une pondération de l'éloignement entre les lois empiriques et les lois prédites.

Soient $\mathbf{q}(\mathbf{n})$ et $\mathbf{z}(\mathbf{w})$, les deux vecteurs de taille K, concaténation respectivement de l'ensemble des lois empiriques $\left\{\mathbf{q}_j^\mathsf{T}(\mathbf{n})\right\}_{j=1}^{J}$ et l'ensemble des lois prédites $\left\{\mathbf{z}_j^\mathsf{T}(\mathbf{w})\right\}_{j=1}^{J}$

$$\mathbf{q}^\mathsf{T}(\mathbf{n}) \triangleq (\mathbf{q}_1^\mathsf{T}(\mathbf{n}), \ldots, \mathbf{q}_J^\mathsf{T}(\mathbf{n})) \quad \text{et} \quad \mathbf{z}^\mathsf{T}(\mathbf{w}) \triangleq (\mathbf{z}_1^\mathsf{T}(\mathbf{w}), \ldots, \mathbf{z}_J^\mathsf{T}(\mathbf{w})). \tag{2.11}$$

Supposons qu'il existe $\mathbf{w}^\star \in \mathcal{S}^{M+}$ tel que $\mathbf{q}(\mathbf{n}) = \mathbf{z}(\mathbf{w}^\star)$. Alors, pour tout j, les lois discrètes $\mathbf{q}_j(\mathbf{n})$ et $\mathbf{z}_j(\mathbf{w}^\star)$ sont identiques et leur divergence de Kullback-Leibler est nulle. Nous en déduisons donc que $\mathcal{L}(\mathbf{w}^\star, \mathbf{X}_n)$ est la valeur maximale de $\mathcal{L}(\mathbf{w}, \mathbf{X}_n)$ et que $\hat{\mathbf{w}}^\mathcal{L} = \mathbf{w}^\star$.

Comme nous l'avons déjà remarqué, l'application définie par la matrice $\mathbf{B}$ n'est généralement pas surjective, et donc il n'existe pas toujours une solution en $\mathbf{w} \in \mathcal{S}^{M+}$ de l'équation $\mathbf{q}(\mathbf{n}) = \mathbf{B}\mathbf{w}$. La détermination de l'ENPMV doit alors faire appel à une optimisation numérique de la fonction $\mathcal{L}(\mathbf{w}, \mathbf{X}_n)$ (section 2.3.2).

2.2 Propriétés de l'ENPMV

La recherche de l'ENPMV d'une densité à partir d'observations censurées a été étudiée dans les travaux pionniers d'Ayer et *al.* [ABE+55]. Les auteurs donnent une expression explicite de l'ENPMV dans le cas de données scalaires ($\Theta = \mathbb{R}$) avec des partitions de taille 2, c'est à dire des régions de censure du type $]-\infty, a)$ et $(a, +\infty[$. Kaplan et Meier [KM58] proposent l'estimateur produit-limite pour des données scalaires censurées par des intervalles dont le nombre peut être supérieur à 2. Efron [Efr67] s'est intéressé au même problème tout en introduisant la notion de l'auto-consistance d'un estimateur. Turnbull [Tur76] a été le premier à caractériser le support de l'ENPMV dans le cas de données scalaires censurées par des intervalles, en donnant un algorithme pour calculer les intervalles (qu'il appelle "*Innermost intervals*") constituant ce support. Gentleman et Vandal [GV01] ont traité le cas d'observations multivariées censurées par des produits d'intervalles en spécifiant les différents types de non-unicité dont souffre l'ENPMV et en mettant en évidence le lien entre support de l'ENPMV et la théorie des graphes.

Nous nous intéressons dans cette section aux propriétés d'unicité et de convergence de l'estimateur $\hat{\pi}^{\mathcal{L}}$ dans le cas de censure par régions. Nous verrons que cet estimateur est sujet en général à deux formes de non-unicité, et que les résultats formels concernant sa convergence sont assez faibles.

2.2.1 Non-unicité

Dans la section 2.1.2 nous avons établi trois écritures de la fonction $\mathcal{L}(\mathbf{w}, \mathbf{X}_n)$ (Lemmes 1, 2 et 3) qui nous ont permis de conclure que l'estimateur $\hat{\pi}^{\mathcal{L}}$ n'est pas unique. Cette limitation est connue depuis les travaux de Turnbull [Tur76] dans le cadre de l'estimation de densités avec des observations univariées censurées. Dans cette section, nous étudions en détail la question de l'unicité de $\hat{\pi}^{\mathcal{L}}$, en faisant un rappel de la littérature sur ce problème.

Exemple 3. Reprenons l'exemple 1, présenté dans les Figures 2.2 et 2.3, page 10. Nous savons déjà (Lemme 1) que toutes les densités dans $\mathscr{P}$ attribuant les mêmes masses aux régions élémentaires $\{\mathcal{Q}_m^J\}_{m=1}^4$ ont la même vraisemblance. Supposons que $\mathbf{n} = \frac{n}{4}\mathbf{1}_4$. La log-vraisemblance (équation (2.8)) est égale à

$$\mathcal{L}(\mathbf{w}, \mathbf{n}, \mathbf{B}) = \frac{1}{4}\left(\log(w_1 + w_2) + \log(w_3 + w_4) + \log(w_1 + w_3) + \log(w_2 + w_4)\right).$$

En remarquant que $w_1 + w_2 = 1 - w_3 - w_4$ et que la fonction $x \mapsto \log x + \log(1 - x)$ atteint son maximum en $x = \frac{1}{2}$, nous déduisons que le vecteur $\mathbf{w}^\star = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})^\mathsf{T}$ maximise la log-vraisemblance.

Nous posons pour toute partie borélienne $\mathcal{A} \subset \Theta$ et tout élément $\theta \in \Theta$, $u_{\mathcal{A}} \in \mathscr{P}$ la distribution de probabilité uniforme sur $\mathcal{A}$, et $\delta_\theta \in \mathscr{P}$ la distribution de Dirac au point θ :

$$u_{\mathcal{A}}(\theta') \triangleq \begin{cases} \frac{1}{\nu(\mathcal{A})} & \text{si } \theta' \in \mathcal{A} \\ 0 & \text{sinon} \end{cases} \qquad \text{pour } \theta' \in \Theta,$$

avec $\nu(\mathcal{A})$ la mesure de Lebesgue de $\mathcal{A}$ et

$$\delta_\theta(\theta') \triangleq \begin{cases} 1 & \text{si } \theta' = \theta \\ 0 & \text{sinon} \end{cases} \qquad \text{pour } \theta' \in \Theta.$$

La Figure 2.4a, représente la densité π_1 définie comme suit

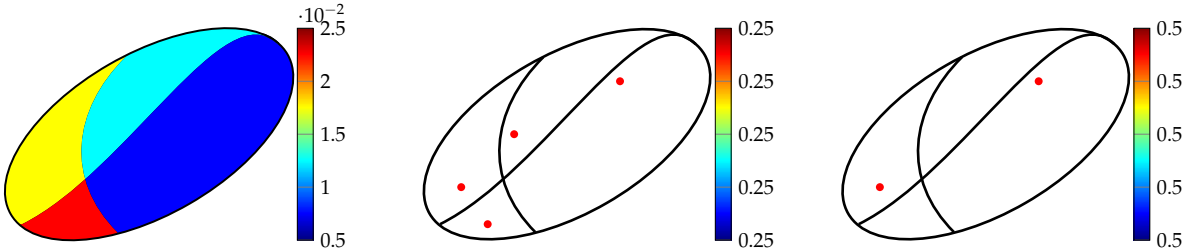
$$\pi_1(\theta) = \sum_{m=1}^{4} \frac{1}{4} u_{\mathcal{Q}_m^J}(\theta).$$

La distribution π_1 attribue à chacune des régions élémentaires la masse de probabilité $\frac{1}{4}$ tout en distribuant cette masse uniformément dans chaque région. Soient θ_1, θ_2, θ_3 et θ_4, 4 éléments dans Θ tels que $\theta_m \in \mathcal{Q}_m^J$ pour $m = 1, \ldots, 4$. La Figure 2.4b présente la densité π_2 attribuant la masse de probabilité $\frac{1}{4}$ à chacun des points θ_m

$$\pi_2(\theta) = \sum_{m=1}^{4} \frac{1}{4} \delta_{\theta_m}(\theta).$$

Les deux densités π_1 et π_2 maximisent $\mathcal{L}(\cdot, \mathbf{X}_n)$ et induisent la même loi de probabilité $\mathbf{w}^\star$ sur la partition $\mathcal{Q}^J$

$$p_{\pi_1, \mathcal{Q}^J} = p_{\pi_2, \mathcal{Q}^J} = \mathbf{w}^\star.$$



(a) π_1. **(b)** π_2. **(c)** π_3.

FIGURE 2.4 – *Trois densités maximisant la vraisemblance dans l'exemple 1.*

Nous avons déjà remarqué que pour cet exemple la matrice $\mathbf{B}$ n'est pas de rang plein. On peut vérifier sans difficulté qu'elle est de rang 3 et que son noyau est engendré par le vecteur $\mathbf{v} =$

$(1,-1,-1,1)^\mathsf{T}$. Ainsi, tous les vecteurs de l'ensemble $\{\mathbf{w}^\star + \alpha\mathbf{v}, \alpha \in \mathbb{R}\} \cap \mathcal{S}^{4+}$ maximisent $\mathcal{L}(\mathbf{w},\mathbf{n},\mathbf{B})$. Nous en déduisons que le vecteur $\mathbf{w}^\star + \frac{1}{4}\mathbf{v} = (\frac{1}{2},0,0,\frac{1}{2})^\mathsf{T}$ maximise également $\mathcal{L}(\mathbf{w},\mathbf{n},\mathbf{B})$. La Figure 2.4c représente la densité π_3

$$\pi_3(\theta) = \frac{1}{2}\left(\delta_{\theta_1}(\theta) + \delta_{\theta_4}(\theta)\right),$$

où nous avons $p_{\pi_3,\mathcal{Q}^J} = \mathbf{w}^\star + \frac{1}{4}\mathbf{v}$, les masses de probabilité $p_{\pi_3,\mathcal{Q}_1^J}$ et $p_{\pi_3,\mathcal{Q}_4^J}$ étant associées aux deux points θ_1 et θ_4. □

L'exemple simple que nous venons de présenter met en évidence les deux types de non-unicité qui peuvent affecter l'ENPMV.

2.2.1.1 Non-unicité représentationnelle (en $\mathscr{P}$)

Nous abordons dans un premier temps le type de non-unicité de l'ENPMV déjà mis en évidence dans le Lemme 1. Cette non-unicité a été qualifiée par Gentleman et Vandal [GV01] de non-unicité *représentationnelle*. Elle est une limitation intrinsèque à l'estimation à partir de données censurées, et a été remarquée pour la première fois dans [Tur76] pour des données scalaires censurées par des intervalles. La non-unicité représentationnelle est généralement contournée en arbitrant un choix sur l'ensemble des estimateurs optimaux, reposant usuellement sur la forme des régions élémentaires. Par exemple dans le cas de censure par des produits d'intervalles pour des données bivariées, il est usuel de placer toute la masse d'une région au coin supérieur droit comme suggéré dans [GGI92].

Définition 4. *(**représentant d'une classe d'équivalence** $\mathscr{P}^{\mathbf{w}}$) Nous contournons la non-unicité représentationnelle en choisissant*

$$\pi_{\mathscr{P}^{\mathbf{w}}} \triangleq \sum_{m=1}^{M} w_m u_{\mathcal{Q}_m^J}.$$

comme représentant de la classe d'équivalence $\mathscr{P}^{\mathbf{w}}$ pour $\mathbf{w} \in \mathcal{S}^{M+}$. □

Selon cette convention, π_1 (Figure 2.4a) est le représentant de la classe $\mathscr{P}^{\mathbf{w}^\star}$.

2.2.1.2 Non-unicité de mélange (en $\mathcal{S}^{M+}$)

Nous nous intéressons maintenant au deuxième type de non-unicité mis en évidence dans l'exemple 3 (Figures 2.4b et 2.4c), où plusieurs vecteurs de probabilité dans $\mathcal{S}^{M+}$ maximisent $\mathcal{L}(\mathbf{w},\mathbf{n},\mathbf{B})$. Ce type de non-unicité a été désignée par Gentleman et Vandal [GV01] comme non-unicité de *mélange*.

Lemme 4. *Soit $\mathcal{W}^{\mathcal{L}} \subset \mathcal{S}^{M+}$ l'ensemble défini par*

$$\mathcal{W}^{\mathcal{L}} \triangleq \left\{\mathbf{w} \in \mathcal{S}^{M+}, \mathbf{B}\mathbf{w} = \mathbf{B}\hat{\mathbf{w}}^{\mathcal{L}}\right\},$$

avec $\hat{\mathbf{w}}^{\mathcal{L}} \in \mathcal{S}^{M+}$ maximisant $\mathcal{L}(\mathbf{w},\mathbf{n},\mathbf{B})$. L'ensemble $\mathcal{W}^{\mathcal{L}}$ est un polytope et contient tous les lois dans $\mathcal{S}^{M+}$ maximisant $\mathcal{L}(\mathbf{w},\mathbf{n},\mathbf{B})$. □

Le Lemme 4 montre que nous pouvons caractériser l'ensemble des vecteurs de probabilité maximisant $\mathcal{L}(\mathbf{w}, \mathbf{n}, \mathbf{B})$ par la somme d'une solution particulière $\hat{\mathbf{w}}^{\mathcal{L}}$ et un élément de l'espace nul de $\mathbf{B}$.

Nous avons clairement unicité de l'ENPMV (au sens mélange) si $\mathbf{B}$ est de rang plein. Ce Lemme a été démontré par Liu [Liu05] dans le cadre de données bivariées censurées par des produits d'intervalles. La démonstration dans le cas de censure par régions reste la même.

Preuve : L'ensemble $\mathcal{W}^{\mathcal{L}}$ est clairement un polytope car défini par les conditions linéaires $\mathbf{Bw} = \mathbf{B}\hat{\mathbf{w}}^{\mathcal{L}}$, $\mathbf{1}_M^{\mathsf{T}}\mathbf{w} = 1$ et $w_m \geqslant 0$ pour $m = 1, \ldots, M$. Tout élément $\mathbf{w}$ dans $\mathcal{W}^{\mathcal{L}}$ vérifie $\mathcal{L}(\mathbf{w}, \mathbf{n}, \mathbf{B}) = \mathcal{L}(\hat{\mathbf{w}}^{\mathcal{L}}, \mathbf{n}, \mathbf{B})$ et donc maximise la log-vraisemblance. Est ce que tout élément $\mathbf{w}$ maximisant $\mathcal{L}(\cdot, \mathbf{n}, \mathbf{B})$ vérifie $\mathbf{Bw} = \mathbf{B}\hat{\mathbf{w}}^{\mathcal{L}}$? Considérons la fonction allant de $\{\mathbf{Bw}; \mathbf{w} \in \mathcal{S}^{M+}\}$ vers $\mathbb{R}$ et liant chaque élément de la forme $\mathbf{Bw}$ à son image $\mathcal{L}(\mathbf{w}, \mathbf{n}, \mathbf{B})$. Cette fonction est strictement concave. Nous posons

$$C_{\mathbf{B}}^{\mathbf{w}_0} = \{\mathbf{Bw}; \mathbf{w} \in \mathcal{S}^{M+}, \mathcal{L}(\mathbf{w}, \mathbf{n}, \mathbf{B}) \geqslant \mathcal{L}(\mathbf{w}_0, \mathbf{n}, \mathbf{B})\},$$

avec $\mathbf{w}_0 = (\frac{1}{M}, \ldots, \frac{1}{M})^{\mathsf{T}}$. L'ensemble $C_{\mathbf{B}}^{\mathbf{w}_0}$ est un compact (fermé borné). Donc la restriction à $C_{\mathbf{B}}^{\mathbf{w}_0}$ est bornée et atteint sa borne à l'unique point $\mathbf{B}\hat{\mathbf{w}}^{\mathcal{L}}$. Nous en déduisons le Lemme. □

Grâce à la définition d'un représentant d'une classe d'équivalence $\mathscr{P}^{\mathbf{w}}$ (Définition 4), nous avons établi une règle de sélection afin d'éviter le problème de la non-unicité représentationnelle. Nous choisissons une deuxième règle de sélection pour éviter la non-unicité de mélange. Motivé par l'objectif de capturer la plus grande diversité possible de la population étudiée, nous sélectionnons dans le polytope $\mathcal{W}^{\mathcal{L}}$, la distribution la moins informative, à savoir, celle d'entropie maximale.

Plusieurs fonctions d'entropie ont été considérées dans la résolution de la non-unicité de l'estimation de densités. L'entropie la plus utilisée dans ce contexte est l'entropie de Shannon en raison de son interprétation simple en termes de quantité d'information. Elle est définie pour une loi de probabilité π comme suit [4]

$$H_1(\pi) \stackrel{\Delta}{=} \mathbb{E}_\pi\left[-\log(\pi)\right]. \tag{2.12}$$

Alfréd Rényi [Ré61] à présenté l'entropie de Rényi comme une généralisation de celle de Shannon. Considérons le cas discret. La fonction $H_1(\cdot)$ dans l'équation (2.12) satisfait les 4 postulats donnés par Fadeev [Fad57]

- $H_1(w_1, \ldots, w_M)$ est symétrique en ses variables (pour $M > 1$).
- $H_1(t, 1-t)$ est une fonction continue en $t \in [0, 1]$.
- $H_1(\frac{1}{2}, \frac{1}{2}) = 1$
- Soient $\mathbf{w} = (w_1, \ldots, w_{M_1})$ et $\mathbf{v} = (v_1, \ldots, v_{M_2})$ deux mesures discrètes de probabilité respectivement dans $\mathcal{S}^{M_1+}$ et $\mathcal{S}^{M_2+}$. Nous notons $\mathbf{w} * \mathbf{v}$ leur produit défini par le vecteur $\{w_{m_1} v_{m_2}\}_{m_1=1, m_2=1}^{M_1, M_2}$ dans $\mathcal{S}^{M_1 M_2+}$. Nous avons $H_1(\mathbf{w} * \mathbf{v}) = H_1(\mathbf{w}) + H_1(\mathbf{v})$.

4. Nous avons donné dans l'équation (2.9) sa définition pour une loi de probabilité discrète.

Plusieurs fonctions peuvent satisfaire ces 4 propriétés. Parmi elles nous trouvons les fonctions de la forme

$$H_\alpha(\mathbf{w}) = H_\alpha(w_1, \ldots, w_M) = \frac{1}{1-\alpha} \log \left(\sum_{m=1}^{M} w_m^\alpha \right), \tag{2.13}$$

pour $\alpha > 0$ et $\alpha \neq 1$. $H_\alpha(\mathbf{w})$ peut aussi être interprété comme une mesure de l'incertitude liée à la distribution discrète $\mathbf{w}$. Nous avons

$$\lim_{\alpha \to 1} H_\alpha(\mathbf{w}) = -\sum_{m=1}^{M} w_i \log w_m = H_1(\mathbf{w}).$$

L'entropie de Rényi d'ordre $\alpha > 0$ est définie comme suit

$$H_\alpha(\mathbf{w}) = \begin{cases} \frac{1}{1-\alpha} \log \left(\sum_{m=1}^{M} w_m^\alpha \right) & \text{si } \alpha \neq 1 \\ -\sum_{m=1}^{M} w_m \log w_m & \text{sinon} \end{cases},$$

pour une mesure discrète de probabilité $\mathbf{w} \in \mathcal{S}^{M+}$. Pour une densité de probabilité $\pi \in \mathscr{P}$, l'entropie de Rényi d'ordre $\alpha > 0$ est définie comme suit

$$H_\alpha(\pi) = \begin{cases} \frac{1}{1-\alpha} \log \int_\Theta \pi^\alpha(\theta) d\theta & \text{si } \alpha \neq 1 \\ -\int_\Theta \pi(\theta) \log \pi(\theta) d\theta & \text{sinon} \end{cases}.$$

Le choix de l'entropie de Rényi d'ordre $\alpha = 2$ permet de déterminer l'ENPMV en résolvant un problème de programmation quadratique sous contraintes linéaires. En effet, si on considère un élément $\mathbf{w}^\star$ de $\mathcal{W}^{\mathcal{L}}$, la loi maximisant l'entropie de Rényi d'ordre 2 est

$$\hat{\mathbf{w}}^{\mathcal{L}} \stackrel{\Delta}{=} \underset{\mathbf{w} \in \mathcal{S}^{M+}}{\operatorname{argmin}} \sum_{m=1}^{M} \frac{w_m^2}{\nu\left(\mathcal{Q}_m^J\right)} \qquad \text{tel que} \qquad \mathbf{B}\mathbf{w} = \mathbf{B}\mathbf{w}^\star,$$

avec $\nu(\cdot)$ la mesure de Lebesgue comme on l'a mentionné auparavant.

Définition 5. *: Nous notons par la suite $\hat{\pi}_\star^{\mathcal{L}}$ la loi de probabilité maximisant la log-vraisemblance telle que $p_{\hat{\pi}_\star^{\mathcal{L}}, \mathcal{Q}^J}$ maximise l'entropie de Rényi d'ordre 2, et la masse de probabilité $p_{\hat{\pi}_\star^{\mathcal{L}}, \mathcal{Q}_m^J}$ est uniformément distribuée sur la région $\mathcal{Q}_m^J$, pour $m = 1, \ldots, M$.* □

Notons qu'à partir de la caractérisation d'un maximum local de $\mathcal{L}(\mathbf{w}, \mathbf{n}, \mathbf{B})$ par les conditions de Kuhn-Tucker, il est possible de déduire une condition pour l'unicité de l'ENPMV (au sens mélange) portant sur le rang d'une sous-matrice de $\mathbf{B}$, comme le démontrent Gentleman et Geyer [GG94]. Nous présentons ce résultat dans l'Annexe 2.A.

2.2.1.3 Unicité de l'estimation des probabilités des régions de censure

Nous venons de voir que $\hat{\pi}^{\mathcal{L}}$ est en général sujet à deux types de non-unicité. Malgré cela, l'estimation des masses de probabilité $\left\{p_{\pi,\mathcal{R}_\ell^j}\right\}_{j=1,\ell=1}^{J,L^j}$ attribuées aux régions de censures est unique.

Lemme 5. *Soit $\mathscr{P}^{\mathcal{L}} \subset \mathscr{P}$ l'ensemble des distributions de probabilité maximisant $\mathcal{L}(\pi, \mathbf{X}_n)$. Nous avons*

$$\mathscr{P}^{\mathcal{L}} = \bigcup_{\mathbf{w}\in\mathcal{W}^{\mathcal{L}}} \mathscr{P}^{\mathbf{w}},$$

et

$$\forall \pi \in \mathscr{P}^{\mathcal{L}} \qquad \pi\left(\mathcal{R}_\ell^j\right) = \hat{\pi}^{\mathcal{L}}\left(\mathcal{R}_\ell^j\right), \qquad \forall j = 1,\ldots,J,\ \ \forall \ell = 1,\ldots,L^j\ .$$

□

Le premier point du Lemme est un résultat immédiat des Lemmes 1 et 4. Le deuxième point découle immédiatement de la concavité stricte de la log-vraisemblance en $\left\{\pi(\mathcal{R}_\ell^j)\right\}_{j=1,\ell=1}^{J,L^j}$.

Preuve : Soient π_1 et π_2 deux lois de probabilité dans $\mathscr{P}^{\mathcal{L}}$. On note $\mathbf{w}^1 = p_{\pi_1,\mathcal{Q}^J}$ et $\mathbf{w}^2 = p_{\pi_2,\mathcal{Q}^J}$. Pour tout $\lambda \in (0,1)$ nous avons $\lambda\mathbf{w}^1 + (1-\lambda)\mathbf{w}^2 \in \mathcal{S}^{M+}$, et puisque le logarithme est strictement concave, nous avons

$$\begin{aligned}
\mathcal{L}(\lambda\mathbf{w}^1 + (1-\lambda)\mathbf{w}^2, \mathbf{n}, \mathbf{B}) &= \frac{1}{n}\sum_{k=1}^{K} n_k \log \mathbf{B}_{k.}(\lambda\mathbf{w}^1 + (1-\lambda)\mathbf{w}^2) \\
&\geqslant \frac{1}{n}\sum_{k=1}^{K} n_k \log \mathbf{B}_{k.}(\lambda\mathbf{w}^1) + \frac{1}{n}\sum_{k=1}^{K} n_k \log \mathbf{B}_{k.}((1-\lambda)\mathbf{w}^2) \\
&= \lambda\mathcal{L}(\mathbf{w}^1, \mathbf{n}, \mathbf{B}) + (1-\lambda)\mathcal{L}(\mathbf{w}^2, \mathbf{n}, \mathbf{B}),
\end{aligned}$$

avec égalité si et seulement si $\mathbf{B}_{k.}\mathbf{w}^1 = \mathbf{B}_{k.}\mathbf{w}^2$ pour tout k, c'est à dire, si les deux densités attribuent la même probabilité à chaque région de censure. □

2.2.2 Convergence ("*Consistency*")

Nous nous intéressons maintenant à la convergence de l'ENPMV, c'est à dire à son comportement limite lorsque nous avons un grand nombre d'observations. Dans le problème étudié dans cette thèse, nous pouvons considérer deux comportements asymptotiques :

1. quand les nombres n_j d'observations censurées par chacune des partitions tendent vers l'infini,
2. quand le nombre des partitions tend vers l'infini.

D'après le modèle d'observation que nous avons défini au paragraphe 2.1.1, page 9, l'ensemble des partitions auxquelles appartiennent les régions de censures est fixe et contient J partitions. Nous

serons donc plus intéressés par le premier comportement asymptotique où le nombre de partitions reste fixe.

Considérons à nouveau l'exemple 1 où nous supposons que $\pi_\theta \equiv \pi_1$ (Figure 2.4a). Nous supposons que toutes les observations sont censurées par les éléments des partitions $\mathcal{R}^1$ et $\mathcal{R}^2$. Le fait de faire tendre à l'infini le nombre d'observations censurées par chacune des deux partitions ne changera en rien la non-unicité représentationnelle dont souffre l'ENPMV. Les densités π_2 et π_3 (Figures 2.4b et 2.4c) ont la même vraisemblance que π_θ et donc maximisent $\mathcal{L}(\pi, \mathbf{X}_n)$ quand n_1 et n_2 tendent vers l'infini. Nous aurons pour la région élémentaire $\mathcal{Q}_1^J$, qui d'ailleurs ne change pas quand le nombre d'observations n tend vers l'infini :

$$\pi_2\left(\left\{\mathcal{Q}_1^J\right\}\right) \xrightarrow[n\to\infty]{} \frac{1}{4} \text{ et } \pi_3\left(\left\{\mathcal{Q}_1^J\right\}\right) \xrightarrow[n\to\infty]{} \frac{1}{2}.$$

Nous concluons à partir de cet exemple qu'on ne peut garantir dans l'absolu, la convergence de l'ENPMV.

Plusieurs auteurs ont traité la question de la convergence de l'ENPMV mais surtout dans le contexte de données censurées par des intervalles. Dans les travaux d'Ayer et al. [ABE+55], les auteurs ont démontré la convergence faible de l'ENPMV dans le cas où toutes les partitions de $\mathbb{R}$ sont de taille 2, c'est à dire des régions de censure de la forme $]-\infty, a_i)$ ou $(a_i, \infty[$. La consistance forte uniforme de l'ENPMV dans le cas de partitions de $\mathbb{R}$ de taille 2 a été démontrée pour une distribution π_θ continue par Groeneboom et Wellner [GW92], Wang et Gardiner [WG96] et Yu et *al.* [YSLW98] en utilisant différentes méthodes. Dans le cas de partitions de $\mathbb{R}$ de taille 3, la consistance forte uniforme de l'ENPMV a été démontré par Groeneboom et Wellner [GW92] et Yu et *al.* [YSLG98].

Schick et Yu [SY00] ont démontré dans le cas où $\Theta \subset \mathbb{R}$, avec des partitions de tailles quelconques, la convergence forte dans la topologie $L_1(\mu)$ pour une certaine mesure μ établie à partir des intervalles de censure. Une généralisation de ce résultat au cas de censure par des régions de formes quelconques dans $\Theta \subset \mathbb{R}^d$ est donnée par van der Vaart et Wellner [vdVW00]. Nous donnons ici un aperçu de cette généralisation.

Les résultats concernant la convergence de l'ENPMV présentés par van der Vaart et Wellner [vdVW00] sont énoncés dans un modèle d'observation différent de celui que nous avons considéré. Dans [vdVW00], une observation est la réalisation de trois variables aléatoires : un vecteur d'ensembles aléatoires constituant une partition de l'espace mesurable considéré, une variable aléatoire à valeurs dans $\mathbb{N}\backslash\{0\}$ décidant de la taille de la partition et un vecteur aléatoire binaire décidant de l'élément de partition selon lequel l'observation est censurée. Dans le cadre de ce modèle d'observation, le nombre des partitions peut lui aussi tendre vers l'infini quand le nombre d'observations est infiniment grand.

En supposant que la classe $\mathcal{C}_\mathcal{R}$ des régions de censure est une classe de Vapnik-Cervonenkis [vDVW96] des sous-ensembles de Θ, van der Vaart et Wellner démontrent la consis-

tance de l'ENPMV au sens de la distance de Hellinger :

$$h^2(\hat{\pi}^{\mathcal{L}}, \pi_\theta) \to 0 \text{ presque surement,} \tag{2.14}$$

où pour tout π_1 et π_2 deux mesures de probabilité absolument continues par rapport à une mesure μ :

$$h^2(\pi_1, \pi_2) = \frac{1}{2} \int \left(\sqrt{\frac{d\pi_1}{d\mu}} - \sqrt{\frac{d\pi_2}{d\mu}} \right)^2 d\mu.$$

Cette définition ne dépend pas de la mesure μ [C+67].

Ce résultat peut s'appliquer dans notre cas si nous arrivons à montrer que l'ensemble des régions de censure $\left\{ \mathcal{R}_\ell^j \right\}_{j=1,\ell=1}^{J,L^j}$ forme une classe de Vapnik-Cervonenkis. Le lemme 2.6.17 dans [vDVW96] affirme que la propriété de Vapnik-Cervonenkis pour une classe de régions se conserve par intersection et que les régions formées par les sous-graphes de fonctions monotones forment une classe de Vapnik-Cervonenkis (Lemme 2.6.18 dans [vDVW96]). En supposant que les régions de censure sont définies comme étant les intersections de sous-graphes de fonctions toutes décroissantes [5], nous pouvons affirmer que l'ensemble des régions de censure forme une classe Vapnik-Cervonenkis, et donc déduire la consistance de $\hat{\pi}^{\mathcal{L}}$ au sens de la distance de Hellinger. Toutefois ce résultat reste d'un faible intérêt dans notre cas, car ne nous intéressons pas au cas où le nombre de partitions tend vers l'infini.

2.3 Calcul de l'ENPMV

Dans cette section nous présentons des méthodes numériques pour la détermination de l'estimateur du maximum de vraisemblance. Nous verrons d'abord comment déterminer le support de l'ENPMV pour nous intéresser ensuite à la résolution du problème d'optimisation de $\mathcal{L}(\mathbf{w}, \mathbf{n}, \mathbf{B})$ par rapport à $\mathbf{w} \in \mathcal{S}^{M+}$. Ces méthodes seront appliquées dans la section 2.4 sur des données simulées.

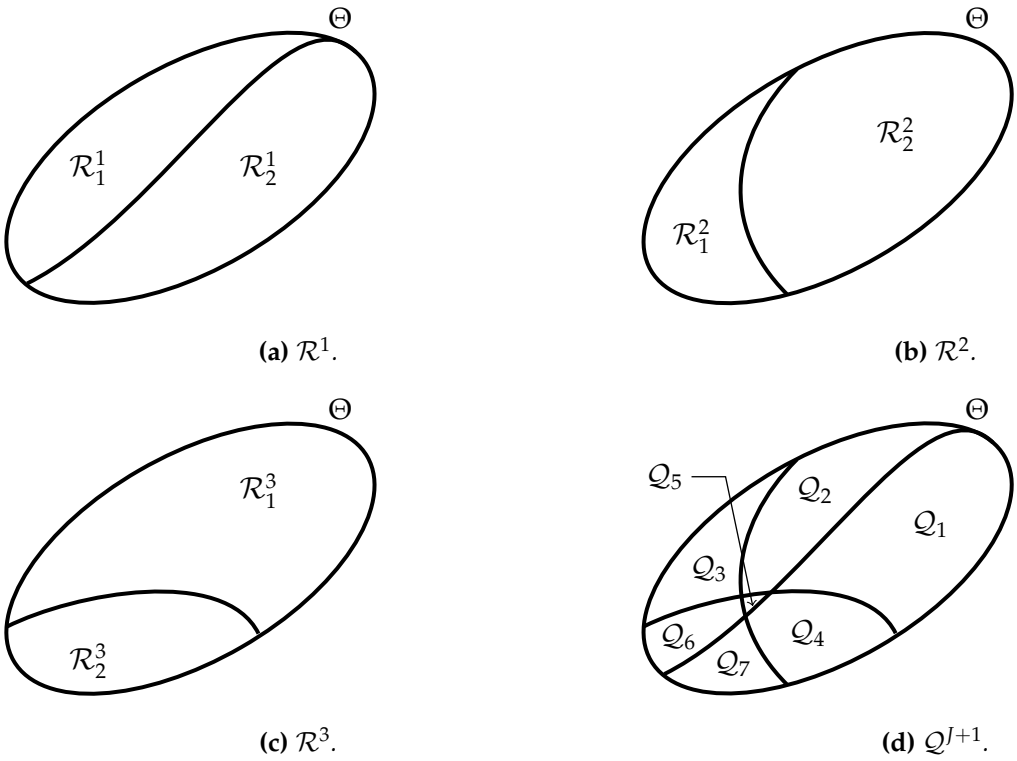
2.3.1 Support de l'ENPMV

Commençons cette section par un exemple qui montre comment le support de l'ENPMV change avec l'ajout d'une observation supplémentaire.

Exemple 4. Reprenons encore l'exemple 1 de la page 9. Admettons que nous avons une observation supplémentaire y_{n+1} censurée selon une nouvelle partition binaire de Θ, $\mathcal{R}^3$. Les Figures 2.5a et 2.5b rappellent les régions de censures issues des partitions $\mathcal{R}^1$ et $\mathcal{R}^2$. La Figure 2.5c montre les régions

5. Sans pouvoir le démontrer analytiquement, nous pensons à partir des simulations effectuées à l'aide du simulateur du modèle de décompression biophysique (Chapitre 4) que les régions de censures dans notre problème sont des intersections de sous-graphes de fonctions décroissantes (voir la forme des régions de censure dans les Figures 5.1c et 5.3, pages 107 et 110).

de censures issues de la nouvelle partition $\mathcal{R}^3$ et la Figure 2.5d montre la partition $\mathcal{Q}^{J+1}$ des régions élémentaires qui en résulte [6].



(a) $\mathcal{R}^1$. **(b)** $\mathcal{R}^2$. **(c)** $\mathcal{R}^3$. **(d)** $\mathcal{Q}^{J+1}$.

FIGURE 2.5 – *Exemple de l'impact d'une observation supplémentaire sur l'ENPMV.*

Nous notons $\mathbf{n}^\star$ et $\mathbf{B}^\star$ les mises à jour de $\mathbf{n}$ et $\mathbf{B}$ suite à l'ajout de l'observation y_{n+1}. Supposons que $y_{n+1} \in \mathcal{R}_1^3$. La log-vraisemblance correspondant au nouvel ensemble de $n+1$ observations pour un $\mathbf{w}$ dans $\mathcal{S}^{7+}$ est

$$\begin{aligned}
\mathcal{L}(\mathbf{w}, \mathbf{n}^\star, \mathbf{B}^\star) =& \frac{n/4}{n+1}\left(\log(p_{\pi_\mathbf{w},\mathcal{R}_1^1}) + \log(p_{\pi_\mathbf{w},\mathcal{R}_2^1}) + \log(p_{\pi_\mathbf{w},\mathcal{R}_1^2}) + \log(p_{\pi_\mathbf{w},\mathcal{R}_2^2})\right) \\
&+ \frac{1}{n+1}\log(p_{\pi_\mathbf{w},\mathcal{R}_1^3}) \\
=& \frac{n/4}{n+1}\left(\log(w_1+w_4+w_7) + \log(w_2+w_3+w_5+w_6)\right) + \\
& \frac{n/4}{n+1}\left(\log(w_1+w_2+w_4+w_5) + \log(w_3+w_6+w_7)\right) + \\
& \frac{1}{n+1}\log(w_1+w_2+w_3)
\end{aligned}$$

Un raisonnement simple nous permet de constater que le maximum de $\mathcal{L}(\mathbf{w}, \mathbf{n}^\star, \mathbf{B}^\star)$ doit nécessairement avoir $w_4 = 0$.

6. Dans la Figure 2.5d, nous notons les éléments $\left\{\mathcal{Q}_m^{J+1}\right\}_{m=1}^7$ par $\{\mathcal{Q}_m\}_{m=1}^7$ pour plus de lisibilité.

Effectivement, nous pouvons remarquer d'une part que $\mathcal{L}(\mathbf{w}, \mathbf{n}^\star, \mathbf{B}^\star)$ ne dépend de w_4 qu'à travers la somme $w_1 + w_4$ et d'autre part que w_1 apparait sans w_4 dans le terme $\frac{1}{n+1}\log(w_1 + w_2 + w_3)$. Donc la log-vraisemblance des solutions avec (w_1, w_4) tels que $w_4 \neq 0$ est strictement inférieure à celle des solutions (w_1', w_4'), avec $w_1' = w_1 + w_4$ et $w_4' = 0$. Le même raisonnement permet de montrer que la solution optimal doit avoir $w_5 = w_6 = 0$ au profit respectivement de w_2 et de w_3. Nous pouvons donc trouver les maxima de $\mathcal{L}(\mathbf{w}, \mathbf{n}^\star, \mathbf{B}^\star)$ en maximisant

$$\begin{aligned}\mathcal{L}(\mathbf{w}, \mathbf{n}^\star, \mathbf{B}^\star)|_{w_4=w_5=w_6=0} =& \frac{n/4}{n+1}\left(\log(w_1 + w_7) + \log(w_2 + w_3) + \log(w_1 + w_2)\right. \\ &\left. + \log(w_3 + w_7)\right) + \frac{1}{n+1}\log(w_1 + w_2 + w_3).\end{aligned}$$

Nous remarquons que les quatre premiers termes de cette expression correspondent, à un facteur multiplicatif près, à $\mathcal{L}(\mathbf{w}, \mathbf{n}, \mathbf{B})$ dans l'exemple 1 avec n observations. Ainsi tout vecteur dans $\{\mathbf{w}^\star + \alpha\mathbf{v}, \alpha \in \mathbb{R}\} \cap \mathcal{S}^{7+}$, avec $\mathbf{w}^\star = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, 0, 0, 0, \frac{1}{4})^\mathsf{T}$ et $\mathbf{v} = (1, -1, 1, 0, 0, 0, -1)^\mathsf{T}$, maximise la somme des quatre premiers termes de $\mathcal{L}(\mathbf{w}, \mathbf{n}^\star, \mathbf{B}^\star)$, notamment le vecteur $(\frac{1}{2}, 0, \frac{1}{2}, 0, 0, 0, 0)^\mathsf{T}$. Ce même vecteur maximise le dernier terme $\frac{1}{4(n+1)}\log(w_1 + w_2 + w_3)$. Nous pouvons donc conclure que $(\frac{1}{2}, 0, \frac{1}{2}, 0, 0, 0, 0)^\mathsf{T}$ maximise $\mathcal{L}(\mathbf{w}, \mathbf{n}^\star, \mathbf{B}^\star)$. □

L'exemple 4 illustre la non-stabilité du support de l'ENPMV, qui a été fortement perturbé suite à l'ajout d'une seule observation. La détermination du support de l'ENPMV pour des observations censurées à été étudiée par plusieurs auteurs. Gentleman et Vandal [GV01] soulignent l'équivalence entre ce problème de détermination du support de l'ENPMV avec censure par intervalles (ou produit d'intervalles pour des données multivariées) et le problème de recherche de cliques maximales dans un graphe. Le paragraphe suivant fait un rappel de cette équivalence, tout en présentant les modifications impliquées par la censure par des régions de forme arbitraire.

2.3.1.1 Rôle de la théorie des graphes dans la détermination du support de l'ENPMV

Nous commençons par présenter quelques notions classiques de la théorie des graphes [Gol04] qui nous seront utiles par la suite.

Définition 6. *[Gol04](**graphe, sous-graphe, graphe complet, clique, clique maximale et hypergraphe**)*

1. *Un graphe $\mathcal{G}(V, E)$ est défini par V, un ensemble fini de sommets, et $E \subset V \times V$, l'ensemble des arêtes entre les éléments de V.*
2. *Le sous-graphe $\mathcal{G}'(V', E')$ de $\mathcal{G}$ induit par $V' \subset V$ est le graphe tel que $E' = \{(u, v) \in E : u, v \in V'\}$.*
3. *Un graphe est dit complet si pour tout $(u, v) \in V \times V, (u, v) \in E$.*
4. *Un ensemble $C \subset V$ est une clique si le sous-graphe de $\mathcal{G}$ induit par C est complet.*
5. *Une clique est maximale si elle n'est pas un sous-graphe d'une autre clique.*
6. *Un hypergraphe $\mathcal{H}(V, E)$ est définit par un ensemble V de sommets et une famille non vide E de sous-ensembles de V appelés hyper-arêtes.* □

La notion d'hypergraphe est une généralisation de la notion de graphe dans le sens où une hyper-arête peut lier plus que deux sommets. Le Lemme suivant, dont la démonstration est immédiate, sera utile par la suite.

Lemme 6. *Soit $\mathcal{G} = (V, E)$ un graphe, et C une clique de $\mathcal{G}$. Alors tout sous-ensemble de C est aussi une clique de $\mathcal{G}$.* □

Nous allégeons les notations en réindexant les régions de censure $\left\{\mathcal{R}_\ell^j\right\}_{j=1,\ell=1}^{J,L^j}$ de 1 à K : $\{\mathcal{R}_k\}_{k=1}^K$ où $K = \sum_{j=1}^J L^j$ comme défini auparavant.

Définition 7. *(**graphe d'intersection**) Le graphe d'intersection associé à $\{\mathcal{R}_k\}_{k=1}^K$, noté $\mathcal{G}(V, E)$ est défini par $V = \{1, \ldots, K\}$ (chaque sommet représente une région $\mathcal{R}_k$) et $E = \{(k_1, k_2) \in V \times V : \mathcal{R}_{k_1} \cap \mathcal{R}_{k_2} \neq \emptyset\}$ (il existe une arête entre deux sommets si et seulement si les régions correspondantes s'intersectent).* □

Définition 8. *(**représentation réelle**) Soit $\mathcal{G}(V, E)$ le graphe d'intersection associé aux régions $\{\mathcal{R}_k\}_{k=1}^K$. Soit $V' \subset V = \{1, \ldots, K\}$. La représentation réelle de V', $\mathscr{R}(V')$, est la région obtenue par intersection de tous les sommets de*
V' : $\mathscr{R}(V') \triangleq \bigcap_{k \in V'} \mathcal{R}_k$. □

Définition 9. *($\mathcal{H}_{\max}(\mathcal{G})$) On note $\mathcal{H}_{\max}(\mathcal{G}) \triangleq (V, E_{\max})$ l'hypergraphe dont les hyper-arêtes E_{max} sont les cliques maximales de $\mathcal{G}(V, E)$.* □

Le Lemme 7 concerne le cas où tous les éléments des partitions $\left\{\mathcal{R}^j\right\}_{j=1}^J$ sont observés.

Lemme 7. *Supposons que toutes les régions de censure $\{\mathcal{R}_\ell^j\}_{j=1,\ell=1}^{J,L^j}$ sont observées. Alors toutes les régions élémentaires $\{\mathcal{Q}_m^J\}_{m=1}^M$ sont des représentations réelles de cliques maximales.* □

Preuve : D'après la définition des régions élémentaires (Définition 2) comme étant la plus petite partition telle que $\sigma(\mathcal{Q}^J)$ contient toutes les régions de censures et la définition des ensembles ξ_ℓ^j (Définition 3), nous déduisons que si $m \in \xi_\ell^j$ alors $\mathcal{Q}_m^J \subset \mathcal{R}_\ell^j$ sinon $\mathcal{Q}_m^J \cap \mathcal{R}_\ell^j = \emptyset$. Puisque les $\left\{\mathcal{R}^j\right\}_{j=1}^J$ sont des partitions, nous déduisons que pour tout $j = 1, \ldots, J$, il existe un seul indice dans $\{1, \ldots, L^j\}$, noté ℓ_m^j, tel que $\mathcal{Q}_m^J \subset \mathcal{R}_{\ell_m^j}^j$, et pour $\ell \neq \ell_m^j$ on a $\mathcal{Q}_m^J \cap \mathcal{R}_\ell^j = \emptyset$. Nous en déduisons que $\mathcal{Q}_m^J = \bigcup_{j=1}^J \mathcal{R}_{\ell_m^j}^j$. Nous notons

$$\mathscr{R}^{-1}\left(\mathcal{Q}_m^J\right) = \left\{k \in \{1, \ldots, K\};\ \exists j :\ \mathcal{R}_k = \mathcal{R}_{\ell_m^j}^j\right\}.$$

Il est clair que $\mathscr{R}^{-1}\left(\mathcal{Q}_m^J\right)$ est une clique (si $k, k' \in \mathscr{R}^{-1}\left(\mathcal{Q}_m^J\right)$ alors $\mathcal{Q}_m^J \subset \mathcal{R}_k \cap \mathcal{R}_{k'}$). La représentation réelle de $\mathscr{R}^{-1}\left(\mathcal{Q}_m^J\right)$ est $\mathcal{Q}_m^J$. Supposons que $\mathscr{R}^{-1}\left(\mathcal{Q}_m^J\right)$ n'est pas maximale. Alors il existe $k \notin \mathscr{R}^{-1}\left(\mathcal{Q}_m^J\right)$ tel que pour tout $k' \in \mathscr{R}^{-1}\left(\mathcal{Q}_m^J\right)$, $\mathcal{R}_k \cap \mathcal{R}_{k'} \neq \emptyset$. Cela implique que deux éléments d'une même partition sont d'intersection non-vide. Donc $\mathscr{R}^{-1}\left(\mathcal{Q}_m^J\right)$ est une clique maximale. □

Gentleman et Vandal [GV01] ont démontré le résultat suivant :

Théorème 1. *([GV01]) Soit $\hat{\pi}^{\mathcal{L}}$ un ENPMV de la distribution π_θ pour des régions de censure de la forme $\{\mathcal{R}_k\}_{k=1}^K = \left\{\prod_{m=1}^d [a_{mk}, b_{mk}]\right\}_{k=1}^K$. Notons $\mathcal{G}(V,E)$ le graphe d'intersection correspondant et $E_{\max}$ l'ensemble des cliques maximales de $\mathcal{G}$. Alors, le support $S_{\mathcal{L}}$ de $\hat{\pi}^{\mathcal{L}}$ est contenu dans l'union des représentations réelles des éléments de $E_{\max}$, c'est à dire $S_{\mathcal{L}} = \bigcup_{C\in E_{\max}} \mathscr{R}(C)$.* □

Ce théorème affirme que pour tout ensemble $\mathcal{A} \subset \Theta$, si $\mathcal{A} \cap S_{\mathcal{L}} = \emptyset$, alors $p_{\hat{\pi}^{\mathcal{L}},\mathcal{A}} = 0$. La démonstration de ce théorème repose sur l'idée qu'en déplaçant toute la masse de probabilité d'une région $\mathcal{R}_k$ dans la représentation réelle d'une clique maximale C contenant le sommet k, cette masse de probabilité figurera dans les termes de $\mathcal{L}(\pi, \mathbf{X}_n)$ impliquant la région $\mathcal{R}_k$ mais aussi dans les termes impliquant les autres régions qui forment la clique maximale C, faisant augmenter la valeur de $\mathcal{L}(\pi, \mathbf{X}_n)$.

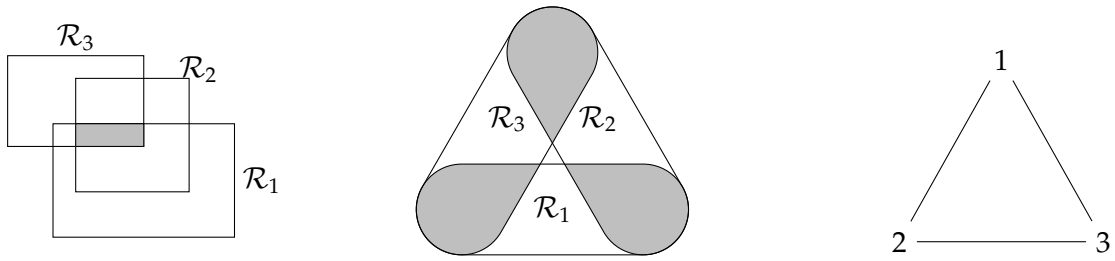
Si nous supposons que les régions de censure $\{\mathcal{R}_k\}_{k=1}^K$ peuvent avoir des formes quelconques dans Θ, le théorème précédent n'est plus valide. En effet, sa démonstration repose sur le fait que pour toute collection d'intervalles qui s'intersectent deux à deux, leur intersection est non-vide. Cela n'est plus vrai dans le cas général où l'on considère des régions de forme arbitraire, ou même dans le cas plus contraint de polyèdres. Cette remarque est mise en évidence dans l'exemple 5.

Exemple 5. La Figure 2.6a montre l'exemple de trois produits d'intervalles dans $\mathbb{R}^2$ qui s'intersectent deux à deux. La Figure 2.6c représente le graphe d'intersection correspondant, où $E_{\max}$ contient une seule clique maximale $\{1,2,3\}$. La région grisée dans 2.6a est la représentation réelle de cette clique maximale $\mathscr{R}(\{1,2,3\})$. C'est cette région qui captera toute la masse de probabilité de l'ENPMV en cas d'observation des régions de censure $\mathcal{R}_1$, $\mathcal{R}_2$ et $\mathcal{R}_3$. La Figure 2.6b montre un exemple de 3 régions dans $\mathbb{R}^2$ qui s'intersectent deux à deux mais dont l'intersection est vide. Même si le graphe d'intersection associé à ces 3 régions est le même que celui des 3 régions de la Figure 2.6c, le support de l'ENPMV n'est pas $\mathscr{R}(\{1,2,3\})$. Comme nous le verrons dans le Théorème 2, le support correspond aux régions grisées dans la Figure 2.6b. □

Pour des régions de censure de forme arbitraire, le support de $\hat{\pi}^{\mathcal{L}}$ est encore déterminé par leur graphe d'intersection, mais il est caractérisé par un hypergraphe plus complexe que celui des cliques maximales. L'Algorithme 1 présenté ci-dessous, construit récursivement, à partir de l'hypergraphe des cliques maximales, un hypergraphe

$$\mathcal{H}'_{\max}(\mathcal{G}) \stackrel{\Delta}{=} (V, E'_{\max}), \tag{2.15}$$

tel que les représentations réelles de ses hyper-arêtes coïncident avec le support de tout ENPMV. L'idée de base de cet algorithme est de remplacer récursivement les cliques maximales dont les représentations réelles sont vides par leurs sous-graphes. L'Algorithme 1 utilise la notion de *nombre de clique* définie ci-dessous.



(a) 3 *intervalles.* **(b)** 3 *regions.* **(c)** *Graphe d'intersection.*

FIGURE 2.6 – *Exemple de la différence entre intersection d'intervalles et intersection de régions de forme quelconque. Les régions grisées correspondent au support de l'ENPMV*

Définition 10. *(**clique doublement maximale et nombre de clique**) Soit $\mathcal{G}$ un graphe et $E_{\max}$ la liste de ses cliques maximales. Une clique $C_{\max} \in E_{\max}$ est dite doublement maximale si son cardinal $\#(C_{\max})$ est le plus grand. Ce cardinal noté $\#(\mathcal{G})$ est appelé nombre de clique du graphe $\mathcal{G}$.* □

Nous remarquons qu'une clique doublement maximale n'est pas forcément unique et est nécessairement maximale.

Algorithme 1 : Construction de $\mathcal{H}'_{\max}(\mathcal{G})$
donnée : $\mathcal{H}_{\max}(\mathcal{G})$
sortie : $\mathcal{H}'_{\max}(\mathcal{G})$

1 : On répartie les cliques contenues dans $E_{\max}$ selon leurs tailles dans au plus $\#(\mathcal{G})$ ensembles : $\mathcal{E}_1, \ldots, \mathcal{E}_{\#(\mathcal{G})}$. On initialise $l = \#(\mathcal{G})$.

2 : Pour toute clique C dans $\mathcal{E}_l$, on calcule $\mathscr{R}(C)$.
-Si $\mathscr{R}(C) \neq \emptyset$, on garde $C \in \mathcal{E}_l$ et toutes les cliques dont les sommets sont contenus dans C, de taille $l' < l$ sont retirées de $\mathcal{E}_{l'}$.
-Si $\mathscr{R}(C) = \emptyset$, on retire C de $\mathcal{E}_l$.

3 : Si $l > 1$, $l \leftarrow l - 1$. Aller à étape 2.

4 : $E'_{\max} = \bigcup_{l=1,\ldots,\#(\mathcal{G})} \mathcal{E}_l$

Nous remarquons que par construction, si C est une clique maximale de $\mathcal{G}$ avec une représentation réelle non vide, alors $C \in E'_{\max}$ et aucun sous-ensemble de C n'est contenu dans $E'_{\max}$. Le Théorème 2 est la généralisation du Théorème 1 au cas où les régions de censure ne sont pas nécessairement des intervalles.

Théorème 2. *Soit $\hat{\pi}^{\mathcal{L}}$ un ENPMV de la distribution π_θ pour des régions de censure de forme quelconque. Notons $\mathcal{G}$ le graphe d'intersection correspondant aux régions de censure $\{\mathcal{R}_k\}_{k=1}^K$. L'estimateur $\hat{\pi}^{\mathcal{L}}$ ne met pas*

7. Bien évidemment s'il n'existe aucune clique de taille l, on a $\mathcal{E}_l = \emptyset$.

de masse de probabilité en dehors de

$$S^\star_{\mathcal{L}} \triangleq \bigcup_{C \in E'_{\max}} \mathscr{R}(C) \ ,$$

où $E'_{\max}$ est la liste des hyper-arrêtes de $\mathcal{H}'_{\max}(\mathcal{G})$, l'hypergraphe construit par l'Algorithme 1. □

Le Théorème 2 exprime une condition nécessaire pour qu'une région appartienne au support d'un ENPMV. Si $\mathcal{A} \subset \Theta$ est telle que $\mathcal{A} \cap S^\star_{\mathcal{L}} = \emptyset$, alors $p_{\hat{\pi}^{\mathcal{L}},\mathcal{A}} = 0$. La démonstration de ce théorème est donnée dans l'Annexe 2.B.

Nous avons vu que l'Algorithme 1 repose sur $E_{\max}$, la liste des cliques maximales du graphe d'intersection. Identifier les cliques maximales d'un graphe est l'un des problèmes fondamentaux en théorie des graphes. Plusieurs algorithmes [BK73, TIAS77, PX94, BBPP99] ont été proposés et évalués expérimentalement ou théoriquement afin de résoudre ce problème générique.

Dans le cadre de cette thèse, il ne s'agit pas de déterminer les cliques maximales d'un graphe quelconque, mais celles d'un graphe d'intersection. Cette restriction peut permettre certaines simplifications, notamment dans le cas où les régions de censures sont des intervalles, et donc d'exploiter les différentes propriétés concernant leurs intersections afin de déduire les cliques maximales.

Plusieurs auteurs se sont intéressés au problème de déduire les cliques maximales d'un graphe d'intersection d'un ensemble d'intervalles (où de produits d'intervalles dans le cas de données multivariées). Gentleman et Vandal [GV01] et Song [Son01] ont proposé deux algorithmes pour calculer les cliques maximales pour des régions rectangulaires, dans le cadre de l'estimation à partir de données bivariées. Ces algorithmes reposent sur des caractéristiques propres aux intervalles (où les produits d'intervalles dans le cas multivarié) et ne sont pas exploitables dans notre cadre. Par exemple, dans le cadre de données univariées censurées par des intervalles, une solution pour déduire les cliques maximales du graphe d'intersection est de classer les points extrêmes des différents intervalles de la forme (a_i, b_i) et de déterminer toutes les paires $a_i < b_j$ de façon à ce qu'il n'y ait pas d'autres points extrêmes à l'intérieur de (a_i, b_j). Un algorithme simple appliqué à ces paires permet de déduire les cliques maximales.

Dans le paragraphe suivant, nous présentons un algorithme pour la détermination de la liste $E'_{\max}$ des éléments du support de $\hat{\pi}^{\mathcal{L}}$ tout en évitant de calculer la liste $E_{\max}$ des cliques maximales du graphe d'intersection. Cela est possible en exploitant la matrice $\mathbf{B}$.

2.3.1.2 Détermination de $E'_{\max}$ à partir de la matrice B

Le Théorème 2 caractérise le support de tout ENPMV en mettant en évidence la structure $E'_{\max}$ déterminée à partir de la liste $E_{\max}$ des cliques maximale par l'Algorithme 1. L'utilisation de la description de toutes les intersections entre les régions de censure nous permet d'éviter de calculer $E_{\max}$.

Alors que la représentation de régions avec une géométrie simple, comme pour le cas des intervalles, est possible avec un nombre fini de paramètres (2 dans le cas d'intervalles de $\mathbb{R}$), la représentation de régions $\mathcal{R} \subset \mathbb{R}^d$ de forme arbitraire peut être complexe au sens du nombre de paramètres

nécessaires. Deux approches sont possibles : soit on représente la frontière des régions, c'est à dire une surface de dimension $d-1$ (la représentation naturelle pour les intervalles), soit on décrit leur intérieur.

Dans notre cas, la géométrie des régions de censure est résumée par la matrice $\mathbf{B}$, définie par les équations (2.5) et (2.7), qui caractérise les intersections entre la collection des régions de censure $\{\mathcal{R}_k\}_{k=1}^{K}$ en précisant l'ensemble des régions élémentaires appartenant à chaque régions de censure. La construction de La matrice $\mathbf{B}$ pour le jeu de données réelles dont nous disposons sera décrite dans le paragraphe 5.1.2, page 110.

L'Algorithme 2 permet une construction efficace de E'_{max} dans le cas où nous disposons de la matrice $\mathbf{B}$.

Algorithme 2 : Identification de $E'_{\max}$

donnée : $\mathbf{B}$

sortie : $E'_{\max}$

$E'_{\max} = \emptyset$,
$\mathrm{y}_{\mathbf{B}} = \mathbf{1}_K^{\mathsf{T}}\mathbf{B}$
tant que $\max(\mathrm{y}_{\mathbf{B}}) > 0$
    $E = \{m;\ \mathrm{y}_{\mathbf{B}}(m) = \max(\mathrm{y}_{\mathbf{B}})\}$
    pour tout $m \in E$
        $\mathrm{y}_{\mathbf{B}}(m) = 0$
        $E'_{\max} \leftarrow E'_{\max} \cup \{k;\ \mathbf{B}_{km} = 1\}$
        pour tout s tel que $\mathrm{y}_{\mathbf{B}}(m) \neq 0$
            si $-1 \notin \mathbf{B}_{.m} - \mathbf{B}_{.s}$
                $\mathrm{y}_{\mathbf{B}}(s) = 0$
            fin si
        fin pour
    fin pour
fin tant que

Théorème 3. *L'Algorithme 2 identifie la liste E'_{max} des hyper-arrêtes de $\mathcal{H}'_{\max}$ (Équation 2.15), dont les représentations réelles de ses éléments constituent le support de $\hat{\pi}^{\mathcal{L}}$.* □

La preuve de Théorème 3 ainsi qu'un exemple illustrant l'application de l'Algorithme 2 sont donnés dans l'Annexe 2.C.

Remarques :

– Dans le Théorème 3, la condition pour appartenir au support de $\hat{\pi}^{\mathcal{L}}$ est nécessaire et non suffisante. L'exemple 7, page 58, montre une situation où toutes les régions élémentaires corres-

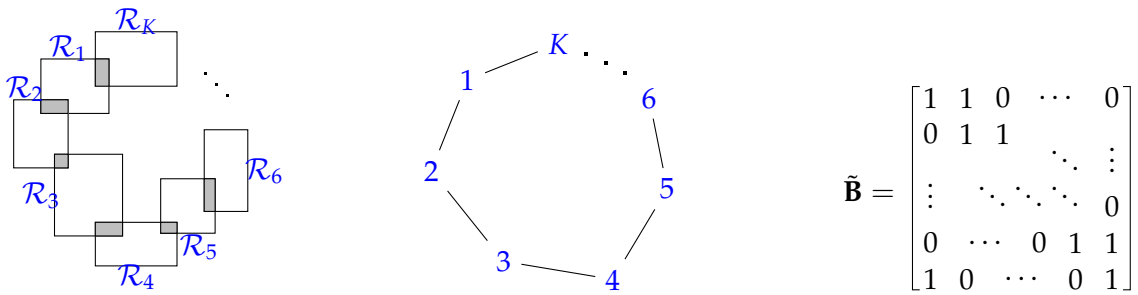
pondent à des éléments de la liste E'_{max} alors qu'il existe une région élémentaire n'appartenant pas au support.

– D'après l'Algorithme 2 et le Lemme 7, page 26, si toutes les régions de censure sont observées, alors les régions élémentaires (forcément toutes différentes de l'ensemble vide) sont les représentations réelles d'éléments de E'_{max}, donc à priori appartiennent au support de l'estimateur $\hat{\pi}^{\mathcal{L}}$.

– Une fois que le support de $\hat{\pi}^{\mathcal{L}}$ est calculé, nous réduisons la taille du problème de l'optimisation de $\mathcal{L}(\mathbf{w}, \mathbf{n}, \mathbf{B})$ par rapport à $\mathbf{w} \in \mathcal{S}^{M+}$, en supprimant les colonnes dans $\mathbf{B}$ correspondant aux régions élémentaires $\mathcal{Q}_m^J$ dans $\mathcal{Q}^J$ qui n'appartiennent pas au support. Ainsi nous obtenons la matrice $\tilde{\mathbf{B}}$ de taille $K \times M'$ avec $M' \triangleq \#\left(E'_{\max}\right)$. Le problème devient pour $\mathbf{w} \in \mathcal{S}^{M'+}$

$$\mathcal{L}(\mathbf{w}, \mathbf{n}, \tilde{\mathbf{B}}) = \frac{1}{n} \sum_{k=1}^{K} n_k \log \tilde{\mathbf{B}}_{k.} \mathbf{w}, \qquad \text{et} \qquad \hat{\mathbf{w}}^{\mathcal{L}} = \underset{\mathbf{w} \in \mathcal{S}^{M'+}}{\operatorname{argmax}} \left(\mathcal{L}(\mathbf{w}, \mathbf{n}, \tilde{\mathbf{B}}) \right).$$

– Nous avons discuté l'unicité de l'ENPMV au paragraphe 2.2.1. Dans certains cas spécifiques l'unicité peut être déduite depuis le graphe d'intersection. Maathuis [Maa03] remarque que si le graphe d'intersection est un cycle contenant un nombre impair de sommets, alors l'ENPMV est unique. En effet, considérons un cycle de taille K impair. Alors nous avons $M' = K$ cliques maximales. La matrice $\tilde{\mathbf{B}}$ correspondante est une matrice de Toeplitz et est de rang plein. Nous en déduisons alors l'unicité. Les Figures 3.1a, 3.1b et 3.1c illustrent cette situation.



(a) *Régions de censures.* **(b)** *Graphe d'intersection.* **(c)** *Matrice* $\tilde{\mathbf{B}}$

FIGURE 2.7 – *Exemple de l'unicité de l'ENPMV dans le cas d'un graphe d'intersection sous forme d'un cycle impair.*

2.3.2 Optimisation

Afin de calculer l'ENPMV, nous sommes amenés à calculer $\hat{\mathbf{w}}^{\mathcal{L}}$ qui maximise $\mathcal{L}(\mathbf{w}, \mathbf{n}, \tilde{\mathbf{B}})$ (ou $\mathcal{L}(\mathbf{w}, \mathbf{n}, \mathbf{B})$ si on n'applique pas la réduction du support (paragraphe 2.3.1.2) et on considère la totalité des éléments de la partition $\mathcal{Q}^J$). Plusieurs résultats relatifs à l'estimation non-paramétrique par maximum de vraisemblance ont des équivalents dans la théorie des plans d'expérience optimaux. Dans cette section, nous soulignons le lien entre le problème d'optimisation de $\mathcal{L}(\mathbf{w}, \mathbf{n}, \tilde{\mathbf{B}})$ et

un problème de construction de plan d'expérience D-optimal. Nous décrivons ensuite les différents algorithmes d'optimisation que nous avons testés. Enfin nous comparons les performances de ces algorithmes en termes de nombre d'itérations et de temps de calcul.

2.3.2.1 Un problème de plan d'expérience D-optimal

Commençons par décrire le contexte des plans d'expériences. On considère une variable dépendante Y, un vecteur de régresseurs $\mathbf{x}^\mathsf{T} = (x_1, \ldots, x_p)^\mathsf{T}$ et un modèle de régression linéaire $\mathbb{E}(y) = \beta^\mathsf{T}\mathbf{x}$ basé sur un échantillon de taille N [8]. Pour des erreurs d'observation i.i.d. de moyenne nulle et variance finie, le meilleur estimateur linéaire et sans biais de β est $\hat{\beta} = \left(\mathbf{X}^\mathsf{T}\mathbf{X}\right)^{-1}\mathbf{X}^\mathsf{T}\mathbf{Y}$, avec $\mathbf{X} = (\mathbf{x}_1^\mathsf{T}, \ldots, \mathbf{x}_N^\mathsf{T})^\mathsf{T}$ la matrice du plan d'expérience et $\mathbf{Y} = (y_1, \ldots, y_N)^\mathsf{T}$. La matrice de covariance de l'estimateur $\hat{\beta}$ est proportionnelle à $\left(\mathbf{X}^\mathsf{T}\mathbf{X}\right)^{-1}$. On peut réécrire $\mathbf{X}^\mathsf{T}\mathbf{X}$ sous la forme $N\sum_{i=1}^N \mathbf{x}_i\mathbf{x}_i^\mathsf{T} p_i$, avec $p_i = \frac{1}{N}$. La question de plan d'expérience optimal se pose lorsqu'on cherche l'ensemble de N points qui minimisent dans un certain sens la matrice de covariance de l'estimateur. On utilise fréquemment le déterminant comme critère d'optimalité. Une simplification consiste à maximiser $\det\sum_{i=1}^N \mathbf{x}_i\mathbf{x}_i^\mathsf{T} p_i$ sous les conditions que $p_i \geq 0$ pour tout $i = 1, \ldots, N$ et $\sum_{i=1}^N p_i = 1$. On parle alors de problème de plan d'expérience D-optimal.

Revenons au problème de recherche de l'ENPMV en optimisant $\mathcal{L}(\mathbf{w}, \mathbf{n}, \tilde{\mathbf{B}})$. Nous introduisons la matrice $\tilde{\mathbf{A}}$ de taille $n \times M'$ constituée des lignes de la matrice $\tilde{\mathbf{B}}$ où la ℓ-ième ligne de $\tilde{\mathbf{B}}_{j\cdot}$ est répétée n_ℓ^j fois. Pour tout $m = 1 \ldots, M'$, nous notons $\mathbf{H}_m(\mathbf{n}, \tilde{\mathbf{B}})$ la matrice de taille $n \times n$

$$\mathbf{H}_m(\mathbf{n}, \tilde{\mathbf{B}}) \stackrel{\Delta}{=} \operatorname{diag}\{\tilde{\mathbf{A}}_{im}, i = 1, \ldots, n\}.$$

Pour tout $i = 1, \ldots, n$, il existe au moins un indice m tel que $\tilde{\mathbf{A}}_{im} = 1$. Alors il existe au moins une matrice $\mathbf{H}_m(\mathbf{n}, \tilde{\mathbf{B}})$ appartenant à $\mathbb{M}_n^{\geq}$ l'ensemble des matrices symétriques définies positives (non strictement) de taille $n \times n$. En posant

$$\mathbf{M}(\mathbf{w}, \mathbf{n}, \tilde{\mathbf{B}}) \stackrel{\Delta}{=} \sum_{m=1}^{M'} w_m \mathbf{H}_m(\mathbf{n}, \tilde{\mathbf{B}}),$$

nous pouvons réécrire la log-vraisemblance

$$\mathcal{L}(\mathbf{w}, \mathbf{n}, \tilde{\mathbf{B}}) = \frac{1}{n}\log\det\mathbf{M}(\mathbf{w}, \mathbf{n}, \tilde{\mathbf{B}}).$$

La détermination du vecteur $\hat{\mathbf{w}}^{\mathcal{L}}$ maximisant $\mathcal{L}(\mathbf{w}, \mathbf{n}, \tilde{\mathbf{B}})$ pour $\mathbf{w} \in \mathcal{S}^{M'+}$ correspond donc à un problème de plan d'expérience D-optimal associé à la matrice $\mathbf{M}(\mathbf{w}, \mathbf{n}, \tilde{\mathbf{B}})$, avec $\mathbf{w}$ considéré comme une mesure de plan d'expérience attribuant le poids w_m à la matrice élémentaire $\mathbf{H}_m(\mathbf{n}, \tilde{\mathbf{B}})$. D'importantes propriétés découlent de cette équivalence avec le problème de recherche de plan d'expérience D-optimal.

8. Les notations dans ce paragraphe sont propres à la description du problème de plan d'expérience D-optimal

Nous définissons pour $\mathbf{w} \in \mathcal{S}^{M'+}$ et $m = 1, \ldots, M'$

$$d(\mathbf{w}, m) \stackrel{\Delta}{=} \text{trace}\left[\mathbf{M}^{-1}(\mathbf{w}, \mathbf{n}, \tilde{\mathbf{B}})\mathbf{H}_m(\mathbf{n}, \tilde{\mathbf{B}})\right] = \sum_{i=1}^{n} \frac{\tilde{\mathbf{A}}_{im}}{\tilde{\mathbf{A}}_{i.}\mathbf{w}} = \sum_{k=1}^{K} \frac{n_k \tilde{\mathbf{B}}_{km}}{\tilde{\mathbf{B}}_{k.}\mathbf{w}}.$$

Pour $\mathbf{w}, \mathbf{w}' \in \mathcal{S}^{M'+}$, nous notons $F(\mathbf{w}, \mathbf{w}')$ la dérivée directionnelle de $\mathcal{L}(\cdot, \mathbf{n}, \tilde{\mathbf{B}})$ en $\mathbf{w}$ dans la direction $\mathbf{w}'$

$$\begin{aligned} F(\mathbf{w}, \mathbf{w}') &\stackrel{\Delta}{=} \lim_{\alpha \to 0^+} \frac{\mathcal{L}((1-\alpha)\mathbf{w} + \alpha\mathbf{w}', \mathbf{n}, \tilde{\mathbf{B}}) - \mathcal{L}(\mathbf{w}, \mathbf{n}, \tilde{\mathbf{B}})}{\alpha} \\ &= \sum_{i=1}^{n} \frac{1}{n} \frac{\tilde{\mathbf{A}}_{i.}\mathbf{w}'}{\tilde{\mathbf{A}}_{i.}\mathbf{w}} - 1 \\ &= \sum_{k=1}^{K} \frac{n_k}{n} \frac{\tilde{\mathbf{B}}_{k.}\mathbf{w}'}{\tilde{\mathbf{B}}_{k.}\mathbf{w}} - 1. \end{aligned}$$

Donc pour $\mathbf{e}_m$ le m-ème vecteur de la base canonique de $\mathbb{R}^{M'}$, nous avons

$$F(\mathbf{w}, \mathbf{e}_m) = \sum_{i=1}^{n} \frac{1}{n} \frac{\tilde{\mathbf{A}}_{im}}{\tilde{\mathbf{A}}_{i.}\mathbf{w}} - 1 = \frac{d(\mathbf{w}, m)}{n} - 1.$$

Le Théorème de Kiefer et Wolfowitz [KW60], appelé aussi Théorème d'équivalence, donne une condition nécessaire et suffisante pour l'optimalité de $\mathbf{w}^\star$ maximisant $\mathcal{L}(\mathbf{w}, \mathbf{n}, \tilde{\mathbf{B}})$.

Théorème 4. *(**Théorème d'équivalence**) : Soit $\mathbf{w}^\star \in \mathcal{S}^{M'+}$; les propriétés suivantes sont équivalentes :*

- *$\mathbf{w}^\star$ est D-optimal (c'est à dire $\mathbf{w}^\star$ maximise $\det \mathbf{M}(\mathbf{w}, \mathbf{n}, \tilde{\mathbf{B}})$) ;*
- *$\mathbf{w}^\star$ minimise $\max_{m=1,\ldots,M'} d(\mathbf{w}, m)$;*
- *$\max_{m=1,\ldots,M'} d(\mathbf{w}^\star, m) = n.$* □

Nous remarquons que la fonction $\log\det(\cdot)$ est strictement concave sur l'ensemble des matrices symétriques définies positives. Nous en déduisons que $\mathbf{M}(\mathbf{w}^\star, \mathbf{n}, \tilde{\mathbf{B}})$ la matrice optimale est unique (mais $\mathbf{w}^\star$ n'est pas nécessairement unique).

Nous pouvons majorer les poids $w_m^\star$ du vecteur $\mathbf{w}^\star$ maximisant la fonction de la log-vraisemblance [HT09].

Théorème 5. *(**Majoration des poids optimaux**) : Soit $\mathbf{w}^\star \in \mathcal{S}^{M'+}$ maximisant $\det \mathbf{M}(\mathbf{w}, \mathbf{n}, \tilde{\mathbf{B}})$. Alors $w_m^\star \leq \frac{rang(\mathbf{H}_m(\mathbf{n}, \tilde{\mathbf{B}}))}{n} = \frac{\mathbf{1}_n^\mathsf{T}\tilde{\mathbf{A}}_{.m}}{n} \quad \forall m = 1, \ldots, M'.$* □

Remarquons que lorsque $\sum_{m=1}^{M'} \text{rang}(\mathbf{H}_m(\mathbf{n}, \tilde{\mathbf{B}})) = \mathbf{1}_n^\mathsf{T}\tilde{\mathbf{A}}\mathbf{1}_{M'} = n$, le théorème précédent implique que $w_m^\star = \frac{\text{rang}(\mathbf{H}_m(\mathbf{n}, \tilde{\mathbf{B}}))}{n}$ pour tout m. Du fait de la concavité de la fonction $\mathcal{L}(\cdot, \mathbf{n}, \tilde{\mathbf{B}})$ sur $\mathcal{S}^{M'+}$, de simples algorithmes peuvent être utilisés pour déterminer $\hat{\pi}^{\mathcal{L}}$.

Algorithme multiplicatif

En utilisant l'équivalence avec un problème de plan d'expérience D-optimal, nous proposons d'utiliser un algorithme multiplicatif, voir [STT78, Tor83] et [HT09], qui correspond aux itérations suivantes

$$w_m^{(t+1)} = \frac{d(\mathbf{w}^{(t)}, m)}{n} = \left(\sum_{k=1}^{K} \frac{n_k}{n} \frac{\tilde{\mathbf{B}}_{km}}{\tilde{\mathbf{B}}_{k.}\mathbf{w}^{(t)}} \right) w_m^{(t)}, \qquad m = 1, \ldots, M', \tag{2.16}$$

avec $\mathbf{w}^{(0)} \in \mathcal{S}^{M'+}$ et $w_m^{(0)} \neq 0, m = 1, \ldots, M'$. Nous remarquons que cet algorithme est équivalent à l'algorithme EM qui a été proposé initialement par Dempster et *al*. [DLR77] et appliqué par Gentleman et Vandal [GV01] au cas des données censurées.

Algorithme de gradient contraint ("Vertex Direction Method : VDM")

L'algorithme de gradient contraint (Algorithme 3) a été proposé par [Fed72] et est basé sur une propriété vérifiée par les dérivées directionnelles $F(\mathbf{w}, \mathbf{e}_m)$. Si $F(\mathbf{w}, \mathbf{e}_m) > 0$ pour un $m = 1, \ldots, M'$, alors la fonction de la log-vraisemblance croît en augmentant la masse de probabilité associée à la composante m. Il existe alors un réel positif α tel que $\mathcal{L}((1-\alpha)\mathbf{w} + \alpha\mathbf{e}_m, \mathbf{n}, \tilde{\mathbf{B}}) > \mathcal{L}(\mathbf{w}, \mathbf{n}, \tilde{\mathbf{B}})$, et on peut choisir α de façon à ce que le gain $\mathcal{L}((1-\alpha)\mathbf{w} + \alpha\mathbf{e}_m, \mathbf{n}, \tilde{\mathbf{B}}) - \mathcal{L}(\mathbf{w}, \mathbf{n}, \tilde{\mathbf{B}})$ soit le plus grand possible à chaque itération. Une approximation du α optimal est donnée par $\alpha_{\max}$ (voir [Böh95]) :

$$\alpha_{\max} \stackrel{\Delta}{=} -\frac{F(\mathbf{w}, \mathbf{e}_m)}{F^{(2)}(\mathbf{w}, \mathbf{e}_m)},$$

tout en s'assurant que $(1-\alpha_{\max})\mathbf{w} + \alpha_{\max}\mathbf{e}_m \in \mathcal{S}^{M'+}$, avec

$$\begin{aligned} F^{(2)}(\mathbf{w}, \mathbf{e}_m) &\stackrel{\Delta}{=} \frac{\partial^2}{(\partial\alpha)^2} \mathcal{L}((1-\alpha)\mathbf{w} + \alpha\mathbf{e}_m, \tilde{\mathbf{B}})|_{\alpha=0} \\ &= -\sum_{k=1}^{K} \frac{n_k}{n} \left(\frac{\tilde{\mathbf{B}}_{km}}{\tilde{\mathbf{B}}_{k.}\mathbf{w}} - 1 \right)^2. \end{aligned}$$

Algorithme 3 : Optimisation de $\mathcal{L}(\cdot, \mathbf{n}, \tilde{\mathbf{B}})$ par gradient contraint

donnée : $\mathbf{n}, \tilde{\mathbf{B}}, \epsilon \ll 1$

sortie : $\hat{\mathbf{w}}^{\mathcal{L}}$

On se fixe un vecteur de départ $\mathbf{w}^{(0)} \in \mathcal{S}^{M'+}$.

Tant que $\max_{m=1,\ldots,M'} F(\mathbf{w}, \mathbf{e}_m) > \epsilon$.

  $m^\star = \operatorname{argmax}_{m=1,\ldots,M'} F(\mathbf{w}, \mathbf{e}_m)$

  $\mathbf{w} = (1-\alpha_{\max})\mathbf{w} + \alpha_{\max}\mathbf{e}_{m^\star}$

fin tant que

Algorithme d'échange de sommets ("Vertex Exchange Method : VEM")

L'algorithme d'échange de sommets (Algorithme 4) proposé par Böhning [Böh95] est basé sur l'idée suivante. Si nous notons

$$m^\star \stackrel{\Delta}{=} \underset{m=1,\ldots,M'}{\operatorname{argmax}} F(\mathbf{w}, \mathbf{e}_m) \text{ et } m_\star \stackrel{\Delta}{=} \underset{m=1,\ldots,M';w_m\neq 0}{\operatorname{argmin}} F(\mathbf{w}, \mathbf{e}_m),$$

il faut alors placer plus de masse de probabilité pour la composante $m^\star$ et moins pour $m_\star$. Explicitement, à chaque itération nous remplaçons $\mathbf{w}$ par $\mathbf{w} + \alpha w_{m_\star}(\mathbf{e}_{m^\star} - \mathbf{e}_{m_\star})$. Là aussi le choix de α à chaque itération est effectué de façon à maximiser le gain $\mathcal{L}(\mathbf{w} + \alpha w_{m_\star}(\mathbf{e}_{m^\star} - \mathbf{e}_{m_\star}), \mathbf{n}, \tilde{\mathbf{B}}) - \mathcal{L}(\mathbf{w}, \mathbf{n}, \tilde{\mathbf{B}})$. Böhning [Böh95] propose $\alpha_{\max}$ comme approximation du α optimal maximisant le gain :

$$\alpha_{\max} \stackrel{\Delta}{=} -\frac{F(\mathbf{w}, \mathbf{e}_{m^\star}) - F(\mathbf{w}, \mathbf{e}_{m_\star})}{w_{m_\star}\tilde{F}^{(2)}(\mathbf{w}, \mathbf{e}_{m^\star}, \mathbf{e}_{m_\star})},$$

avec

$$\begin{aligned}\tilde{F}^{(2)}(\mathbf{w}, \mathbf{e}_m, \mathbf{e}_{m'}) &\stackrel{\Delta}{=} \frac{\partial^2}{(\partial\alpha)^2}\mathcal{L}(\mathbf{w} + \alpha(\mathbf{e}_m - \mathbf{e}_{m'}), \mathbf{w}, \tilde{\mathbf{B}})|_{\alpha=0} \\ &= -\sum_{k=1}^{K}\frac{n_k}{n}\left(\frac{\tilde{\mathbf{B}}_{km} - \tilde{\mathbf{B}}_{km'}}{\tilde{\mathbf{B}}_{k.}\mathbf{w}}\right)^2,\end{aligned}$$

tout en s'assurant que $\mathbf{w} + \alpha_{\max} w_{m_\star}(\mathbf{e}_{m^\star} - \mathbf{e}_{m_\star}) \in \mathcal{S}^{M'+}$. Si ce n'est pas le cas, la masse $w_{m_\star}$ devient nulle.

Algorithme 4 : Optimisation de $\mathcal{L}(\cdot, \mathbf{n}, \tilde{\mathbf{B}})$ par échange de sommets

donnée : $\mathbf{n}, \tilde{\mathbf{B}}, \epsilon \ll 1$

sortie : $\hat{\mathbf{w}}^{\mathcal{L}}$

1 : On se fixe un vecteur de départ $\mathbf{w}^{(0)} \in \mathcal{S}^{M'+}$.

2 : Tant que $\max_{m=1,\ldots,M'} F(\mathbf{w}, \mathbf{e}_m) > \epsilon$.

3 : $\quad m^\star = \underset{m=1,\ldots,M'}{\operatorname{argmax}} F(\mathbf{w}, \mathbf{e}_m)$ et $m_\star = \underset{m=1,\ldots,M';w_m>0}{\operatorname{argmin}} F(\mathbf{w}, \mathbf{e}_m)$

4 : $\quad \mathbf{w} = \mathbf{w} + \alpha_{\max} w_{m_\star}(\mathbf{e}_{m^\star} - \mathbf{e}_{m_\star})$

5 : fin tant que

Harman et Pronzato [HP07] remarquent qu'il existe aussi une propriété concernant le support de $\mathbf{w}^\star$ (c'est à dire les indices m tels que $w^\star_m > 0$) permettant de réduire progressivement la dimension du problème.

Théorème 6. *[HP07](**Support d'un plan optimal**) : Pour tout $\mathbf{w} \in \mathcal{S}^{M'+}$ tel que $\det \mathbf{M}(\mathbf{w}, \mathbf{n}, \tilde{\mathbf{B}}) > 0$, pour tout $m \in \{1, \ldots, M'\}$ tel que $d(\mathbf{w}, m) < n\left[1 + \frac{\epsilon}{2} - \frac{\sqrt{\epsilon(4+\epsilon-4/n)}}{2}\right]$, avec $\epsilon = \max_{m'=1,\ldots,M'} d(\mathbf{w}, m') - n$, on a alors $w^\star_m = 0$.* □

Le Théorème 6 nous permet de proposer pour les algorithmes d'optimisation présentés auparavant, des variantes (notées par $^+$) consistant à éliminer à chaque itération des composantes susceptibles de ne pas appartenir au support de l'optimum.

2.3.2.2 Comparaison des performances des différents algorithmes d'optimisation

Dans ce paragraphe nous comparons les performances des trois algorithmes d'optimisation déjà présentés. Nous comparons ces algorithmes en rajoutant à toutes les itérations la vérification permettant de réduire le support de l'optimum cité dans le Théorème 6 (les algorithmes EM$^+$, VDM$^+$ et VEM$^+$) et en la rajoutant toutes les x itérations (les algorithmes EM^{x+}, VDM^{x+} et VEM^{x+}). Nous présentons une première comparaison des différents algorithmes à partir de leurs vitesses de convergence pour l'exemple 4, page 23 avec $n = 30$ et $M = 7$.

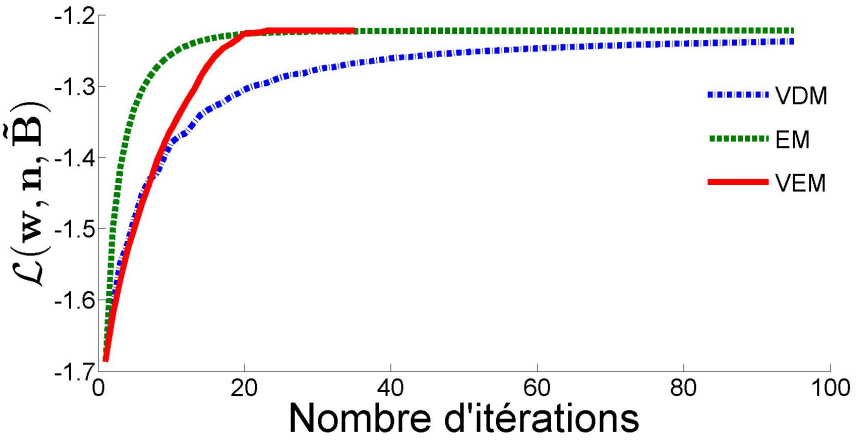


FIGURE 2.8 – *Évolution de la log-vraisemblance en fonction du nombre d'itérations des différents algorithmes d'optimisation appliqués à l'exemple 6.*

La Figure 2.8 montre l'évolution de $\mathcal{L}(\mathbf{w}, \mathbf{n}, \tilde{\mathbf{B}})$ en fonction du nombre d'itérations dans les 3 algorithmes d'optimisation déjà présentés, pour $\epsilon = 10^{-4}$. L'algorithme de gradient contraint (VDM) converge (c'est à dire $\max_{m=1,\ldots,M'} F(\mathbf{w}, \mathbf{e}_m) \leqslant \epsilon$) au bout de 21349 itérations, tandis que l'algorithme multiplicatif (EM) converge en 95 itérations et l'algorithme d'échange de sommets (VEM) en seulement 35 itérations.

Nous jugeons par la suite les performances de chaque algorithme en nous basant sur le nombre d'itérations et le temps de calcul nécessaires pour le calcul de $\hat{\mathbf{w}}^{\mathcal{L}}$ à partir d'une collection d'exemples simulés. Nous simulons aléatoirement 1000 matrices $\mathbf{B}$ de taille 30×700[9]. Cette taille des matrices $\mathbf{B}$ est du même ordre de grandeur que la matrice obtenue lors de l'application aux données réelles de grades de plongées (Chapitre 5). La simulation consiste à placer aléatoirement les valeurs 0 et 1 des matrices $\mathbf{B}$ tout en vérifiant que nous obtenons des matrices où lignes et colonnes ne se répètent pas.

9. Même si nous avons défini les algorithmes d'optimisation pour notre problème après réduction du support, c'est à dire pour optimiser $\mathcal{L}(\mathbf{w}, \mathbf{n}, \tilde{\mathbf{B}})$, ces algorithmes peuvent s'appliquer directement pour l'optimisation de $\mathcal{L}(\mathbf{w}, \mathbf{n}, \mathbf{B})$.

Nous simulons aussi des fréquences d'observation $\{\frac{n_1}{n}, \ldots, \frac{n_{30}}{n}\}$ (les mêmes fréquences pour toutes les matrices $\mathbf{B}$). La table 2.1 compare les moyennes calculées sur des échantillons de taille 1000, des temps de calcul (en secondes) et des nombres d'itérations des différents algorithmes d'optimisation, et ce pour plusieurs valeurs de ϵ.

	$\epsilon = 10^{-3}$		$\epsilon = 10^{-6}$		$\epsilon = 10^{-12}$	
Algorithmes	CPU	Itérations	CPU	Itérations	CPU	Itérations
EM	0.174	202.2	2.945	2411.4	7.846	5273.4
VDM	1.193	1198.3				
VEM	0.709	696.6	0.820	768.4	1.225	1247.7
EM^{+}	0.285	198.3	3.642	2409.8	5.518	4178.1
VDM^{+}	2.049	1337.2				
VEM^{+}	0.977	705.9	1.113	767.6	1.322	1074.5
EM^{10+}	0.162	200.1	2.864	2410.1	3.428	4165.7
VDM^{10+}	1.459	1289.3				
VEM^{10+}	0.769	701.1	0.816	768.3	1.044	1182.2

TABLE 2.1 – *Comparaison des moyennes des temps de calcul et des nombres d'itérations des différents algorithmes d'optimisation sur une collection simulée d'exemples.*

Dans la table 2.1, les résultats concernant l'algorithme de gradient contraint (VDM) et ses variantes (VDM^{+}, VDM^{10+}) ne sont donnés que pour la précision $\epsilon = 10^{-3}$ car pour les deux autres précisions testées, la convergence est beaucoup plus lente que pour les méthodes EM et VEM. L'algorithme de gradient contraint (VDM, VDM^{+} et VDM^{10+}) est nettement le moins performant.

Naturellement, plus le coefficient de précision ϵ diminue, plus les moyennes des temps de calcul et des nombres d'itérations augmentent. Pour les deux méthodes EM et VEM, le fait de rajouter la vérification du support (Théorème 6) à toutes les itérations (EM^{+} et VEM^{+}) permet de réduire le nombre d'itérations mais la moyenne de temps de calcul ne diminue pas forcément (sauf pour EM avec la précision 10^{-12}). Rajouter cette vérification toutes les 10 itérations (EM^{10+} et VEM^{10+}) certes donne les moyennes de temps de calcul les plus basses pour les 3 valeurs testées de précision, mais ces moyennes ne sont pas très inférieures à celles des algorithmes EM et VEM.

Pour $\epsilon = 10^{-3}$ l'algorithme le plus rapide est l'algorithme multiplicatif EM^{10+}. Pour les précisions $\epsilon = 10^{-6}$ et $\epsilon = 10^{-12}$, l'algorithme d'échange de sommets VEM^{10+} est le plus rapide. D'après cette étude nous choisissons de garder l'algorithme d'échange de sommets VEM^{10+} pour sa rapidité même lorsqu'une grande précision est requise.

Dans le paragraphe suivant nous testons s'il est plus judicieux de réduire la taille du problème en appliquant l'Algorithme 2 (page 30) ou d'appliquer directement l'algorithme d'optimisation VEM^{10+}.

2.3.2.3 Réduction du support versus optimisation

Nous avons vu dans la section 2.3.1 que nous pouvons réduire le support de l'ENPMV en utilisant des outils de la théorie des graphes grâce à l'Algorithme 2. Dans un même temps, les algorithmes

d'optimisation permettent également de retrouver le support de l'ENPMV en attribuant la masse zéro aux régions en dehors du support. Dans ce paragraphe, nous nous intéressons à la comparaison de ces deux approches. Plus précisément, nous voulons répondre à la question suivante : vaut-il mieux calculer la matrice $\tilde{\mathbf{B}}$ de taille $(K \times M')$ avant l'optimisation ou bien appliquer l'optimisation directement au problème défini par la matrice $\mathbf{B}$? Nous menons cette comparaison sur la même collection de matrices $\mathbf{B}$ utilisée pour comparer les algorithmes d'optimisation, en utilisant l'algorithme d'échange de sommets VEM^{10+} avec une précision $\epsilon = 10^{-6}$.

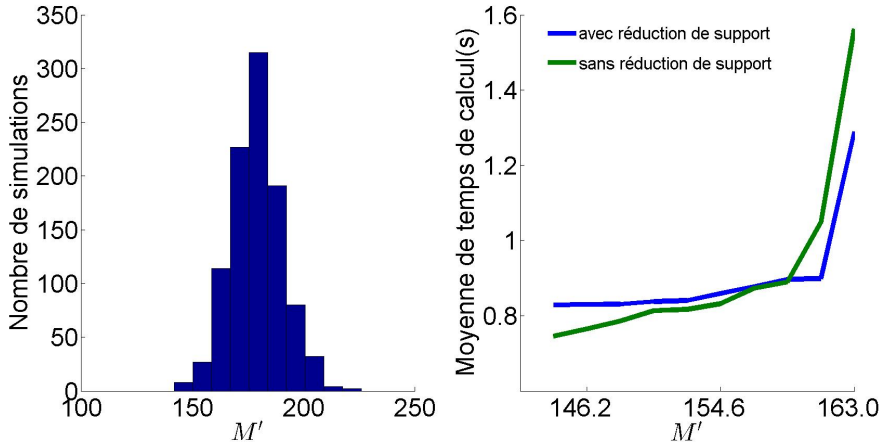


FIGURE 2.9 – *Histogramme classant les simulations par taille de M' (nombre des colonnes de la matrice $\tilde{\mathbf{B}}$) et comparaison du temps de calcul (en secondes) de l'optimisation avec et sans réduction du support.*

La Figure 2.9 montre que plus M', le nombre de colonne de la matrice $\tilde{\mathbf{B}}$ est grand, plus il est intéressant d'un point de vue computationnel de passer par l'étape de réduction de support. En effet, une valeur petite de M' implique que les matrices $\mathbf{B}$ et $\tilde{\mathbf{B}}$ ne sont pas très différentes l'une de l'autre. Donc les temps de calcul de l'optimisation en utilisant $\mathbf{B}$ ou $\tilde{\mathbf{B}}$ seront proches. Dans ce cas il n'est pas intéressant de calculer $\tilde{\mathbf{B}}$ à partir de $\mathbf{B}$. Par contre il est intéressant de passer par la réduction du support quand elle permet de diminuer considérablement la taille du problème (parfois la réduction du support permet d'éliminer plus que 80% des régions élémentaires). En moyenne, il est plus avantageux de ne pas faire appel à l'algorithme de réduction de support qui dure en moyenne $0.27s$. La moyenne de temps de calcul sur tout l'échantillon sans réduction de support est de $0.86s$ et la même moyenne avec réduction de support est de $0.89s$.

2.4 CARACTÉRISATION DE LA PERFORMANCE DE L'ENPMV

Avant de présenter les résultats de l'estimation non-paramétrique par maximum de vraisemblance sur les données réelles disponibles (Chapitre 5), nous nous intéressons dans cette section au calcul de $\hat{\pi}^{\mathcal{L}}_{\star}$ à partir de données simulées selon une loi π_θ connue $y_i \overset{iid}{\sim} \pi_\theta, i = 1, \ldots, n$. afin d'illustrer le comportement de l'ENPMV sur des jeux de données simulées, mettant en évidence ses limitations pour le problème motivant cette thèse. Nous allons alors pouvoir comparer $\hat{\pi}^{\mathcal{L}}_{\star}$ avec la

"vraie" loi π_θ.

Le domaine paramétrique considéré est rectangulaire, $\Theta = \Theta^1 \times \Theta^2 = \left[\theta^1_{\min}, \theta^1_{\max}\right] \times \left[\theta^2_{\min}, \theta^2_{\max}\right] \subset \mathbb{R}^2$, où nous avons implicitement défini les intervalles $\Theta^i, i = 1, 2$. Nous considérons sans perdre en généralité que $\Theta = [0, 1]^2$. La loi de simulation π_θ utilisée est une loi gaussienne tronquée au compact Θ

$$\pi_\theta = \mathcal{N}(\mu_\theta, \Sigma_\theta)$$

où $\mu_\theta = \left(\frac{\theta^1_{\min}+\theta^1_{\max}}{2}, \frac{\theta^2_{\min}+\theta^2_{\max}}{2}\right)^\mathsf{T} = (0.5, 0.5)^\mathsf{T}$ (le milieu du domaine Θ) et $\Sigma_\theta = 0.2^2 I_2$ avec I_2 la matrice identité de taille 2 (voir Figure 2.11a).

Nous commençons par considérer une géométrie simple des régions de censure (paragraphe 2.4.1) donnant lieu à une partition $\mathcal{Q}^J$ régulière de Θ avant de générer dans le paragraphe 2.4.2 des régions de censure à partir d'un diagramme de Voronoi où les germes (centres des cellules) sont générés aléatoirement dans le domaine Θ.

2.4.1 Génération de régions de censure à géométrie simple

Soient J et L dans $\mathbb{N}\backslash\{0\}$. Nous posons pour $i = 1, 2$, pour $j = 1, \ldots, J$ et pour $\ell = 1, \ldots, L-1$

$$\Delta^i = \frac{\theta^i_{\max} - \theta^i_{\min}}{J(L-1)+1}, \quad \text{et} \quad \theta^i_{j,\ell} = \theta^i_{\min} + (j + \ell - 1)\Delta^i.$$

Soient $\{\mathcal{I}^i_j\}_{j=1}^J$, J partitions de $\Theta^i, i = 1, 2$ respectivement, telles que

$$\mathcal{I}^i_j = \left\{ \left[\theta^i_{j,0}, \theta^i_{j,1}\right[, \ldots, \left[\theta^i_{j,L-1}, \theta^i_{j,L}\right] \right\},$$

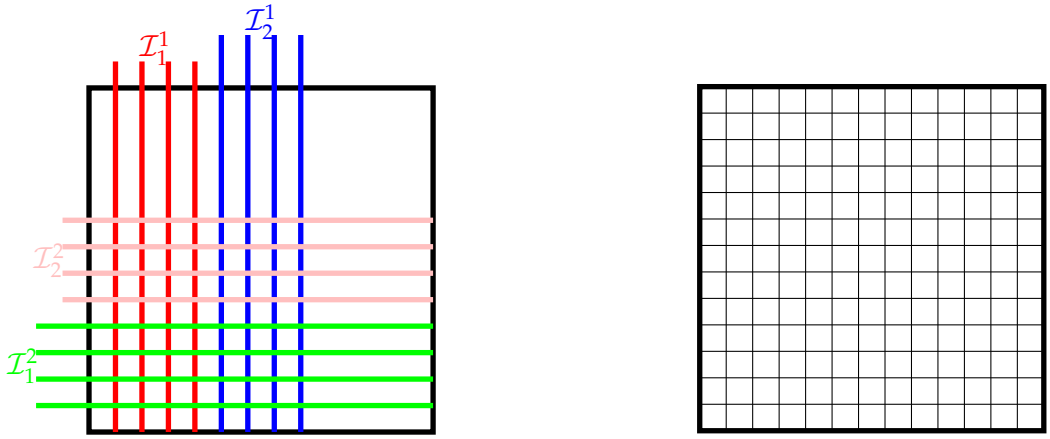
avec $\theta^i_{j,0} = \theta^i_{\min}$ et $\theta^i_{j,L} = \theta^i_{\max}$ pour $j = 1, \ldots, J$. Nous avons dans cet exemple $J = J_1 + J_2$ partitions "régulières" du domaine paramétrique Θ de la forme

$$\mathcal{R}^j = \mathcal{I}^1_j \times \Theta^2 \text{ pour } j = 1, \ldots, J_1, \quad \text{et} \quad \mathcal{R}^j = \Theta^1 \times \mathcal{I}^2_{j-J_1} \text{ pour } j = J_1 + 1, \ldots, J_1 + J_2.$$

La Figure 2.10a montre 4 exemples de ces partitions, où chacune est formée de 5 régions ($L = 5$). Dans la Figure 2.10b, nous montrons la partition $\mathcal{Q}^J$ (Définition 2), obtenue pour $J_1 = J_2 = 3$.

Pour le jeu de données simulées, nous avons choisi $J = 6$ ($J_1 = J_2 = 3$) et $L = 5$, et donc le nombre total de régions de censure distinctes est $J \times L = 30$. La partition $\mathcal{Q}^J$ est dans ce cas un partitionnement régulier du domaine Θ formée de 169 régions carrées correspondant aux régions élémentaires (Figure 2.10b). Nous remarquons que toute région élémentaire de $\mathcal{Q}^J$ est définie par l'intersection de 6 régions de censure (une région de chacune des 6 partitions $\mathcal{R}^j$). Dans notre étude, chacune des 6 différentes partitions est observée le même nombre de fois $n^1 = \cdots = n^6 = n/6$.

La Figure 2.11a montre π_θ sur $\Theta = [0, 1]^2$ et la Figure 2.11b montre le vecteur de probabilité $p_{\pi_\theta, \mathcal{Q}^J}$ associé aux régions élémentaires de $\mathcal{Q}^J$ (Définition 1).



(a) *Exemple de 4 partitions de censure.* **(b)** *La partition $\mathcal{Q}^J$ résultante.*

FIGURE 2.10 – *Les 4 partitions $\mathcal{I}_1^1, \mathcal{I}_2^1, \mathcal{I}_1^2$ et $\mathcal{I}_2^2$ de Θ (à gauche) et la partition $\mathcal{Q}^J$ qui résulte de $\left\{\mathcal{I}_j^i\right\}_{i=1,j=1}^{2,3}$ (à droite).*

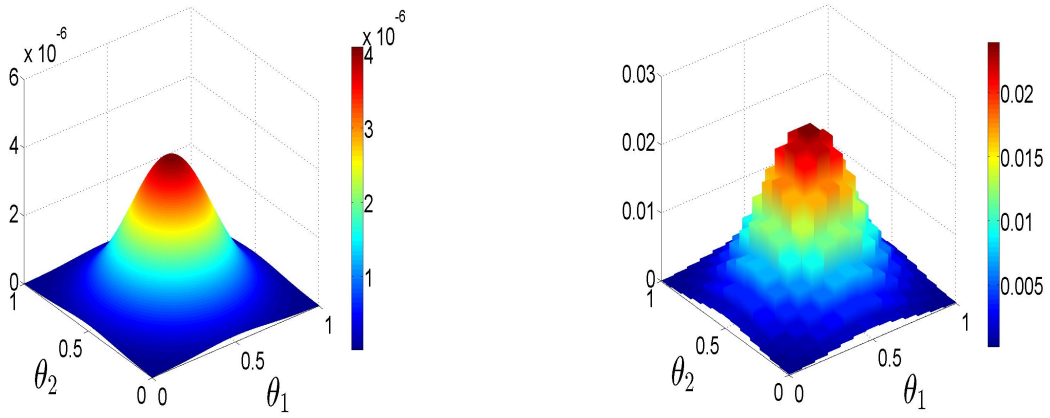
Nous admettons dans un premier temps que la fréquence d'observation de chaque région de censure $\mathcal{R}_\ell^j$ est égale à sa probabilité induite par π_θ

$$n_\ell^j = n^j p_{\pi_\theta, \mathcal{R}_\ell^j} = \frac{n}{6} p_{\pi_\theta, \mathcal{R}_\ell^j}.$$

Cela correspond à supposer que $n^j \to \infty$ pour tout $j = 1, \ldots, J$. Compte tenu du choix de π_θ effectué, toutes les régions de censure $\{\mathcal{R}_\ell^j, j = 1, \ldots, 6, \ell = 1, \ldots, 5\}$ sont observées. D'après le Lemme 7, page 26, toutes les régions élémentaires, dont le nombre s'élève à $13^2 = 169$ correspondent à des cliques maximales avec des représentations réelles non-vides, et peuvent donc appartenir au support de l'ENPMV.

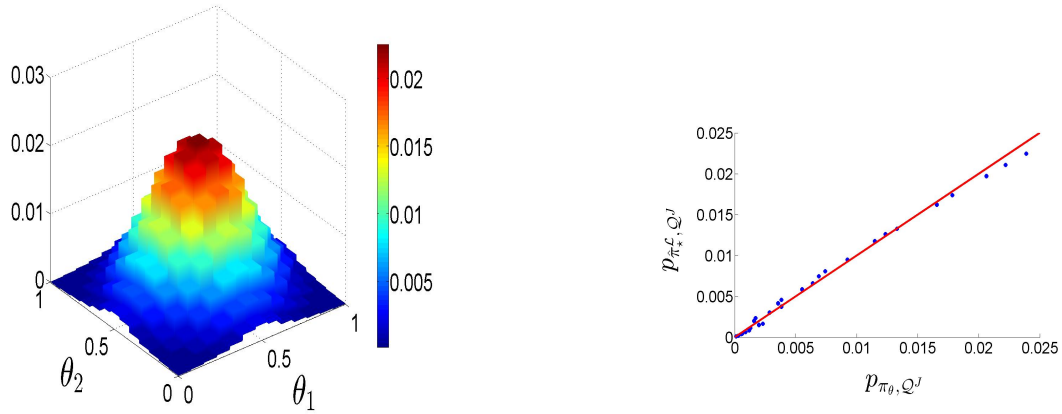
La matrice $\mathbf{B}$ est de taille 30×169 et de rang 25. L'optimisation qui détermine l'ENPMV pour ce problème simple est par conséquent effectuée dans le simplexe $\mathcal{S}^{169+}$. L'estimateur $\hat{\pi}_\star^{\mathcal{L}}$ est illustré dans la Figure 2.12a. Nous remarquons que L'ENPMV ressemble bien à la loi induite par π_θ sur la partition $\mathcal{Q}^J$ (Figures 2.11b et 2.12a). Cela est confirmé par la Figure 2.12b, qui représente les paires $\left(p_{\pi_\theta, \mathcal{Q}^J}, p_{\hat{\pi}_\star^{\mathcal{L}}, \mathcal{Q}^J}\right)$. Le trait en rouge correspond à la droite d'équation $y = x$. Les différences observées entre cette droite et les points $\left(p_{\pi_\theta, \mathcal{Q}^J}, p_{\hat{\pi}_\star^{\mathcal{L}}, \mathcal{Q}^J}\right)$ sont une conséquence directe de la non-unicité de mélange (rang ligne incomplet de $\mathbf{B}$). En effet, les deux lois $p_{\pi_\theta, \mathcal{Q}^J}$ et $p_{\hat{\pi}_\star^{\mathcal{L}}, \mathcal{Q}^J}$ ont la même vraisemblance mais elle n'ont pas la même entropie de Rényi d'ordre 2.

Nous supposons maintenant que l'échantillon de données est de taille finie $n = 6000$ et que chaque partition $\mathcal{R}^j$ est observée $n^j = \frac{n}{6}$ fois. Le résultat de l'estimation est montré dans les Figures 2.13a et 2.13b. Nous remarquons que la qualité de l'estimation dans le cas de n fini s'est beaucoup dégradée par rapport au cas d'un échantillon de taille infinie.



(a) π_θ. **(b)** $p_{\pi_\theta,\mathcal{Q}^J}$.

FIGURE 2.11 – *Loi π_θ (à gauche) et vecteur de probabilité correspondant associé aux régions élémentaires de $\mathcal{Q}^J$ (à droite).*



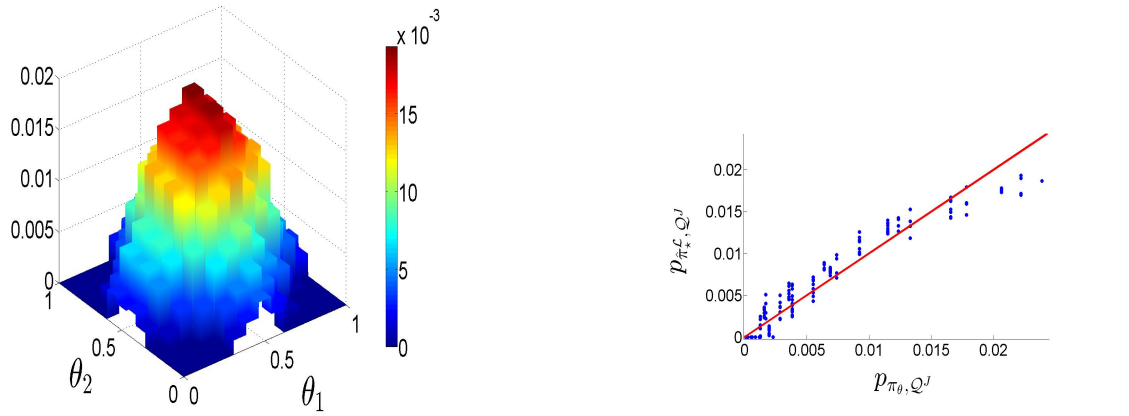
(a) $p_{\hat{\pi}^{\mathcal{L}}_\star,\mathcal{Q}^J}$. **(b)** *Les paires* $\left(p_{\pi_\theta,\mathcal{Q}^J}, p_{\hat{\pi}^{\mathcal{L}}_\star,\mathcal{Q}^J}\right)$.

FIGURE 2.12 – *Loi de probabilité $p_{\hat{\pi}^{\mathcal{L}}_\star,\mathcal{Q}^J}$ induite par $\hat{\pi}^{\mathcal{L}}_\star$ sur $\mathcal{Q}^J$ (à gauche) et paires $\left(p_{\pi_\theta,\mathcal{Q}^J}, p_{\hat{\pi}^{\mathcal{L}}_\star,\mathcal{Q}^J}\right)$ (à droite) avec $n^j = \infty$.*

Nous faisons à présent augmenter J, le nombre de partitions. Cela conduit à des régions élémentaires de plus en plus petites. Nous gardons l'hypothèse selon laquelle nous observons pour chaque expérience, c'est à dire pour chaque J différent, un échantillon de taille infinie. Cela implique que la fréquence d'observation de chaque région de censure $\mathcal{R}^j_\ell$ est égale à

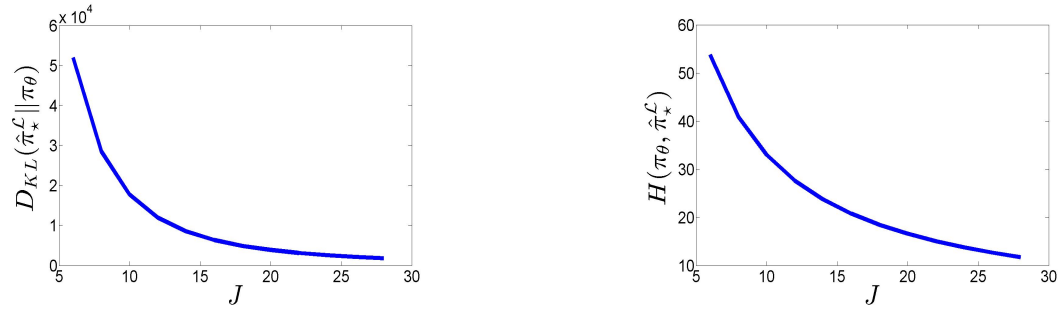
$$n^j_\ell = n^j p_{\pi_\theta,\mathcal{R}^j_\ell} = \frac{n}{J} p_{\pi_\theta,\mathcal{R}^j_\ell}.$$

Nous évaluons la qualité de l'estimateur en calculant pour chaque valeur de J la divergence de Kullback-Leibler et la distance de Hellinger entre π_θ et $\hat{\pi}^{\mathcal{L}}_\star$. Les Figures 2.14a et 2.14b montrent le comportement de la divergence de Kullback-Leibler et de la distance de Hellinger, calculées par inté-



(a) $p_{\hat{\pi}^{\mathcal{L}}_\star,\mathcal{Q}^J}$. **(b)** *Les paires* $\left(p_{\pi_\theta,\mathcal{Q}^J}, p_{\hat{\pi}^{\mathcal{L}}_\star,\mathcal{Q}^J}\right)$.

FIGURE 2.13 – *Loi de probabilité* $p_{\hat{\pi}^{\mathcal{L}}_\star,\mathcal{Q}^J}$ *induite par* $\hat{\pi}^{\mathcal{L}}_\star$ *sur* $\mathcal{Q}^J$ *(à gauche) et paires* $(p_{\pi_\theta,\mathcal{Q}^J}, p_{\hat{\pi}^{\mathcal{L}}_\star,\mathcal{Q}^J})$ *(à droite) avec* $n^j = 1000$.



(a) *Divergence de Kullback-Leibler.* **(b)** *Distance de Hellinger.*

FIGURE 2.14 – *Divergence de Kullback-Leibler (à gauche) et distance de Hellinger (à droite) entre* π_θ *et* $\hat{\pi}^{\mathcal{L}}_\star$ *en fonction de* J.

gration numérique sur chaque région élémentaire, quand le nombre des partitions augmente. Nous remarquons que ces deux critères baissent en augmentant le nombre des partitions, attestant d'une augmentation de la qualité de l'estimation. Cependant, cette amélioration de la qualité de l'estimation est due à la géométrie simple que nous avons considérée dans cet exemple et à l'hypothèse d'observation d'échantillons de taille infinie. Dans le paragraphe suivant nous considérons un mécanisme de génération de partitions à géométrie plus complexe et des échantillons de taille finie.

2.4.2 Régions de censure générées à partir d'un diagramme de Voronoi

Dans ce paragraphe nous considérons des partitions $\left\{\mathcal{R}^j\right\}_{j=1}^J$ générées aléatoirement, et qui sont des unions aléatoires de cellules voisines dans un diagramme de Voronoi.

Soit $\{\theta_s\}_{s=1}^S$ un ensemble de points dans $\mathbb{R}^2$, appelés *germes*. Le diagramme de Voronoi [Aur91] associé aux S germes est une partition de $\mathbb{R}^2$ de taille S où chaque élément de la partition, appelé

cellule, contient les points de $\mathbb{R}^2$ les plus proches du germe correspondant.

Nous avons considéré un diagramme de Voronoi avec 50 germes uniformément distribués dans le domaine Θ. La Figure 2.15 montre l'exemple de deux partitions de taille 2 générées aléatoirement par ce mécanisme sur $\Theta = [0,1]^2$.

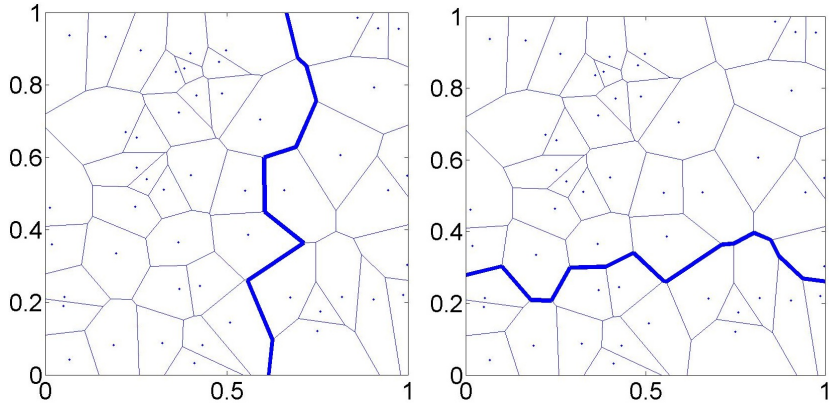
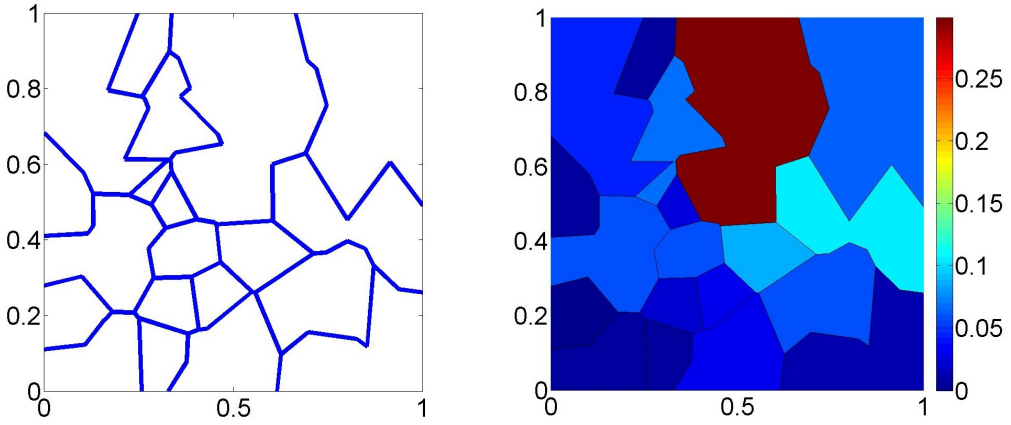


FIGURE 2.15 – *Deux partitions générées aléatoirement dans le domaine* $\Theta = [0,1]^2$.

La Figure 2.16a montre la partition $\mathcal{Q}^J$ obtenue à partir de $J = 10$ partitions de Θ où chacune est de taille $L = 2$. La Figure 2.16b représente la loi de probabilité $p_{\pi_\theta,\mathcal{Q}^J}$ induite par π_θ où les régions élémentaires sont aisément reconnaissables (les polygones délimités par des traits noirs). La taille des partitions générées par ce mécanisme suit approximativement une loi de gamma avec les deux paramètres (de forme et d'intensité) égaux à $(7/2)\lambda^{-2}$, où λ est l'intensité du processus de Poisson [FN07] ($\lambda = 1/50$ dans notre cas).

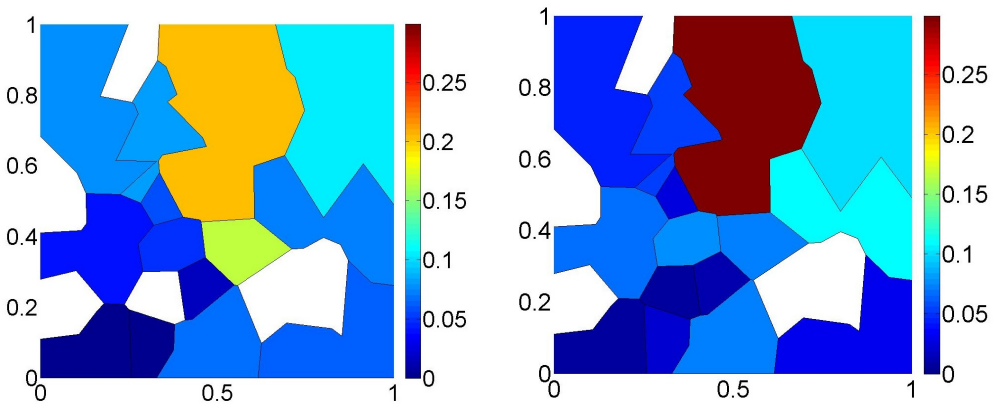


(a) $\mathcal{Q}^J$ **(b)** $p_{\pi_\theta,\mathcal{Q}^J}$

FIGURE 2.16 – *Partition* $\mathcal{Q}^J$ *déterminée par* $J = 10$ *partitions aléatoires of* Θ *de taille* 2 *(à gauche) et loi de probabilité* $p_{\pi_\theta,\mathcal{Q}^J}$ *induite sur* $\mathcal{Q}^J$ *(à droite).*

Des échantillons i.i.d. de taille n^j sont générés selon les lois de probabilité $p_{\pi_\theta,\mathcal{R}^j}$ induites par π_θ

sur les J partitions $\{\mathcal{R}^j\}_{j=1}^{J=10}$. Dans le paragraphe 2.4.1, toutes les partitions avaient une probabilité égale d'être choisies. Dans cet exemple nous supposons que les partitions sont subdivisées en deux groupes. La probabilité de choisir une partition dans le premier groupe est égale à 10^{-3}, et à l'intérieur de chaque groupe les partitions sont choisies uniformément. L'échantillon ainsi simulé ne contient pas tous les éléments de toutes les partitions et par conséquent les hypothèses du Lemme 7 ne tiennent plus. Ce choix favorise le fait que le support de L'ENPMV ne recouvre pas tout le domaine paramétrique. Nous simulons $n = 10^4$ observations.



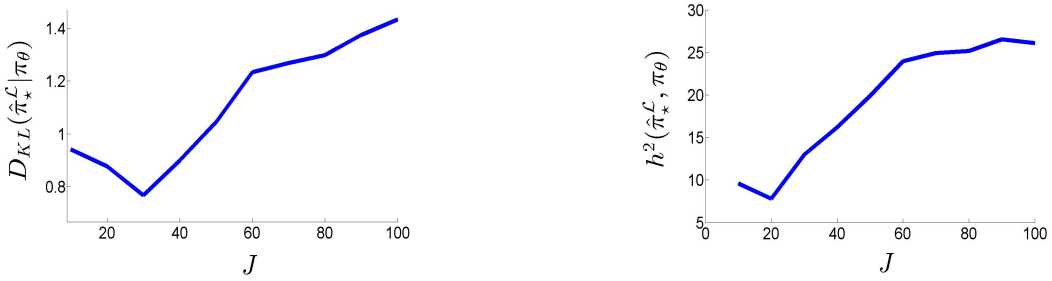
(a) *Une loi maximisant $\mathcal{L}$.* **(b)** $\hat{\pi}^{\mathcal{L}}_{\star}$.

FIGURE 2.17 – *Une loi de probabilité maximisant $\mathcal{L}$ (à gauche) et l'ENPMV à maximum d'entropie de Rényi (à droite). Les régions en blanc ont une masse de probabilité nulle.*

Nous présentons dans la Figure 2.17a une loi de probabilité maximisant la log-vraisemblance, obtenue en utilisant l'algorithme VEM^{10+} (Algorithme 4, page 35), ceci pour les régions élémentaires de la Figure 2.16a. La Figure 2.17b représente $\hat{\pi}^{\mathcal{L}}_{\star}$, l'ENPMV maximisant l'entropie de Rényi (parmi toutes les solutions optimales au sens du maximum de vraisemblance).

Cet exemple illustre les singularités dont souffre l'ENPMV que nous avons présentées dans ce chapitre. Bien que π_θ soit strictement positive sur tout le domaine Θ, son estimation par maximum de vraisemblance non-paramétrique attribue une masse nulle à de nombreuses régions élémentaires (les régions en blanc dans les Figures 2.17a et 2.17b). Ce fait est intrinsèquement lié au critère de vraisemblance, qui favorise les densités les plus concentrées expliquant les données observées. C'est ce que nous avons mis en évidence dans l'exemple 4 où l'ajout d'une seule observation modifiait considérablement le support de l'ENPMV.

Les Figures 2.18a et 2.18b montrent la variation, en fonction du nombre J des partitions, des moyennes empiriques de la divergence de Kullback-Leibler et de la distance de Hellinger entre la distribution estimé $\hat{\pi}^{\mathcal{L}}_{\star}$ et la vraie distribution π_θ. Là aussi, divergence de Kullback-Leibler et distance de Hellinger sont calculées par intégration numérique sur chaque région élémentaire. Le nombre d'observations augmente en fonction du nombre des partitions $n = 100J$ et J varie de 10 à 100 par



(a) *Divergence de Kullback-Leibler.* **(b)** *Distance de Hellinger.*

FIGURE 2.18 – *Moyennes de la divergence de Kullback-Leibler et de la distance de Hellinger entre* π_θ *et* $\hat{\pi}_\star^{\mathcal{L}}$ *calculées sur* 100 *simulations.*

pas de 10. Pour toute valeur de J, nous avons effectué 100 simulations. Les deux critères considérés montrent un comportement inconsistant de l'ENPMV.

Afin d'éviter le comportement singulier de l'ENPMV mis en évidence par ces résultats, nous devons estimer π_θ en utilisant sur un critère autre que la vraisemblance. En considérant le lien entre notre problème et le problème d'estimation de densité sous contraintes de moments, nous proposons d'estimer π_θ par le principe du maximum d'entropie. Ceci fait l'objet du Chapitre 3.

2.5 CONCLUSION

Ce chapitre est consacré à l'estimation par maximum de vraisemblance non-paramétrique d'une densité de probabilité à partir de données censurées par régions. Le maximum de vraisemblance est largement utilisée en estimation de densités du fait de ses propriétés d'efficacité et de convergence sous certaines conditions de régularité. Nous nous sommes intéressés aux propriétés de l'ENPMV ainsi qu'aux méthodes de construction. Nous avons donné des algorithmes permettant de calculer le support de l'ENPMV et comparé plusieurs méthodes d'optimisation de la vraisemblance.

Les études présentées dans ce chapitre permettent d'affirmer que la censure des observations induit en général plusieurs formes de non-unicités de cet estimateur, ainsi que la concentration de la masse de probabilité de l'ENPMV sur un sous-ensemble du support de π_θ, déterminé par le graphe d'intersection des régions observés. Nous avons montré à l'aide de simulations les problèmes dont souffre cet estimateur, confirmant qu'il ne conduit pas dans les conditions de notre étude, à une estimation consistante de π_θ. Le chapitre suivant propose un nouvel estimateur qui fait appel au principe du maximum d'entropie.

ANNEXE DU CHAPITRE 2

2.A LES CONDITIONS DE KUHN-TUCKER POUR L'ENPMV

Nous présentons ici les conditions de Kuhn-Tucker caractérisant un maximum local de la log-vraisemblance $\mathcal{L}(\mathbf{w}, \mathbf{n}, \mathbf{B})$.

La recherche de l'ENPMV consiste à maximiser la fonction $\mathcal{L} : \mathbf{w} \in \mathbb{R}^{M+} \mapsto \mathcal{L}(\mathbf{w}, \mathbf{n}, \mathbf{B})$, continuement dérivable, sous les $(M+1)$ contraintes : $g_m(\mathbf{w}) \leq 0, h(\mathbf{w}) = 0$ avec $g_m(\mathbf{w}) = -w_m$ et $h(\mathbf{w}) = \sum_{m=1}^{M} w_m - 1$, avec M le nombre des colonnes de $\mathbf{B}$. On peut donc caractériser un maximum local $\mathbf{w}^\star$ par les conditions de Kuhn-Tucker :

$$\mathbf{w}^\star \text{ est un maximum local de } \mathcal{L} \Leftrightarrow \exists \mu_1, \ldots, \mu_M \text{ et } \lambda \text{ tels que}$$

$$\begin{cases} -\nabla\mathcal{L}(\mathbf{w}^\star, \mathbf{n}, \mathbf{B}) = \sum_{m=1}^{M} \mu_m \nabla g_m(\mathbf{w}^\star) + \lambda \nabla h(\mathbf{w}^\star) \\ g_m(\mathbf{w}^\star) \leq 0, \qquad \forall m = 1, \ldots, M \\ h(\mathbf{w}^\star) = 0, \\ \mu_m \geq 0, \qquad \forall m = 1, \ldots, M \\ \mu_m g_m(\mathbf{w}^\star) = 0, \qquad \forall m = 1, \ldots, M, \end{cases}$$

avec $\nabla f(\mathbf{w}) = \left(\frac{\partial}{\partial w_1} f(\mathbf{w}), \ldots, \frac{\partial}{\partial w_M} f(\mathbf{w})\right)^\mathsf{T}$ le vecteur des dérivées partielles de la fonction f au point $\mathbf{w}$. Soit $\mathbf{w}^\star = (w_1^\star, \ldots, w_M^\star)^\mathsf{T}$ une solution des conditions de Kuhn-Tucker avec $\mu_1, \ldots, \mu_M, \lambda$ les multiplicateurs de Lagrange associés.

Nous avons alors :

$$\begin{cases} \sum_{m=1}^{M} w_m^\star = 1, \\ w_m^\star \geq 0, \\ \mu_m w_m^\star = 0, \\ \sum_{k=1}^{K} \frac{n_k}{n} \frac{\mathbf{B}_{km}}{\mathbf{B}_{k.}\mathbf{w}} + \mu_m - \lambda = 0, \ \ \forall m \end{cases}$$

En multipliant la dernière équation par $w_m^\star$ et en sommant sur tous les m en obtient :

$$\lambda = \sum_{m=1}^{M} \lambda w_m^\star = \sum_{m=1}^{M} \sum_{k=1}^{K} \frac{n_k}{n} \frac{\mathbf{B}_{km}}{\mathbf{B}_{k.}\mathbf{w}} w_m^\star = \sum_{k=1}^{K} \frac{n_k}{n\mathbf{B}_{k.}\mathbf{w}} \mathbf{B}_{k.}\mathbf{w} = 1.$$

Étant donné $\mu_m w_m^\star = 0$, si $w_m^\star > 0$, alors $\mu_m = 0$ et par conséquent

$$\frac{\partial}{\partial w_m}\mathcal{L}(\mathbf{w}^\star, \mathbf{n}, \mathbf{B}) = \lambda = 1.$$

A partir de ces conditions, Gentleman et Geyer [GG94] ont dérivé une condition faible pour l'unicité de l'ENPMV. Soit $\mathbf{w}^\star$ une solution des conditions de Kuhn-Tucker avec $\mu_1, \ldots, \mu_M, \lambda$ les multiplicateurs de Lagrange associés. On définit l'ensemble

$$\mathcal{W}' = \{\mathbf{w} \in \mathbb{R}^M : w_m = 0 \text{ si } \mu_m > 0; w_m \geq 0 \text{ si } \mu_m = 0; \sum_{m=1}^{M} w_m = 1\}.$$

On a vu précédemment que l'unicité de l'ENPMV au sens du mélange (paragraphe 2.2.1.2), est garantie si la matrice $\mathbf{B}$ est de rang plein. Gentleman et Geyer [GG94] montrent à partir des conditions de Kuhn-Tucker qu'il suffit que la matrice $\mathbf{B}_2$, la sous-matrice constituée à partir de $\mathbf{B}$ en ne gardant que les lignes correspondant aux $\mu_m = 0$, soit de rang plein.

2.B Démonstration du théorème 2 (page : 28)

Dans cet annexe, nous présentons une démonstration du Théorème 2.

Preuve : Pour tout sommet $k \in V$, on note M_k la liste des cliques $C \in E'_{\max}$ contenant k. Supposons qu'il existe un ensemble $\mathcal{A}$ tel que $\mathcal{A} \cap S_{\mathcal{L}}^\star = \emptyset$, avec $\hat{\pi}^{\mathcal{L}}(\mathcal{A}) > 0$. Nous supposons que $\mathcal{A} \subset \cup_{k=1}^{K} \mathcal{R}_k$ (sinon on augmentera la valeur de la vraisemblance en affectant arbitrairement la masse $\hat{\pi}^{\mathcal{L}}(\mathcal{A} \cap \overline{\cup_{k=1}^{K} \mathcal{R}_k})$ à l'une des hyper-arrêtes dans $E'_{\max}$).

Soit π une distribution de probabilité. On a :

$$\pi(\mathcal{R}_k) = \sum_{C \in M_k} \pi(\mathscr{R}(C)) + \pi(\mathcal{A} \cap \mathcal{R}_k).$$

La log-vraisemblance de π est donc :

$$\mathcal{L}(\pi, \mathbf{X}_n) = \sum_{k=1}^{K} \frac{n_k}{n} \log \pi(\mathcal{R}_k) = \sum_{k=1}^{K} \frac{n_k}{n} \log \left[\sum_{C \in M_k} \pi(\mathscr{R}(C)) + \pi(A \cap \mathcal{R}_k) \right]$$

Pour tout sommet k et toute clique $C \in E'_{\max}$, considérons la décomposition suivante :

$$\mathcal{R}_k = \left(\mathcal{R}_k \cap \left(\cup_{i \notin C} \mathcal{R}_i\right)\right) \cup \left(\mathcal{R}_k \cap \left(\overline{\cup_{i \notin C} \mathcal{R}_i}\right)\right) = \mathcal{R}_k^C \cup \overline{\mathcal{R}_k^C}.$$

Il s'agit d'une décomposition de $\mathcal{R}_k$ en deux sous-ensembles disjoints, avec $\overline{\mathcal{R}_k^C}$ le sous-ensemble de $\mathcal{R}_k$ qui intersecte les régions n'appartenant pas à $\mathscr{R}(C)$, et donc si $k \notin C, \mathcal{R}_k \cap \mathcal{R}_k^C = \emptyset$. Puisque ces sous-ensembles sont disjoints, on peut modifier la mesure d'un ensemble sans modifier la mesure

d'un autre. Choisissons une clique C^* et considérons la distribution suivante :

$$\begin{cases} \pi'(\mathscr{R}(C)) & = \pi(\mathscr{R}(C)), C \neq C^*, \\ \pi'(\mathscr{R}(C^*)) & = \pi'(\mathscr{R}(C^*)) + \pi\left(\mathcal{A} \cap (\cup_{l \in C^*} \mathcal{R}_l^{C^*})\right), \\ \pi'\left(\mathcal{A} \cap \mathcal{R}_i^{C^*}\right) & = 0, i \in C^*, \\ \pi'\left(\mathcal{A} \cap \overline{\mathcal{R}_i^{C^*}}\right) & = \pi\left(\mathcal{A} \cap \overline{\mathcal{R}_i^{C^*}}\right), i \in C^* \\ \pi'(\mathcal{A} \cap \mathcal{R}_i) & = \pi(\mathcal{A} \cap \mathcal{R}_i), i \notin C^* \end{cases}$$

Remarquons que $\forall i \in C^*, \pi'(\mathcal{A} \cap \mathcal{R}_i) = \pi(\mathcal{A} \cap \overline{\mathcal{R}_i^{C^*}}) \leqslant \pi(\mathcal{A} \cap \mathcal{R}_i)$. La log-vraisemblance de π est

$$\begin{aligned} \mathcal{L}(\pi, \mathbf{X}_n) &= \sum_{k=1}^{K} \frac{n_k}{n} \log \pi(\mathcal{R}_k) \\ &= \sum_{k \in C^*} \frac{n_k}{n} \log \left[\left(\sum_{C \in M_k} \pi(\mathscr{R}(C)) \right) + \pi(\mathcal{A} \cap \mathcal{R}_k) \right] + \\ &\quad \sum_{k \notin C^*} \frac{n_k}{n} \log \left[\left(\sum_{C \in M_k} \pi(\mathscr{R}(C)) \right) + \pi(\mathcal{A} \cap \mathcal{R}_k) \right] \\ &= \sum_{k \in C^*} \frac{n_k}{n} \log \left[\left(\sum_{C \in M_k, C \neq C^*} \pi(\mathscr{R}(C)) \right) + \pi(\mathscr{R}(C^*)) + \pi(\mathcal{A} \cap \mathcal{R}_k) \right] + \\ &\quad \sum_{k \notin C^*} \frac{n_k}{n} \log \left[\left(\sum_{C \in M_k} \pi'(\mathscr{R}(C)) \right) + \pi'(\mathcal{A} \cap \mathcal{R}_k) \right] \\ &= \sum_{k \in C^*} \frac{n_k}{n} \log \left[\sum_{C \in M_k, C \neq C^*} \pi'(\mathscr{R}(C)) + \pi'(\mathscr{R}(C^*)) - \pi(\mathcal{A} \cap \cup_{l \in C^*} \mathcal{R}_l^{C^*}) + \right. \\ &\quad \left. \pi(\mathcal{A} \cap \mathcal{R}_k) \right] + \sum_{k \notin C^*} \frac{n_k}{n} \log \pi'(\mathcal{R}_k) \\ &= \sum_{k \in C^*} \frac{n_k}{n} \log \left[\sum_{C \in M_k, C \neq C^*} \pi'(\mathscr{R}(C)) - \pi(\mathcal{A} \cap \cup_{l \in C^*} \mathcal{R}_l^{C^*}) + \pi(\mathcal{A} \cap \mathcal{R}_k) \right] + \\ &\quad \sum_{k \notin C^*} \frac{n_k}{n} \log \pi'(\mathcal{R}_k). \end{aligned}$$

Puisque

$$\begin{aligned} \pi(\mathcal{A} \cap \mathcal{R}_k) &= \pi(\mathcal{A} \cap \overline{\mathcal{R}_k^{C^*}}) + \pi(\mathcal{A} \cap \mathcal{R}_k^{C^*}) \\ &= \pi'(\mathcal{A} \cap \overline{\mathcal{R}_k^{C^*}}) + \pi'(\mathcal{A} \cap \mathcal{R}_k^{C^*}) \\ &\leqslant \pi(\mathcal{A} \cap \mathcal{R}_k^{C^*}) + \pi(\mathcal{A} \cap \cup_{l \in C^*} \mathcal{R}_l^{C^*}), \end{aligned}$$

on conclut que

$$\pi(\mathcal{A} \cap \mathcal{R}_k) - \pi(\mathcal{A} \cap \cup_{l \in C^*} \mathcal{R}_l^{C^*}) \leqslant \pi(\mathcal{A} \cap \mathcal{R}_k)$$

On obtient donc

$$\mathcal{L}(\pi, \mathbf{X}_n) \leqslant \mathcal{L}(\pi', \mathbf{X}_n).$$

Nous avons modifié π de telle sorte que la probabilité augmente, et que toutes les régions $\mathcal{A}$ appartenant à la clique C^* ont une mesure nulle pour la nouvelle distribution π'. Nous pouvons maintenant choisir une autre clique $C \neq C^*$ et procéder à la même modification, en mettant une masse de probabilité nulle dans toutes les parties de $\mathcal{A}$ qui ne croisent pas d'autres cliques, jusqu'à ce que la masse de $\mathcal{A}$ devienne nulle. □

2.C Démonstration du théorème 3 (page : 30)

Dans cet annexe, nous présentons une démonstration du Théorème 3, ainsi qu'un exemple d'application.

Preuve : D'abord nous remarquons que les représentations réelles des cliques maximales sont incluses dans $\mathcal{Q}^J$. Pour cela nous notons $\mathcal{C} = \{\mathscr{R}(C) : C \in E'_{\max}\}$. Soient $\mathcal{C}_1 \in \mathcal{C}$ et $\delta_{\mathcal{C}_1} \subset \{1, \ldots, K\}$ tel que $\mathcal{C}_1 = \cap_{k \in \delta_{\mathcal{C}_1}} \mathcal{R}_k$. La partition $\mathcal{Q}^J$ est défini comme étant la plus petite partition de Θ telle que $\{\mathcal{R}_k\}_{k=1}^K \subset \sigma(\mathcal{Q}^J)$. Nous en déduisons donc que $\mathcal{C}_1 \in \mathcal{Q}^J$. Il est évident qu'un élément $\mathcal{C}_i = \cap_{k \in \delta_{\mathcal{C}_i}} \mathcal{R}_k$ de $\mathcal{Q}^J$ qui est défini par l'intersection du plus grand nombre possible des $\mathcal{R}_k$ ($\#(\delta_{\mathcal{C}_i})$ est maximale) correspond à une clique maximale. Nous en déduisons que tout élément $\mathcal{C}'_i = \cap_{k \in \delta_{\mathcal{C}'_i}} \mathcal{R}_k$ de $\mathcal{Q}^J$ tel que $\delta_{\mathcal{C}'_i} \subset \delta_{\mathcal{C}_i}$ n'est pas maximale. Nous réitérons ce même raisonnement pour toutes les tailles possibles des ensembles $\delta_{\mathcal{C}_i}$ tout en excluant de la liste des éléments de $\mathcal{Q}^J$ les éléments qui ne correspondent pas aux cliques maximales (la ligne 10 de l'algorithme). □

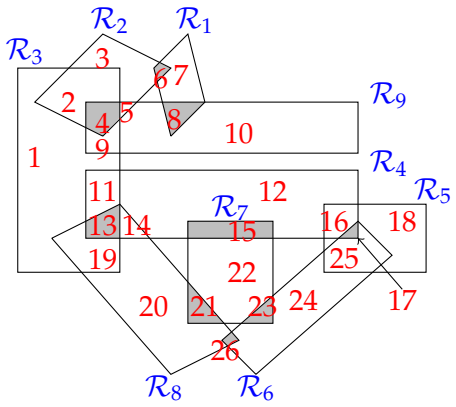
L'exemple 6 montre une application de l'Algorithme 2 à un cas concret.

Exemple 6. Dans cet exemple, dont le graphe d'intersection (Figure 2.C.1b) est le même que celui cité par Tomita [TTT06], nous supposons avoir observé les régions de censure $\{\mathcal{R}_k\}_{k=1}^9$ (Figure 2.C.1a).

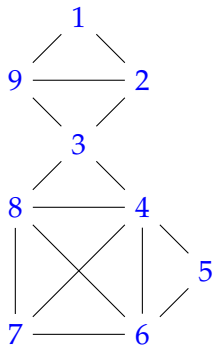
Les 9 régions de censure donnent lieu à la partition des régions élémentaire $\mathcal{Q}^J = \{\mathcal{Q}_1^J, \ldots, \mathcal{Q}_{26}^J\}$ dont les indices sont indiqués en rouge dans la Figure 2.C.1a. La matrice $\mathbf{B}$ (2.C.1c) décrit les dépendances entre régions de censure et régions élémentaires. Puisqu'il n'y a pas d'ambiguïté, nous confondons les éléments de la liste $E'_{\max}$ avec leurs représentations réelles. Appliquons l'Algorithme 2 à cette matrice. Nous avons

$$\mathrm{y}_{\mathbf{B}} = (1,2,1,3,2,2,1,2,2,1,2,1,3,2,2,2,3,1,2,1,2,1,2,1,2,2)\,.$$

Donc les éléments de $\mathcal{Q}^J$ correspondant aux maxima de $\mathrm{y}_{\mathbf{B}}$, c'est à dire $\left\{\mathcal{Q}_4^J, \mathcal{Q}_{13}^J, \mathcal{Q}_{17}^J\right\}$ sont des éléments de $E'_{\max}$. Commençons par la région élémentaire $\mathcal{Q}_4^J$. C'est l'intersection des 3 régions de censures $\mathcal{R}_2$, $\mathcal{R}_3$ et $\mathcal{R}_9$. Puisque nous retenons $\mathcal{Q}_4^J = \{2,3,9\}$ dans $E'_{\max}$, toutes les régions élémen-



(a) *Les régions de censures.*



(b) *Le graphe d'intersection.*

$$\mathbf{B} = \begin{bmatrix} 00000111000000000000000000 \\ 01111100000000000000000000 \\ 11010000101010000010000000 \\ 00000000001111111000000000 \\ 00000000000000011100000010 \\ 00000000000000001000001111 \\ 00000000000000100000111000 \\ 00000000000011000011100001 \\ 00011001110000000000000000 \end{bmatrix}$$

(c) *La matrice* $\mathbf{B}$.

FIGURE 2.C.1 – *Un exemple d'observation avec censure pour* 9 *régions donnant lieu à* 26 *régions élémentaires. Le support de l'ENPMV est formé de* 9 *régions élémentaires (les régions grisées).*

taires autres que $\mathcal{Q}_4^J$, définies par les régions de censure $\mathcal{R}_2$, $\mathcal{R}_3$ ou $\mathcal{R}_9$, ne feront pas partie de $E'_{\max}$. Ainsi, les régions élémentaires $\mathcal{Q}_1^J$, $\mathcal{Q}_2^J$, $\mathcal{Q}_3^J$, $\mathcal{Q}_5^J$, $\mathcal{Q}_9^J$ et $\mathcal{Q}_{10}^J$ ne font pas partie de $E'_{\max}$. De même, le fait que $\mathcal{Q}_{13}^J \in E'_{\max}$ conduit à ce que $\mathcal{Q}_{11}^J$, $\mathcal{Q}_{12}^J$, $\mathcal{Q}_{14}^J$, $\mathcal{Q}_{19}^J$, $\mathcal{Q}_{20}^J \notin E'_{\max}$ et le fait que $\mathcal{Q}_{17}^J \in E'_{\max}$ conduit à ce que $\mathcal{Q}_{16}^J$, $\mathcal{Q}_{18}^J$, $\mathcal{Q}_{24}^J$, $\mathcal{Q}_{25}^J \notin E'_{\max}$. La mise à jour du vecteur $\mathrm{y}_{\mathbf{B}}$ (les lignes 6 et 10 de l'algorithme) donne

$$\mathrm{y}_{\mathbf{B}} = (0,0,0,0,0,2,1,2,0,0,0,0,0,0,2,0,0,0,0,0,2,1,2,0,0,2)\,.$$

Ainsi, les régions élémentaires $\mathcal{Q}_6^J$, $\mathcal{Q}_8^J$, $\mathcal{Q}_{15}^J$, $\mathcal{Q}_{21}^J$, $\mathcal{Q}_{23}^J$ et $\mathcal{Q}_{26}^J$, correspondant à la valeur 2, la valeur maximale de $\mathrm{y}_{\mathbf{B}}$, font partie de $E'_{\max}$. Les régions élémentaires $\mathcal{Q}_7^J$ et $\mathcal{Q}_{22}^J$ ne sont pas retenues car $\mathcal{Q}_7^J = \{1\} \subset \{1,2\} = \mathcal{Q}_6^J$ et $\mathcal{Q}_{22}^J = \{7\} \subset \{4,7\} = \mathcal{Q}_{15}^J$.

2.D Extension de l'algorithme multiplicatif

Dans cette annexe nous présentons une extension de l'algorithme multiplicatif présenté dans le Chapitre 2, page 34.

Algorithme multiplicatif

Nous nous plaçons dans le cadre général où l'on note $\phi(\mathbf{w})$ la fonction à maximiser par rapport à $\mathbf{w}$ dans le simplexe probabiliste $\mathcal{S}^{M+}$. Nous désignons par $\mathbf{g}(\mathbf{w})$ le gradient $\frac{\partial \phi(\mathbf{w})}{\partial \mathbf{w}}$. Nous supposons que $\phi(\cdot)$ est deux fois continuement dérivable, qu'elle est concave et que toutes les composantes $\{\mathbf{g}(\mathbf{w})\}_m$ du gradient sont positives pour tout $\mathbf{w} \in \mathcal{S}^{M+}$.

Considérons un algorithme de gradient projeté, où à l'itération t, le gradient $\mathbf{g}^{(t)} = \mathbf{g}(\mathbf{w}^{(t)})$ est projeté sur l'hyperplan orthogonal à $\mathbf{1}$ en utilisant la métrique induite par une matrice $\mathbf{\Lambda}$ définie

positive. La direction de recherche de l'optimum est donnée par

$$\mathbf{d}^{(t)} = \left(\mathbf{\Lambda} - \frac{\mathbf{\Lambda 1 1}^\mathsf{T}\mathbf{\Lambda}}{\mathbf{1}^\mathsf{T}\mathbf{\Lambda 1}}\right)\mathbf{g}^{(t)} = \mathbf{\Lambda g}^{(t)} - \frac{\mathbf{1}^\mathsf{T}\mathbf{\Lambda g}^{(t)}}{\mathbf{1}^\mathsf{T}\mathbf{\Lambda 1}}\mathbf{\Lambda 1}.$$

Nous remarquons que $\mathbf{1}^\mathsf{T}\mathbf{d}^{(t)} = 0$. La mise à jour du vecteur des poids est donnée par

$$\mathbf{w}^{(t+1)} = \mathbf{w}^{(t)} + \alpha^{(t)}\mathbf{d}^{(t)},$$

avec $\alpha^{(t)}$ un réel à définir. En posant $\mathbf{\Lambda} = \text{diag}\left\{\{\mathbf{w}^{(t)}\}_m, m = 1, \ldots, M\right\}$, nous avons

$$\mathbf{d}^{(t)} = \left[\mathbf{D}\left(\mathbf{g}^{(t)}\right) - \left(\left(\mathbf{w}^{(t)}\right)^\mathsf{T}\mathbf{g}^{(t)}\right)\mathbf{I}\right]\mathbf{w}^{(t)}, \tag{2.17}$$

et

$$\mathbf{w}^{(t+1)} = \mathbf{w}^{(t+1)} + \alpha^{(t)}\left[\mathbf{D}\left(\mathbf{g}^{(t)}\right) - \left(\left(\mathbf{w}^{(t)}\right)^\mathsf{T}\mathbf{g}^{(t)}\right)\mathbf{I}\right]\mathbf{w}^{(t)},$$

où $\mathbf{D}\left(\mathbf{g}^{(t)}\right)$ est la matrice diagonale $\text{diag}\left\{\{\mathbf{g}^{(t)}\}_m, m = 1, \ldots, M\right\}$.

Pour tout vecteur $\mathbf{v} \in \mathbb{R}^M$, nous posons

$$\mathbb{E}^{(t)}\{\mathbf{v}\} = \left(\mathbf{w}^{(t)}\right)^\mathsf{T}\mathbf{v} \text{ et } \mathbb{V}^{(t)}(\mathbf{v}) = \sum_{m=1}^{M}\{\mathbf{w}^{(t)}\}_m\left[\{\mathbf{v}\}_m - \mathbb{E}^{(t)}\{\mathbf{v}\}\right]^2.$$

Des calculs directs donnent

$$\left(\mathbf{d}^{(t)}\right)^\mathsf{T}\mathbf{g}^{(t)} = \sum_{m=1}^{M}\{\mathbf{w}^{(t)}\}_m\left[\{\mathbf{g}^{(t)}\}_m - \left(\left(\mathbf{w}^{(t)}\right)^\mathsf{T}\mathbf{g}^{(t)}\right)\right]^2 = \mathbb{V}^{(t)}(\mathbf{g}^{(t)}),$$

et

$$\left(\mathbf{d}^{(t)}\right)^\mathsf{T}\mathbf{d}^{(t)} = \sum_{m=1}^{M}\{\mathbf{w}^{(t)}\}_m^2\left[\{\mathbf{g}^{(t)}\}_m - \left(\left(\mathbf{w}^{(t)}\right)^\mathsf{T}\mathbf{g}^{(t)}\right)\right]^2 \leqslant \left(\mathbf{d}^{(t)}\right)^\mathsf{T}\mathbf{g}^{(t)}.$$

La taille de pas

$$\alpha^{(t)} = \alpha_\star^{(t)} = \frac{1}{\left(\mathbf{w}^{(t)}\right)^\mathsf{T}\mathbf{g}^{(t)}}, \tag{2.18}$$

conduit à des itérations de la forme

$$\mathbf{w}^{(t+1)} = \frac{\mathbf{D}\left(\mathbf{g}^{(t)}\right)\mathbf{w}^{(t)}}{\left(\mathbf{w}^{(t)}\right)^\mathsf{T}\mathbf{g}^{(t)}}. \tag{2.19}$$

Remarquons que lorsque $\phi(\cdot)$ est strictement positive, appliquer cet algorithme à $\log[\phi(\cdot)]$ donne les mêmes itérations que l'équation (2.19). Nous pouvons donc remplacer l'hypothèse de concavité par

les hypothèses de positivité et de log-concavité. Lorsque $\phi(\mathbf{w}) = \Phi\left[\mathbf{M}(\mathbf{w})\right]$ avec

$$\mathbf{M}(\mathbf{w}) = \sum_{m=1}^{M} w_m \mathbf{H}_m, \tag{2.20}$$

où pour tout $m = 1, \ldots, M$, $\mathbf{H}_m$ est une matrice de taille $n \times n$ symétrique définie positive, $n \geqslant 1$, et $\Phi(\cdot)$ est une fonction d'information (voir le chapitre 5 de [Puk93]), alors les itérations de l'équation(2.19) correspondent à un algorithme multiplicatif pour un problème de plan d'expérience pour lequel il existe une large littérature [Tit76, STT78, Tor83, Fel89, DPZ08, Yu10a, Yu10b]. Comme nous l'avons déjà remarqué dans le paragraphe 2.3.2.1, page 32, la recherche de l'ENPMV est équivalente à un problème de plan d'expérience D-optimal, avec $\Phi(\mathbf{M}) = \log\det(\mathbf{M})$, ce qui offre la possibilité d'utiliser un algorithme multiplicatif similaire à celui de l'équation (2.19), voir [Lin83, Böh89]. La preuve de convergence monotone de ces itérations pour le problème de plan d'expérience D-optimal est donnée dans [Tit76, Páz86].

Les itérations de la forme

$$\mathbf{w}^{(t+1)} = \frac{\mathbf{D}\left(\left(\mathbf{g}^{(t)}\right)^{\odot\lambda}\right)\mathbf{w}^{(t)}}{\left(\mathbf{w}^{(t)}\right)^{\mathsf{T}}\left(\mathbf{g}^{(t)}\right)^{\odot\lambda}}, \tag{2.21}$$

où $\left(\mathbf{g}^{(t)}\right)^{\odot\lambda}$ est le vecteur dont les composantes sont $\left\{\mathbf{g}^{(t)}\right\}_m^{\lambda}$, ont été considérées pour le problème de construction d'un plan d'expérience A-optimal, avec $\Phi\left(\mathbf{M}\right) = -\text{trace}\left(\mathbf{M}^{-1}\right)$ et c-optimal, avec $\Phi\left(\mathbf{M}\right) = -\mathbf{c}^{\mathsf{T}}\mathbf{M}^{-1}\mathbf{c}$ pour $\mathbf{c} \in \mathbb{R}^n$. La convergence monotone de ces itérations est prouvée dans [Fel89] pour $\lambda = \frac{1}{2}$. Dans [Yu10a], l'auteur donne une preuve de cette monotonie pour une classe de critères incluant les D- et A-optimalités et pour $\lambda \in]0, 1]$.

Des extensions de des itérations de l'équation (2.19), basées sur la taille de pas $\alpha^{(t)} = \frac{1}{\left(\mathbf{w}^{(t)}\right)^{\mathsf{T}}\mathbf{g}^{(t)} - \beta^{(t)}}$ ont été proposées par [DPZ08]. Elles sont de la forme

$$\mathbf{w}^{(t+1)} = \frac{\left(\mathbf{D}\left(\mathbf{g}^{(t)}\right) - \beta^{(t)}\mathbf{I}\right)\mathbf{w}^{(t)}}{\left(\mathbf{w}^{(t)}\right)^{\mathsf{T}}\mathbf{g}^{(t)} - \beta^{(t)}}, \tag{2.22}$$

Dette et *al.* donnent dans le même article une preuve de convergence monotone pour $-\infty < \beta^{(t)} \leqslant \frac{1}{2}\min_{m=1,\ldots,M}\left\{\mathbf{g}^{(t)}\right\}_m$.

La monotonicité est usuellement l'argument utilisé pour prouver la convergence d'une séquence $\left\{\mathbf{w}^{(t)}\right\}_t$ vers un optimum $\mathbf{w}^*$. Remarquons que pour tout m où $\left\{\mathbf{w}^{(t)}\right\}_m > 0$, alors $\left\{\mathbf{w}^{(t+1)}\right\}_m > 0$ pour toutes les itérations des équations (2.19), (2.21) et (2.22). Pour tout point fixe $\mathbf{w}^{(t+1)} = \mathbf{w}^{(t)}$ tel que $\left\{\mathbf{w}^{(t)}\right\}_m > 0$, on a $\left\{\mathbf{g}^{(t)}\right\}_m = c$ où c est un réel positif. La dérivée directionnelle

$$F_\phi(\mathbf{w}, \mathbf{z}) = \lim_{\lambda \to 0^+} \frac{\phi((1-\lambda)\mathbf{w} + \lambda\mathbf{z}) - \lambda(\mathbf{w})}{\lambda} = \mathbf{g}^{\mathsf{T}}(\mathbf{w})(\mathbf{z} - \mathbf{w}), \quad \mathbf{w}, \mathbf{z} \in \mathcal{S}^{M+},$$

satisfait $F_\phi(\mathbf{w}^{(t)}, \mathbf{z}) = 0$ pour tout $\mathbf{z} \in \mathcal{S}^{M+}$ et $\mathbf{w}^{(t)}$ est ϕ-optimal. Prendre en compte la possibilité d'avoir des composantes nulles dans le vecteur optimal nécessite des arguments plus subtils (voir [Yu10a]).

Nous remarquons que l'interprétation des itérations (2.19) comme un algorithme de gradient projeté avec une taille de pas égale à $\alpha_\star^{(t)}$ permet de conclure à la convergence monotone des itérations (2.22) pour $\beta^{(t)} \leqslant 0$. D'autre part, des tailles de pas plus grands que $\alpha_\star^{(t)}$ assurent la convergence monotone quand

$$\alpha_\star^{(t)} \leqslant \bar{\alpha}^{(t)} \stackrel{\Delta}{=} \underset{\alpha \in \mathbb{R}}{\operatorname{argmax}} \ \phi\left(\mathbf{w}^{(t)} + \alpha \mathbf{d}^{(t)}\right).$$

Dans la suite, nous utilisons la preuve de convergence monotone de [Yu10a] pour la D et A-optimalité afin de montrer que $\alpha_\star^{(t)} \leqslant \bar{\alpha}^{(t)}$ en D-optimalité.

Convergence monotone pour les plans d'expérience D et A-optimaux

Considérons l'itération $t \geqslant 0$ de l'une des variantes de l'algorithme multiplicatif présentées dans le paragraphe précédent. On note $\phi = \phi(\mathbf{w}^{(t)})$, $\mathbf{g} = \mathbf{g}(\mathbf{w}^{(t)})$, $\mathbf{w}' = \mathbf{w}^{(t+1)}$ et $\phi' = \phi(\mathbf{w}')$, $\mathbf{g}' = \mathbf{g}(\mathbf{w}')$. On suppose $w_m > 0$ pour tout $m = 1, \ldots, M$.

A-optimalité : Nous considérons le critère

$$\Phi_A(\mathbf{M}) = -\text{trace}\left[\mathbf{K}^\mathsf{T}\mathbf{M}^{-1}\mathbf{K}\right],$$

avec $\mathbf{K}$ une matrice de rang plein et de taille $n \times n$. Une extension à des matrices de taille $n \times n'$ avec $n' < n$ est possible (voir [Yu10a]). Nous avons $\phi_A(\mathbf{w}) = \Phi_A(\mathbf{M}[\mathbf{w}])$. Nous supposons que les matrices $\mathbf{H}_m$ sont de rang 1. Nous posons $\mathbf{H}_m = \mathbf{x}_m\mathbf{x}_m^\mathsf{T}$, $m = 1, \ldots, M$ où $\mathbf{x}_m \in \mathbb{R}^n$. Comme dans [Yu10a], nous définissons

$$f(\mathbf{w}, \mathbf{Q}) = -\text{trace}\left[\mathbf{K}^\mathsf{T}\mathbf{Q}\mathbf{W}^{-1}\mathbf{Q}^\mathsf{T}\mathbf{K}\right],$$

avec $\mathbf{W} = \text{diag}\{w_m, m = 1, \ldots, M\}$ et $\mathbf{Q}$ une matrice de taille $n \times M$. La matrice $\mathbf{Q}\mathbf{W}^{-1}\mathbf{Q}^\mathsf{T}$ correspond à la matrice variance-covariance de l'estimateur $\hat{\theta} = \mathbf{Q}\mathbf{y}$ dans le modèle de la régression linéaire $\mathbf{y} = \mathbf{X}\theta + \epsilon$ où θ est le vecteur des paramètres et les erreurs sont supposées de moyenne nulle et de matrice de variance-covariance $\mathbf{W}^{-1}$. Le meilleur estimateur linéaire sans biais est dans ce cas

$$\mathbf{Q}_*(\mathbf{w}) = \left(\mathbf{X}^\mathsf{T}\mathbf{W}\mathbf{X}\right)^{-1}\mathbf{X}^\mathsf{T}\mathbf{W},$$

et $\mathbf{Q} = \mathbf{Q}_*(\mathbf{w})$ maximise $f(\mathbf{w}, \mathbf{Q})$ pour tout $\mathbf{w}$ de composantes toutes strictement positives et à condition que $\mathbf{Q}\mathbf{X} = \mathbf{I}_n$. Si la m-ième ligne de $\mathbf{X}$ est donnée par le vecteur $\mathbf{x}_m^\mathsf{T}$ alors $\mathbf{Q}_*(\mathbf{w})\mathbf{W}^{-1}\mathbf{Q}_*^\mathsf{T}(\mathbf{w}) = \mathbf{M}^{-1}(\mathbf{w})$ et

$$f[\mathbf{w}, \mathbf{Q}_*(\mathbf{w})] = \phi_A(\mathbf{w}) = \max_{\mathbf{Q}:\mathbf{Q}\mathbf{X}=\mathbf{I}_n} f(\mathbf{w}, \mathbf{Q}).$$

Des calculs directs donnent

$$g_m = \{\mathbf{g}(\mathbf{w})\}_m = \mathbf{x}_m^\mathsf{T}\mathbf{M}^{-1}(\mathbf{w})\mathbf{K}\mathbf{K}^\mathsf{T}\mathbf{M}^{-1}(\mathbf{w})\mathbf{x}_m,$$

et les itérations de l'équation(2.19) correspondent à

$$w'_m = w_m \frac{\mathbf{x}_m^\mathsf{T}\mathbf{M}^{-1}(\mathbf{w})\mathbf{K}\mathbf{K}^\mathsf{T}\mathbf{M}^{-1}(\mathbf{w})\mathbf{x}_m}{\text{trace}\left[\mathbf{K}^\mathsf{T}\mathbf{M}^{-1}(\mathbf{w})\mathbf{K}\right]}. \tag{2.23}$$

En notant $a_m = \left\{\mathbf{Q}_*^\mathsf{T}(\mathbf{w})\mathbf{K}\mathbf{K}^\mathsf{T}\mathbf{Q}_*(\mathbf{w})\right\}_{mm} = w_m^2\left(\mathbf{x}_m^\mathsf{T}\mathbf{M}^{-1}(\mathbf{w})\mathbf{K}\mathbf{K}^\mathsf{T}\mathbf{M}^{-1}(\mathbf{w})\mathbf{x}_m\right)$, on a

$$\phi_A(\mathbf{w}) = -\sum_{m=1}^{M}\frac{a_m}{w_m}, \quad w'_m = \frac{a_m/w_m}{\sum_{m=1}^{M} a_m/w_m}$$

et

$$f\left[\mathbf{w}', \mathbf{Q}_*(\mathbf{w})\right] = -\sum_{m=1}^{M}\frac{a_m}{w'_m} = \left(\sum_{m=1}^{M} w_m\right)\left(\sum_{m=1}^{M}\frac{a_m}{w_m}\right) = \phi_A(\mathbf{w}).$$

Cela implique que

$$\phi_A(\mathbf{w}') = f\left[\mathbf{w}', \mathbf{Q}_*(\mathbf{w}')\right] \geqslant f\left[\mathbf{w}', \mathbf{Q}_*(\mathbf{w})\right] = \phi_A(\mathbf{w}),$$

et donc les itérations (2.19) produisent une séquence $\phi_A(\mathbf{w}^{(t)})$ croissante.

D-optimalité : Nous considérons le critère

$$\phi_D(\mathbf{w}) = \log\det\left[\mathbf{M}(\mathbf{w})\right].$$

Nous avons $g_m = \mathbf{x}_m^\mathsf{T}\mathbf{M}^{-1}(\mathbf{w})\mathbf{x}_m$ et l'itération (2.19) correspond à

$$w'_m = w_m \frac{\mathbf{x}_m^\mathsf{T}\mathbf{M}^{-1}(\mathbf{w})\mathbf{x}_m}{\sum_{m=1}^{M} w_m\mathbf{x}_m^\mathsf{T}\mathbf{M}^{-1}(\mathbf{w})\mathbf{x}_m} = w_m\frac{\mathbf{x}_m^\mathsf{T}\mathbf{M}^{-1}(\mathbf{w})\mathbf{x}_m}{n}. \tag{2.24}$$

les composantes du gradient au $\mathbf{w}' = \mathbf{w}^{(t+1)}$ sont $g'_m = \mathbf{x}_m^\mathsf{T}\mathbf{M}^{-1}(\mathbf{w}')\mathbf{x}_m$. Puisque l'itération (2.19) correspond à un algorithme de gradient projeté avec une taille de pas $\alpha^{(t)} = \alpha_\star^{(t)}$, la convergence sera monotone, en particulier si $\alpha_\star^{(t)}$ est telle que $(\mathbf{w}^{(t+1)} - \mathbf{w}^{(t)})^\mathsf{T} \geqslant 0$, c'est à dire $\sum_{m=1}^{M} w_m g'_m \leqslant \sum_{m=1}^{M} w'_m g'_m = n$. Considérons maintenant le critère $\Phi_A(\cdot)$ de la section précédente pour $\mathbf{K} = \mathbf{M}^{1/2}(\mathbf{w})$. L'itération (2.23) coïncide avec celle exprimée par l'équation (2.24) et la monotonicité de cette dernière implique que

$$\phi_A(\mathbf{w}') = -\text{trace}\left[\mathbf{M}(\mathbf{w})\mathbf{M}^{-1}(\mathbf{w}')\right] = -\sum_{m=1}^{M} w_m g'_m \geqslant \phi_A(\mathbf{w}) = n. \tag{2.25}$$

Nous en déduisons que la convergence est monotone, c'est à dire que

$$\phi_D(\mathbf{w}^{(t+1)}) \geqslant \phi_D(\mathbf{w}^{(t)}).$$

Pas optimal en gradient projeté : Considérons à nouveau l'itération (2.17), avec une taille de pas $\alpha^{(t)} = \lambda \alpha_\star^{(t)}$ où $\alpha_\star^{(t)}$ est donné par l'équation (2.18). On peut écrire

$$\mathbf{w}^{(t+1)} = (1-\gamma)\mathbf{w}^{(t)} + \gamma \bar{\mathbf{w}}^{(t)}, \tag{2.26}$$

avec $\bar{\mathbf{w}}^{(t)}$ donnée par l'équation (2.19). Afin d'avoir toutes les composantes w_m positives il faut que

$$\gamma \leqslant \bar{\gamma}^{(t)} = \frac{n}{n - \min_m \{g_m\}},$$

avec $g_m = \left\{\mathbf{g}^{(t)}\right\}_m$, la m-ième composante du gradient de $\phi_D(\mathbf{w}) = \log\det[\mathbf{M}(\mathbf{w})]$ au point $\mathbf{w}^{(t)}$. Nous remarquons que l'équation (2.26) est équivalente à l'équation (2.22) pour

$$\gamma = \frac{\left(\mathbf{w}^{(t)}\right)^\mathsf{T} \mathbf{g}^{(t)}}{\left(\mathbf{w}^{(t)}\right)^\mathsf{T} \mathbf{g}^{(t)} - \beta} = \frac{n}{n-\beta}.$$

Il est indiqué dans [DPZ08] que l'itération (2.26) est monotone pour $\gamma \leqslant \hat{\gamma}^{(t)} = \frac{n}{n-(1/2)\min_m\{g_m\}}$. Nous posons $h^{(t)}(\gamma) = \log\det\left[\mathbf{M}(\mathbf{w}^{(t+1)})\right]$ considérée comme une fonction de γ, et $h'^{(t)}(\gamma)$ sa dérivée. Nous avons

$$h'^{(t)}(\gamma) = \operatorname{trace}\left\{\left[\mathbf{M}^{(t)} + \gamma(\mathbf{M}'^{(t)} - \mathbf{M}^{(t)})\right]^{-1}\left(\mathbf{M}'^{(t)} - \mathbf{M}^{(t)}\right)\right\},$$

avec $\mathbf{M}^{(t)} = \mathbf{M}(\mathbf{w}^{(t)})$ et $\mathbf{M}'^{(t)} = \mathbf{M}(\mathbf{w}'^{(t)})$. Des calculs directs donnent

$$\begin{aligned} h'^{(t)}(0) &= \operatorname{trace}\left(\left(\mathbf{M}^{(t)}\right)^{-1}\mathbf{M}'^{(t)}\right) - n \\ &= \frac{\sum_{m=1}^M w_m g_m^2 - n^2}{n} \\ &= \frac{\sum_{m=1}^M w_m g_m^2 - \left(\sum_{m=1}^M w_m g_m\right)^2}{n} \\ &\geqslant 0, \end{aligned}$$

et

$$h'^{(t)}(1) = n - \operatorname{trace}\left(\left(\mathbf{M}'^{(t)}\right)^{-1}\mathbf{M}^{(t)}\right) \geqslant 0.$$

Donc $\gamma_\star^{(t)} \stackrel{\Delta}{=} \underset{\gamma \leqslant \bar{\gamma}^{(t)}}{\text{argmax}}\ h^{(t)}(\gamma) \geqslant 1$. On peut écrire

$$h'^{(t)}(\gamma) = \text{trace}\left\{\left[(1-\gamma)\mathbf{I} + \gamma\mathbf{A}\right]^{-1}(\mathbf{A}-\mathbf{I})\right\} = \sum_{m=1}^{M} \frac{D_m - 1}{1+\gamma(D_m-1)},$$

où $\mathbf{A} = \left(\mathbf{M}^{(t)}\right)^{-1}\mathbf{M}'^{(t)}$ et D_m est la m-ième valeur propre de $\mathbf{A}$. Le pas optimal $\gamma_\star^{(t)}$ peut être déterminée par dichotomie. Nous fixons un paramètre $\epsilon > 0$ qui définit la précision avec laquelle nous cherchons à déterminer $\gamma_\star^{(t)}$.

Algorithme 5 : Recherche de $\gamma_\star^{(t)}$ par dichotomie

donnée : ϵ

sortie : $\gamma_\star^{(t)}$

1 : Calculer $h'^{(t)}\left(\bar{\gamma}^{(t)}\right)$.

2 : Si $h'^{(t)}\left(\bar{\gamma}^{(t)}\right) \geqslant 0$ alors $\gamma_\star^{(t)} = \bar{\gamma}^{(t)}$. fin.

3 : Sinon on pose $i = 1$, $\gamma_L^{(1)} = 1$, $\gamma_R^{(1)} = \bar{\gamma}^{(t)}$, $\hat{\gamma} = (\bar{\gamma}^{(t)} + 1)/2$ et $p = \left\lfloor \log\left[(\bar{\gamma}^{(t)} - 1)/\epsilon\right] / \log(2) \right\rfloor$, avec $\lfloor k \rfloor$ le plus petit entier supérieur ou égale à k.

4 Tant que $i \leqslant p$

5 : Si $h'^{(t)}(\hat{\gamma}) > 0$, mettre $\gamma_L^{(i+1)} = \hat{\gamma}$ et $\gamma_R^{(i+1)} = \gamma_R^{(i)}$.

6 : Si $h'^{(t)}(\hat{\gamma}) < 0$, mettre $\gamma_L^{(i+1)} = \gamma_L^{(i)}$ et $\gamma_R^{(i+1)} = \hat{\gamma}$.

7 : Si $h'^{(t)}(\hat{\gamma}) = 0$, mettre $\gamma_\star^{(t)} = \hat{\gamma}$. fin.

8 : Mettre $\hat{\gamma} = \left(\gamma_L^{(i+1)} + \gamma_R^{(i+1)}\right)/2$ et $i \leftarrow i + 1$.

9 : fin tant que

10 : Mettre $\gamma_\star^{(t)} = \hat{\gamma}$.

11 : fin.

- CHAPITRE 3 -

ESTIMATION PAR MAXENT

Dans le chapitre 2, nous avons pu constater que l'estimateur non paramétrique par maximum de vraisemblance construit à partir de données censurées conduit souvent à des modèles de population peu réalistes. En particulier, la concentration de la masse de probabilité sur des zones de petite mesure de Lebesgue, l'affectation de masse nulle à de larges zones du domaine paramétrique et l'instabilité de la distribution résultante par rapport à l'ajout de nouvelles données rendent ce critère peu adapté à notre problème. Nous proposons dans ce chapitre une autre approche pour estimer π_θ, reposant à la fois sur la maximisation de la vraisemblance et sur le principe d'estimation par Maximum d'Entropie (MaxEnt).

Le principe d'entropie maximale a été largement utilisé dans l'estimation de densités de probabilité depuis les travaux pionniers de Jaynes [Jay57]. Étant donné un ensemble de contraintes que doit satisfaire une densité, ce principe suggère de choisir parmi toutes les distributions satisfaisant ces contraintes celle contenant le moins d'information, et donc la distribution la moins arbitraire. Il est dans ce sens dual du maximum de vraisemblance, qui tend à favoriser les densités les plus concentrées expliquant les observations, comme nous l'avons vu. Il se trouve que parfois aucune distribution ne satisfait à l'ensemble des contraintes qu'il faut alors relaxer. Le degré de relaxation des contraintes est usuellement un hyper-paramètre de l'estimateur, dont le choix est laissé à l'utilisateur. Nous proposons ici de déterminer la relaxation des contraintes en fonction de la vraisemblance de la distribution qui en résulte, exploitant ainsi les propriétés des deux critères, l'entropie et la vraisemblance.

Nous commençons ce chapitre par la section 3.1 où nous nous introduisons la notion de compatibilité des lois empiriques définies par les observations. Dans la section 3.2 nous présentons quelques propriétés de l'estimateur MaxEnt. Nous discutons dans la section 3.3 de l'adaptation du critère MaxEnt à notre problème d'estimation à partir d'observations censurées tout en présentant un cas spécial d'application au paragraphe 3.3.2 où il est possible de faire appel à certains résultats relatifs au MaxEnt. Dans la section 3.4, nous présentons le critère présenté dans cette thèse reposant sur le

maximum d'entropie et le maximum de vraisemblance. La section 3.5 caractérise les performances de la solution proposée, à travers des études de simulation.

3.1 Compatibilité des lois empiriques

Comme nous avons vu dans le Chapitre 2, page 15, les données de notre problème sont des réalisations de J lois multinomiales différentes, $\left\{\mathbf{q}_j(\mathbf{n})\right\}_{j=1}^{J}$. Souvent il n'existe pas d'un vecteur de probabilité $\mathbf{w} \in \mathcal{S}^{M+}$ tel que

$$\mathbf{q}_j(\mathbf{n}) = \mathbf{B}^{(j)}\mathbf{w} \qquad \forall j = 1, \ldots, J, \tag{3.1}$$

c'est à dire il n'existe pas de loi de probabilité qui conduise à l'ensemble des lois observées. Lorsque cela se produit, le critère du maximum de vraisemblance peut attribuer une masse de probabilité nulle à des régions de l'espace paramétrique Θ qui sont parfois à l'origine d'une partie des données disponibles, comme nous le montrons par la suite dans l'exemple 7.

Reprenons le vecteur $\mathbf{q}(\mathbf{n})$, de taille K, qui est la concaténation des vecteurs $\left\{\mathbf{q}_j(\mathbf{n})\right\}_{j=1}^{J}$ selon sa définition dans l'équation (2.11). Admettons que $n_\ell^j > 0$ pour tout $j = 1, \ldots, J$ et $\ell = 1, \ldots, L^j$, c'est à dire que tous les événements possibles sont observés [1]. Soit $\mathcal{C}(\mathbf{n}, \mathbf{B})$ l'ensemble :

$$\mathcal{C}(\mathbf{n}, \mathbf{B}) = \left\{\mathbf{w} \in \mathcal{S}^{M+};\ \ \mathbf{B}\mathbf{w} = \mathbf{q}(\mathbf{n})\right\}. \tag{3.2}$$

Notons que $\mathcal{C}(\mathbf{n}, \mathbf{B})$ peut parfois être vide.

Définition 11. ***(lois empiriques compatibles)*** *: Nous disons que les lois empiriques $\left\{\mathbf{q}_j(\mathbf{n})\right\}_{j=1}^{J}$ sont compatibles si l'ensemble $\mathcal{C}(\mathbf{n}, \mathbf{B})$ n'est pas vide.* □

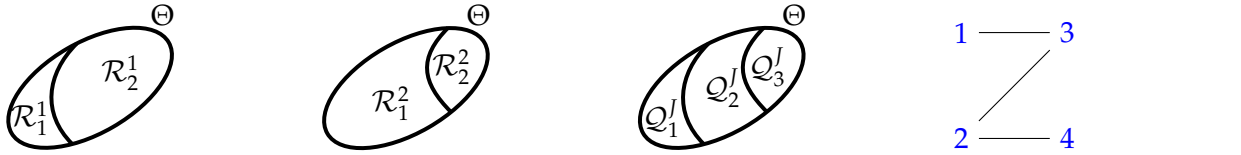
Comme nous l'avons déjà vu (Chapitre 2, page 15), si $\mathbf{w} \in \mathcal{C}(\mathbf{n}, \mathbf{B})$, alors $\mathbf{w}$ maximise la log-vraisemblance (égale dans ce cas à $-\sum_{j=1}^{J} \frac{n^j}{n}\left[H_1(\mathbf{q}_j(\mathbf{n}))\right]$). Toutefois la probabilité que les lois empiriques soient compatibles est d'autant plus petite que le nombre des partitions J augmente, et que les n^j sont petits. Nous présentons un exemple de l'impact de l'incompatibilité des lois empirique sur l'ENPMV.

Exemple 7. Considérons n observations, notées $\mathbf{X}_n$, issues des éléments des deux partitions $\mathcal{R}^1 = \left\{\mathcal{R}_1^1, \mathcal{R}_2^1\right\}$ et $\mathcal{R}^2 = \left\{\mathcal{R}_1^2, \mathcal{R}_2^2\right\}$ illustrées dans la Figure 3.1. La partition $\mathcal{Q}^J$ (Définition 2, page 10) contient 3 régions élémentaires $\left\{\mathcal{Q}_1^J, \mathcal{Q}_2^J, \mathcal{Q}_3^J\right\}$, voir la Figure 3.1c. En notant

$$\left\{\mathcal{R}_1^1, \mathcal{R}_2^1, \mathcal{R}_1^2, \mathcal{R}_2^2\right\} = \left\{\mathcal{R}_1, \mathcal{R}_2, \mathcal{R}_3, \mathcal{R}_4\right\},$$

le graphe d'intersection des régions $\left\{\mathcal{R}_1^1, \mathcal{R}_2^1, \mathcal{R}_1^2, \mathcal{R}_2^2\right\}$ est donné dans la Figure 3.1d.

1. Si ce n'est pas le cas, nous pouvons tout simplement supprimer les fréquences nulles du vecteur $\mathbf{q}_j(\mathbf{n})$ de façon à avoir toutes fréquences strictement positives.



(a) $\mathcal{R}^1$. **(b)** $\mathcal{R}^2$. **(c)** $\mathcal{Q}^J$. **(d)** *Graphe d'intersection.*

FIGURE 3.1 – *Régions de censure, régions élémentaires et graphe d'intersection de l'Exemple 7.*

Admettons que $n_\ell^j > 0$, $j = 1, 2$ et $\ell = 1, 2$. Nous avons $n^1 = n_1^1 + n_2^1$, $n^2 = n_1^2 + n_2^2$ et $n = n^1 + n^2$. Puisque les 3 régions élémentaires $\mathcal{Q}_1^J$, $\mathcal{Q}_2^J$ et $\mathcal{Q}_3^J$ ne sont pas vides, alors elles appartiennent à la liste $E'_{\max}$ (calculée par l'Algorithme 2, page 30) et appartiennent donc a priori au support de l'ENPMV. Si les lois empiriques sont compatibles alors il existe $\mathbf{w} \in \mathcal{S}^{3+}$ tel que

$$w_1 = \frac{n_1^1}{n_1^1 + n_2^1}, \quad w_2 + w_3 = \frac{n_2^1}{n_1^1 + n_2^1}, \quad w_1 + w_2 = \frac{n_1^2}{n_1^2 + n_2^2} \text{ et } w_3 = \frac{n_2^2}{n_1^2 + n_2^2}.$$

Les conditions $w_1 + w_2 \geqslant w_1$ et $w_2 + w_3 \geqslant w_3$ impliquent que $n_2^1 n_1^2 \geqslant n_1^1 n_2^2$. D'autre part, si nous supposons que $n_2^1 n_1^2 \geqslant n_1^1 n_2^2$ alors le vecteur $\mathbf{w} = (w_1, w_2, w_3)^\mathsf{T}$ avec

$$w_1 = \frac{n_1^1}{n_1^1 + n_2^1}, \quad w_3 = \frac{n_2^1}{n_1^1 + n_2^1}, \quad w_2 = 1 - w_1 - w_3 = \frac{n_2^1 n_1^2 - n_1^1 n_2^2}{(n_1^1 + n_2^1)(n_1^2 + n_2^2)},$$

appartient à $\mathcal{C}(\mathbf{n}, \mathbf{B})$. Nous concluons que

$$\mathcal{C}(\mathbf{n}, \mathbf{B}) \neq \emptyset \Leftrightarrow n_2^1 n_1^2 \geqslant n_1^1 n_2^2. \tag{3.3}$$

La log-vraisemblance d'un vecteur de probabilité $\mathbf{w} = (w_1, w_2, w_3)^\mathsf{T} \in \mathcal{S}^{3+}$ est

$$\mathcal{L}(\mathbf{w}, \mathbf{X}_n) = \frac{n_1^1}{n} \log(w_1) + \frac{n_2^1}{n} \log(w_2 + w_3) + \frac{n_1^2}{n} \log(w_1 + w_2) + \frac{n_2^2}{n} \log(w_3).$$

Les conditions de premier ordre impliquent que si $\mathbf{w}^\star$ maximise $\mathcal{L}(\cdot, \mathbf{X}_n)$, alors $\forall m = 1, 2, 3$

$$w_m^\star \left. \frac{\partial \mathrm{L}_{\mathcal{S}^{3+}}^{\mathcal{L}(\cdot, \mathbf{X}_n)}}{\partial w_m} \right|_{\mathbf{w}^\star} = 0,$$

où

$$\mathrm{L}_{\mathcal{S}^{3+}}^{\mathcal{L}(\cdot, \mathbf{X}_n)}(\mathbf{w}, \lambda) = \mathcal{L}(\mathbf{w}, \mathbf{X}_n) - \lambda \left(\sum_{m=1}^{3} w_m - 1 \right),$$

le Lagrangien associé à la maximisation de la log-vraisemblance et λ le multiplicateur de Lagrange. Nous remarquons que $w_1^\star \neq 0$ et $w_3^\star \neq 0$ car sinon la log-vraisemblance serait égale à $-\infty$. Si de plus

$w_2^\star \neq 0$ les conditions de premier ordre entrainent que

$$\left.\frac{\partial \mathcal{L}(\mathbf{w}, \mathbf{X}_n)}{\partial w_1}\right|_{\mathbf{w}^\star} = \left.\frac{\partial \mathcal{L}(\mathbf{w}, \mathbf{X}_n)}{\partial w_2}\right|_{\mathbf{w}^\star} = \left.\frac{\partial \mathcal{L}(\mathbf{w}, \mathbf{X}_n)}{\partial w_3}\right|_{\mathbf{w}^\star} = \lambda,$$

Un calcul direct implique que $w_2^\star = \frac{n_2^1 n_1^2 - n_1^1 n_2^2}{(n_1^1+n_2^1)(n_1^2+n_2^2)}$. Donc si $n_2^1 n_1^2 < n_1^1 n_2^2$, $w_2^\star$ est forcément égal à 0. Nous concluons d'après l'équation 3.3 que $\mathcal{C}(\mathbf{n}, \mathbf{B}) = \emptyset \Rightarrow w_2^\star = 0$. □

Dans l'Exemple 7, le critère de maximum de vraisemblance, appliqué à lois empiriques incompatibles, attribue une masse nulle à la région élémentaire $\mathcal{Q}_2^J$ alors qu'il est tout à fait plausible que les observations censurées par les régions $\mathcal{R}_2^1$ et $\mathcal{R}_1^2$ proviennent de cette région (puisque $\mathcal{Q}_2^J \subset \mathcal{R}_2^1$ et $\mathcal{Q}_2^J \subset \mathcal{R}_1^2$). L'incompatibilité des lois empiriques s'est traduite par une solution de l'ENPMV sur le bord du simplexe. Cela se produira dans le Chapitre 5 quand on appliquera l'ENPMV au jeu de données réelles où une grande partie des composantes w_m reçoit une masse nulle à cause de l'incompatibilité des lois empiriques (voir Figure 5.3a, page 112).

Le critère MaxEnt permet de dépasser ce problème. En effet, comme expliqué dans [Jay57], l'estimateur MaxEnt attribue une masse positive à tout événement qui n'est pas absolument exclu par les données. La section suivante présente le principe MaxEnt et quelques propriétés intéressantes de l'estimateur qui en résulte.

3.2 Estimation par MaxEnt

3.2.1 Principe de l'estimation d'une densité par MaxEnt

Étant donné un ensemble de contraintes que doit satisfaire la densité à estimer, le principe de l'estimation par MaxEnt suggère de choisir parmi toutes les densités qui répondent aux contraintes, celle d'entropie maximale. Ce choix s'explique par le fait que de toutes les distributions satisfaisant les contraintes, c'est celle d'entropie maximale qui apporte le moins d'information supplémentaire. Elle est donc pour cette raison la moins arbitraire de toutes celles que l'on pourrait choisir. Le principe d'estimation par MaxEnt a été introduit la première fois en 1957 par Jaynes [Jay57] en précisant que la maximisation de l'entropie en mécanique statistique utilisée par Boltzmann (1871) et Gibbs (1902) n'est qu'une application de ce principe général d'inférence issu de la théorie de l'information.

3.2.2 Quelques propriétés de l'estimateur MaxEnt

Notons $\mathscr{C} \subset \mathscr{P}$ l'ensemble des densités satisfaisant les contraintes de l'équation 3.1. C'est l'ensemble des densités *admissibles* pour le MaxEnt. Considérons H une fonction d'entropie. L'estimateur par MaxEnt $\hat{\pi}_{\mathscr{C}}^H$ est solution du problème $\mathrm{P}(H, \mathscr{C})$:

$$\mathrm{P}(H, \mathscr{C}) : \qquad \hat{\pi}_{\mathscr{C}}^H \stackrel{\Delta}{=} \underset{\pi \in \mathscr{C}}{\operatorname{argmax}} \; H(\pi). \tag{3.4}$$

On utilisera par la suite la notation $\mathrm{P}(H, \mathscr{C})$ pour désigner un problème de maximisation de la fonction H sous les contraintes $\mathscr{C}$. Souvent, les contraintes sont exprimées sous forme d'égalités portant sur les moments d'un ensemble de fonctions $\mathbf{g} = \{g_k\}_{k=1}^K$ où

$$g_k : \ \Theta \to \mathbb{R}, \qquad k = 1, \dots, K.$$

Les Lemmes 8 et 9 précisent la forme que prend l'estimateur MaxEnt dans des cas spécifiques.

Lemme 8. *[Jay57] Si $H = H_1$, l'entropie de Shannon, et si les contraintes sont des égalités portant sur les espérances des fonctions $\{g_k\}_{k=1}^K$ telles que l'ensemble des densités admissibles est*

$$\mathscr{C}^{eg}(\mathbf{g}, \mathbf{a}) \triangleq \{\pi \in \mathscr{P}; \ \mathbb{E}_\pi[g_k] = a_k, \ k = 1, \dots, K\}, \tag{3.5}$$

avec $\mathbf{a} = (a_1, \dots, a_K)^\mathsf{T}$ dans $\mathbb{R}^K$, alors l'estimateur MaxEnt $\hat{\pi}^{H_1}_{\mathscr{C}^{eg}(\mathbf{g},\mathbf{a})}$ appartient à la famille de Gibbs $Q(\mathbf{g})$, appelée aussi famille des densités exponentielles :

$$Q(\mathbf{g}) = \left\{\pi_{\mathbf{\Lambda},\mathbf{g}}; \ \Lambda \in \mathbb{R}^K\right\},$$

où

$$\pi_{\mathbf{\Lambda},\mathbf{g}}(\theta) \triangleq \frac{e^{\mathbf{\Lambda}.\mathbf{g}(\theta)}}{Z_{\mathbf{\Lambda},\mathbf{g}}}, \qquad \mathbf{\Lambda} \in \mathbb{R}^K, \tag{3.6}$$

avec $\mathbf{g}(\theta) = (g_1(\theta), \dots, g_K(\theta))^\mathsf{T}$ et $Z_{\mathbf{\Lambda},\mathbf{g}} \triangleq \int_\Theta e^{\mathbf{\Lambda}.\mathbf{g}(\theta)} d\theta$ la constante de normalisation. □

Lemme 9. *[GMZ09] Si $H = H_\alpha$, l'entropie de Rényi, avec α strictement positif et différent de 1, et si l'ensemble des densités admissibles est $\mathscr{C}^{eg}(\mathbf{g}, \mathbf{a})$ comme défini dans l'équation (3.5), alors l'estimateur MaxEnt $\hat{\pi}^{H_\alpha}_{\mathscr{C}^{eg}(\mathbf{g},\mathbf{a})}$ est de la forme*

$$\pi_\alpha^{\mathbf{\Lambda},\mathbf{g}}(\theta) \triangleq \frac{1}{\alpha}\left[\mathbf{\Lambda}.\bar{\mathbf{g}}(\theta)\right]_+^{\frac{1}{\alpha-1}}, \qquad \mathbf{\Lambda} \in \mathbb{R}^{K+1}, \tag{3.7}$$

avec

$$\int_\Theta \pi_\alpha^{\mathbf{\Lambda},\mathbf{g}}(\theta) d\theta = 1,$$

$[\cdot]_+ \equiv \max(\cdot, 0)$, $\bar{\mathbf{g}}(\theta) = (g_0(\theta), \dots, g_K(\theta))^\mathsf{T}$ *et* $g_0 \equiv 1$. □

Les Lemmes 8 et 9 sont des conséquences directes des conditions de premier ordre des Lagrangiens des problèmes $\mathrm{P}(H_1, \mathscr{C}^{eg}(\mathbf{g}, \mathbf{a}))$ et $\mathrm{P}(H_\alpha, \mathscr{C}^{eg}(\mathbf{g}, \mathbf{a}))$ où $\alpha \neq 1$.

Dans certains cas, l'estimation par MaxEnt est équivalente à l'estimation par maximum de vraisemblance dans certaines familles paramétriques. Cependant, les deux approches MaxEnt et maximum de vraisemblance différent par plusieurs aspects. Pour le maximum de vraisemblance paramétrique, nous supposons que la vraie distribution appartient à la famille paramétrique à travers laquelle la vraisemblance est maximisée, alors que le maximum d'entropie n'impose aucune hypothèse de ce genre. Le Théorème 7 ci-après, connu sous le nom de théorème de Boltzmann, souligne l'équivalence entre MaxEnt et maximum de vraisemblance paramétrique dans la famille de Gibbs.

Supposons observer directement $\theta_1, \dots, \theta_n \in \Theta$. Pour toute densité $\pi_{\mathbf{\Lambda},\mathbf{g}} \in Q(\mathbf{g})$, la vraisemblance des observations $\theta_1, \dots, \theta_n$ est

$$\mathscr{L}(\theta_1, \dots, \theta_n; \pi_{\mathbf{\Lambda},\mathbf{g}}) = \prod_{i=1}^{n} \pi_{\mathbf{\Lambda},\mathbf{g}}(\theta_i),$$

et la log-vraisemblance (normalisée) s'écrit

$$\begin{aligned} \mathcal{L}(\theta_1, \dots, \theta_n; \pi_{\mathbf{\Lambda},\mathbf{g}}) &= \mathbb{E}_{\tilde{\pi}} \left[\log \pi_{\mathbf{\Lambda},\mathbf{g}}\right] \\ &= -\frac{1}{n} \left(D_{KL}(\tilde{\pi} \| \pi_{\mathbf{\Lambda},\mathbf{g}}) - H_1(\tilde{\pi})\right). \end{aligned} \tag{3.8}$$

Le terme $H_1(\tilde{\pi})$ est une constante qui ne dépend pas de $\pi_{\mathbf{\Lambda},\mathbf{g}}$. Le problème de maximisation de la vraisemblance est donc équivalent à :

$$P1 : \quad \min_{\pi_{\mathbf{\Lambda},\mathbf{g}} \in \overline{Q}(\mathbf{g})} D_{KL}(\tilde{\pi} \| \pi_{\mathbf{\Lambda},\mathbf{g}}), \tag{3.9}$$

avec $\overline{Q}(\mathbf{g})$ la fermeture de $Q(\mathbf{g})$.

Théorème 7. *[DPDPL97](**Théorème de Boltzmann**) : Les deux problèmes d'optimisation $P(H_1, \mathscr{C}^{eg}(\mathbf{g}, \mathbf{a}))$ et $P1$, exprimés respectivement par les équations (3.4) et (3.9) sont équivalents.* □

Preuve : [DPDPL97] La fonction $\pi \mapsto D_{KL}(\tilde{\pi} \| \pi)$ est continue et strictement convexe, alors elle est bornée sur $\overline{Q}(\mathbf{g})$ et atteint ses bornes. Soit $\pi^\star$ son unique minimum sur $\overline{Q}(\mathbf{g})$. Puisque $\pi^\star \in \overline{Q}(\mathbf{g})$, alors pour tout $\mathbf{\Lambda} \in \mathbb{R}^M$, on a $\pi^\star \circ \pi_{\mathbf{\Lambda},\mathbf{g}} \in \overline{Q}(\mathbf{g})$, où

$$\pi^\star \circ \pi_{\mathbf{\Lambda},\mathbf{g}}(\theta) = \pi^\star(\theta) \frac{e^{\mathbf{\Lambda}.\mathbf{g}(\theta)}}{Z^{\pi^\star}_{\mathbf{\Lambda},\mathbf{g}}},$$

avec $Z^{\pi^\star}_{\mathbf{\Lambda},\mathbf{g}}$ la constante de normalisation. Par définition de $\pi^\star$, $\mathbf{\Lambda} = 0$ est un minimum de la fonction $\mathbf{\Lambda} \mapsto D_{KL}(\tilde{\pi} \| \pi^\star \circ \pi_{\mathbf{\Lambda},\mathbf{g}})$. Puisque

$$\left.\frac{d}{dt}\right|_{t=0} D_{KL}(\tilde{\pi} \| \pi \circ \pi_{t\mathbf{\Lambda},\mathbf{g}}) = \mathbf{\Lambda} \cdot (\mathbb{E}_{\tilde{\pi}} [\mathbf{g}] - \mathbb{E}_{\pi} [\mathbf{g}]),$$

alors $\mathbb{E}_{\tilde{\pi}} [\mathbf{g}] = \mathbb{E}_{\pi^\star} [\mathbf{g}]$ et donc $\pi^\star \in \mathcal{P}(\mathbf{g}, \tilde{\pi})$. Donc $\pi^\star \in \mathcal{P}(\mathbf{g}, \tilde{\pi}) \cap \overline{Q}(\mathbf{g})$. Soient $\pi^1, \pi^2, \pi^3 \in \mathscr{P}$ et $\mathbf{\Lambda} \in \mathbb{R}^M$. Un simple calcul montre que

$$D_{KL}(\pi^1 \| \pi^2) - D_{KL}(\pi^1 \| \pi^2 \circ \pi_{\mathbf{\Lambda},\mathbf{g}}) - D_{KL}(\pi^3 \| \pi^2) + D_{KL}(q_3 \| \pi^2 \circ \pi_{\mathbf{\Lambda},\mathbf{g}}) = \mathbf{\Lambda} \cdot (\mathbb{E}_{\pi^1} [\mathbf{g}] - \mathbb{E}_{\pi^3} [\mathbf{g}]).$$

Donc pour tout π^1 dans $\mathcal{P}(\mathbf{g}, \tilde{\pi})$ et tout π^2 dans $\overline{Q}(\mathbf{g})$, on a

$$D_{KL}(\pi^1 \| \pi^2) = D_{KL}(\pi^1 \| \pi^\star) + D_{KL}(\pi^\star \| \pi^2).$$

Puisque $(\pi^1, \pi^2) \mapsto D_{KL}(\pi^1 \| \pi^2)$ est une fonction positive sur $\mathscr{P} \times \mathscr{P}$, alors pour tout $\pi^1 \in \mathcal{P}(\mathbf{g}, \tilde{\pi})$ et $\pi^2 \in \overline{Q}(\mathbf{g})$

$$D_{KL}(\pi^1 \| \pi_0) = D_{KL}(\pi^1 \| \pi^\star) + D_{KL}(\pi^\star \| \pi_0) \geqslant D_{KL}(\pi^\star \| \pi_0),$$

et

$$D_{KL}(\tilde{\pi} \| \pi^2) = D_{KL}(\tilde{\pi} \| \pi^\star) + D_{KL}(\pi^\star \| \pi_2) \geqslant D_{KL}(\tilde{\pi} \| \pi^\star).$$

Donc

$$\pi^\star = \underset{\pi \in \mathcal{P}(\mathbf{g}, \tilde{\pi})}{\operatorname{argmin}} D_{KL}(\pi \| \pi_0) = \underset{\pi_{\mathbf{\Lambda}, \mathbf{g}} \in \overline{Q}(\mathbf{g})}{\operatorname{argmin}} D_{KL}(\tilde{\pi} \| \pi_{\mathbf{\Lambda}, \mathbf{g}}).$$

Les problèmes de MaxEnt avec entropie de Shannon et de Maximum de vraisemblance paramétrique dans la famille de Gibbs sont équivalents dans le cas précis où l'observation est complète. □

En général, ce théorème n'est plus valable quand on n'observe pas directement les θ_i, mais uniquement des versions censurées des ces réalisations. En effet, en comparant les fonctions de la log-vraisemblance dans les cas d'observation complète (équation (3.8)) et observation avec censure (équation (2.10)), nous remarquons que la divergence de Kullback-Leibler devient une pondération de plusieurs divergences dans le cas d'observations censurées. Nous présenterons dans le paragraphe 3.3.2, un cas spécial de données censurées où ce théorème reste valable. Dans la section suivante, nous présentons l'ensemble des contraintes déterminées par les données qui seront utilisées pour définir l'estimateur MaxEnt.

3.3 Estimateur MaxEnt pour observations censurées par des éléments d'un ensemble fini de partitions

Dans notre problème les données sont censurées par les éléments des partitions $\left\{\mathcal{R}^j\right\}_{j=1}^J$ selon le mécanisme expliqué dans le Chapitre 2, page 8, et peuvent être considérées comme des réalisations de J lois multinomiales : $\left\{p_{\pi_\theta, \mathcal{R}^j}\right\}_{j=1}^J$ (Définition 1, page 10). En posant

$$g_\ell^j(\cdot) \equiv \mathbb{1}_{\mathcal{R}_\ell^j}, \qquad \forall j = 1, \ldots, J, \forall \ell = 1, \ldots, L^j,$$

nous pouvons déduire des observations un ensemble de contraintes que doit vérifier la densité estimée π :

$$\mathbb{E}_\pi \left[g_\ell^j\right] = \mathbb{E}_{\tilde{\pi}} \left[g_\ell^j\right] = \frac{n_\ell^j}{n^j}, \qquad \forall j = 1, \ldots, J, \forall \ell = 1, \ldots, L^j.$$

Le fait que $\mathcal{R}^j = \left\{\mathcal{R}_\ell^j\right\}_{\ell=1}^{L^j}$ soit une partition de Θ pour tout $j = 1, \ldots, J$ implique que $1 = \sum_{\ell=1}^{L^j} g_\ell^j$. Donc les fonctions définissant les contraintes ne sont pas linéairement indépendantes. Il suffit alors d'imposer $L^j - 1$ contraintes parmi les L^j possibles. Nous choisissons sans perte de généralité les

$L^j - 1$ premières contraintes. Nous notons

$$\check{\mathbf{g}} = \left\{ g_1^1, \ldots, g_{L^1-1}^1, \ldots, g_1^J, \ldots, g_{L^J-1}^J \right\},$$

et

$$\check{\mathbf{q}}(\mathbf{n}) = \left\{ q_1^1, \ldots, q_{L^1-1}^1, \ldots, q_1^J, \ldots, q_{L^J-1}^J \right\}.$$

Ainsi, l'ensemble des densités admissibles est

$$\mathscr{C}^{eg}(\check{\mathbf{g}}, \check{\mathbf{q}}(\mathbf{n})) = \{\pi \in \mathscr{P};\ \mathbb{E}_\pi[\check{\mathbf{g}}] = \mathbb{E}_{\tilde{\pi}}[\check{\mathbf{g}}]\}. \tag{3.10}$$

3.3.1 Relaxation des contraintes

Nous avons déjà vu (section 3.1) que l'ensemble des distributions admissibles peut être vide. Nous élargissons cet ensemble en remplaçant les contraintes d'égalité par des contraintes d'inégalité. C'est la *Relaxation des contraintes*. Des travaux précédents se sont intéressés au problème de MaxEnt avec des contraintes relaxées. Chen, Rosenfeld [CR00], Kazama et Tsuji [KT03] ont démontré l'équivalence entre MaxEnt sous contraintes relaxées et maximum de vraisemblance régularisé dans certains contextes spécifiques. Dudík et *al.* [DPS07] montrent dans un cas précis que le résultat de l'approche MaxEnt avec des contraintes relaxées est presque aussi performant en terme de vraisemblance que la meilleure solution possible.

La relaxation des contraintes conduit à des inégalités du type

$$-a_\ell^j \leqslant \mathbb{E}_\pi\left[g_\ell^j\right] - \mathbb{E}_{\tilde{\pi}}\left[g_\ell^j\right] \leqslant b_\ell^j, \qquad \forall j = 1, \ldots, J, \forall \ell = 1, \ldots, L^j - 1,$$

avec $a_\ell^j \geqslant 0$ et $b_\ell^j \geqslant 0$ pour $j = 1, \ldots, J$ et $\ell = 1, \ldots, L^j - 1$. Plus les largeurs $a_\ell^j + b_\ell^j$ sont grands, plus l'ensemble des densités admissibles défini par

$$\mathscr{C}(\check{\mathbf{g}}, \check{\mathbf{a}}, \check{\mathbf{b}}) \triangleq \left\{ \pi \in \mathscr{P};\ -\check{\mathbf{a}} \leqslant \mathbb{E}_\pi[\check{\mathbf{g}}] - \mathbb{E}_{\tilde{\pi}}[\check{\mathbf{g}}] \leqslant \check{\mathbf{b}} \right\}, \tag{3.11}$$

avec

$$\check{\mathbf{a}} = \left\{ a_1^1, \ldots, a_{L^1-1}^1, \ldots, a_1^J, \ldots, a_{L^J-1}^J \right\}\ ,\ \check{\mathbf{b}} = \left\{ b_1^1, \ldots, b_{L^1-1}^1, \ldots, b_1^J, \ldots, b_{L^J-1}^J \right\},$$

devient grand. Les auteurs de [KT03] démontrent en dérivant la condition de premier ordre que si l'entropie à maximiser est celle de Shannon, alors l'estimateur MaxEnt est une densité $\pi_{\boldsymbol{\Lambda}, \check{\mathbf{g}}}$ (équation (3.6)) dans la famille de Gibbs où $\boldsymbol{\Lambda} \in \mathbb{R}^{K-J}$. Si l'on maximise l'entropie de Rényi d'ordre α, l'estimateur MaxEnt prend la forme d'une densité $\pi_\alpha^{\boldsymbol{\Lambda}, \check{\mathbf{g}}}$ (équation (3.7)) telle que $\boldsymbol{\Lambda} \in \mathbb{R}^{K-J+1}$. Cela découle immédiatement en dérivant le Lagrangien associé à ce problème de l'optimisation.

A partir des définitions de $\pi_{\boldsymbol{\Lambda}, \check{\mathbf{g}}}$ et de $\pi_\alpha^{\boldsymbol{\Lambda}, \check{\mathbf{g}}}$, et puisque $g_\ell^j(\cdot) \equiv \mathbb{1}_{\mathcal{R}_\ell^j}$ pour tout $j = 1, \ldots, J$ et tout $\ell = 1, \ldots, L^j - 1$, nous déduisons que l'estimateur MaxEnt, $\hat{\pi}^{H_\alpha}_{\mathscr{C}(\check{\mathbf{g}}, \check{\mathbf{a}}, \check{\mathbf{b}})}$, solution du problème

P $\left(H_\alpha, \mathscr{C}(\check{\mathbf{g}}, \check{\mathbf{a}}, \check{\mathbf{b}})\right)$ avec $\alpha > 0$, attribue la même masse de probabilité pour les points d'une même région élémentaire $\mathcal{Q}_m^J, \forall m \in \{1, \ldots, M\}$. Cet estimateur est donc complètement identifié en connaissant la loi induite par $\hat{\pi}^{H_\alpha}_{\mathscr{C}(\check{\mathbf{g}},\check{\mathbf{a}},\check{\mathbf{b}})}$ sur $\mathcal{Q}^J$, et le problème de MaxEnt devient donc un problème discret. Il suffit de distribuer uniformément sur une région élémentaire $\mathcal{Q}_m^J$ sa masse de probabilité. Ainsi nous obtenons pour $\alpha > 0$

$$\hat{\pi}^{H_\alpha}_{\mathscr{C}(\check{\mathbf{g}},\check{\mathbf{a}},\check{\mathbf{b}})}(\theta) = \frac{p_{\hat{\pi}^{H_\alpha}_{\mathscr{C}(\check{\mathbf{g}},\check{\mathbf{a}},\check{\mathbf{b}})}, \mathcal{Q}_m^J}}{\nu(\mathcal{Q}_m^J)}, \quad \forall \theta \in \mathcal{Q}_m^J, \quad \forall m \in \{1, \ldots, M\}. \tag{3.12}$$

Nous nous limitons donc à l'ensemble des vecteurs de probabilité $\mathbf{w} = (w_1, \ldots, w_M)^\mathsf{T} \in \mathcal{S}^{M+}$ tels que $\pi_\mathbf{w} \in \mathscr{C}(\check{\mathbf{g}}, \check{\mathbf{a}}, \check{\mathbf{b}})$ où

$$\pi_\mathbf{w} = \sum_{m=1}^{M} w_m \delta_{\mathcal{Q}_m^J},$$

et $\delta_{\mathcal{Q}_m^J}$ une mesure de probabilité attribuant toute la masse de probabilité à la région élémentaire $\mathcal{Q}_m^J$. D'après les définitions des entropies de Shannon et de Rényi, nous avons pour des densités de mélange $\pi_\mathbf{w}$

$$H_1(\pi_\mathbf{w}) = -\sum_{m=1}^{M} w_m \log \frac{w_m}{\nu(\mathcal{Q}_m^J)},$$

et

$$H_\alpha(\pi_\mathbf{w}) = \frac{1}{1-\alpha} \log \int_\Theta \pi_\mathbf{w}^\alpha(\theta) d\theta = \frac{1}{1-\alpha} \log \sum_{m=1}^{M} \frac{w_m^\alpha}{(\nu(\mathcal{Q}_m^J))^{\alpha-1}}.$$

Le choix de l'entropie de Rényi d'ordre $\alpha = 2$ permettra de déterminer l'estimateur de π_θ en résolvant un problème quadratique.

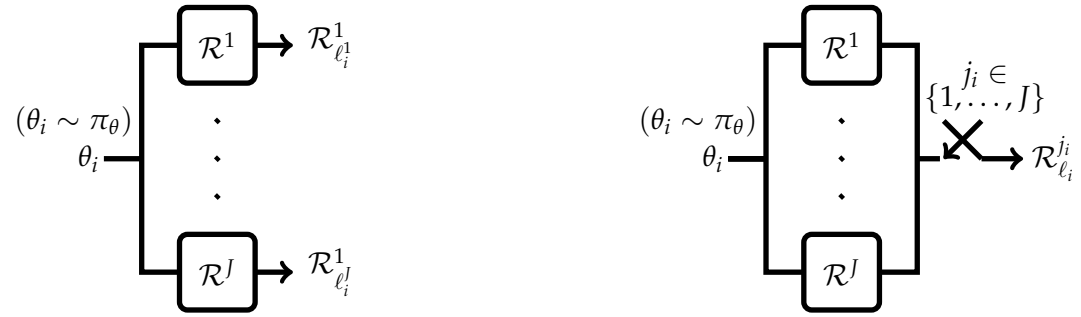
Le paragraphe suivant met en évidence un modèle d'observation différent de celui considéré dans notre étude, et qui permet d'établir des liens forts entre MaxEnt régularisé et maximum de vraisemblance dans la famille de Gibbs.

3.3.2 Observations exhaustives

Supposons dans ce paragraphe[2] que pour toutes les réalisations $\theta_i \in \Theta$, nous disposons des J régions de censure correspondant aux éléments des partitions $\{\mathcal{R}^j\}_{j=1}^J$ auxquels appartient θ_i. La Figure 3.1 montre les deux modèles d'observation : la Figure 3.1a motre le modèle d'observation spécifique à l'observation exhaustive et la Figure 3.1b montre celui considéré dans le cadre de cette thèse.

Nous appelons ces observations *observations exhaustives* dans le sens où, étant donnée une liste de J partitions, pour tout individu i nous disposons des J éléments dans les J partitions, qui contiennent

2. Les notations introduites dans ce paragraphes sont propres à ce paragraphe.



(a) *Modèle d'observation exhaustive.* **(b)** *Modèle d'observation général.*

FIGURE 3.1 – *Différence entre les deux modèles d'observation, celui des observations exhaustives (à gauche) et celui considéré dans le cadre de cette thèse (à droite).*

la valeur de son paramètre θ_i. Les observations issues de ce nouveau mécanisme sont aussi des observations censurées par régions.

Pour $i = 1, \ldots, n$, nous notons $\mathbf{x}_i = \left(x_i^1, \ldots, x_i^J\right)^\mathsf{T}$ le vecteur de taille J tel que

$$x_i^j = \sum_{\ell=1}^{L^j} \ell \mathbb{1}_{\left\{\theta_i \in \mathcal{R}_\ell^j\right\}}, \qquad j = 1, \ldots, J.$$

La vraisemblance d'une densité $\pi \in \mathscr{P}$ dans ce cas s'écrit

$$\mathscr{L}(\mathbf{x}_1, \ldots, \mathbf{x}_n; \pi) = \prod_{i=1}^{n} \pi \left(\bigcap_{j=1}^{J} \mathcal{R}^j_{x_i^j} \right) = \prod_{i=1}^{n} \pi \left(\mathcal{A}_i \right).$$

avec $\mathcal{A}_i = \bigcap_{j=1}^{J} \mathcal{R}^j_{x_i^j}$. Nous remarquons que pour tout $i = 1, \ldots, n$, $\mathcal{A}_i \in \mathcal{Q}^J$ et donc pour tout i, i' dans $\{1, \ldots, n\}$, ou bien $\mathcal{A}_i = \mathcal{A}_{i'}$, ou bien $\mathcal{A}_i \cap \mathcal{A}_{i'} = \emptyset$. Nous remarquons aussi que le support de la densité maximisant la vraisemblance est forcément contenu dans $\bigcup_{i=1}^{n} \mathcal{A}_i$ et que la vraisemblance ne dépend que des masses attribuées aux régions $\mathcal{A}_i$.

Soient n' le nombre des régions $\mathcal{A}_i$ uniques et n_i le nombre de répétition de chaque région $\mathcal{A}_i$ de façon à ce que $\sum_{i=1}^{n'} n_i = n$. La log-vraisemblance (normalisée) est dans ce cas

$$\mathcal{L}(\mathbf{x}_1, \ldots, \mathbf{x}_n; \pi) = \sum_{i=1}^{n'} \frac{n_i}{n} \log \pi \left(\mathcal{A}_i \right),$$

avec $\mathcal{A} = \bigcup_{i=1}^{n'} \mathcal{A}_i$. La distribution empirique $\tilde{\pi}$ est définie sur $\mathcal{A}$ telle que

$$\tilde{\pi}\left(\theta\right) = \sum_{i=1}^{n'} \frac{n_i}{n} \mathbb{1}_{\{\theta \in \mathcal{A}_i\}}.$$

En posant pour tout $\theta \in \Theta$, $\mathbf{g}(\theta) = (g_1(\theta), \ldots, g_{n'}(\theta))^\mathsf{T} = (\mathbb{1}_{\mathcal{A}_1}(\theta), \ldots, \mathbb{1}_{\mathcal{A}_{n'}}(\theta))^\mathsf{T}$ et en se limitant aux densités de la famille de Gibbs $Q(\mathbf{g})$, la log-vraisemblance s'écrit

$$\mathcal{L}(\mathbf{x}_1, \ldots, \mathbf{x}_n; \pi_{\boldsymbol{\Lambda},\mathbf{g}}) = \sum_{i=1}^{n'} \frac{n_i}{n} \log \int_{\mathcal{A}_i} \frac{e^{\boldsymbol{\Lambda}_i}}{\mathbf{Z}_{\boldsymbol{\Lambda},\mathbf{g}}} d\theta = \mathbb{E}_{\tilde{\pi}} \left[\log \pi_{\boldsymbol{\Lambda},\mathbf{g}}\right]. \tag{3.13}$$

Le Théorème 7 affirme que l'estimateur de maximum de vraisemblance coïncide avec l'estimateur MaxEnt, solution du problème

$$\mathrm{P}\left(H_1, \mathscr{C}^{eg}\left(\mathbf{g}, \mathbb{E}_{\tilde{\pi}}\left[\mathbf{g}\right]\right)\right).$$

L'ensemble $\mathscr{C}^{eg}\left(\mathbf{g}, \mathbb{E}_{\tilde{\pi}}\left[\mathbf{g}\right]\right)$ des estimateurs admissibles avec des contraintes d'égalité peut être vide. Dans ces cas il peut être élargi à l'ensemble $\mathscr{C}\left(\mathbf{g}, -\boldsymbol{\varepsilon}, \boldsymbol{\varepsilon}\right)$ avec $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_{n'})^\mathsf{T}$ un vecteur de réels positifs. Kazama et Tsuji [KT03] ont démontré l'équivalence entre MaxEnt avec contraintes relaxées et maximum de vraisemblance avec régularisation de type ℓ_1, dite aussi régularisation Lasso, quand H est l'entropie de Shannon.

Définition 12. *(**régularisation de type** ℓ_1) : L'estimateur par maximum de vraisemblance paramétrique dans la famille $Q(\mathbf{g})$ avec régularisation de type ℓ_1 est la solution de problème*

$$P2: \quad \max_{\boldsymbol{\Lambda} \in \mathbb{R}^{n'}} \left(\mathcal{L}(\mathbf{x}_1, \ldots, \mathbf{x}_n, \pi_{\boldsymbol{\Lambda},\mathbf{g}}) - \sum_{i=1}^{n'} \varepsilon_i |\boldsymbol{\Lambda}_i| \right).$$

avec $\varepsilon_i > 0$ *pour tout* $i = 1, \ldots, n'$. □

Le Lagrangien du problème $\mathrm{P}\left(H_1, \mathscr{C}\left(\mathbf{g}, -\boldsymbol{\varepsilon}, \boldsymbol{\varepsilon}\right)\right)$ s'écrit

$$\begin{aligned} \mathrm{L}^{H_1}_{\mathscr{C}(\mathbf{g},-\boldsymbol{\varepsilon},\boldsymbol{\varepsilon})}(\pi, \boldsymbol{\lambda}) =& H_1(\pi) - \lambda_0 \left(\int_{\mathcal{A}} \pi(\theta) d\theta - 1 \right) + \\ & \sum_{i=1}^{n'} \left(\lambda_i^+ - \lambda_i^-\right) \left(\mathbb{E}_\pi\left[g_i\right] - \mathbb{E}_{\tilde{\pi}}\left[g_i\right]\right) + \sum_{i=1}^{n'} \varepsilon_i \left(\lambda_i^+ + \lambda_i^-\right), \end{aligned}$$

où $\boldsymbol{\lambda} = \left(\lambda_0, \lambda_1^+, \ldots, \lambda_{n'}^+, \lambda_1^-, \ldots, \lambda_{n'}^-\right)^\mathsf{T}$. En notant $\boldsymbol{\Lambda}_i = \lambda_i^+ - \lambda_i^-$ et $\boldsymbol{\Lambda} = (\boldsymbol{\Lambda}_1, \ldots, \boldsymbol{\Lambda}_{n'})^\mathsf{T}$, on a

$$\begin{aligned} \mathrm{L}^{H_1}_{\mathscr{C}(\mathbf{g},-\boldsymbol{\varepsilon},\boldsymbol{\varepsilon})}(\pi, \lambda_0, \boldsymbol{\Lambda}) =& H_1(\pi) - \lambda_0 \left(\int_{\mathcal{A}} \pi(\theta) d\theta - 1 \right) + \\ & \sum_{i=1}^{n'} \boldsymbol{\Lambda}_i \left(\mathbb{E}_\pi\left[g_i\right] - \mathbb{E}_{\tilde{\pi}}\left[g_i\right]\right) + \sum_{i=1}^{n'} \varepsilon_i |\boldsymbol{\Lambda}_i|. \end{aligned}$$

Cette expression est différentiable et concave en $\pi(\theta)$. En contraignant la dérivée partielle de $\mathrm{L}^{H_1}_{\mathscr{C}(\mathbf{g},-\boldsymbol{\varepsilon},\boldsymbol{\varepsilon})}$ par rapport à π à être nulle, le problème devient

$$\min_{\boldsymbol{\Lambda} \in \mathbb{R}^{n'}} \left[H_1(\pi_{\boldsymbol{\Lambda},\mathbf{g}}) + \boldsymbol{\Lambda}. \left(\mathbb{E}_{\pi_{\boldsymbol{\Lambda},\mathbf{g}}}\left[\mathbf{g}\right] - \mathbb{E}_{\tilde{\pi}}\left[\mathbf{g}\right] \right) + \sum_{i=1}^{n'} \varepsilon_i |\boldsymbol{\Lambda}_i| \right].$$

Puisque $H_1(\pi_{\mathbf{\Lambda},\mathbf{g}}) = -\int_{\mathcal{A}} \pi_{\mathbf{\Lambda},\mathbf{g}}(\theta)\log\pi_{\mathbf{\Lambda},\mathbf{g}}(\theta)d\theta = -\mathbf{\Lambda}.\mathbb{E}_{\pi_{\mathbf{\Lambda},\mathbf{g}}}[\mathbf{g}] + \log Z_{\mathbf{\Lambda},\mathbf{g}}$, et que d'après l'équation (3.13) on a

$$\mathcal{L}\left(\mathbf{x}_1,\ldots,\mathbf{x}_n;\pi_{\mathbf{\Lambda},\mathbf{g}}\right) = \mathbf{\Lambda}.\mathbb{E}_{\tilde{\pi}}[\mathbf{g}] - \log Z_{\mathbf{\Lambda},\mathbf{g}} \ , \tag{3.14}$$

ce qui établit le Lemme 10.

Lemme 10. *[KT03]* $\quad P2 \Leftrightarrow P\left(H_1, \mathscr{C}\left(\mathbf{g}, -\boldsymbol{\varepsilon}, \boldsymbol{\varepsilon}\right)\right)$. □

Dudíck remarque dans sa thèse [Dud07] le Lemme suivant.

Lemme 11. *[Dud07] Pour tout $\pi \in \mathscr{P}$ on a*

$$\left|\mathbb{E}_{\tilde{\pi}}\left[\log\pi_{\mathbf{\Lambda},\mathbf{g}}\right] - \mathbb{E}_{\pi}\left[\log\pi_{\mathbf{\Lambda},\mathbf{g}}\right]\right| \leqslant \sum_{i=1}^{n'} |\mathbf{\Lambda}_i| \left|\mathbb{E}_{\tilde{\pi}}[g_i] - E_{\pi_{\mathbf{\Lambda},\mathbf{g}}}[g_i]\right|,$$

et on déduit que pour toute distribution $\pi \in \mathscr{C}\left(\mathbf{g}, -\boldsymbol{\varepsilon}, \boldsymbol{\varepsilon}\right)$

$$\mathbb{E}_{\pi}\left[\log\pi_{\mathbf{\Lambda},\mathbf{g}}\right] - \mathbb{E}_{\pi}\left[\log\pi_{\mathbf{\Lambda}^\star,\mathbf{g}}\right] \leqslant 2\sum_{i=1}^{n'} \varepsilon_i|\mathbf{\Lambda}_i|,$$

où $\mathbf{\Lambda}^\star$ est la solution du programme P2. □

Ce Lemme montre que l'écart en termes de vraisemblance entre deux éléments de l'ensemble $\mathscr{C}\left(\mathbf{g}, -\boldsymbol{\varepsilon}, \boldsymbol{\varepsilon}\right)$ des densités admissibles est majoré par une fonction du paramètre $\boldsymbol{\varepsilon}$. Les résultats cités dans ce paragraphe concernent le cas de l'observation exhaustive.

Les données que nous utilisons dans le cadre de cette étude sont issues d'un modèle d'observation plus général représenté dans la Figure 3.1b. En effet, nous ne disposons pas d'informations indiquant à quel individu les observations sont relatives, en plus du fait que nous n'avons pas la garantie que les observations de tous les individus ont été censurées par la même liste de partitions. Dans la section 3.4, nous proposons une solution qui prend en compte l'incertitude autour des lois empiriques dans la détermination de l'ensemble des densités admissibles et fait appel au critère de vraisemblance pour choisir le niveau de relaxation le plus adéquat.

3.4 Nouveau critère

3.4.1 Relaxation dépendant des lois empiriques $\left\{\mathbf{q}_j(\mathbf{n})\right\}_{j=1}^{J}$

Pour tout $j = 1,\ldots,J$, $\mathbf{q}_j(\mathbf{n})$ est la réalisation d'une loi multinomiale prenant L^j modalités. Notons pour tout $j \in \{1,\ldots,J\}$ et tout $\ell \in \left\{1,\ldots,L^j\right\}$,

$$\mathbb{V}\left[q_\ell^j\right] = \mathbb{V}\left[\frac{n_\ell^j}{n^j}\right] \triangleq \frac{n_\ell^j(n^j - n_\ell^j)}{\left(n^j\right)^3},$$

la variance de l'estimateur empirique $q_\ell^j = \frac{n_\ell^j}{n^j}$ de $p_{\pi_\theta,\mathcal{R}^j}(\ell)$ (voir la Définition 1, page 10). Soient j_1 et j_2 les indices de deux partitions tels que $n^{j_1} \gg n^{j_2}$, c'est à dire que les observations censurées par la partition $\mathcal{R}^{j_1}$ sont beaucoup plus nombreuses que celles censurées par la partition $\mathcal{R}^{j_2}$. Nous avons pour tout $\ell_1 \in \{1,\dots,L^{j_1}\}$ et $\ell_2 \in \{1,\dots,L^{j_2}\}$

$$\mathbb{V}\left[q_\ell^{j_1}\right] = \frac{n_\ell^{j_1}(n^{j_1}-n_\ell^{j_1})}{\left(n^{j_1}\right)^3} \ll \frac{n_{\ell_2}^{j_2}(n^{j_2}-n_{\ell_2}^{j_2})}{\left(n^{j_2}\right)^3} = \mathbb{V}\left[q_{\ell_2}^{j_2}\right],$$

c'est à dire que l'estimation empirique $\mathbf{q}_{j_1}(\mathbf{n}) = \left(q_1^{j_1},\dots,q_{L^{j_1}}^{j_1}\right)^\mathsf{T}$ de la loi multinomiale $p_{\pi_\theta,\mathcal{R}^{j_1}}$ est plus précise que l'estimation empirique $\mathbf{q}_{j_2}(\mathbf{n})$ de $p_{\pi_\theta,\mathcal{R}^{j_2}}$. L'élargissement de l'ensemble des densités admissibles pour le MaxEnt par relaxation des contraintes doit tenir en compte de l'incertitude associée $\left\{\mathbf{q}_j(\mathbf{n})\right\}_{j=1}^J$. Nous notons $\Sigma^{(j)}$ pour $j = 1,\dots,J$, la matrice de covariance de l'estimateur empirique $\left(\mathbb{E}_{\tilde\pi}\left[g_1^j\right],\dots,\mathbb{E}_{\tilde\pi}\left[g_{L^j-1}^j\right]\right)^\mathsf{T}$. En prenant en compte les incertitudes exprimées par les matrices $\left\{\Sigma^{(j)}\right\}_{j=1}^J$, nous définissons l'ensemble des candidats au MaxEnt pour un niveau de relaxation $\epsilon \geqslant 0$ donné Afin de simplifier les notations, nous notons $\mathbb{E}_{\pi-\tilde\pi}[\cdot] = \mathbb{E}_\pi[\cdot] - \mathbb{E}_{\tilde\pi}[\cdot]$.

Définition 13. *(**ensemble des candidats au MaxEnt**) : L'ensemble des candidats au MaxEnt avec relaxation d'un niveau égal à $\epsilon \geqslant 0$ est*

$$\mathscr{C}(\check{\mathbf{q}},\epsilon) \triangleq \Big\{\pi \in \mathscr{P};\ \forall j \in \{1,\dots,J\} : \left\|\Sigma^{(j)^{-\frac{1}{2}}}\left(\mathbb{E}_{\pi-\tilde\pi}\left[g_1^j\right],\dots,\mathbb{E}_{\pi-\tilde\pi}\left[g_{L^j-1}^j\right]\right)^\mathsf{T}\right\|_\infty \leqslant \epsilon\Big\}. \tag{3.15}$$

□

Nous définissons le niveau de relaxation minimale associé à l'estimateur MaxEnt comme suit

Définition 14. *(**niveau de relaxation minimale**) : Le niveau de relaxation minimale, noté $\epsilon_\star$ est le plus petit réel positif assurant que l'ensemble des candidats au MaxEnt est non-vide*

$$\epsilon_\star \triangleq \min_{\epsilon\geqslant 0}\{\epsilon;\ \mathscr{C}(\check{\mathbf{q}},\epsilon) \neq \emptyset\}. \tag{3.16}$$

□

Définition 15. *(**estimateur MaxEnt avec relaxation**) : L'estimateur MaxEnt avec niveau de relaxation $\epsilon \geqslant \epsilon_\star$ est solution du problème $P(H,\mathscr{C}(\check{\mathbf{q}},\epsilon))$*

$$\hat\pi^H_{\mathscr{C}(\check{\mathbf{q}},\epsilon)} \triangleq \operatorname*{argmax}_{\pi\in\mathscr{C}(\check{\mathbf{q}},\epsilon)} H(\pi). \tag{3.17}$$

Nous notons $\hat\pi^H_\epsilon$ au lieu de $\hat\pi^H_{\mathscr{C}(\check{\mathbf{q}},\epsilon)}$ pour alléger les notations. □

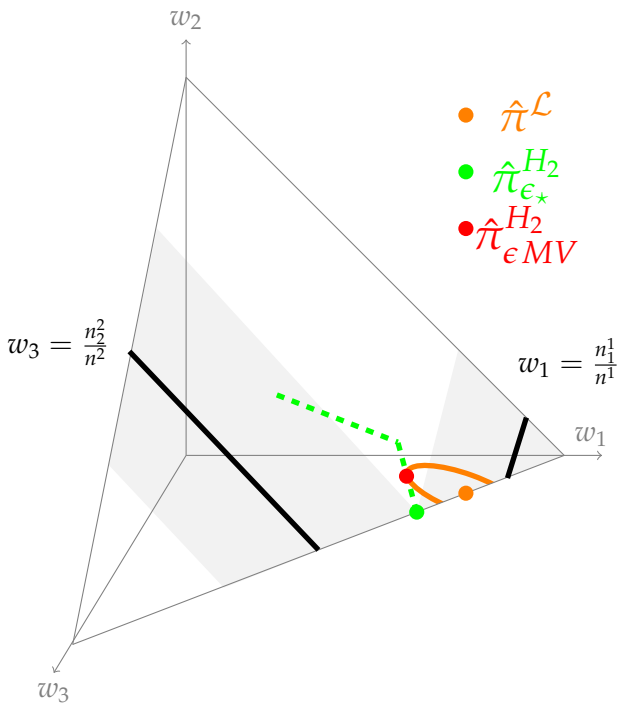


FIGURE 3.1 – *Illustration des 3 estimateurs $\hat{\pi}^{\mathcal{L}}$, $\hat{\pi}^{H_2}_{\epsilon_\star}$ et $\hat{\pi}^{H_2}_{\epsilon^{MV}}$ dans l'Exemple 7.*

3.4.2 Estimateur MaxEnt-MV

Dans le paragraphe précédent, nous avons défini, pour chaque niveau de relaxation ϵ, l'estimateur MaxEnt correspondant. La question qui se pose est comment choisir ϵ ? Nous posons

$$\epsilon^\star \stackrel{\Delta}{=} \min_{\epsilon \geqslant \epsilon_\star} \left\{ \epsilon : \ \hat{\pi}^H_\epsilon = \hat{\pi}^H_{\mathscr{P}} \right\},$$

où $\hat{\pi}^H_{\mathscr{P}}$ est la distribution dans $\mathscr{P}$ maximisant l'entropie H. Nous nous intéressons donc aux valeurs de ϵ dans l'intervalle $[\epsilon_\star, \epsilon^\star]$. La définition suivante définit l'estimateur proposé dans cette thèse.

Définition 16. ***(Estimateur MaxEnt-MV)*** *: L'estimateur MaxEnt-MV, $\hat{\pi}^H_{\epsilon^{MV}}$, est l'estimateur MaxEnt correspondant au niveau de relaxation ϵ^{MV} tel que*

$$\epsilon^{MV} \stackrel{\Delta}{=} \operatorname*{argmax}_{\epsilon \in [\epsilon_\star, \epsilon^\star]} \mathscr{L}\left(\hat{\pi}^H_\epsilon, \mathbf{X}_n\right), \tag{3.18}$$

le niveau de relaxation correspondant à l'estimateur MaxEnt ayant la vraisemblance la plus élevée □

Le paragraphe 3.4.3 présente le résultat de l'estimation appliquée à l'Exemple 7, page 58.

3.4.3 Résultat de l'estimation dans l'Exemple 7

Nous reprenons l'Exemple 7 pour donner une illustration géométrique de l'estimateur proposé. Nous avons $M = 3$ ce qui permet de représenter les estimateurs dans le simplexe $\mathcal{S}^{3+}$ (Figure 3.1). Les nombres d'observation des différentes régions, n_1^1, n_2^1, n_1^2 et n_2^2, sont choisis tels que les lois empiriques ne sont pas compatibles (Définition 11), c'est à dire $n_1^1 n_2^2 > n_2^1 n_1^2$ comme nous l'avons déjà vu (équation (3.3)). Le fait que $\mathcal{C}(\mathbf{n}, \mathbf{B}) = \emptyset$ évite le cas trivial où les estimateurs maximum de vraisemblance et MaxEnt-MV coïncident.

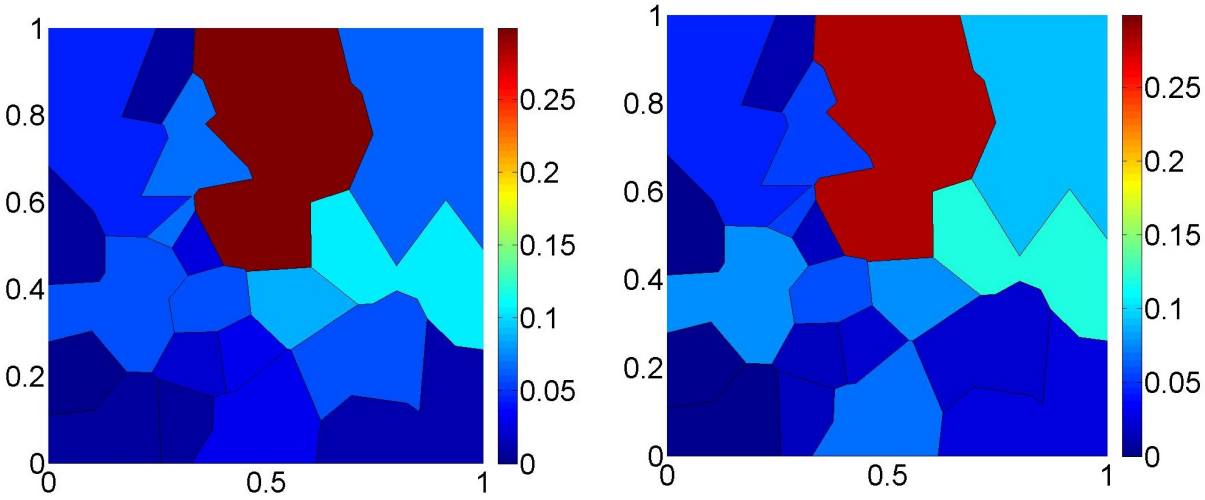
La Figure 3.1 donne une illustration graphique de cet exemple. Les lignes noires correspondent aux contraintes d'égalité (équation (3.10)). L'incompatibilité de ces contraintes se traduit graphiquement par le fait que ces deux lignes ne se croisent pas. Comme déjà mentionné dans l'Exemple 7, l'ENPMV $\hat{\pi}^{\mathcal{L}}$, représenté par le point orange est unique et sa deuxième composante est nulle. Les deux régions en gris autour de chacune des lignes noires correspondent à la relaxation minimale des contraintes, avec le niveau $\epsilon_\star$. L'ensemble $\mathscr{C}(\check{\mathbf{q}}, \epsilon_\star)$ contient un seul point (point en vert sur la Figure) correspondant à l'estimateur $\hat{\pi}^{H_2}_{\epsilon_\star}$. La ligne verte pointillée correspond aux estimateurs MaxEnt avec relaxation des contraintes pour des valeurs croissantes du ϵ jusqu'à atteindre le maximum global de l'entropie. La courbe quadratique (en orange) est à la première courbe de niveau de la vraisemblance qui se croise avec la ligne verte pointillée. Le point d'intersection de ces deux courbes est l'estimateur MaxEnt-MV $\hat{\pi}^{H_2}_{\epsilon MV}$ proposé.

La section 3.5 étudie la performance de la solution proposée sur des données simulées.

3.5 Caractérisation de la performance de MaxEnt sur données simulées

Pour pouvoir évaluer la performance de l'estimateur MaxEnt-MV, nous reprenons l'expérience du paragraphe 2.4.2, où des données simulées sont observées avec censure par les éléments de partitions, générées aléatoirement comme des unions de cellules voisines d'un diagramme de Voronoi.

La Figure 3.1b montre l'estimateur proposé $\hat{\pi}^{H_2}_{\epsilon MV}$ pour la même loi $p_{\pi_\theta, \mathcal{Q}^l}$, générée aléatoirement et représentée dans la Figure 2.16b, page 43, et reprise dans la Figure 3.1a. Nous rappelons que la masse attribuée par π_θ à chacune des cellules est strictement positive. Remarquons que la répartition de la masse de probabilité est beaucoup plus lisse que celle correspondant à l'estimateur $\hat{\pi}^{\mathcal{L}}$ dans la Figure 2.17b, page 44, et que le support du $\hat{\pi}^{H_2}_{\epsilon MV}$ est maintenant tout le domaine paramétrique Θ. Cet exemple montre que le nouvel estimateur $\hat{\pi}^{H_2}_{\epsilon MV}$ est en mesure d'exploiter les deux critères d'estimation MaxEnt et maximum de vraisemblance afin de bien modéliser les données observées tout en produisant une estimation qui n'est pas trop informative, en particulier que tout le support de π_θ reçoit une masse strictement positive.



(a) $p_{\pi_\theta,\mathcal{Q}^J}$. **(b)** $\hat{\pi}^{H_2}_{\epsilon MV}$.

FIGURE 3.1 – *Solution de l'estimation par MaxEnt-MV de la loi $p_{\pi_\theta,\mathcal{Q}^J}$ générée aléatoirement.*

La Figure 3.2 montre la variation de la log-vraisemblance en fonction du ratio $\epsilon/\epsilon_\star$. Nous voyons que la vraisemblance de l'estimateur $\hat{\pi}^{H_2}_{\epsilon MV}$ est quasiment confondue avec la meilleure vraisemblance possible (ligne rouge dans la Figure).

Puisque la définition de l'estimateur $\hat{\pi}^{H_2}_{\epsilon MV}$ (équation (3.15)) repose sur la métrique ℓ_∞ pour évaluer la déviation d'une densité par rapport aux moments empiriques, et que cette métrique n'est pas équivalente à celle induite par le maximum de vraisemblance sur le simplexe $\mathcal{S}^{M+}$, nous ne pouvons pas garantir que la vraisemblance décroit de façon monotone quand le niveau de relaxation ϵ augmente. Comme le montre la Figure 3.2, la vraisemblance croit pour des valeurs de ϵ inférieurs à ϵ^{MV}. Plus important encore, cette figure montre que le choix approprié du niveau de la relaxation des contraintes peut conduire à une perte de vraisemblance qui est très réduite, tout en améliorant l'ajustement aux données. Les remarques observés consolident le choix du nouvel estimateur proposé dans cette thèse.

La Figure 3.3 caractérise la performance de l'estimation de la vrai loi de probabilité, montrant les boites à moustaches (box-plot) des distances de Kolmogorov-Smirnov (à gauche) et de variation totale (à droite) entre $\pi_{\theta,\mathcal{Q}^J}$ et les estimateurs $\hat{\pi}^{\mathcal{L}}$ et $\hat{\pi}^{H_2}_{\epsilon MV}$ calculés à partir de 200 simulations où pour chaque simulation nous considérons $n = 10^3$ observations. Dans la Figure 3.3, les boites à moustaches à gauche correspondent à l'estimateur $\hat{\pi}^{\mathcal{L}}$ quant à celles à droite correspondent à l'estimateur $\hat{\pi}^{H_2}_{\epsilon MV}$. Cette Figure montre clairement la supériorité de la solution proposée par rapport à l'estimateur de maximum de vraisemblance. Notons que la différence est plus prononcée pour la distance de la variation totale, qui est le critère le plus adapté pour évaluer le pouvoir prédictif de la densité estimée comme mentionné dans [Rya11].

Enfin, la Figure 3.4 montre le comportement des estimateurs $\hat{\pi}^{H_2}_{\epsilon MV}$ et $\hat{\pi}^{\mathcal{L}}$ quand le nombre J des partitions générés aléatoirement augmente. Le nombre J varie de 10 à 100 par pas de 10 et le nombre total d'observations n augmente avec J : $n = 100J$. Les Figures 3.4a et 3.4b montrent les

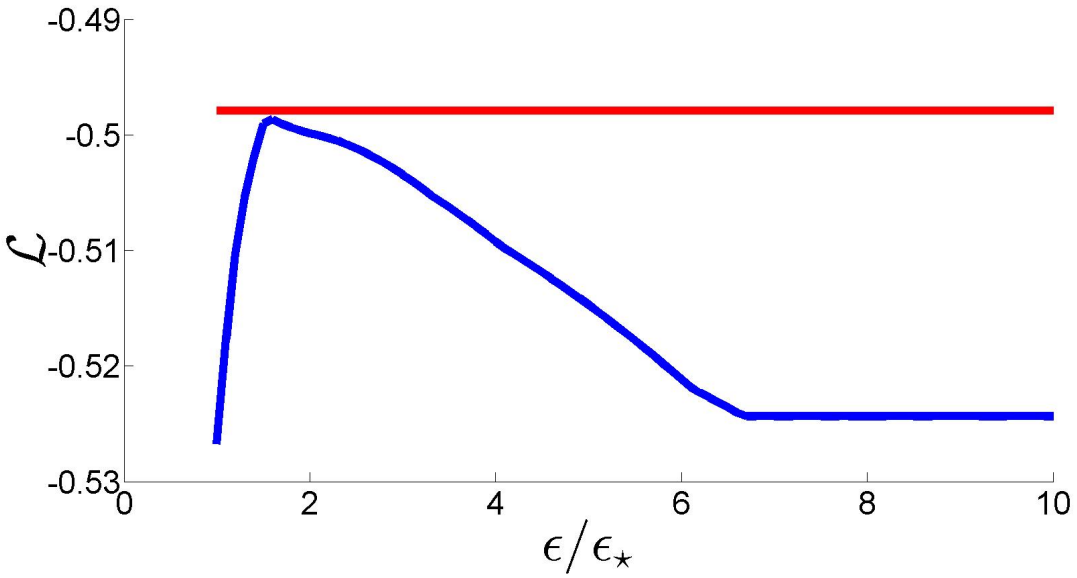
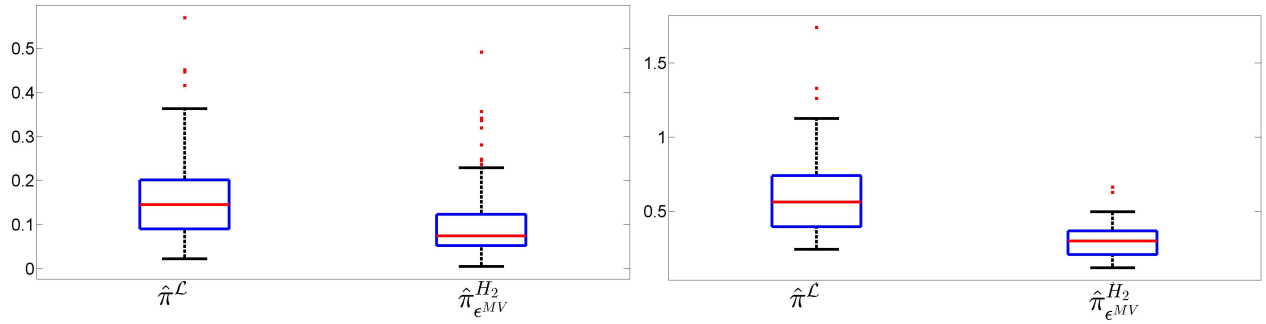


FIGURE 3.2 – *Log-vraisemblance de $\hat{\pi}_\epsilon^{H_2}$ comme fonction de $\epsilon/\epsilon_\star$. La ligne rouge correspond à $\mathcal{L}(\hat{\pi}^{\mathcal{L}}, \mathbf{X}_n)$.*



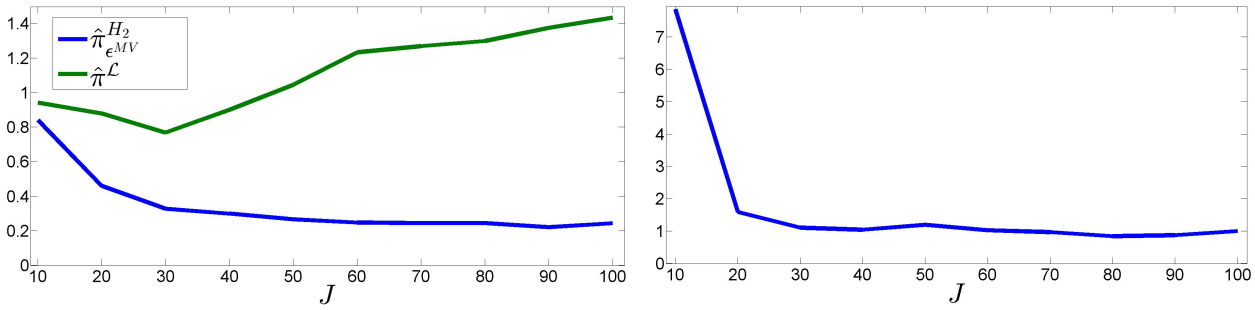
(a) **(b)**

FIGURE 3.3 – *Les boites à moustaches des distances de Kolmogorov (à gauche) et de la variation totale (à droite) entre les la vraie loi de probabilité et les lois estimées $\hat{\pi}^{\mathcal{L}}$ et $\hat{\pi}^{H_2}_{\epsilon MV}$.*

moyennes empiriques des divergences de Kullback-Leibler $D_{KL}(\cdot\|\pi_\theta)$ et $D(\pi_\theta\|\cdot)$. La Figure 3.4a suggère que l'estimateur $\hat{\pi}^{H_2}_{\epsilon MV}$ peut être convergent, tout en soulignant le comportement divergent observé pour l'estimateur $\hat{\pi}^{\mathcal{L}}$. La divergence $D(\pi_\theta|\ \hat{\pi}^{\mathcal{L}})$ prend une valeur infinie pour toutes les simulations en raison de l'attribution de masses nulles de probabilité à certaines régions élémentaires, par conséquent, elle ne peut pas être représentée dans la Figure 3.4b.

3.6 CONCLUSION

Nous avons constaté lors de cette étude que l'incompatibilité des lois empiriques peut mener à des solutions de l'ENPMV qui ne sont pas plausibles pour la modélisation d'une population naturelle. Nous avons donc proposé de baser l'estimateur de la densité cible sur le critère de MaxEnt. Nous



(a) $D_{KL}(\hat{\pi}^{H_2}_{\epsilon MV} \| \pi_\theta)$ *et* $D_{KL}(\hat{\pi}^{\mathcal{L}} \| \pi_\theta)$. **(b)** $D_{KL}(\pi_\theta \| \hat{\pi}^{H_2}_{\epsilon MV})$.

FIGURE 3.4 – *Divergences de Kullback-Leibler entre la vraie loi* π_θ *et ses estimations* $\hat{\pi}^{H_2}_{\epsilon MV}$ *et* $\hat{\pi}^{\mathcal{L}}$ *en fonction d'un nombre croissant de partitions.*

avons adapté cet estimateur pour notre problème en faisant dépendre la relaxation des contraintes du niveau de confiance sur les lois empiriques qui définissent les contraintes de l'estimateur MaxEnt. Le résultat principal de ce chapitre est la définition d'un nouveau critère d'estimation par MaxEnt où l'hyper-paramètre définissant le niveau de relaxation est fixé par maximum de vraisemblance. La solution proposée a démontré de meilleures performances comparée à la solution de l'ENPMV sur des données simulées. Les chapitres suivant s'intéressent à l'application de cette solution au jeu de données réelles dont nous disposons.

- CHAPITRE 4 -

MODÈLE ET OBSERVATIONS

Les accidents de décompression en plongée sous-marine surviennent quand la pression ambiante qui environne le corps du plongeur diminue rapidement. En effet, une réduction de cette pression peut induire l'apparition de bulles gazeuses au sein de certains tissus considérées comme étant la raison principale des accidents de décompression. L'estimateur de densité de probabilité construit dans notre étude s'appuie sur le modèle biophysique présenté par J. Hugon dans ses travaux de thèse [Hug10] et qui caractérise les échanges de gaz entre bulles et tissus pendant et après la décompression Nous avons construit un simulateur numérique de ce modèle et apporté quelques modifications et approximations pour optimiser le temps de calcul. Ce chapitre détaille ces travaux.

Les données disponibles pour la caractérisation de la population observée sont des mesures quantifiées, appelées grades de plongée, reflétant la sévérité de la production des bulles gazeuses libérées suite à un ensemble d'expositions hyperbares, aussi appelées profils de plongée. Les grades de plongée sont le résultat d'une classification réalisée dans notre cas par une interprétation humaine en 5 niveaux de la quantité de bulles perçue au travers de signaux acoustiques produits par un système de détection Doppler. Pour chaque plongée, une seule mesure de grade a été renseignée, sans indication de l'instant de mesure. Nous supposons que les grades enregistrés correspondent à la quantification de la valeur maximale du volume de gaz contenu dans les bulles produites suite à la décompression. La Figure 4.1 schématise les deux systèmes mis en jeu dans la modélisation de la décompression et des grades observés.

La section 4.1 est dédiée à la modélisation du phénomène de décompression. Le paragraphe 4.1.1 décrit les différentes hypothèses du modèle biophysique de décompression sur lequel est basée l'étude. Dans le paragraphe 4.1.2, nous présentons la mise en équation des mécanismes biophysiques mis en jeu par ce modèle dans la production de bulles gazeuses suite à une exposition hyperbare. La réalisation du simulateur du volume de gaz instantané libéré sous forme de bulles ainsi que les choix et hypothèses retenus dans cette réalisation – dont une présentation détaillée figure dans l'annexe 4.A – sont brièvement présentés dans le paragraphe 4.1.3. La section 4.2, est dédiée à la modélisation des observations.

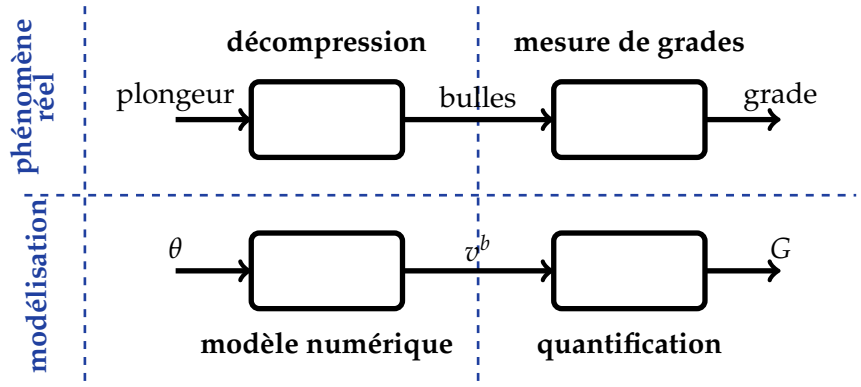


FIGURE 4.1 – *Schéma précisant les modélisations proposées au phénomène réel de la décompression et de l'observation des grades. v^b désigne le volume de gaz de bulle calculé par le modèle numérique et G le grade "théorique" issu de la quantification.*

4.1 MODÈLE BIOPHYSIQUE DE DÉCOMPRESSION

4.1.1 Mécanismes biophysiques

Cette étude doit contribuer à la prévention les accidents de décompression. Ces accidents vont de simples troubles de respiration ou des perturbations locales de la circulation sanguine jusqu'à des accidents graves atteignant la moelle épinière ou le cerveau. De nombreuses études ont montré la forte corrélation qui existe entre l'apparition de symptômes d'accidents de décompression et la quantité de bulles qui circulent dans le sang veineux [Spe76, KMG78, Nis90, SN90]. Le modèle biophysique de décompression décrit les mécanismes de formation, d'amplification et de résorption des bulles de gaz inerte que l'on trouve dans le sang du corps humain pendant la décompression. Il est donc la pierre angulaire de tout système de modélisation du risque d'accidents de décompression. Nous allons en particulier, admettre que le volume de gaz global des bulles piégées dans le filtre pulmonaire est le critère décisif pour la prédiction du risque d'accident de décompression. La probabilité de développer un accident de décompression doit être d'autant plus grande que ce volume est élevé.

La modélisation de la décompression met en jeu l'interaction entre plusieurs phénomènes.

A) **Tension tissulaire du gaz inerte** :

Comme nous l'avons déjà dit, la formation des bulles est due à une baisse de la pression ambiante dans les tissus du plongeur. Comme expliqué dans [Hug10], et depuis les travaux pionniers de John Scott Haldane en 1908, la plupart des modèles biophysiques décrivent l'hétérogénéité des tissus quant à leur échanges gazeux en modélisant la tension tissulaire du gaz inerte au sein de compartiments contenant plusieurs tissus. Afin de décrire les échanges gazeux des tissus de l'organisme humain avec leur milieu, le modèle biophysique de décompression décompose les tissus du corps humain en N compartiments distincts. Les tissus composant chaque compartiment sont supposés avoir un comportement assez proche concernant leur

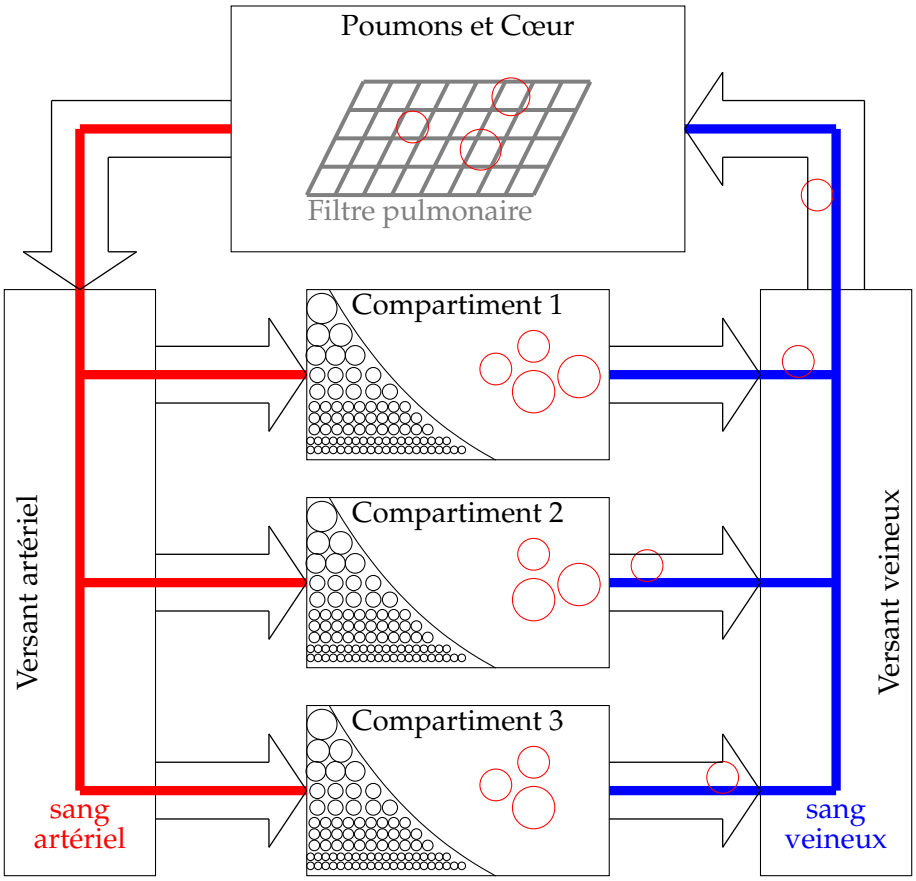


FIGURE 4.1 – *Schéma du modèle biophysique de décompression en subdivisant les tissus du corps humain en trois compartiments plus poumons et cœur.*

échange gazeux avec leur milieu. Les N compartiments du modèle biophysique de décompression sont reliés entre eux par la circulation sanguine et sont supposés être indépendants pour ce qui est de leurs échanges gazeux avec le sang, voir Figure 4.1 où $N = 3$.

B) **Bulles gazeuses** :

Une population préexistante de micro-noyaux gazeux : Le modèle utilisé suppose, conformément aux travaux de David Yount [You79], que les bulles de gaz inerte que l'on retrouve dans le sang ont une origine intra-tissulaire. Chacun des N compartiments est supposé abriter, avant toute exposition hyperbare, une population de micro-noyaux gazeux. Ces populations de micro-noyaux gazeux sont représentées dans la Figure 4.1 par des cercles noirs. Yount suppose que les micro-noyaux gazeux sont sphériques de dimension variable, et propose une distribution exponentielle de leurs rayons. Selon cette loi les micro-noyaux gazeux sont d'autant plus nombreux que leur rayon initial est petit.

Formation des bulles gazeuses : Les bulles de gaz inerte se forment dans les compartiments tissulaires lors de la décompression (les cercles rouges dans la Figure 4.1). Ces bulles naissent

d'une activation des micro-noyaux gazeux lorsqu'un certain écart (appelé *sursaturation locale*) entre la pression du gaz inerte au sein du compartiment et la pression ambiante est atteint. Selon le modèle admis, les micro-noyaux gazeux les plus grands sont les premiers à être activés et la population de micro-noyaux gazeux activée est d'autant plus nombreuse que le niveau de sursaturation locale est important.

Croissance et résorption des bulles gazeuses : Le modèle biophysique [Hug10] reprend les travaux de Van Liew [LB93] afin de modéliser la croissance des bulles gazeuses. Après l'activation des micro-noyaux gazeux, une partie de la quantité de gaz inerte dissous dans le milieu tissulaire environnant est transférée dans les bulles gazeuses. Suite aux changements de pression ambiante, ces bulles gazeuses changent de taille jusqu'à leur réabsorption, qui peut survenir plusieurs heures après la fin de la décompression.

C) **Transfert des bulles vers le sang et blocage au niveau du filtre pulmonaire** : Les bulles gazeuses transitent ensuite vers le sang veineux et sont acheminées jusqu'aux poumons et le cœur (circuit en bleu dans la Figure 4.1). Même si le mécanisme biophysique du passage sur le versant veineux n'est pas identifié avec certitude, on peut supposer que toutes les bulles gazeuses finissent par rejoindre le sang veineux. Lors de ce trajet les conduisant aux poumons et au cœur, les bulles continuent d'échanger du gaz inerte avec le sang et donc de changer de volume selon la même dynamique dépendant de la pression ambiante.

Les bulles gazeuses sont ensuite piégées dans le filtre pulmonaire (bloc central en haut de la Figure 4.1). La sortie du modèle biophysique de décompression est le volume total des bulles circulant dans le versant veineux (avant le filtre pulmonaire).

La mise en équation des trois phénomènes A), B) et C) est présentée dans la section suivante.

4.1.2 Mise en équation du modèle biophysique de décompression

Nous présentons la mise en équation des trois mécanismes biophysiques :
- l'évolution de la tension du gaz inerte au sein de chaque compartiment,
- la genèse des bulles gazeuses,
- l'évolution du volume total des bulles de gaz inerte.

Évolution de la tension tissulaire

Lors d'une exposition hyperbare, le gaz inerte respiré (l'azote pour les expositions à l'air) se dissout progressivement dans les tissus des différents compartiments. Désignons par C l'un des compartiments du modèle biophysique de décompression. Le modèle suppose qu'avant l'activation des premiers micro-noyaux gazeux, le gaz inerte présent dans les tissus est entièrement dissous. L'évolution de la tension du gaz inerte dans le compartiment C est régie par l'équation différentielle suivante :

$$\frac{dP_C}{dt} = k_C \left[f \left(P_{amb} - P_{H_2O} \right) - P_C \right], \tag{4.1}$$

avec :

- P_C la tension de gaz inerte dans le compartiment C (bar),
- k_C la constante d'échange du gaz inerte entre le sang et le compartiment C (min^{-1}), décrivant comment est irrigué le compartiment par le sang,
- f la fraction de gaz inerte dans le mélange respiré (environ 0.78 pour l'air),
- P_{amb} la pression ambiante (bar),
- P_{H_2O} la pression de vapeur saturante d'eau à la temperature 37° Celsius (bar).

Après l'activation des premiers micro-noyaux gazeux dans les tissus du compartiment C suite à une forte sursaturation locale, des bulles gazeuses se forment à l'intérieur de C. Désormais la présence de ces bulles doit être prise en compte lorsqu'on considère l'équilibre des pressions de gaz inerte et l'évolution de la tension du gaz inerte dans C dépend aussi de l'évolution de la quantité de gaz contenue dans ces bulles. L'équation (4.1) est alors remplacée par :

$$\frac{dP_C}{dt} = k_C \left[f \left(P_{amb} - P_{H_2o} \right) - P_C \right] - \frac{1}{V_C S_C} \frac{dn_C^b}{dt}, \tag{4.2}$$

avec :

- n_C^b le nombre de môles de gaz inerte que renferme l'ensemble des bulles dans le compartiment C (mol),
- S_C la solubilité du gaz inerte dans le compartiment C $(mol/m^3/bar)$,
- V_C le volume du compartiment C (m^3).

En résumé, la pression tissulaire est régie par l'équation (4.1) avant l'apparition des bulles gazeuses et par l'équation (4.2) dès l'apparition de la première bulle de gaz.

Genèse des bulles tissulaires

Comme nous l'avons déjà indiqué, le modèle biophysique de décompression s'appuie sur les travaux de David Yount [You79].Afin d'expliquer l'apparition des bulles gazeuses suite à une décompression, Yount considère une population préexistante de micro-noyaux gazeux dans les tissus. Les micro-noyaux sont des bulles gazeuses de taille très petites qui sont générées au sein des tissus suite à des phénomènes de cavitation par frottement de surfaces. Lors de la décompression, une partie de cette population de micro-noyaux commence à recevoir du gaz inerte libéré par le milieu tissulaire environnant. On dit que ces micro-noyaux sont activés et on les appelle alors *bulles gazeuses*. Tout au long de la décompression, ces bulles continueront à échanger du gaz inerte avec leur milieu.

Une première approximation considérée par [You79] consiste à supposer que les micro-noyaux gazeux sont sphériques. Selon Yount, l'organisme contient des micro-noyaux gazeux dont les rayons sont distribués suivant une loi exponentielle :

$$\mu(r) = N^{max} e^{(-rA)}, \tag{4.3}$$

avec N^{max} le nombre total de micro-noyaux gazeux par unité de volume (m^{-3}) et A un coefficient positif (m^{-1}). Yount propose une modélisation où N^{max} et A changent de valeur d'un tissu à l'autre. La loi $\mu(\cdot)$ suppose que plus les micro-noyaux sont petits, plus ils sont nombreux. Cette loi a été validée à travers des études de décompression de blocs de gélatine [You79].

Le nombre de micro-noyaux activés dans un compartiment dépend de l'écart entre la pression ambiante et la pression au sein du compartiment du gaz inerte. A un instant t donné, le niveau maximal de sursaturation atteint dans le compartiment C est défini par

$$P_C^{ss\_max}(t) = \max_{0 \leqslant s \leqslant t} \left[P_C(s) - P_{amb}(s) + \beta \right], \tag{4.4}$$

avec β la somme des tensions des gaz métaboliques et de la pression saturante en vapeur d'eau (bar). Par définition, $t \mapsto P_C^{ss\_max}(t)$ est une fonction croissante dans le temps. L'instant t_{min} de l'activation du premier noyau-gazeux est le premier instant où $t \mapsto P_C^{ss\_max}(t)$ devient strictement positive. Pour tout instant $t \geqslant t_{min}$, le rayon $r_C^0(t)$ du plus petit micro-noyau gazeux activé jusqu'à l'instant t dans le compartiment C est donné par

$$r_C^0(t) = \frac{2\gamma_C}{P_C^{ss\_max}(t)} \quad \text{pour } t \geqslant t_{min}, \tag{4.5}$$

avec γ_C la tension superficielle dans le compartiment C (N/m). Nous en déduisons que les premiers micro-noyaux gazeux activés sont les plus grands. Tant que $t \mapsto P_C^{ss\_max}(t)$ est strictement croissante, de nouveaux micro-noyaux gazeux de plus en plus petits sont activés. Le nombre total $N_C^{tot}(t)$ des noyaux-gazeux activés jusqu'à un instant $t \geqslant t_{min}$ est

$$N_C^{tot}(t) = \frac{N^{max}}{A} e^{-A r_C^0(t)}.$$

Le rayon r_C^{min} du plus petit micro-noyau gazeux activé lors d'une décompression est donc

$$r_C^{min} = \frac{2\gamma_C}{\max\limits_{0 \leqslant t} \left[P_C(t) - P_{amb}(t) + \beta \right]}.$$

Le nombre total $\bar{N}_C^{tot}$ de bulles générées dans le compartiment C lors de la décompression est ainsi d'autant plus grand que r_C^{min} est petit. En tenant compte de la loi de distribution μ des micro-noyaux gazeux dans le compartiment C, nous avons

$$\bar{N}_C^{tot} = \frac{N^{max}}{A} e^{-A r_C^{min}}.$$

L'exemple de la Figure 4.2 illustre le lien entre $P_C^{ss\_max}(t)$ et $r_C^0(t)$. Nous supposons dans cet exemple que $t \mapsto P_C^{ss\_max}(t)$ est constante par morceaux [1]. Nous indiquons à chaque changement de

1. Dans la réalité, $t \mapsto P_C^{ss\_max}(t)$ n'est pas constante par morceaux, nous choisissons ce cas de figure simplifié pour

$P_C^{ss\_max}$ la population des micro-noyaux activés. Ainsi, à l'instant t_{min} les micro-noyaux activés sont ceux de rayons supérieurs à $r_C^0(t_{min}) = \frac{2\gamma_C}{P_C^{ss\_max}(t_{min})}$ (les bulles de couleur bleue). Ces micro-noyaux restent les seuls activés jusqu'à l'instant t_1. A l'instant t_1, les micro-noyaux de rayons appartenant à $\left[r_C^0(t_1), r_C^0(t_{min})\right]$ (bulles rouges) sont activés. Elles échangent du gaz inerte avec leur milieu environnant, aux cotés des bulles en bleue, jusqu'à l'instant t_2 où elles sont rejointes par les bulles vertes. $P_C^{ss\_max}$ ne dépasse pas $P_C^{ss\_max}(t_2)$, et donc les micro-noyaux de rayon inférieur à $r_C^0(t_2)$ (les petites bulles en noir) ne sont pas activés.

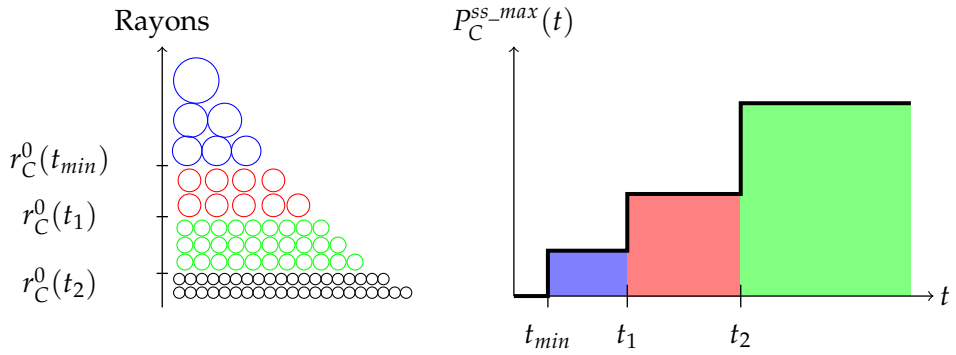


FIGURE 4.2 – *Exemple du lien entre $P_C^{ss\_max}(t)$ et $r_C^0(t)$.*

Évolution du rayon d'une bulle

Nous avons vu dans le paragraphe précédent comment, lors d'une décompression, le fait d'atteindre des niveaux de sursaturation locale de plus en plus élevés dans les différents compartiments induit l'apparition des bulles gazeuses de plus en plus petites. Ces bulles interagissent avec leur milieu environnant en échangeant du gaz inerte. Lorsqu'elles sont formées dans les tissus, les bulles gazeuses reçoivent de leur milieu environnant sursaturé le surplus de gaz inerte, ce qui les amplifie. Dans le cas inverse, si les bulles gazeuses deviennent sursaturées par rapport à leur milieu, le gaz inerte sort des bulles, ce qui provoque leur résorption progressive. Van Liew et Burkard [LB93] proposent une expression analytique de l'évolution du rayon d'une bulle dans un milieu sursaturé ou sous-saturé. Le modèle [Hug10] reprend leurs travaux. Soit une bulle gazeuse de rayon R dans le compartiment C. L'évolution de R est donnée par :

$$\frac{dR}{dt} = \frac{\mathfrak{R}T_C D_C S_C (P_C - P_C^b)\frac{1}{R} - \frac{R}{3}\frac{dP_{amb}}{dt}}{P_{amb} + \frac{4\gamma_C}{3R} - P_{H_2O}}, \tag{4.6}$$

avec

- $\mathfrak{R}$ la constante des gaz parfaits $(bar \times m^3/mol/K)$,
- T_C la température dans le compartiment C (K),
- D_C le coefficient de diffusion du gaz inerte dans le compartiment C (m^2/s).

illustrer le phénomène d'activation des micro-noyaux gazeux.

– P_C^b la pression de gaz inerte dans les bulles du compartiment C (bar).

En supposant que l'équilibre mécanique est réalisé à chaque instant entre la bulle gazeuse de rayon R et son milieu, l'expression de la pression P_C^b de gaz inerte dans cette bulle est :

$$P_C^b = P_{amb} + \frac{2\gamma_C}{R} - \beta. \tag{4.7}$$

L'exploitation des équations (4.6) et (4.7) permettra de déduire la dynamique du rayon de chaque bulle gazeuse générée lors de la décompression. Ces équations, plus les équations de (4.1) à (4.5) déterminent le volume de gaz libéré sous forme de bulles.

Volume des bulles gazeuses libérées

Considérons un profil de plongée P défini par $t \mapsto P_{amb}(t)$. Nous notons $t \mapsto v_C^b(t, \theta, P)$ le volume de gaz inerte instantané libéré par le compartiment C d'un organisme caractérisé par le vecteur de paramètres θ et qui a effectué le profil de plongée P. Notons $t \mapsto v^b(t, \theta, P)$ le volume total de gaz inerte instantané libéré par tous les tissus (qui a été déjà mentionné dans la Figure 4.1). Nous admettons [Hug10] pour chaque instant t le modèle simple suivant :

$$v^b(t, \theta, P) = \sum_{C \in \mathscr{C}} v_C^b(t, \theta, P), \tag{4.8}$$

avec $\mathscr{C}$ l'ensemble des compartiments.

Une modélisation plus complexe du volume des bulles libérées prendrait en compte les temps de transit dans le sang caractérisant la participation de chaque compartiment, ainsi que le blocage de certaines bulles dans le filtre pulmonaire. Le modèle implémenté considère que le temps de transit des bulles est nul pour tous les compartiments et néglige la proportion des bulles gazeuses bloquées dans le filtre pulmonaire.

Afin de pouvoir implémenter notre estimateur, nous construisons, à partir des équations (4.1) – (4.7), un simulateur numérique du volume instantané v_C^b pour chaque compartiment C dans $\mathscr{C}$. Ces équations ne peuvent pas être résolues analytiquement à cause de la complexité de la dynamique non-linéaire des rayons et de la condition d'activation de nouveaux micro-noyaux portant sur $P_C^{ss\_max}$, nous avons donc dû les résoudre numériquement. Le paragraphe suivant traite de la construction du modèle numérique du volume de gaz libéré à partir des équations du modèle biophysique.

4.1.3 Simulateur numérique du volume de gaz des bulles

La construction du simulateur est présentée plus en détail dans l'annexe 4.A. Nous nous contentons ici de décrire la simulation numérique des trois mécanismes biophysiques (schéma de la Figure 4.3) et les approximations retenues.

Les équations présentés dans le paragraphe précédent pour l'ensemble des dynamiques des trois mécanismes biophysiques pour déduire le volume de gaz instantané libéré sont des équations diffé-

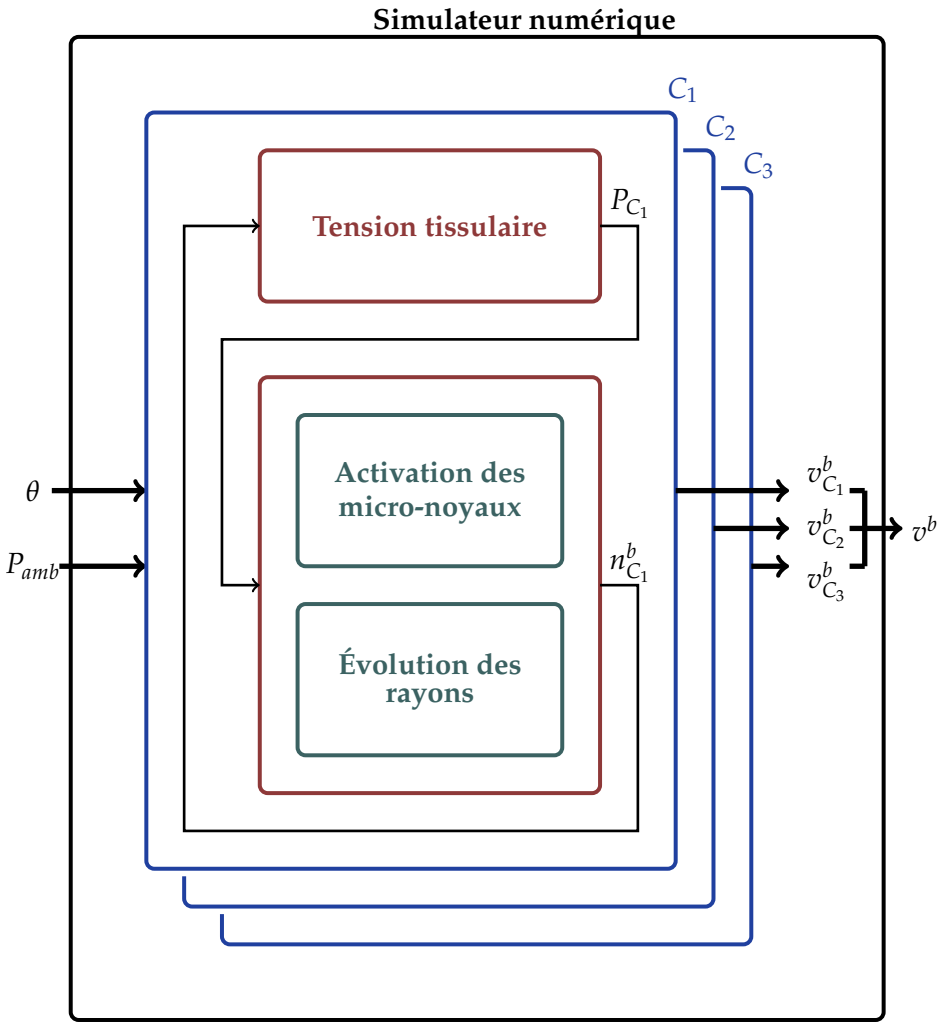


FIGURE 4.3 – *Schéma de l'algorithme ayant servi à la simulation de volume de gaz libéré suite à la décompression.*

rentielles qui ne possèdent pas de solution analytique. Nous procédons à leur résolution numérique par schéma d'Euler. Le choix d'un pas "optimal" de discrétisation, assurant des résultats pertinents tout en permettant une réduction du temps de calcul des simulations a fait l'objet d'une étude, présentée dans l'annexe 4.B, page 99. Ce choix a été validé sur un grand nombre de simulations, faisant appel à plusieurs profils standards de plongée et de larges plages des domaines paramétriques.

Avant déclenchement du processus d'activation des micro-noyaux, la pression tissulaire est régie par l'équation (4.1) qui peut être résolue analytiquement. Au premier instant où la quantité $P_C^{ss\_max}$ devient strictement positive, les plus grands micro-noyaux sont activés et la pression tissulaire commence à être régie par l'équation (4.2). A chaque instant où $P_C^{ss\_max}$ est strictement croissante, de nouveaux micro-noyaux sont activés et leur nombre est déterminé par les équations (4.4) et (4.5). A chaque instant de discrétisation, la condition d'activation de nouveaux micro-noyaux est vérifiée afin

de mettre à jour la population des bulles gazeuses actives dans les échanges de gaz avec les tissus. Nous avons apporté essentiellement deux hypothèses dans la réalisation du simulateur de volume de gaz contenu dans les bulles.

Un seul rayon par recrutement : Nous admettons que tous les micro-noyaux activés à un certain instant ont le même rayon. Cette hypothèse, bien détaillée dans l'annexe 4.A, page 94, nous évite la lourdeur computationnelle d'un schéma d'Euler discrétisant aussi l'axe des rayons. Une fois que la population des bulles gazeuses est mise à jour, nous résolvons numériquement l'évolution du rayon des bulles actives, décrite par l'équation (4.6).

Quasi-stationnarité de certains termes dans l'équation (4.6) : Dû au caractère fortement non-linéaire de cette équation, sa résolution numérique est très coûteuse en temps de calcul. Nous avons apporté une approximation à cette équation, présentée dans l'annexe 4.B, page 98, afin de diminuer le temps d'exécution de la simulation. La résolution de la dynamique du rayon nous permet de déduire la quantité $\frac{dn_C^b}{dt}$ à l'instant de discrétisation considéré et de l'injecter dans l'équation (4.2). Le schéma de la Figure 4.3 représente l'algorithme retenu pour la simulation du volume instantané de gaz libéré sous forme de bulles lors des expositions hyperbares.

Paramétrisation : Certains paramètres des équations du modèle biophysique sont connus ou ont été fixés à leurs valeurs nominales en se référant à [Hug10]. La table 4.1 donne les valeurs numériques de ces paramètres.

Paramètre	Valeur numérique	Unité
f	0.78	$-$
P_{H_2O}	0.06	bar
S_{C_i}	0.605	$mol/m^3/bar$
β	0.175	bar
$\mathfrak{R}$	$8.31.10^{-5}$	$bar \times m^3/mol/K$
T_{C_i}	310	K
V_{C_i}	1	m^3
D_{C_i}	$1.28.10^{-8}$	m^2/min
γ_{C_i}	5.10^{-7}	$bar \times m$
k_{C_1}	0.002	min^{-1}
k_{C_2}	0.02	min^{-1}
k_{C_3}	0.2	min^{-1}

TABLE 4.1 – *Valeurs numériques et unités des paramètres connus du modèle biophysique de décompression. Quand on mentionne l'indice i dans la notation du paramètre, c'est qu'on a supposé que sa valeur est commune au différents tissus de l'organisme.*

Dans la suite de cette étude, nous caractérisons la variabilité de la population des plongeurs par leur variabilité au niveau des paramètres $N_{\max}$ et A. Il s'agit ici des deux paramètres auxquels nous nous intéresserons, dans le Chapitre 5, afin d'estimer leur distribution chez une population de

plongeurs. Même s'il est évident que les micro-noyaux gazeux n'ont pas une distribution homogène pour tous les tissus, nous avons choisi dans cette étude, après concertation avec Julien Hugon, de considérer des valeurs de N^{max} et A communes à tous les compartiments. Ceci est dû au fait que l'on s'intéressera par la suite au volume total de gaz libéré par l'organisme et non pas à la participation de chaque tissu dans ce volume. D'après [Hug10], les plages de valeurs envisagées pour N^{max} et A sont respectivement $\left[5.10^6, 5.10^{10}\right]$ et $\left[10^5, 10^7\right]$.

4.1.4 Illustration et analyse du comportement du simulateur numérique

4.1.4.1 Exemple d'une simulation numérique

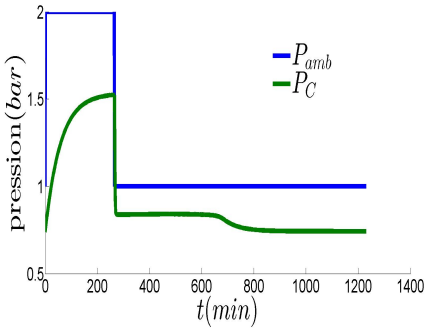
Afin d'illustrer le comportement du simulateur de v_C^b nous considérons l'exemple d'une plongée simple. Il s'agit d'une plongée de 10m/265min, sans palier, avec une vitesse de descente de 15m/min et une vitesse de remontée de 20m/min. Ce même profil de plongée a été utilisé par les auteurs de [LCB94]. Dans cet exemple, nous nous intéressons en particulier à un compartiment C dont la constante k_C est de $0.0173 min^{-1}$ (valeur choisie dans [LCB94], et proche de la valeur de k_{C_2} du compartiment des tissus moyens présenté dans l'annexe 4.A, page 93). Nous nous plaçons dans le cas où le volume du compartiment est de $0.015 m^3$ avec $N^{max} = 4.10^{10} m^{-3}$ et $A = 10^6 m^{-1}$. La table 4.2 donne les valeurs numériques des paramètres utilisées pour la simulation de cet exemple, autres que ceux mentionnés dans la table 4.1.

Paramètre	Valeur numérique	Unité
k_C	0.0173	min^{-1}
N^{max}	4.10^{10}	m^{-3}
A	10^6	m^{-1}

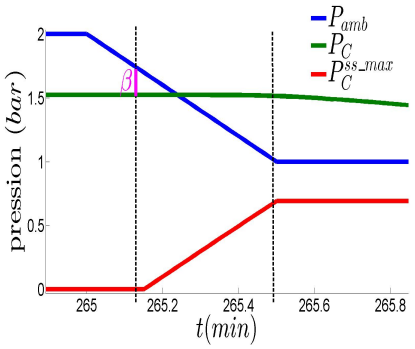
TABLE 4.2 – *Valeurs numériques et unités de certains paramètres du modèle biophysique de décompression utilisées pour l'exemple 4.1.4.*

Lors de cet exemple, pour un pas de discrétisation d'une seconde, $9,38.10^9$ micro-noyaux gazeux sont recrutés, distribués en 21 paquets de bulles (21 instants de recrutement). La Figure 4.4 montre quelques résultats pour cette plongée, produits par le simulateur développé.

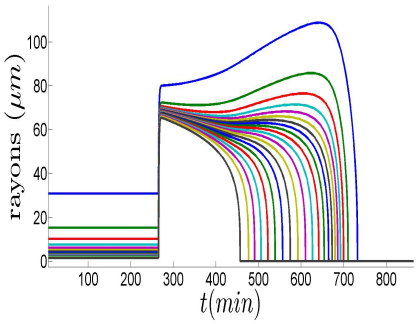
Dans la Figure 4.4a nous voyons comment les pressions P_{amb} et P_C varient au cours du temps. Si la variation de P_C avant la phase de la remontée (jusqu'à la minute 265) peut être totalement expliquée par la variation de la pression ambiante, nous remarquons que le changement de P_C observé vers la minute 700 ne peut pas être expliqué par un changement de la pression ambiante. La Figure 4.4b met l'accent sur le phénomène de recrutement des micro-noyaux gazeux (pendant l'intervalle de temps entre les deux lignes en pointillé) lors de la phase de décompression. Les micro-noyaux gazeux commencent à être activés une fois que la pression au sein du compartiment devient plus grande que la pression ambiante moins β (en rose). Tant que $P_C^{ss\_max}$ est strictement croissante, le recrutement des bulles gazeuses continue en activant des micro-noyaux de plus en plus petits et de plus en plus nombreux. Le recrutement, dans cet exemple, dure 20 secondes. La Figure 4.4c présente



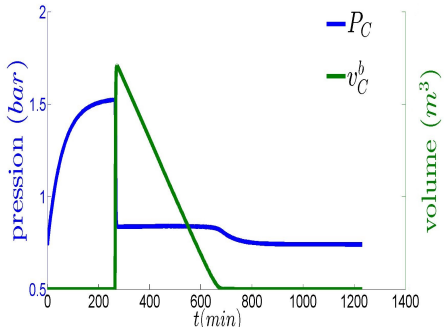
(a) *Pressions P_{amb} et P_C.*



(b) *Phase de recrutement.*



(c) *Évolution des rayons.*



(d) *Volume de gaz inerte libéré.*

FIGURE 4.4 – *Résultats de la simulation de v_C^b appliqué au profil de plongée 10m/265min, avec $A = 10^6 m^{-1}$, $N^{max} = 4.10^{10} m^{-3}$ et $k_C = 0.0173 min^{-1}$.*

la dynamique des rayons des bulles dans le compartiment considéré. Nous remarquons que les premières bulles recrutées au premier instant de recrutement et qui sont aussi les plus grandes mais les moins nombreuses, mettent presque 8 heures après la remontée en surface pour être complètement absorbées. Les plus petites bulles, activées en dernier, mettent environ 3 heures. Le graphique 4.4d montre l'évolution du volume du gaz inerte que renferment toutes les bulles dans le compartiment considéré. Nous remarquons que ce volume diminue quasi-linéairement en fonction du temps. Ce phénomène est habituellement désigné par le terme de *clamping* [Hug10].

Nous supposons par la suite que le lien entre une valeur θ des paramètres biophysiques et une observation impliquant le profil de plongée P dépend uniquement du pic de volume de gaz inerte libéré au cours du temps

$$\max_{0 \leqslant t} \{v^b(t, \theta, P)\} = \|v^b(\theta, P)\|_\infty.$$

Nous quantifions donc l'impact du pas de discrétisation des équations différentielles sur v_C^b en utilisant l'erreur relative ER par rapport au pic de volume calculé avec le pas de discrétisation $0.1s$ en deçà duquel nous constatons que les résultats deviennent stables. L'erreur relative par rapport à une sortie f du simulateur numérique calculée avec les deux pas de discrétisation Δt_0 et Δt_1 est définie

par :

$$ER^{f}_{\Delta t_0}(\Delta t_1) = \frac{|\|f_{\Delta t_1}\|_\infty - \|f_{\Delta t_0}\|_\infty|}{\|f_{\Delta t_0}\|_\infty}. \tag{4.9}$$

$\Delta t(s)$	**temps de calcul (s)**	**nb. de recrutements**	$ER^{v^b_C}_{0.1s}(\Delta t)$
30	1.18	1	0.3657
10	7.49	3	0.0358
1	83.86	21	0.0018
0.5	186.58	42	6.575e-04
0.1	1898.46	210	0

TABLE 4.3 – *Impact du pas de discrétisation sur un exemple de simulation.*

Nous avons vu dans l'exemple de simulation présenté dans la Figure 4.4 que la phase d'activation des micro-noyaux gazeux dure environ 20 secondes. Cette activation s'effectue de façon continue tant que $P_C^{ss\_max}$ est strictement croissante. Une discrétisation fine de l'axe de temps permet une meilleure estimation du volume de gaz inerte libéré lors de la décompression. Par contre cette discrétisation fine peut être très coûteuse en terme de temps de calcul. La table 4.3 compare l'impact du choix du pas sur le temps de calcul, le nombre d'activations des micro-noyaux et l'erreur relative sur le pic du volume v^b_C. Nous remarquons que pour le pas de discrétisation $\Delta t = 1s$ nous réduisons le temps de calcul de plus de 20 fois par rapport au pas de discrétisation $\Delta t = 0.1s$ tout en ayant une erreur relative ne dépassant pas 0.2%. La table 4.3 montre qu'un choix judicieux du pas de discrétisation nous permettra de réduire le temps de calcul d'une simulation (voir l'annexe 4.B où l'on établit une règle de choix du pas de discrétisation en utilisant l'erreur relative).

4.1.4.2 Validation qualitative du modèle

Nous mettons l'accent dans ce paragraphe sur la sensibilité de $\|v^b(\cdot,\cdot)\|_\infty$, le maximum de volume de gaz libéré par les 3 compartiments de l'organisme, aux variations de certains paramètres. Nous nous intéressons évidemment à la sensibilité par rapport aux paramètres N^{max}, A et k_C, mais aussi à l'impact d'un accroissement de la profondeur, la durée d'une plongée ou de la vitesse de remontée du plongeur. Nous nous basons pour cette illustration sur des profils de plongée sans paliers appelés aussi *profils carrés*. La Figure 4.5 présente $\|v^b(\cdot,\cdot)\|_\infty$, obtenu avec la version simplifiée du simulateur et le pas de discrétisation $0.25s$. Nous remarquons, comme attendu d'après les équations du modèle biophysique de décompression, que le maximum du volume libéré par un compartiment croit avec N^{max} et diminue avec A. La figure 4.5 montre que le pic du volume de gaz est d'autant plus grand que la profondeur de la plongée, sa durée ou la vitesse de remontée sont plus grandes. Toutes ces sensibilités sont conformes à ce qui est observé en pratique en plongée sous-marine. En effet, plus la profondeur de la plongée, sa durée ou sa vitesse de remontée augmentent, plus le risque d'avoir un accident de décompression augmente. Ceci conforte donc la corrélation qui semble exister entre risque d'accident de décompression et volume de gaz inerte libéré dans le corps.

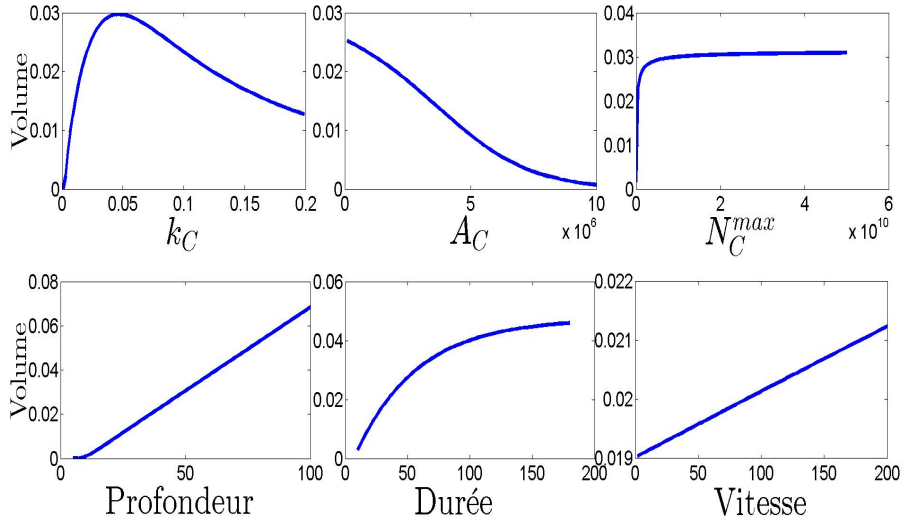


FIGURE 4.5 – *Maximum du volume de gaz libéré en fonction de $k_C(min^{-1})$, $A(m^{-1})$, $N^{max}(m^{-3})$, profondeur (m), durée (min) et vitesse (m/min) de remontée dans le cas d'un profil de plongée carré.*

Le simulateur discuté lors de cette section nous sera utile pour pouvoir estimer la densité des paramètres N^{max} et A chez une population donnée de plongeurs. En effet, afin d'établir le lien entre les mesures disponibles et les paramètres N^{max} et A, nous devrons pour tous les profils de plongée observés, pouvoir comparer les observations avec la sortie du simulateur pour un grand nombre de valeurs de N^{max} et A. Nous serons donc amenés à faire appel à ce simulateur un grand nombre de fois. Par conséquent, il faut que le simulateur soit codé de manière efficace afin d'assurer une exécution rapide sans que cela n'affecte la précision des résultats. L'annexe 4.B traite les choix retenus pour accélérer l'exécution du simulateur numérique. La section suivante est consacrée à la modélisation des observations.

4.2 OBSERVATIONS

Les observations disponibles dans le cadre de cette étude, appelées *grades* de plongée, donnent le niveau de sévérité de production des bulles gazeuses suite à la décompression. Dans la suite nous présentons la méthode utilisée dans la mesure de ces grades, la modélisation proposée pour cette mesure et le choix des seuils de quantification auquel nous avons procédé.

4.2.1 Instruments de détection et codage

L'apparition des bulles gazeuses dans le sang est mise en évidence essentiellement par deux méthodes de détection : la méthode de Doppler et la méthode d'échographie.

Le système de détection Doppler est basé sur la réflexion de faisceaux d'ultrasons et l'effet Dop-

pler (décalage de fréquence des ondes entre la mesure à l'émission et la mesure à la réception) pour détecter des bulles dans le sang circulant. Le détecteur Doppler permet de recevoir un signal acoustique du flot de sang et des bulles gazeuses. Ces signaux acoustiques recueillis sont classés selon principalement 2 codes : le code de Spencer [Spe76] et le code Kisman-Masurel [KMG78]. Ces deux codes classent la sévérité de la production des bulles de gaz en $L = 5$ catégories, appelées grades, allant de 0 à 4.

En échographie, un transducteur émet une courte impulsion d'ultrasons puis "écoute" l'écho des réflexions, d'une façon qui ressemble au système Doppler. La distance entre le transducteur et la matière réfléchissant les ultrasons est déterminée par le retard entre la transmission et la réception de l'écho et l'intensité de l'écho détermine la luminosité avec laquelle le réflecteur est affiché dans l'image. Les images produites par l'échographie permettent ensuite de compter les bulles gazeuses par unité de surfaces et déduire le grade correspondant selon le code Eftedal-Brubbak [EB96]. Selon les auteurs de [BGMB14], il n'y a pas de différence significative entre les mesures de grades issues de l'échographie et celles issues du système Doppler. Cette remarque nous sera utile dans le choix des seuils dans notre modélisation de la quantification de volume de bulles gazeuses que nous présentons dans le paragraphe 4.2.3.

4.2.2 Base de données

Les données dans cette thèse sont sous forme d'une liste de grades observés suite à l'exécution d'un ensemble de profils de plongée par une population de plongeurs. Il s'agit d'un ensemble de $n = 433$ grades $\mathbf{G}_n = \{G_i\}_{i=1}^n$ de plongée observés sur n profils de plongée $\mathbf{P}_n = \{P_i\}_{i=1}^n$ effectués par des plongeurs supposés tirés au hasard (tirages indépendants) dans la population étudiée. Nous notons ces données $\mathbf{X}_n = \{G_i, P_i\}_{i=1}^n$. Un profil de plongée peut être exécuté plusieurs fois, ce qui implique que les éléments de la liste $\{P_i\}_{i=1}^n$ ne sont pas uniques. Soit J le nombre des profils uniques dans $\{P_i\}_{i=1}^n$ et $\{P^j\}_{j=1}^J$ la liste de ces profils. Nous avons $J = 19$ profils de plongée distincts, répétés n_j fois allant de 12 à 41 (voir Table 4.1). Les profils les plus dangereux ont été exécutés moins souvent.

Profil	1	2	3	4	5	6	7	8	9	10
n^j	31	41	24	31	28	12	18	14	14	17
Profil	11	12	13	14	15	16	17	18	19	
n^j	16	26	14	16	18	30	12	41	30	

TABLE 4.1 – *Nombre d'observations par profil de plongée dans le jeu de données réelles.*

Les profils de plongée sont indiqués par des ensembles de couples durée/profondeur, où la variation de la pression ambiante entre deux couples successifs est supposée être linéaire. La base de données disponibles contient une seule mesure de grade par plongée et n'indique pas l'instant de mesure de grade par rapport au profil de plongée effectué. Nous supposons que le grade enregistré est la quantification du maximum de volume de gaz contenu dans les bulles. Avec cette hypothèse,

les mesures de grade sont une quantification du scalaire $\|v^b(\theta, P)\|_\infty$, voir la Figure 4.1.

Ces grades correspondent à des seuils $\tau = \{\tau_\ell\}_{\ell=0}^{L}$ agissant sur la variable $\|v^b(\theta, P)\|_\infty$, avec $\tau_0 = 0 < \tau_1 < \cdots < \tau_{L-1} < \tau_L = \infty$ pour les codes Spencer et Kisman-Masurel.

Le grade G associé à une valeur θ du paramètre biophysique et au profil de plongée P satisfait

$$G(\theta, P) = \ell \Leftrightarrow \tau_\ell \leq \|v^b(\theta, P)\|_\infty < \tau_{\ell+1}, \quad \forall \ell = 0, \ldots, L-1, \tag{4.10}$$

voir les Figures 4.1 et 4.2.

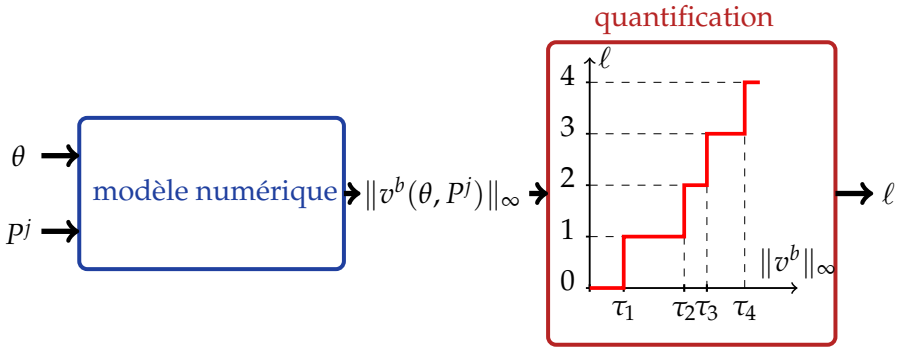


FIGURE 4.1 – *Mécanisme de censure dans le cas précis d'observations de grades de plongée.*

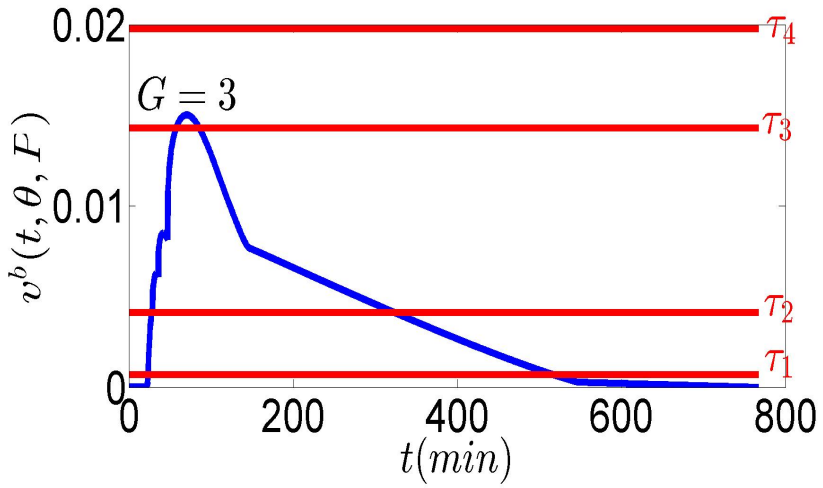


FIGURE 4.2 – *Lien entre le pic de volume v^b et les seuils $\tau = \{\tau_\ell\}_{\ell=1}^{L}$ pour $L = 4$.*

4.2.3 Seuils de quantification

Idéalement, nous aurions dû estimer simultanément la distribution des paramètres biophysiques et l'ensemble des seuils sur le volume maximum de bulles qui définissent les 5 grades observés. C'est un problème encore plus complexe que la seule identification de la distribution des paramètres biophysiques. Compte tenu de la pauvreté des données à disposition, nous avons procédé à une détermination des seuils de quantification par une procédure ad-hoc.

Estimation des seuils de quantification relatifs au code Eftedal-Brubbak (échographie) : Dans ses travaux de thèse [Eft07], Eftedal a développé un programme permettant un comptage automatique des bulles de gaz à partir des images d'échographie. En se basant sur des échographies issues de plusieurs centaines d'expériences de décompression sur des animaux, ce programme de comptage de bulles a permis à Eftedal d'estimer les seuils de quantification $\{\tau'_1, \ldots, \tau'_4\}$ du code Eftedal-Brubbak qui font le lien entre les grades de plongée et le nombre de bulles par cm^2 dans les images d'échographie. Malgré le fait que les deux systèmes de codage de grades Kisman-Masurel et Eftedal-Brubbak sont appliqués sur des signaux de natures différentes (des signaux Doppler pour le premier code et des images d'échographie pour le deuxième), les auteurs de [BGMB14] concluent, comme déjà mentionné, qu'il n'y a pas de différence significative entre les deux systèmes de codage.

Nous avons remarqué au le paragraphe 4.1.4.2 en faisant appel au simulateur numérique que le volume maximum de gaz des bulles est d'autant plus grand que $N^{\max}$ est grand et que A est petit. En considérant un domaine paramétrique de la forme $\Theta = [N_1^{\max}, N_2^{\max}] \times [A_1, A_2]$, nous avons pour chaque profil de plongée P, et pour chaque $\theta \in \Theta$:

$$\|v^b((N_1^{\max}, A_2), P)\|_\infty \leqslant \|v^b(\theta, P)\|_\infty \leqslant \|v^b((N_2^{\max}, A_1), P)\|_\infty.$$

Ainsi, nous déterminons un encadrement des seuils du quantificateur du code Kisman-Masurel en considérant les ensembles $\mathcal{P}_0$ et $\mathcal{P}_4$ des profils de plongée pour lesquels nous avons observé respectivement les grades 0 et 4 :

$$\tau_\star \leqslant \tau_1 < \cdots < \tau_4 \leqslant \tau^\star, \tag{4.11}$$

avec

$$\tau_\star \triangleq \max_{P \in \mathcal{P}_0} \|v^b((N_1^{\max}, A_2), P)\|_\infty \quad \text{et} \quad \tau^\star \triangleq \min_{P \in \mathcal{P}_4} \|v^b((N_2^{\max}, A_1), P)\|_\infty.$$

En faisant un grand nombre de simulations numériques utilisant les profils de notre base de données ainsi que 121 points dans Θ choisis selon une grille régulière, nous avons pu déterminer un jeu de seuils garantissant une distribution uniforme des maximums de volume de gaz simulés sur les 5 grades possibles. Nous avons pu constaté que les seuils ainsi obtenus vérifient

$$\frac{\tau_{\ell+1}}{\tau_\ell} \approx \frac{\tau'_{\ell+1}}{\tau'_\ell}, \quad \ell = 1, \ldots, 3. \tag{4.12}$$

Ce choix des seuils a été menée en collaboration avec J. Hugon, expert en plongée sous-marine. Cette

étude ad-hoc nous a permis de séparer l'étude de l'identification de la densité des paramètres du modèle de décompression du problème de caractérisation du mécanisme d'observation.

4.3 CONCLUSION

Ce chapitre a été consacré au modèle biophysique de décompression. Nous nous sommes appuyé sur ce modèle afin de réaliser un simulateur du volume de gaz libéré par l'organisme suite à une décompression. Ainsi, les fonctions scalaires $F_j(\cdot)$ mentionnées dans le Chapitre 1, constituant la sortie du modèle numérique et faisant l'objet de l'observation censurée sont les fonctions $\|v^b(\theta, P)\|_\infty$ pour chaque profil de plongée P. Nous utilisons le simulateur numérique pour lier les observations sous forme de grades de plongée établis pour une population donnée de plongeurs, aux paramètres biophysiques (N^{max}, A). Un travail de réduction du temps de calcul des simulations a été nécessaire à cause du grand nombre de fois où nous ferons appel au simulateur. Au chapitre suivant, nous allons estimer la distribution de (N^{max}, A) à partir des données de grades.

ANNEXE DU CHAPITRE 4

4.A CONSTRUCTION D'UN SIMULATEUR DU VOLUME DE GAZ CONTENU DANS LES BULLES

Pour simuler $t \mapsto v^b(t,\cdot,\cdot)$, nous devons déterminer à tout instant la contribution de chaque compartiment $v^b_C(t,\cdot,\cdot)$. Il faudra donc sommer à chaque instant les volumes des bulles gazeuses générées en déterminant le rayon R et la pression P^b_C de gaz inerte dans chaque bulle recrutée. D'une manière équivalente, nous pouvons déterminer $t \mapsto n^b_C(t,\cdot,\cdot)$, la quantité de matière de gaz inerte libéré par C. Le lien entre volume et nombre de môles se fait par l'intermédiaire de la loi des gaz parfaits. Commençons par présenter les hypothèses retenues pour la construction du simulateur.

Hypothèses pour la construction du simulateur de v^b_C

Les hypothèses présentées ci-dessous, retenues pour la construction du simulateur de v^b_C, ont été adoptées en collaboration avec Julien Hugon.

Trois compartiments

La plupart des approches modélisant le phénomène de la décompression ont adopté la subdivision des tissus de l'organisme humain en plusieurs compartiments. La subdivision des tissus en compartiments consiste à mettre dans un même compartiment les tissus qui ont des comportements similaires en terme d'échange gazeux lors de la décompression. Dans le modèle biophysique de décompression, un tissu ts est caractérisé par la constante k_{ts} d'échange du gaz inerte entre le sang et le tissu. La valeur de k_{ts} est connue pour chaque tissu et est défini par l'équation suivante :

$$k_{ts} = \frac{\dot{q}_{ts} S_{sang}}{V_{ts} S_{ts}},$$

avec $\dot{q}_{ts}$ le débit sanguin total dans le tissu ts (m^3/s), S_{sang} la solubilité du gaz inerte dans le sang $(mol/m^3/bar)$, V_{ts} le volume de ts (m^3) et S_{ts} la solubilité du gaz inerte dans ts $(mol/m^3/bar)$. La constante k_{ts} traduit la capacité du tissu à évacuer le gaz qui lui est transmis. nous mettons dans un même compartiment des tissus qui ont des constantes k_{ts} similaires. En se basant sur les valeurs

de k_{ts} pour les différents tissus de l'organisme humain, citées dans [Hug10], nous distinguons 3 compartiments : le compartiment des tissus lents ($k_{C_1} = 0.002min^{-1}$), le compartiment des tissus moyens ($k_{C_2} = 0.02min^{-1}$) et le compartiment des tissus rapides ($k_{C_3} = 0.2min^{-1}$).

Un seul rayon par recrutement

Nous avons vu dans le paragraphe 4.1.2 qu'à chaque instant $t \geqslant t_{min}$ où $P_C^{ss\_max}$ atteint un nouveau maximum, de nouveaux noyaux-gazeux sont activés. Afin de pouvoir déterminer le volume total de gaz inerte libéré par l'organisme, nous devrons déterminer le nombre et les rayons des bulles gazeuses générées. La détermination de l'évolution des rayons des différents bulles gazeuses à l'instant t revient à résoudre un système de taille $N_C^{tot}(t)$ d'équations différentielles du type de l'équation (4.6). Sachant que le nombre $N_C^{tot}(t)$ peut être très grand ($N_C^{tot}(t)$ dépasse parfois 10^8), il est impossible de traiter l'évolution du rayon des bulles gazeuses une par une.

Il est suggéré dans [Hug10] de considérer que les micro-noyaux activés à un instant t ont tous le même rayon initial. Nous notons ce rayon $r_C^*(t)$. Une première approximation consiste à confondre $r_C^*(t)$ avec $r_C^0(t)$ (rayon du plus petit micro-noyaux activé à l'instant t). Cette approximation repose sur le fait que les plus petits micro-noyaux sont les plus nombreux, mais cela implique un biais concernant le volume total des micro-noyaux forcément inférieur à sa vraie valeur. Nous proposons donc de choisir $r_C^*(t)$ de façon à ce que ce volume soit initialisé à sa vraie valeur.

Reprenons l'exemple de la Figure 4.2. Les premiers micro-noyaux gazeux activés à l'instant t_{min} sont représentés par des bulles bleues. Le volume total de ces micro-noyaux est $\int_{r_C^0(t_1)}^{+\infty} \frac{4\pi}{3} r^3 \mu(r) dr$. Nous choisissons donc $r_C^*(t)$ de façon à ce que

$$\frac{4\pi}{3} r_C^*(t_{min})^3 \int_{r_C^0(t_1)}^{+\infty} \mu(r) dr = \int_{r_C^0(t_1)}^{+\infty} \frac{4\pi}{3} r^3 \mu(r) dr. \tag{4.13}$$

Les micro-noyaux qui viennent d'être activés à l'instant t_1 sont représentés dans la Figure 4.2 par des bulles rouges. De même, le choix de $r_C^*(t_1)$ repose sur :

$$\frac{4\pi}{3} r_C^*(t_1)^3 \int_{r_C^0(t_1)}^{r_C^0(t_{min})} \mu(r) dr = \int_{r_C^0(t_1)}^{r_C^0(t_{min})} \frac{4\pi}{3} r^3 \mu(r) dr.$$

Cela nous permet de traiter une seule dynamique pour toutes les bulles recrutées à un instant donnée.

Algorithme du simulateur

Les paramètres du simulateur de v_C^b sont : le paramètre k_C caractérisant le compartiment C plus les deux paramètres $N^{\max}$ et A de la distribution des rayons des micro-noyaux. La variable d'entrée est la pression ambiante instantanée P_{amb}, commune aux différents compartiments. A chaque instant t, l'état d'un compartiment C est caractérisé par $P_C(t)$ la pression de gaz inerte dans ce compartiment, $P_C^{ss\_max}(t)$ le niveau maximal de sursaturation et $n_C^b(t)$ le nombre de môles de gaz inerte

que renferment les bulles dans C. Le diagramme de la Figure 4.A.1 montre comment sont utilisées les équations du modèle biophysique de décompression afin d'obtenir $v_C^b(t,\cdot,\cdot)$. Faire appel au simulateur pour les différents compartiments et tout au long de la phase de décompression jusqu'à résorption des bulles gazeuses nous permet de construire le profil de volume $t \mapsto v^b(t,\cdot,\cdot)$.

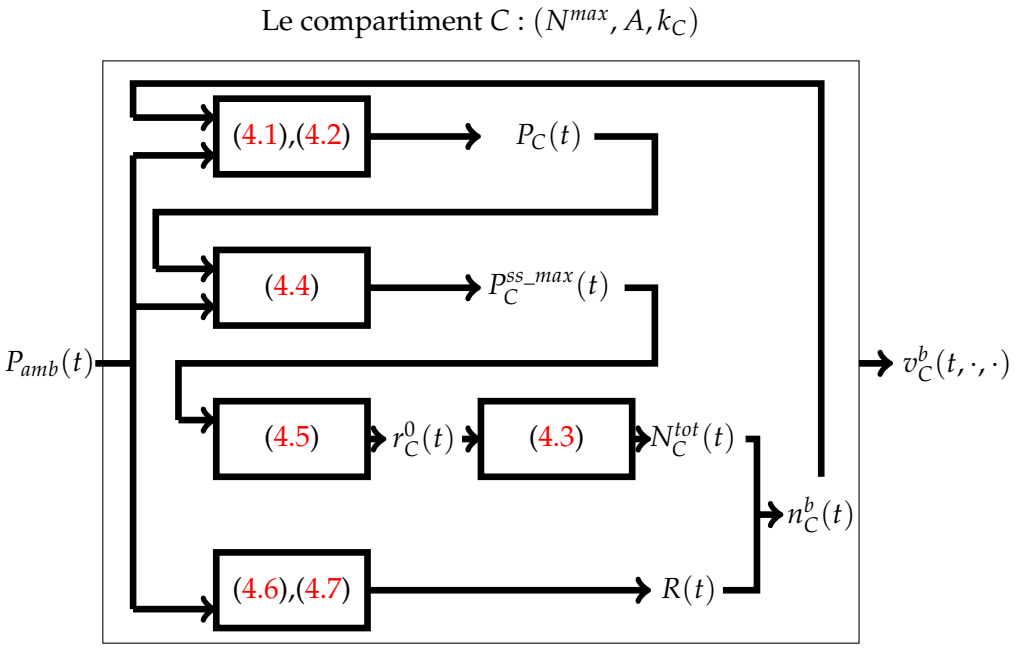


FIGURE 4.A.1 – *Diagramme de calcul de $v_C^b(t,\cdot,\cdot)$ à partir des équations du modèle biophysique de décompression pour un compartiment C.*

Décrivons l'algorithme du simulateur de v_C^b. Nous fixons un pas de discrétisation de temps Δt. Nous résolvons analytiquement la cinétique des échanges gazeux (4.1). Tant que $P_C^{ss\_max}(t) \leqslant 0$, la pression tissulaire $P_C(t)$ est confondue avec la solution analytique de (4.1). Sachant que la pression ambiante est supposée linéaire par morceaux en fonction du temps, plaçons nous sur un intervalle $[s_1, s_2]$ où elle est linéaire :$P_{amb}(t) = at + b$. Nous avons donc :

$$\frac{dP_C(t)}{dt} = -k_C P_C(t) + k_C fat + k_C f(b - P_{H_2O}).$$

Il s'agit d'une équation différentielle linéaire du premier ordre dont la solution pour $t \in [s_1, s_2]$ est la suivante :

$$P_C(t) = \left[P_C(s_1) - fas_1 - f(b - P_{H_2o}) + \frac{fa}{k_C}\right] e^{-k_C(t-s_1)} + fat + f(b - P_{H_2o}) - \frac{fa}{k_C}.$$

Cette solution reste valable jusqu'à t_{min} le premier instant où $P_C^{ss\_max}$ devient strictement positive. Après cet instant, il y a activation de noyaux gazeux, et donc la pression tissulaire devient solution de l'équation (4.2).

D'après l'équation (4.13), les micro-noyaux recrutés à t_{min} ont tous le même rayon initial $r^*(t_{min})$. Le nombre de ces micro-noyaux est $N_C^{tot}(t_{min})$. Ainsi le volume de gaz séparé par unité de volume à l'instant t_{min} est :

$$v_C^b(t_{min}) = \frac{4\pi}{3} N_C^{tot}(t_{min}) r_C^*(t_{min})^3.$$

Selon la loi des gaz parfaits, le nombre de môles de gaz inerte que renferme l'ensemble de ces bulles gazeuses à l'instant t_{min} est :

$$n_C^b(t_{min}) = \frac{V_C}{\mathfrak{R} T_C} P_C^b(t_{min}) v_C^b(t_{min}).$$

Notons $\frac{dR_{t_{min}}}{dt}$ la valeur à t_{min} de la dérivée du rayon des bulles recrutées à t_{min}. Par dérivation des deux équations précédentes, nous déduisons la quantité :

$$\begin{aligned}\frac{dn_C^b}{dt}(t_0) =& \frac{4\pi V_C}{3\mathfrak{R} T_C} N_C^{tot}(t_{min}) r_C^*(t_{min})^2 \times \left[3\frac{dR_{t_{min}}}{dt} \left(P_{amb}(t_{min}) + \frac{2\gamma_C}{r_C^*(t_{min})} - \beta \right) \right. \\ &\left. + r_C^*(t_{min}) \left(\frac{dP_{amb}}{dt} - \frac{2\gamma_C}{r_C^*(t_{min})^2} \frac{dR_{t_{min}}}{dt} \right) \right].\end{aligned}$$

Juste avant l'instant $t_1 = t_{min} + \Delta t$, toutes les bulles ont le même rayon $R_{t_{min}}^{t_1}$. Nous intégrons numériquement le système d'équations différentielles issu des équations (4.2) et (4.6) sur l'intervalle $[t_{min}, t_1]$ afin d'obtenir $R_{t_{min}}^{t_1}$. Le volume de gaz séparé par unité de volume dans le compartiment C jusqu'à l'instant t_1 est :

$$v_C^b(t_1) = \frac{4\pi}{3} N_C^{tot}(t_{min}) (R_{t_{min}}^{t_1})^3.$$

Nous déduisons le nombre de môles de gaz inerte que renferme l'ensemble des bulles gazeuses dans le compartiment C juste avant l'instant t_1 :

$$n_C^b(t_1) = \frac{V_C}{\mathfrak{R} T_C} P_C^b(t_1) v_C^b(t_1).$$

Nous notons par $\frac{dR_{t_{min}}^{t_1}}{dt}$ la valeur à t_1 de la dérivée du rayon des bulles gazeuses recrutées à t_{min}. Par dérivation des deux équations précédentes, nous déduisons

$$\begin{aligned}\frac{dn_C^b}{dt}(t_1) =& \frac{4\pi V_C}{3\mathfrak{R} T_C} N_C^{tot}(t_{min}) (R_{t_{min}}^{t_1})^2 \times \\ &\left[3\frac{dR_{t_{min}}^{t_1}}{dt} \left(P_{amb}(t_1) + \frac{2\gamma_C}{R_{t_{min}}^{t_1}} - \beta \right) + R_{t_{min}}^{t_1} \left(\frac{dP_{amb}}{dt} - \frac{2\gamma_C}{(R_{t_{min}}^{t_1})^2} \frac{dR_{t_{min}}^{t_1}}{dt} \right) \right].\end{aligned}$$

Nous évaluons $P_C^{ss\_max}$ à l'instant t_1. Si $P_C^{ss\_max}(t_1) \leqslant P_C^{ss\_max}(t_{min})$, il n'y a pas de nouveaux micro-noyaux gazeux activés à t_1. Le rayon $R_{t_{min}}^{t_1}$ des bulles déjà existantes continue à évoluer et prend la

valeur $R_{t_{min}}^{t_2}$ à l'instant $t_2 = t_1 + \Delta t$.

Si maintenant $P_C^{ss\_max}(t_1) > P_C^{ss\_max}(t_{min})$, de nouveaux micro-noyaux gazeux sont activés. Le nombre de ces nouveaux micro-noyaux est $N_C^{tot}(t_1) - N_C^{tot}(t_{min})$. Là encore nous supposons que toutes les bulles gazeuses issues de ces nouveaux micro-noyaux auront la même dynamique et que leur rayon initial est $r_C^*(t_1)$. A l'instant t_2, nous avons deux ensembles de bulles, celles de rayon $R_{t_{min}}^{t_2}$ (recrutées à t_{min}) et celles de rayon $R_{t_1}^{t_2}$ (recrutées à t_1). Les volumes de gaz séparé à l'instant t_2 par unité de volume, correspondant à ces deux ensembles de bulles, sont respectivement :

$$v_C^{b,t_{min}}(t_2) = \frac{4\pi}{3} N_C^{tot}(t_{min})(R_{t_{min}}^{t_2})^3,$$

et

$$v_C^{b,t_1}(t_2) = \frac{4\pi}{3}(N_C^{tot}(t_1) - N_C^{tot}(t_{min}))(R_{t_1}^{t_2})^3.$$

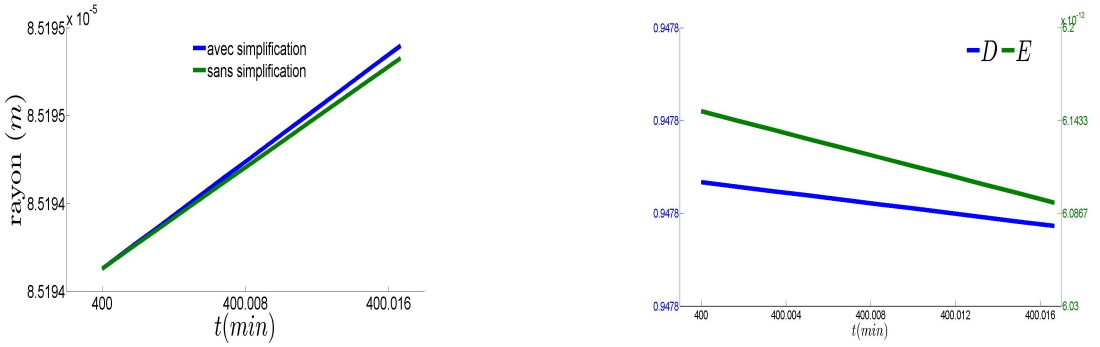
Le nombre de môles que renferme l'ensemble des microbulles à l'instant t_2 est donc :

$$n_C^b(t_2) = \frac{4\pi V_C}{3\mathfrak{R}T_C} \times \left[\left(P_{amb}(t_2) + \frac{2\gamma_C}{R_{t_{min}}^{t_2}} - \beta \right) v_C^{b,t_{min}}(t_2) + \left(P_{amb}(t_2) + \frac{2\gamma_C}{R_{t_1}^{t_2}} - \beta \right) v_C^{b,t_1}(t_2) \right].$$

Nous dérivons la dernière équation et nous déduisons de la même façon la pression tissulaire $P_C(t_2)$. Ainsi, nous continuons à résoudre pas à pas la dynamique décrite par les équations (4.2) et (4.6) jusqu'à la résorption de toutes les bulles gazeuses.

4.B RÉDUCTION DU TEMPS DE CALCUL DU SIMULATEUR

Afin de réduire le temps de calcul nécessaire pour la simulation du volume de gaz libéré sous forme de bulles, nous proposons deux démarches : la simplification des équations du modèle bio-physique et le choix d'un pas de discrétisation optimal. L'équation (4.6) décrit la dynamique des rayons des bulles de gaz inerte libéré. Il s'avère que la simulation de cette équation est très couteuse en termes de temps de calcul. Nous en proposons ici une simplification qui réduira sensiblement le temps d'exécution des simulations. Nous avons vu dans la table 4.3 l'impact du pas de discrétisation Δt sur le temps de calcul pour l'exemple de plongée 10m/265min avec un seul jeu de paramètres. Nous pouvons donc choisir un pas de discrétisation de simulation Δt^*, garantissant la pertinence des résultats tout en permettant un temps de calcul raisonnable. Afin de valider l'approximation apportée à l'équation (4.6) et le choix d'un pas de discrétisation optimal, nous définissons une grille de calcul composée de plusieurs expositions.



(a) *Impact sur un rayon.* **(b)** *Variation de D et de E.*

FIGURE 4.B.1 – *Impact de l'approximation de la dynamique du rayon sur la plus grande bulle activée lors de l'exemple de la plongée 10m/265min, sur un intervalle d'une seconde.*

Approximation de la dynamique des rayons

La simulation de l'équation (4.6) est très couteuse en terme de temps de calcul à cause de son caractère fortement non-linéaire. Afin d'accélérer la simulation de cette équation, nous allégeons cette dynamique en apportant 2 approximations. Nous supposons qu'entre 2 instants t_1 et t_2, et en partant d'une valeur initiale $R(t_1)$, nous avons :

- le terme $P_{amb} + \frac{4\gamma_C}{3R} - P_{H_2o}$ reste constant entre t_1 et t_2, et vaut : $P_{amb}(t_1) + \frac{4\gamma_C}{3R(t_1)} - P_{H_2o} = D$,
- le terme $\Re T_C D_C S_C (P_C - P_{amb} + \beta - \frac{2\gamma_C}{R})$ reste constant entre t_1 et t_2, et vaut : $\Re T_C D_C S_C (P_C(t_1) - P_{amb}(t_1) + \beta - \frac{2\gamma_C}{R(t_1)}) = E$.

La Figure 4.B.1a montre l'impact de cette approximation sur le rayon de la plus grande bulle activée lors de l'exemple traité dans le paragraphe 4.1.4 sur un intervalle de temps d'une seconde. Dans cette exemple, la différence entre la valeur du rayon selon la dynamique de l'équation (4.6) à l'instant $t_2 = t_1 + 1s$ et la valeur du rayon en considérant l'approximation présentée ci-dessus au même instant est inférieur à $10^{-9}m$. La Figure 4.B.1b montre de petites variations des quantités D et E sur le même intervalle de temps d'une seconde selon la dynamique (4.6), ce qui appuie l'approximation en question. Nous avons ainsi

$$\frac{D(t_2) - D(t_1)}{D(t_1)} = -5.10^{-7} \text{ et } \frac{E(t_2) - E(t_1)}{E(t_1)} = -0.009.$$

En adoptant cette approximation, la dynamique du carré du rayon vérifie une équation différentielle linéaire du premier ordre :

$$\frac{dR^2}{dt} = \frac{2(E - \frac{R^2}{3}\frac{dP_{amb}}{dt})}{D}. \tag{4.14}$$

Pour $t \in [t_1, t_2]$, la solution est :

$$R(t) = \begin{cases} \sqrt{\frac{2E}{D}(t-t_1) + R^2(t_1)} & \text{si } \dot{p} = 0 \\ \sqrt{(R^2(t_1) - \frac{3E}{\dot{p}})e^{-\frac{2\dot{p}}{3D}(t-t_1)} + \frac{3E}{\dot{p}}} & \text{sinon} \end{cases},$$

avec $\dot{p} = \frac{dP_{amb}}{dt}$. Nous appelons dans la suite *version détaillée* du simulateur celle utilisant l'équation (4.6) pour la dynamique de rayon des bulles et *version simplifiée* celle utilisant l'approximation donnée par l'équation (4.14). Nous vérifions ensuite que cette approximation permet d'accélérer les calculs sans nuire à la pertinence des résultats en testant un grand nombre d'exemple que nous définissons ultérieurement.

Choix du pas de discrétisation optimal

Le pic du volume du gaz libéré est la quantité qui nous permettra par la suite de lier les observations aux valeurs des paramètres du modèle biophysique. Nous nous basons donc sur $\|v_C^b\|_\infty$ le pic de volume de gaz libéré par un compartiment C pour choisir le pas de discrétisation optimal. Afin d'établir une règle pour le choix d'un pas de discrétisation optimal de l'axe des temps, nous définissons un ensemble d'exemples, que nous appelons *grille de calcul*, composé de plusieurs valeurs des paramètres biophysiques N^{max}, A, k_C, d'un ensemble J de profils de plongée et de plusieurs pas de discrétisation $\Delta t_0 < \cdots < \Delta t_n$. Nous reprenons la définition de l'erreur relative $ER^{v_C^b}_{\Delta t_0}(\Delta t_1)$ du volume v_C^b calculé avec un pas de discrétisation Δt_1 par rapport à celui calculé avec Δt_0 donnée par l'équation (4.9). Le choix du pas de discrétisation optimal Δt^* ne doit dépendre que des valeurs des paramètres biophysiques (N^{max}, A, k_C) afin de pouvoir appliquer ce choix à des profils de plongée autres que ceux retenus dans la grille de calcul. Pour une valeur du paramètre k_C et des paramètre $\theta = (N^{max}, A)$, et pour un profil de plongée j, nous définissons l'ensemble $D^{5\%}_{j,\theta,C}$ des pas de discrétisation assurant une erreur relative inférieure à 5% par rapport au pas de discrétisation le plus petit

$$D^{5\%}_{j,\theta,C} = \{\Delta t_i : ER^{v_C^b(\cdot,\theta,j)}_{\Delta t_0}(\Delta t_i) \leqslant 5\%\}.$$

En se limitant à l'ensemble des profils de plongée J dans la grille de calcul, et en testant les pas de discrétisation de la grille $\Delta t_1 < \cdots < \Delta t_n$ par rapport à Δt_0, le pas de discrétisation optimal retenu pour chaque valeur de θ et pour chaque compartiment, est définit par

$$\Delta t^*(\theta, C) = \min_{j \in J}\{\max(D^{5\%}_{j,\theta,C})\}. \tag{4.15}$$

Nous notons $G^{j,\theta,C}_{\Delta t_0}(\Delta t_i)$ le gain en temps de calcul pour le profil j, le paramètre θ et le compartiment C, de Δt_i par rapport à Δt_0

$$G^{j,\theta,C}_{\Delta t_0}(\Delta t_i) = 1 - \frac{Cout(v_C^b(\cdot,\theta,j), \Delta t_i)}{Cout(v_C^b(\cdot,\theta,j), \Delta t_0)},$$

avec $Cout(v_C^b(\cdot,\theta,j),\Delta t_i)$ le temps de calcul qu'a mis le simulateur pour simuler $v_C^b(\cdot,\theta,j)$ avec le pas de discrétisation Δt_i.

Validation

Nous définissons une grille de calcul afin de valider d'une part le choix de pas de discrétisation optimal Δt^* et d'autre part l'approximation apportée à la dynamique du rayon d'une bulle gazeuse donnée dans le paragraphe 4.B. La grille choisie est composée de

- **8 pas de discrétisation** : $\Delta t_0 = 0.25s$ et $\Delta t_i = 0.5s$, $1s$, $2s$, $5s$, $10s$, 30s et $1min$,
- **6 expositions** :
  - 6m/360min, Pas de palier
  - 18m/90min, un palier de 23min à 3m
  - 32m/40min, deux paliers : 1min à 6m et 29min à 3m
  - 48m/30min, trois paliers : 1min à 9m, 12min à 6m et 37min à 3m
  - 65m/10min, deux paliers : 3min à 6m et 8min à 3m
  - 60m/55min, 5 paliers : 5min à 15m, 15min à 12m, 24min à 9m, 40min à 6m, 88min à 3m
- **3 valeurs pour chacun des paramètres** N^{max} et A couvrant leurs valeurs possibles
  - 3 valeurs de N^{max} : 5.10^6, 5.10^8 et 5.10^{10}
  - 3 valeurs de A : 10^5, 10^6 et 10^7
- **3 compartiments** comme définis auparavant.

Les simulations des cas définis par cette grille de calcul ont été réalisées en utilisant les versions détaillée et simplifiée du simulateur. Au total, 2592 simulations ont été effectuées.

Les tables 4.B.1 et 4.B.2 donnent respectivement pour la version détaillée et la version simplifiée du simulateur, et pour chaque valeur du triplet (N^{max}, A, k_C) de la grille de calcul, le pas de discrétisation optimal Δt^*, ER^* la pire erreur relative du pas de discrétisation optimal par rapport au plus petit pas donnée par

$$ER^* = \max_{j\in J} ER_{\Delta t_0}^{v_C^b(\cdot,\theta,j)}(\Delta t^*),$$

et G^* le gain moyen en temps de calcul de Δt^* par rapport à Δt_0 :

$$G^* = \frac{1}{\#J}\sum_{j=1}^{\#J} G_{\Delta t_0}^{j,\theta,C}(\Delta t^*).$$

Les tables 4.B.1 et 4.B.2 indiquent que le paramètre le plus influant dans le choix de Δt^* est la constante d'échange de gaz k_C. Plus le compartiment est rapide pour évacuer le gaz inerte qu'il contient (k_C grand), plus le pas optimal Δt^* est petit. Cela nous semble logique car pour un compartiment dont le k_C est grand, les changements de pression sont rapides et donc le phénomène d'activation des noyaux-gazeux peut aller très vite. Avec un pas de discrétisation qui n'est pas suffisamment fin, le risque d'omettre une importante partie du gaz libéré est grand.

N^{max}		5.10^6			5.10^8			5.10^{10}		
A		10^5	10^6	10^7	10^5	10^6	10^7	10^5	10^6	10^7
$k_C = 0.2$	$\Delta t^*(s)$	2	2	2	2	2	2	1	1	2
	$ER^*(\%)$	3	5	3	4	4	3	3	3	2
	$G^*(\%)$	96	96	96	96	96	96	90	90	96
$k_C = 0.02$	$\Delta t^*(s)$	2	2	2	2	2	2	2	1	2
	$ER^*(\%)$	3	3	3	4	4	3	4	3	4
	$G^*(\%)$	94	94	93	94	94	94	93	85	94
$k_C = 0.002$	$\Delta t^*(s)$	30	30	30	10	30	30	5	10	30
	$ER^*(\%)$	5	1	1	2	1	1	4	1	1
	$G^*(\%)$	99	99	99	98	99	99	64	98	99

TABLE 4.B.1 – *Pas de discrétisation optimal, pire erreur relative et gain moyen en temps de calcul pour les 27 triplets des paramètres N^{max}, A et k_C avec la version détaillée du simulateur.*

La moyenne de temps de calcul nécessaire pour une simulation avec la version détaillée et le pas de discrétisation le plus petit de la grille de calcul $\Delta t_0 = 0.25s$ est de $11min$. En appliquant le choix du pas optimal de discrétisation cette moyenne passe à $51s$ tout en assurant une erreur relative qui ne dépasse pas 5%. Concernant la version simplifiée, l'application du pas de discrétisation optimal permet de réduire le temps moyen d'une simulation de $52s$ à $1.2s$.

Afin de mesurer la pertinence de l'approximation de la dynamique du rayon d'une bulle (version simplifiée du simulateur) et le gain en temps de calcul qu'elle apporte, nous comparons les sorties des versions simplifiée et détaillée du simulateur sur la base des cas définis par la grille de calcul. Nous notons $v^b_{\Delta t_0}(\cdot,\theta,j)^{(s)}$ et $v^b_{\Delta t^*}(\cdot,\theta,j)^{(s)}$ les volumes de gaz libéré par le compartiment C, simulé par la version simplifiée du simulateur respectivement avec les pas de discrétisation $\Delta t_0 = 0.25s$ et Δt^* (défini par la table 4.B.2) et $v^b_{\Delta t_0}(\cdot,\theta,j)^{(d)}$ le volume de gaz simulé par la version détaillée du simulateur avec le pas de discrétisation $\Delta t_0 = 0.25s$. Pour les 27 valeurs des triplets de paramètres N^{max}, A, k_C, nous calculons

$$ER1 = \max_{j\in J} \frac{|\|v^b_{\Delta t_0}(\cdot,\theta,j)^{(s)}\|_\infty - \|v^b_{\Delta t_0}(\cdot,\theta,j)^{(d)}\|_\infty|}{\|v^b_{\Delta t_0}(\cdot,\theta,j)^{(d)}\|_\infty},$$

$$ER2 = \max_{j\in J} \frac{|\|v^b_{\Delta t^*}(\cdot,\theta,j)^{(s)}\|_\infty - \|v^b_{\Delta t_0}(\cdot,\theta,j)^{(d)}\|_\infty|}{\|v^b_{\Delta t_0}(\cdot,\theta,j)^{(d)}\|_\infty},$$

$$G1 = 1 - \frac{1}{\#J}\sum_{j=1}^{\#J} \frac{Cout(v^b_C(\cdot,\theta,j)^{(s)},\Delta t_0)}{Cout(v^b_C(\cdot,\theta,j)^{(d)},\Delta t_0)},$$

N^{max}		5.10^6			5.10^8			5.10^{10}		
A		10^5	10^6	10^7	10^5	10^6	10^7	10^5	10^6	10^7
$k_C = 0.2$	$\Delta t^*(s)$	1	2	1	2	2	1	2	2	1
	$ER^*(\%)$	2	5	2	5	5	3	3	3	3
	$G^*(\%)$	92	97	92	97	97	93	97	97	93
$k_C = 0.02$	$\Delta t^*(s)$	5	5	5	10	5	5	10	10	5
	$ER^*(\%)$	2	3	3	3	2	4	2	3	4
	$G^*(\%)$	99	99	99	99	99	99	99	99	99
$k_C = 0.002$	$\Delta t^*(s)$	10	10	30	10	10	30	10	10	30
	$ER^*(\%)$	3	1	1	3	1	1	2	1	1
	$G^*(\%)$	99	99	99	99	99	99	99	99	99

TABLE 4.B.2 – *Pas de discrétisation optimal, pire erreur relative et gain moyen en temps de calcul pour les 27 triplets des paramètres N^{max}, A et k_C avec la version simplifiée du simulateur.*

$$G2 = 1 - \frac{1}{\#J}\sum_{j=1}^{\#J}\frac{Cout(v_C^b(\cdot,\theta,j)^{(s)},\Delta t^*)}{Cout(v_C^b(\cdot,\theta,j)^{(d)},\Delta t_0)},$$

Les résultats sont présentés dans la table 4.B.3.

D'après la table 4.B.3, l'erreur relative $ER2$ ne dépasse pas 5% tout en assurant un gain moyen $G2$ supérieur à 99% pour tous les triplets de la grille de calcul. Nous concluons qu'avec la version simplifiée du simulateur, nous réussissons à prédire correctement le volume de gaz inerte libéré après décompression tout en réduisant considérablement le temps de calcul.

4.C IDENTIFIABILITÉ

Identifiabilité

Avant d'estimer la distribution des paramètres biophysiques $(N^{\max}, A) = \theta$, nous nous intéressons à l'étude de l'identifiabilité de θ. L'objectif ici est de savoir si les mesures envisagées contiendront assez d'information pour l'estimation de θ.

Pour cette étude d'identifiabilité, nous nous plaçons dans un cadre idéalisé où nous pouvons effectuer, sous des conditions expérimentales arbitraires (ici des profils de plongées), un nombre d'expériences arbitrairement grand. Il est clair que dans les vraies expériences nous sommes limités par des contraintes liées à la dangerosité des profils de plongée.

N^{max}		5.10^6			5.10^8			5.10^{10}		
A		10^5	10^6	10^7	10^5	10^6	10^7	10^5	10^6	10^7
$k_C = 0.2$	$ER1(\%)$	2	1	1	1	1	1	1	2	1
	$G1(\%)$	88	87	87	87	86	88	86	86	86
	$ER2(\%)$	3	2	1	4	2	1	2	1	1
	$G2(\%)$	99	99	99	99	99	99	99	99	99
$k_C = 0.02$	$ER1(\%)$	2	2	1	1	1	1	1	2	1
	$G1(\%)$	93	94	94	94	94	94	92	93	94
	$ER2(\%)$	3	3	5	1	1	5	1	2	2
	$G2(\%)$	99	99	99	99	99	99	99	99	99
$k_C = 0.002$	$ER1(\%)$	1	4	5	3	1	5	3	5	5
	$GC1(\%)$	92	92	92	91	91	92	90	92	92
	$ER2(\%)$	1	4	3	3	1	3	3	5	3
	$GC2(\%)$	99	99	99	99	99	99	99	99	99

TABLE 4.B.3 – *Comparaison du volume de gaz libéré calculé par les versions détaillée et simplifiée du simulateur*

D'après les définitions relatives à l'identifiabilité d'un modèle paramétrique citées dans [WP94], θ est dit *globalement identifiable* si pour tout θ', il existe un profil de plongée P tel que $G(\theta', P) \neq G(\theta, P)$. De même, nous disons que θ est *localement identifiable* s'il existe un voisinage $\mathcal{V}(\theta)$ de θ tel que pour tout $\theta' \in \mathcal{V}(\theta), \theta' \neq \theta$, il existe un profil de plongée P pour lequel on a $G(\theta', P) \neq G(\theta, P)$. Lorsque ces propriétés sont vraies pour presque tout $\theta \in \Theta$, nous disons que l'identifiabilité est satisfaite de façon structurelle globalement ou localement sur Θ.

Seuils connus

Supposons que les seuils τ définissant les mesures de grades observés (équation (4.10)) sont connus. Pour un seuil τ_ℓ et pour tout $\theta \in \Theta$, nous construisons à l'aide du simulateur deux familles de profils de plongées $h_\theta^{\tau_\ell}$ et H_θ^ℓ telles que $\|v^b(\theta, P)\|_\infty = \tau_\ell$ pour tout $P \in h_\theta^{\tau_\ell}$ et $G(\theta, P) = \ell$ pour tout $P \in H_\theta^\ell$. Nous devons vérifier ensuite que pour tout $\theta' \neq \theta$, il existe $P' \in H_\theta^\ell$ tel que $G(\theta', P) \neq \ell$.

Nous illustrons cela par une famille paramétrée de profils de plongées : $H = \{P(T, Z) : 5min \leqslant T \leqslant 120min, 6m \leqslant Z \leqslant 220m\}$, où $P(T, Z)$ correspond à un profil de plongé carré (sans paliers), de durée T, de profondeur Z et avec des vitesses de descente et de remontée fixes. Pour tout θ et pour tout seuil τ_ℓ, nous pouvons ajuster la profondeur Z en fonction de T afin de maintenir $\|v^b(\theta, P(T, Z))\|_\infty$ à la valeur τ_ℓ. Nous obtenons alors la famille $h_\theta^{\tau_\ell} = \{P(T, Z) \in H : \|v^b(\theta, P(T, Z))\|_\infty =$

$\tau_\ell\}$.

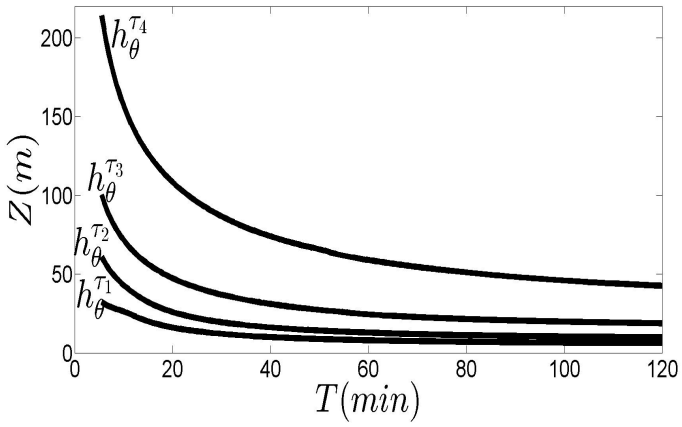


FIGURE 4.C.1 – *Représentation dans le plan (T, Z) de la famille paramétrique h_θ^τ pour 4 valeurs de τ, $A = 10^6 m^{-1}$ et $N^{\max} = 5.10^8 m^{-3}$.*

Dans la figure 4.C.1, les 4 courbes correspondent aux 4 familles de profils : $h_\theta^{\tau_\ell}, \ell = 1, \ldots, 4$. A partir de la relation qui lie les seuils aux grades (4.10), nous déduisons que

$$G(\theta, h_\theta^{\tau_\ell}) = \ell.$$

Même si nous n'observons pas $\|v^b(\theta, \cdot)\|_\infty$ pour tous les profils carrés des familles H_θ^ℓ, nous pouvons déduire leurs grades en les situant par rapport aux courbes $h_\theta^{\tau_\ell}, \ell = 1, \ldots, 4$. Cette déduction est faisable en se basant sur la croissance stricte de $\|v^b(\theta, \cdot)\|_\infty$ par rapport à T et Z que nous avons remarqué dans le paragraphe sur la sensibilité (4.1.4.2). Ainsi nous construisons $\{H_\theta^0, \ldots, H_\theta^4\}$ une partition de H telle que :

$$H_\theta^0 = \{P(T, Z) \in H : Z < h_\theta^{\tau_1}(T)\},$$

$$H_\theta^\ell = \{P(T, Z) \in H : h_\theta^{\tau_\ell}(T) \leqslant Z < h_\theta^{\tau_{\ell+1}}(T)\}, \ell = 1, 2, 3$$

$$H_\theta^4 = \{P(T, Z) \in H : Z \geqslant h_\theta^{\tau_4}(T)\},$$

Nous considérons maintenant une valeur $\theta' \neq \theta$ du paramètre biophysique. Si nous montrons que $\{H_{\theta'}^0, \ldots, H_{\theta'}^4\} \neq \{H_\theta^0, \ldots, H_\theta^4\}$ alors θ sera identifiable. D'après le paragraphe (4.1.4.2), $\|v^b(\cdot, P)\|_\infty$ est strictement croissant par rapport à $N^{\max}$ et strictement décroissante par rapport à A. Donc si $\theta' = (N'^{\max}, A) \neq (N^{\max}, A) = \theta$ tel que $A \geqslant A'$ et $N'^{\max} \geqslant N^{\max}$, nous avons forcément $H_\theta'^0 \subsetneq H_\theta^0$. De même si $\theta' \neq \theta$ avec $A' \geqslant A$ et $N^{\max} \geqslant N'^{\max}$, nous avons $H_\theta{}^0 \subsetneq H_\theta'^0$. Il reste donc les cas où $\theta' \neq \theta$ avec $A \geqslant A'$ et $N^{\max} \geqslant N'^{\max}$ ou $A \leqslant A'$ et $N^{\max} \leqslant N'^{\max}$. La Figure 4.C.2 montre les familles $\{h_\theta^{\tau_\ell}\}$ et $\{h_{\theta'}^{\tau_\ell}\}$ pour deux valeurs différentes de θ. Nous voyons dans cet exemple que deux courbes $h_\theta^{\tau_\ell}$ et $h_{\theta'}^{\tau_\ell}$ se croisent au plus en un seul point. Pour que le modèle ne soit pas identifiable en un point θ, il faudrait qu'il existe un $\theta' \neq \theta$ tel que les deux courbes $h_{\theta'}^{\tau_\ell}$ et $h_\theta^{\tau_\ell}$ coïncident pour

$\ell = 1, \ldots, L$ (ici, L=4). Compte tenu de la complexité des équations du modèle biophysique, nous pouvons légitimement penser que cela ne se produit pas.

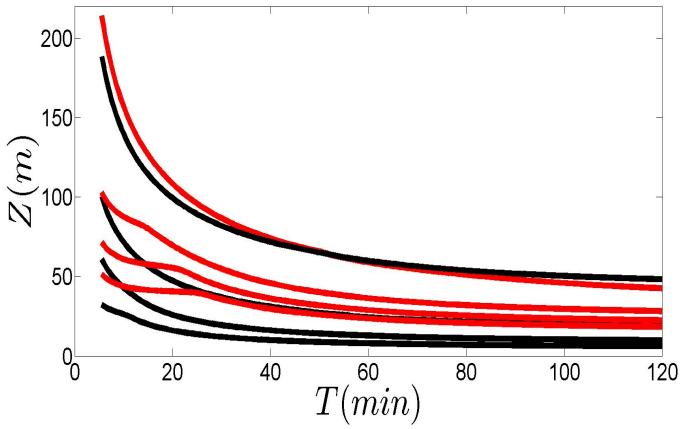


FIGURE 4.C.2 – *Représentation dans le plan (T, Z) des familles paramétriques $\{h_\theta^{\tau_\ell}\}$ en noir et $\{h_{\theta'}^{\tau_\ell}\}$ en rouge avec $\theta = (10^6, 5.10^8)$ et $\theta' = (10^7, 10^{10})$.*

Seuils inconnus

Supposons maintenant que les seuils τ sont inconnus. Nous fixons une valeur de θ. Nous ne pouvons plus alors choisir P afin d'ajuster $\|v^b(\theta, P)\|_\infty$ à la valeur d'un seuil τ_l puisque τ_l est inconnu. Néanmoins, même si nous n'observons pas $\|v^b(\theta, P)\|_\infty$, nous pouvons construire pour θ une famille de profils h_θ telle que $G(\theta, P)$ est maintenu sur le bord entre deux niveaux de grades ℓ et $\ell + 1$. Par exemple, nous choisissons un profil carré $P(T, Z)$ tel que $G(\theta, P(T, Z)) = \ell$. Nous fixons alors T et nous augmentons petit à petit Z jusqu'à obtenir le grade $\ell + 1$. La non-identifiabilité du modèle en θ se traduirait par le fait qu'il existe $\theta' \neq \theta$ tel que pour tout $P \in h_\theta$, $\|v^b(\theta', P)\|_\infty$ reste constant et $G(\theta, P) = G(\theta', P)$. La conclusion est la même qu'au paragraphe précédent.

Nous précisons que nous nous sommes limités dans cette étude d'identifiabilité aux profils carrés qui forment une famille extrêmement simple de profils. La considération de tous les profils possibles donne encore plus de variabilité et favorise donc l'identifiabilité de θ.

- CHAPITRE 5 -

ESTIMATION DE LA DENSITÉ DE $\theta = (N^{max}, A)$ À PARTIR DE DONNÉES DE GRADE DE PLONGÉE

Dans ce chapitre, nous appliquons les méthodes d'estimatoin de π_θ présentées dans les Chapitres 2 et 3 au problème de caractérisation d'une population de plongeurs hyperbares. Plus précisément nous allons nous intéresser à l'estimation de la densité des paramètres $\theta = (\theta_1, \theta_2) = (N^{max}, A)$ à partir des données de grades de plongées.

Comme nous l'avons déjà vu dans le chapitre précédent, il est connu que l'apparition des accidents de décompression est fortement corrélée à la présence de bulles de gaz dans le sang du plongeur. La capacité de prévoir correctement la probabilité que ce volume devienne excessivement élevé peut donc être exploitée pour établir des règles de sécurité en plongée, en évitant les profils de plongée (durée/profondeur) qui peuvent être dangereux pour une partie non négligeable de la population. C'est dans cette optique que l'estimation de la densité π_θ des paramètres biophysiques $\theta = (N^{max}, A)$ pour une population de plongeurs est importante.

Dans la section 5.1 nous discutons de la détermination des régions élémentaires, éléments de la partition $\mathcal{Q}^J$. La section 5.2 présente les résultats des méthodes d'estimation appliquées au jeu de données réelles dont nous disposons, montrant la suprématie de l'estimateur proposé dans cette étude par rapport aux estimateurs classiques. Nous précisons que les lois empiriques définies par le jeu de données dons nous disposons ne sont pas compatibles. Cela implique que l'estimateur MaxEnt-MV peut être différent des estimateurs du maximum de vraisemblance et MaxEnt avec le plus petit niveau de relaxation.

5.1 DÉTERMINATION DE LA PARTITION $\mathcal{Q}^J$

Nous sommes intéressés par l'estimation de la distribution de π_θ du paramètre biophysique $\theta = (\theta_1, \theta_2)$ du modèle de décompression [Hug10] présenté dans le Chapitre 4. Ce modèle biophysique

décrit l'évolution temporelle du volume de gaz instantané $v^b(t,\theta,P)$ contenu dans les bulles gazeuses circulant dans l'organisme d'un plongeur caractérisé par le paramètre θ, durant et après l'exécution d'un profil de décompression $(t \mapsto P(t))$ (voir la Figure 5.1a) :

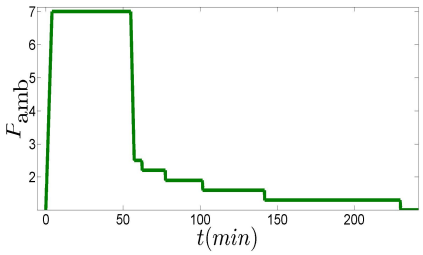
$$(\theta, (t \mapsto P(t))) \to \Big(t \mapsto v^b(t,\theta,P)\Big)\,. \tag{5.1}$$

La présence de gaz dans l'organisme du plongeur est observée par l'intermédiaire des grades de bulles qui sont une quantification du maximum du volume $\|v^b(\theta,P)\|_\infty$, comme exprimé par l'équation (4.10). Cette quantification implique que l'observation du paramètre θ suite à l'exécution d'un profil de plongée P est censurée par les régions de censure $\{\mathcal{R}_\ell^P\}_{\ell=1}^L$ où

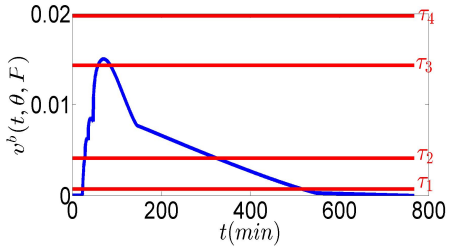
$$\mathcal{R}_\ell^P \triangleq \Big\{\theta \in \Theta \,:\, \|v^b(\theta,P)\|_\infty \in [\tau_{\ell-1},\tau_\ell[\Big\}\,,\quad \forall \ell = 1,\dots,L, \tag{5.2}$$

avec L le nombre des grades possibles. Les régions de censure issues d'un même profil P constituent une partition $\mathcal{R}^P = \{\mathcal{R}_\ell^P\}_{\ell=1}^L$ du domaine paramétrique Θ

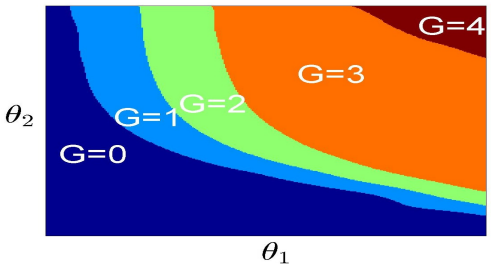
$$\bigcup_{\ell=1}^{L} \mathcal{R}_\ell^P = \Theta \qquad \text{et} \qquad \mathcal{R}_{\ell_1}^P \cap \mathcal{R}_{\ell_2}^P = \emptyset, \ell_1 \neq \ell_2\ .$$



(a) P_{amb}.



(b) $v^b(\cdot,\theta,P)$.



(c) $\{\mathcal{R}_\ell^P\}_{\ell=1}^{L=5}$.

FIGURE 5.1 – *Définition des grades de bulles G et des régions de censure $\mathcal{R}_\ell^P$. (**a**) un profil de plongée P ; (**b**) en bleu : volume de gaz $v^b(\cdot,\theta,P)$; en rouge : seuils $\{\tau_\ell\}_{\ell=1}^4$; (**c**) régions de censure $\{\mathcal{R}_\ell^P\}_{\ell=1}^{L=5}$ correspondant aux 5 grades de bulles.*

La Figure 5.1a affiche un profil de plongée indiquant la variation de la pression ambiante P_{amb} subie par le corps du plongeur en fonction du temps. La Figure 5.1b affiche le volume de gaz $t \mapsto$

$v^b(t, \theta, P)$ qu'a libéré l'organisme du plongeur, représenté par la courbe bleue. Les lignes horizontales en rouge indiquent les seuils de quantification $\{\tau_1, \ldots, \tau_4\}$ appliqués à $\|v^b(\theta, P)\|_\infty$. Dans notre cas $L = 5$, et les régions dans l'espace paramétrique Θ correspondant aux 5 valeurs possibles de grades pour le profil P sont représentées dans la Figure 5.1c. Dans cet exemple, nous observons un grade $G = 3$ puisque $\tau_3 \leqslant \|v^b(\theta, P)\|_\infty < \tau_4$, indiquant que la valeur du paramètre biophysique θ du plongeur en question appartient à la région orange.

La détermination pour chaque profil de plongée P^j dans $\left\{P^j\right\}_{j=1}^J$ des régions $\mathcal{R}_\ell^j \equiv \mathcal{R}_\ell^{P^j}$ pour $\ell = 1, \ldots, L$, nécessite le calcul de la sortie du modèle biophysique pour un grand nombre de valeurs de paramètres. Cependant, même après la simplification du simulateur du modèle biophysique (section 4.B), chaque calcul prend environ 30 secondes, ce qui reste très couteux au vu du nombre de valeurs de θ pour lesquelles nous voulons calculer la sortie du modèle. Pour éviter une utilisation excessive du simulateur du modèle biophysique, nous utilisons un méta-modèle basé sur le krigeage.

5.1.1 Métamodèle

Nous choisissons le krigeage simple car cela nous permet de prédire rapidement de nouvelles réponses du modèle biophysique avec une bonne précision. Fixons P^j un profil de plongée. Nous considérons que la sortie du simulateur $\|v^b(\cdot, P^j)\|_\infty$ est la réalisation d'une variable aléatoire Z :

$$Z(\theta) = m + W(\theta),$$

où m est une moyenne inconnue et $W(\cdot)$ un processus gaussien centré avec une variance σ^2 et une fonction de corrélation R :

$$Cov(W(\theta), W(\theta')) = \sigma^2 R(\theta, \theta').$$

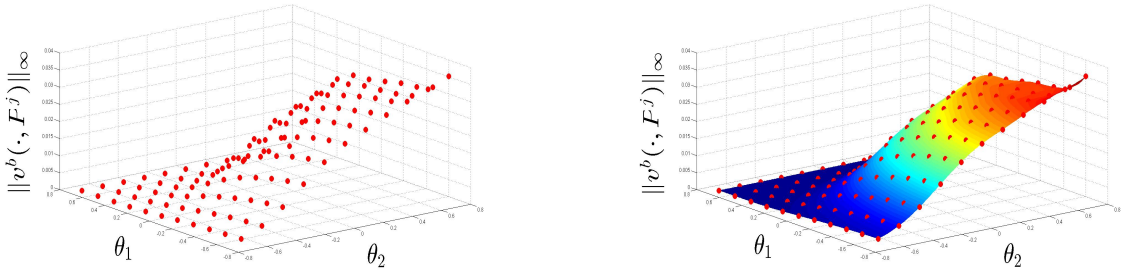
Considérons $\Theta_s = \left\{\theta^1, \ldots, \theta^N\right\}$, un ensemble de N points dans Θ où l'on calcule $\|v^b(\cdot, P^j)\|_\infty$ par le simulateur du modèle biophysique. Nous avons donc N réalisations de la variable aléatoire associée

$$Z(\Theta_s) = \left(Z(\theta^1), \ldots, Z(\theta^N)\right)^\mathsf{T}.$$

Nous avons : $Z(\Theta_s) \sim \mathcal{N}(m\mathbf{1}_N^\mathsf{T}, \Sigma_s)$ avec Σ_s la matrice de covariance : $(\Sigma_s)_{i,j} = \sigma^2 R(\theta^i, \theta^j)$. Considérons un nouveau point $\theta^\star \notin \Theta_s$.Pour prédire $\|v^b(\theta^\star, P^j)\|_\infty$ à partir de $Z(\Theta_s)$, nous utilisons la formule analytique de la loi de $Z(\theta^\star)|Z(\Theta_s)$ basée sur la théorie des processus gaussiens. Nous déduisons $\hat{Z}$ le prédicteur de Z du krigeage simple, qui interpole parfaitement les données $\hat{Z}(\Theta_s) = Z(\Theta_s)$. Il s'agit du meilleur prédicteur linéaire sans biais.

Nous choisissons une fonction de covariance de la classe Matérn

$$R(\theta, \theta') = R_\lambda(|\theta_1 - \theta_1'|, \rho_1) R_\lambda(|\theta_2 - \theta_2'|, \rho_2),$$



(a) **(b)**

FIGURE 5.2 – *Prédiction par krigeage simple de la réponse du simulateur du modèle biophysique pour un profil de plongée. (**a**) La réponse $\|v^b(\cdot, P^j)\|_\infty$ pour les points de Θ_s de taille 11×11. (**b**) Prédiction par krigeage simple de $\|v^b(\cdot, P^j)\|_\infty$ pour les points de $\hat{\Theta}$ de taille 295×295.*

avec

$$R_\lambda(|x|, \rho) = \frac{2^{1-\lambda}}{\Gamma(\lambda)} \left(\frac{\sqrt{2\lambda}}{\rho} |x| \right)^\lambda K_\lambda \left(\frac{\sqrt{2\lambda}}{\rho} |x| \right), \quad x \in \mathbb{R}, \lambda > 0, \rho > 0,$$

où Γ est la fonction gamma, K_λ la fonction de Bessel modifiée du second ordre [AS64], ρ est un paramètre d'échelle et λ est un paramètre positif qui définit la régularité du processus gaussien associé. En effet il est q fois dérivable presque sûrement si, et seulement si, $\lambda > q$ [Ste12]. D'après [Ste12], il est suggéré d'utiliser cette famille de fonctions de corrélation à cause de la flexibilité qu'offre le paramètre λ qui permet d'ajuster la régularité du processus.

La Figure 5.2a montre pour un profil de plongée P^j, la réponse $\|v^b(\cdot, P^j)\|_\infty$ du modèle biophysique sur les points de Θ_s, grille régulière de points de taille 11×11 (les points rouges) tandis que la Figure 5.2b montre la prédiction par krigeage simple de $\|v^b(\cdot, P^j)\|_\infty$ pour les points de $\hat{\Theta}$, grille régulière de points de Θ de taille 295×295 [1]. Nous choisissons $\lambda = \frac{5}{2}$; les résultats pour $\lambda = \frac{3}{2}$ ou $\frac{1}{2}$ restent quasiment les mêmes du fait du nombre relativement grand de points dans la grille de départ Θ_s.

5.1.2 Représentation des régions $\mathcal{R}_\ell^j$

Grâce à cette prédiction par krigeage simple, nous pouvons représenter les régions $\left\{\mathcal{R}_\ell^j\right\}_{j=1,\ell=1}^{J,L}$ par $\left\{\hat{\mathcal{R}}_\ell^j\right\}_{j=1,\ell=1}^{J,L}$ où $\hat{\mathcal{R}}_\ell^j \triangleq \left\{\theta \in \hat{\Theta};\ \theta \in \mathcal{R}_\ell^j\right\}$. A partir des régions $\left\{\hat{\mathcal{R}}_\ell^j\right\}_{j=1,\ell=1}^{J,L}$ nous pouvons déduire $\hat{\mathcal{Q}}^J = \left\{\hat{\mathcal{Q}}_m^J\right\}_{m=1}^{\hat{M}}$, une approximation des régions élémentaires $\left\{\mathcal{Q}_m^J\right\}_{m=1}^{M}$, éléments de la partition $\mathcal{Q}^J$ (Définition 2, page 10).

En effet, après avoir estimé $\|v^b(\cdot, P^j)\|_\infty$ pour les $\hat{N} \triangleq \#\left(\hat{\Theta}\right)$ points de $\hat{\Theta}$ et ce pour les J profils de plongée, nous disposons pour chaque point $\theta \in \hat{\Theta}$ de la liste $S(\theta)$ des grades "théoriques" des J

1. Le choix de la taille 295 pour $\hat{\Theta}$ est expliqué plus loin.

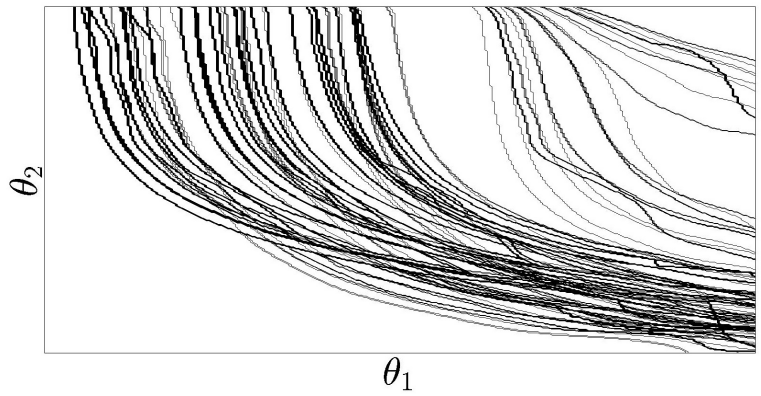


FIGURE 5.3 – *Partition $\hat{\mathcal{Q}}^J$ induite par les* 19 *profils de plongée du jeu de données réelles, contenant $\hat{M} = 665$ éléments.*

profils : $S(\theta) = \left\{\ell_\theta^1, \ldots, \ell_\theta^J\right\}$, avec $\ell_\theta^j = G(\theta, P^j)$ pour $j = 1, \ldots, J$ (voir l'équation (4.10)).

Pour déduire les régions élémentaires, il suffit de déterminer les ensembles de mots uniques au sein des $\hat{N}$ mots disponibles. Nous définissons donc $\hat{\mathcal{Q}}^J$ la partition de $\hat{\Theta}$ dont les éléments sont définis par $\hat{\mathcal{Q}}_m^J \triangleq \left\{\theta \in \hat{\Theta};\ \theta \in \mathcal{Q}_m^J\right\}$ pour $m = 1, \ldots, \hat{M}$. La Figure 5.3 représente les éléments de la partition $\hat{\mathcal{Q}}^J$ calculée à partir des 19 profils de plongée du jeu de données réelles. Nous obtenons $\hat{M} = 665$. La taille de $\hat{\Theta}$ a été fixée à 295×295 après l'avoir augmentée petit à petit jusqu'à 500×500 et remarqué que le nombre $\hat{M}$ reste le même à partir de la taille 295×295 [2].

Nous remarquons une forte dispersion des tailles des éléments de $\hat{\mathcal{Q}}^J$ dans le cas de données réelles, en particulier, une présence d'un grand nombre de régions à très faible surface. Remarquons que les éléments de $\hat{\mathcal{Q}}^J$ ont des formes fortement allongées qui sont sensiblement différentes des formes construites à partir de cellules de Voronoi, utilisées dans les simulations des chapitres 2 et 3.

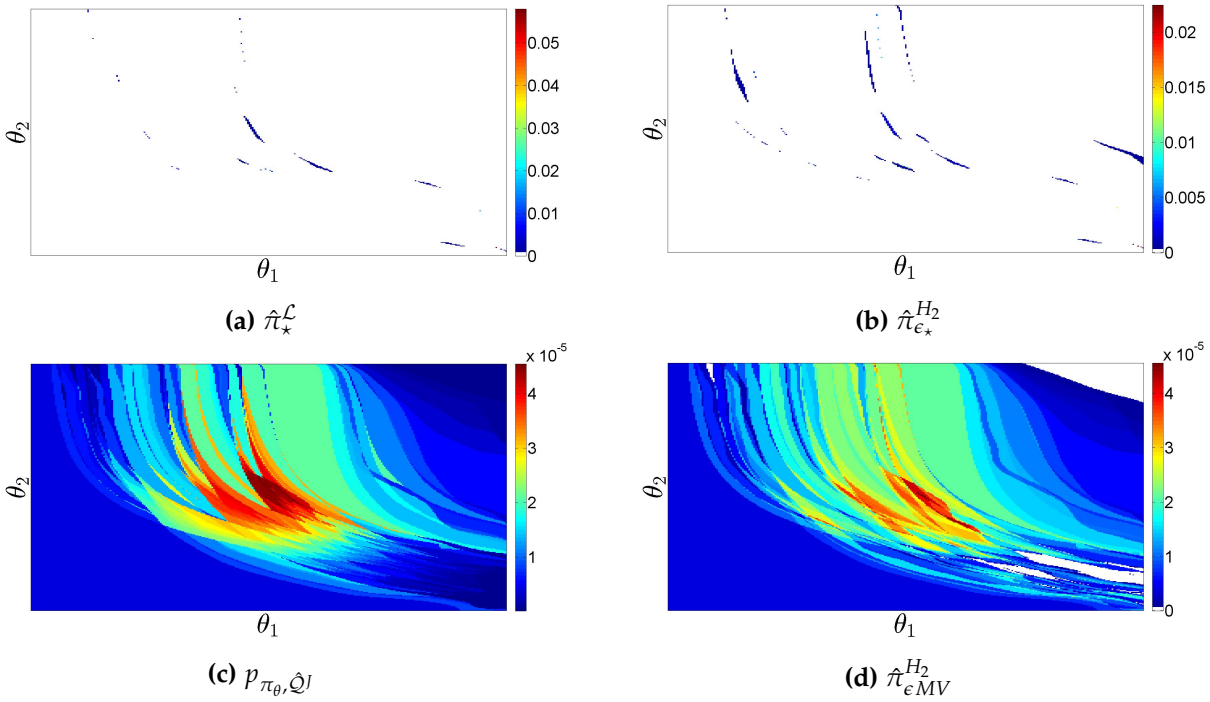
Construction de la matrice B : Nous rappelons la définition de la matrice $\mathbf{B}$ de taille $K \times M$ (équation (2.7)) comme étant la concaténation des matrices $\mathbf{B}^{(j)}$ dont la définition (équation (2.5)) dépend des régions de censure $\left\{\mathcal{R}_\ell^j\right\}_{\ell=1}^L$ et des régions élémentaires $\left\{\mathcal{Q}_m^J\right\}_{m=1}^M$. En utilisant l'approximation de Θ par $\hat{\Theta}$, nous posons $\mathbf{B}_{\ell m}^{(j)} = \mathbb{1}_{\hat{\mathcal{Q}}_m^J \subset \hat{\mathcal{R}}_\ell^j}$. La construction de la matrice $\mathbf{B}$ est ensuite possible en déterminant pour chaque région $\hat{\mathcal{Q}}_m^J$, la liste des régions $\left\{\hat{\mathcal{R}}_\ell^j\right\}_{j=1,\ell=1}^{J,L}$ dans lesquelles elle est contenue.

2. Cependant, la différence entre les deux grilles (de tailles 295×295 et 500×500) est importante pour l'estimation du volume des régions élémentaires.

5.2 Caractérisation de la population de plongeurs étudiée

5.2.1 Données simulées

Avant de discuter les résultats obtenus pour le jeu de données $\mathbf{X}_n$ des mesures de grades, nous présentons une application à des données simulées, notées $\mathbf{X}_n^s$, utilisant la partition $\hat{\mathcal{Q}}^J$ induite par les profils de plongée. Cela nous permettra de comparer la densité estimée avec la "vraie" densité dans le cas où l'on garde les mêmes nombre d'observation n^j et la même partition $\mathcal{Q}^J$ issue du jeu de données réelles. Nous considérons π_θ la restriction de la loi $\mathcal{N}(\theta^m, \sigma^2\mathbf{I}_2)$ au domaine Θ où θ^m est le centre de Θ et $\sigma^2 = 0.2$. La Figure 5.1c représente $p_{\pi_\theta,\hat{\mathcal{Q}}^J}$.



(a) $\hat{\pi}_\star^{\mathcal{L}}$ **(b)** $\hat{\pi}_{\epsilon_\star}^{H_2}$

(c) $p_{\pi_\theta,\hat{\mathcal{Q}}^J}$ **(d)** $\hat{\pi}_{\epsilon MV}^{H_2}$

Figure 5.1 – *Estimateurs $\hat{\pi}_\star^{\mathcal{L}}$, $\hat{\pi}_{\epsilon_\star}^{H_2}$ et $\hat{\pi}_{\epsilon MV}^{H_2}$ d'une loi simulée $p_{\pi_\theta,\hat{\mathcal{Q}}^J}$ où $\hat{\mathcal{Q}}$ est l'estimation par krigeage simple de la partition des régions élémentaires induites par les profils de plongée du jeu de données des grades.*

Pour la simulation des données, nous avons gardé les mêmes nombres n^j d'observations par profil, indiqués dans le tableau 4.1, et ainsi, le même total $n = 433$. Les Figures 5.1a, 5.1b et 5.1d montrent respectivement les estimateurs $\hat{\pi}_\star^{\mathcal{L}}$ l'ENPMV maximisant l'entropie de Rényi d'ordre 2, $\hat{\pi}_{\epsilon_\star}^{H_2}$ l'estimateur MaxEnt correspondant au plus petit niveau de relaxation et $\hat{\pi}_{\epsilon MV}^{H_2}$ le MaxEnt-MV. La singularité des estimateurs $\hat{\pi}_\star^{\mathcal{L}}$ et $\hat{\pi}_{\epsilon_\star}^{H_2}$ est très forte. En effet, la masse de probabilité est concentrée dans des sous-ensembles de Θ de mesure de Lebesgue très petite. Au contraire, même pour une partition de géométrie complexe comme celle obtenue dans notre cas, l'estimateur MaxEnt-MV est capable de surmonter les insuffisances des deux premiers estimateurs en produisant une estimation qui ressemble à la vraie loi. Ces résultats indiquent donc que le pouvoir prédictif de l'estimateur

$\hat{\pi}^{H_2}_{\epsilon MV}$ doit être bien meilleur que celui des estimateurs $\hat{\pi}^{\mathcal{L}}$ et $\hat{\pi}^{H_2}_{\epsilon_\star}$.

La Figure 5.2 montre l'évolution de la log-vraisemblance de $\hat{\pi}^{H_2}_{\epsilon}$ en fonction de $\epsilon \geqslant \epsilon_\star$. L'estimateur $\hat{\pi}^{H_2}_{\epsilon MV}$ est obtenu pour $\epsilon \simeq \epsilon_\star$. La perte en vraisemblance par rapport à l'exemple des partitions simulées, voir la Figure 3.2, peut être expliquée par le petit nombre d'observations (ici $n = 433$ contre $n = 10^4$ pour l'étude avec partitions simulées) et également par la géométrie plus irrégulière de la partition $\hat{\mathcal{Q}}^J$ contenant un grand nombre de régions de petite mesure de Lebesgue.

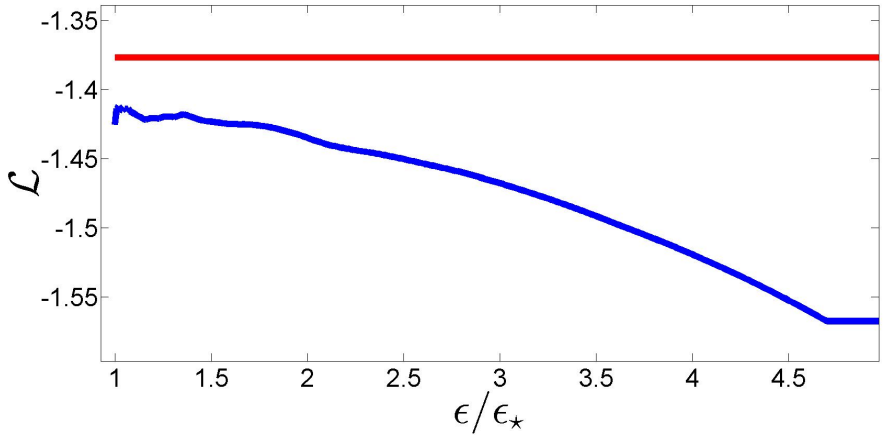


FIGURE 5.2 – *Évolution de* $\mathcal{L}(\hat{\pi}^{H_2}_{\epsilon}, \mathbf{X}^s_n)$ *en fonction de* $\epsilon/\epsilon_\star$. *La ligne rouge correspond à* $\mathcal{L}(\hat{\pi}^{\mathcal{L}}, \mathbf{X}^s_n)$.

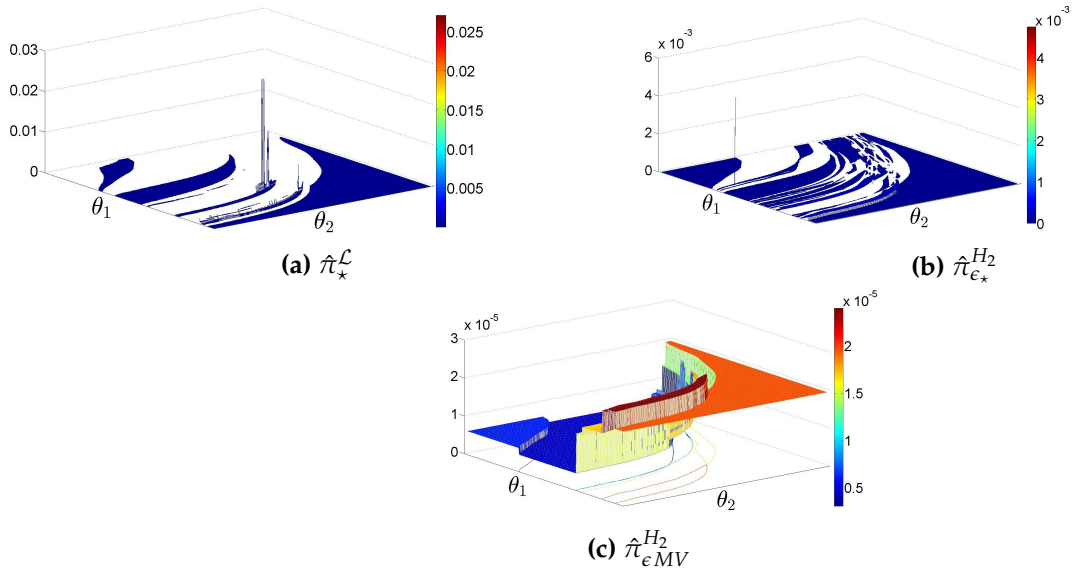
5.2.2 Données réelles

La Figure 5.3 montre les densités obtenues pour le jeu de données réelles $\mathbf{X}_n$ avec les trois estimateurs $\hat{\pi}^{\mathcal{L}}_\star$, $\hat{\pi}^{H_2}_{\epsilon_\star}$ et $\hat{\pi}^{H_2}_{\epsilon MV}$. Comme déjà mentionné, les profils de plongée conduisent à une partition $\hat{\mathcal{Q}}^J$ de taille $\hat{M} = 665$. Nous observons une convergence très rapide de l'algorithme VEM (Algorithme 4, page 35) avec 35 itérations pour $\delta = 10^{-4}$.

La Figure 5.3 [3] révèle la singularité des deux estimations $\hat{\pi}^{\mathcal{L}}$ et $\hat{\pi}^{H_2}_{\epsilon_\star}$, dont les masses de probabilité sont fortement concentrées dans des régions de petite mesure de Lebesgue. L'estimateur $\hat{\pi}^{H_2}_{\epsilon MV}$ proposé conduit à une solution beaucoup plus lisse couvrant la totalité du domaine Θ. Il semble offrir un modèle plus adéquat que les solutions trouvées par les deux premiers estimateurs.

La Figure 5.4 montre la variation de $\mathcal{L}(\hat{\pi}^{H_2}_{\epsilon}, \mathbf{X}_n)$ en fonction de $\epsilon/\epsilon_\star$ pour le jeu de données réelles. Par rapport à ce que nous avons observé avec des partitions aléatoires dans la figure 3.2, page 73, la perte en termes de log-vraisemblance est plus importante dans le cas réel que dans le cas de données simulées sur la figure 5.2 : la courbe bleue reste bien en dessous de la valeur du maximum de

3. Nous avons préféré une présentation en relief afin de mieux visualiser les pics de masse de probabilité dans les deux premières figures.



(a) $\hat{\pi}^{\mathcal{L}}_{\star}$ **(b)** $\hat{\pi}^{H_2}_{\epsilon_\star}$

(c) $\hat{\pi}^{H_2}_{\epsilon MV}$

FIGURE 5.3 – *Estimations de π_θ pour le jeu de données réelles. (**a**) $\hat{\pi}^{\mathcal{L}}_{\star}$. (**b**) $\hat{\pi}^{H_2}_{\epsilon_\star}$. (**c**) $\hat{\pi}^{H_2}_{\epsilon MV}$. Les régions en blanc ont une masse nulle de probabilité.*

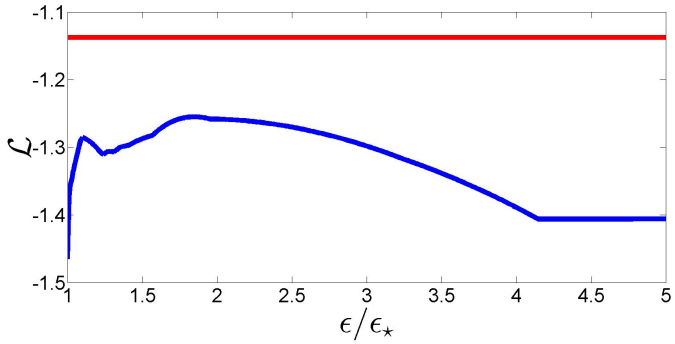


FIGURE 5.4 – *Évolution de $\mathcal{L}(\hat{\pi}^{H_2}_{\epsilon}, \mathbf{X}_n)$ en fonction de $\epsilon/\epsilon_\star$. La ligne rouge : $\mathcal{L}(\hat{\pi}^{\mathcal{L}}, \mathbf{X}_n)$.*

vraisemblance pour toutes les valeurs du paramètre de régularisation ϵ. Cela semble être une conséquence naturelle de la définition des lois empiriques induites par la quantification de $\|v^b(\cdot, P^j)\|_\infty$ et qui impliquent des erreurs compromettant la capacité des estimateurs à bien répondre aux données.

Finalement, nous montrons pour le jeu de données réelles, l'impact de la prise en compte des corrélations des lois empiriques dans la relaxation des contraintes (Définition 13). La Figure 5.5 montre les estimateurs $\hat{\pi}^{H_2}_{\epsilon_\star}$ ((5.5a) et (5.5b)) et $\hat{\pi}^{H_2}_{\epsilon MV}$ ((5.5c) et (5.5d)) obtenus en utilisant les matrices ${\Sigma^{(j)}}^{-1/2}$ (à gauche) et les mêmes estimateurs en utilisant seulement les éléments diagonaux de ces matrices (à droite). Nous pouvons remarquer que l'utilisation des dépendances entre les éléments des lois empiriques permet de mieux définir la relaxation des contraintes que doit satisfaire l'estimateur MaxEnt.

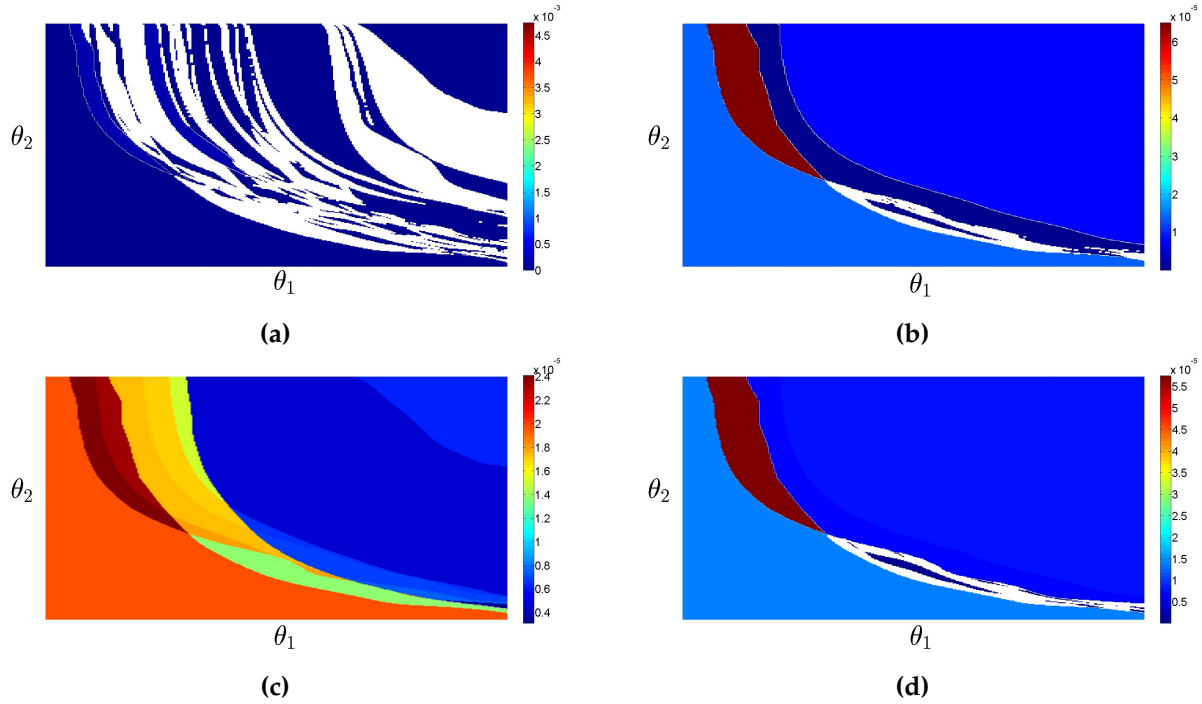


FIGURE 5.5 – *(a)* $\hat{\pi}^{H_2}_{\epsilon_\star}$ *avec* $\Sigma^{(j)^{-1/2}}$ *; (b)* $\hat{\pi}^{H_2}_{\epsilon_\star}$*, avec* $diag(\Sigma^{(j)^{-1/2}})$ *; (c)* $\hat{\pi}^{H_2}_{\epsilon MV}$*, avec* $\Sigma^{(j)^{-1/2}}$ *; (d)* $\hat{\pi}^{H_2}_{\epsilon MV}$*, avec* $diag(\Sigma^{(j)^{-1/2}})$.

5.2.3 Évaluation du pouvoir prédictif des trois estimateurs $\hat{\pi}^{\mathcal{L}}_\star$, $\hat{\pi}^{H_2}_{\epsilon_\star}$ et $\hat{\pi}^{H_2}_{\epsilon MV}$

Afin d'évaluer le pouvoir prédictif des trois estimateurs $\hat{\pi}^{\mathcal{L}}_\star$, $\hat{\pi}^{H_2}_{\epsilon_\star}$ et $\hat{\pi}^{H_2}_{\epsilon MV}$, où la relaxation des contraintes dans les estimateurs MaxEnt est définie en utilisant toutes les dépendances entre les éléments des lois empiriques, nous avons effectué une validation croisée du type *'leave-one-out'*, retirant à chaque fois toutes les observations relatives à l'un des 19 profils de plongées du jeu de données réelles P^j, et calculant les trois estimateurs en utilisant les observations issues des 18 profils restants. Nous comparons ensuite les fréquences $\mathbf{q}^j(\mathbf{n})$ de grades observées et estimées relatives au profil de plongée écarté. La Figure 5.6 représente les boites à moustaches de la distance de la variation totale pour chacun des 19 profils. Le résultat de la validation croisée confirme la supériorité de l'estimateur proposé pour des fins de prédiction. Au sens de la distance de la variation totale, l'estimateur $\hat{\pi}^{H_2}_{\epsilon MV}$ propose la distribution la plus proche, en moyenne, des fréquences empiriques sur l'ensemble des données.

5.3 CONCLUSION

Ce chapitre a appliqué à un jeu de données réelles de grades de plongée les approches d'estimation d'une densité de probabilité à partir d'observations censurées présentées dans cette thèse. Nous avons vu que l'estimation non-paramétrique par maximum de vraisemblance est intrinsèque-

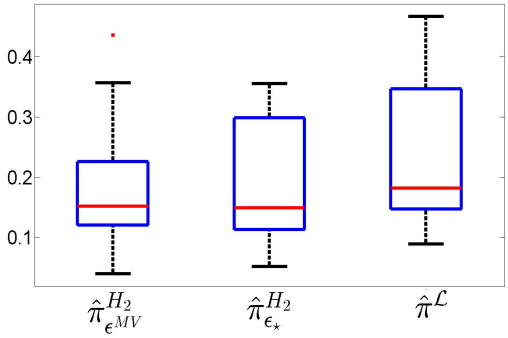


FIGURE 5.6 – *Boites à moustaches de la distance de la variation totale pour les données relatives aux 19 profils dans l'étude de validation croisée.*

ment mal-posée, conduisant à une solution instable et peu plausible d'un point de vue biophysique. Confronté au problème de l'incompatibilité des lois empiriques issues des données, nous proposons une estimation par MaxEnt avec relaxation des contraintes. L'estimateur MaxEnt-MV proposé détermine le niveau de relaxation en choisissant la solution ayant la plus grande vraisemblance. L'application de cette approche d'estimation à des jeux de données (simulé et réel) avec des partitions de géométrie complexe conduit à une solution se rapprochant des observations (pour le jeu de données simulées), tout en proposant une description plausible d'un point de vu biophysique de la population étudiée. En particulier, cette approche aboutit à une distribution avec de bonnes propriétés telle que sa capacité à prédire les probabilités de grades pour des profils de plongée non-expérimentés. Elle peut donc être utilisée pour détecter les profils avec un risque élevé d'accident de décompression pour une proportion non-négligeable d'une population de plongeurs.

- CHAPITRE 6 -

CONCLUSION

Cette étude a porté sur la caractérisation de la diversité d'une population à partir de mesures quantifiées d'un modèle non-linéaire. Elle avait comme cadre applicatif la plongée hyperbare.

6.1 CONTRIBUTIONS

La principale contribution de cette étude est la proposition d'un nouvel estimateur de densité à partir de données censurées. Ce nouvel estimateur se base simultanément sur les critères de maximum de vraisemblance et de maximum d'entropie. Tout en optimisant l'exploitation de l'information disponible, il prend en compte les dépendances statistiques entre l'ensemble des contraintes définissant l'estimateur du maximum d'entropie. Une étude des estimateurs classiques sur données simulées a été menée et elle a montré que la solution proposée conduit à des densités estimées offrant une bonne capacité de généralisation, tout en gardant une forte attache aux données.

Nous avons pu gérer le cas où les régions de censure ne sont plus de simples intervalles comme dans les travaux classiques sur données censurées, mais des régions de forme arbitraire. Il s'agit ici d'un cadre d'application beaucoup plus général que celui de la censure par intervalles. En effet, dans de nombreux domaines des modèles numérique d'une grande complexité ont été développés afin de décrire des phénomènes pour lesquels on ne dispose que d'observations ordinales. Notre étude permet une généralisation de la solution proposée pour l'estimation de densités multivariées à partir d'observations fortement quantifiées d'un ensemble de fonctions scalaires des variables d'intérêt issues d'un modèle numérique. Nous avons pu montrer grâce à cette étude que le modèle numérique n'intervient dans le problème d'estimation de la densité qu'à travers la définition des régions de censure, plus précisément, que dans la détermination de la partition de régions élémentaires du domaine paramétrique.

La mise en œuvre de l'estimateur proposé nous à demandé le développement d'outils pour le calcul des estimateur classiques, à savoir l'estimateur non-paramétrique du maximum de vraisem-

blance (ENPMV) et l'estimateur par maximum d'entropie. Ainsi, nous avons pu proposer plusieurs algorithmes permettant de calculer ces deux estimateurs, notamment en précisant le lien avec la théorie des graphes et avec le problème de détermination des plans D-optimaux (pour l'ENPMV). Nous avons pu mener une étude computationnelle comparant les différentes méthodes proposées.

Afin de pouvoir appliquer les méthodes développées au domaine de la plongée sous-marine, nous avons implémenté un modèle numérique décrivant le comportement de l'organisme d'un plongeur suite à une exposition hyperbare. Nous avons présenté d'une part les mécanismes biophysiques supposées conduire à un accident de décompression, et d'autre part les choix et approximations retenus dans l'implémentation de ce modèle numérique. Nous avons enfin pu caractériser la diversité de la population étudiée de plongeur en se basant sur le jeu de données réelles disponibles.

6.2 PERSPECTIVES

Dans notre cas, la détermination des régions de censure à partir du modèle numérique a été relativement simple dû au fait que le paramètre dont nous cherchons à estimer la densité est de dimension 2. Nous avions fait appel à la technique de krigeage simple pour construire une interpolation des variables d'intérêt issues du modèle numérique, permettant un calcul rapide de leurs valeurs pour tout $\theta \in \Theta$. Nous imaginons que pour des dimensions plus élevées, le problème de la détermination de ces régions est beaucoup plus complexe et nécessitera des techniques plus avancées.

L'estimateur MaxEnt-MV que nous avons proposé dans cette étude a montré de bonnes performances, comparé aux estimateurs classiques sur des jeux de données simulées. Il serait intéressant de pouvoir dériver analytiquement des encadrements à ces performances.

BIBLIOGRAPHIE

[ABE+55] M. Ayer, H.D. Brunk, G.M. Ewing, W. Reid, and E. Silverman. An empirical distribution function for sampling with incomplete information. *The Annals of Mathematical Statistics*, 26(4) :641–647, 1955.

[AS64] M. Abramowitz and I.A. Stegun. *Handbook of mathematical functions : with formulas, graphs, and mathematical tables*. Number 55. Courier Corporation, 1964.

[Aur91] F. Aurenhammer. Voronoi diagrams—a survey of a fundamental geometric data structure. *ACM Computing Surveys (CSUR)*, 23(3) :345–405, 1991.

[BBPP99] I.M. Bomze, M. Budinich, P.M. Pardalos, and M. Pelillo. The maximum clique problem. In *Handbook of Combinatorial Optimization*, pages 1–74. Springer, 1999.

[BGMB14] S.L. Blogg, M. Gennser, A. Møllerløkken, and A.O. Brubakk. Ultrasound detection of vascular decompression bubbles : the influence of new technology and considerations on bubble load. *Ultrasound for bubble detection*, 44(1) :35–44, 2014.

[BK73] C. Bron and J. Kerbosch. Algorithm 457 : Finding all cliques of an undirected graph. *Communications of the ACM*, 16(9) :575–577, 1973.

[Böh89] D. Böhning. Likelihood inference for mixtures : geometrical and other constructions of monotone step-length algorithms. *Biometrika*, 76(2) :375–383, 1989.

[Böh95] D. Böhning. A review of reliable maximum likelihood algorithms for semiparametric mixture models. *Journal of Statistical Planning and Inference*, 47(1) :5–28, 1995.

[BPR14a] Y. Bennani, L. Pronzato, and M.J. Rendas. Maximum de vraisemblance non-paramétrique avec régions de censure. Application à un modèle biophysique de décompression. In *46 èmes Journées de Statistique de la SFdS*, Rennes, 2014.

[BPR14b] Y. Bennani, L. Pronzato, and M.J. Rendas. Nonparametric density estimation with region-censored data. In *Signal Processing Conference (EUSIPCO 2014) Proceedings of the 22nd European*, pages 1098–1102. IEEE, 2014.

[BPR15a] Y. Bennani, L. Pronzato, and M.J. Rendas. Most likely maximum entropy for population analysis : A case study in decompression sickness prevention. In *MAXENT 2014*, volume 1641, pages 414–421. AIP Publishing, 2015.

[BPR15b] Y. Bennani, L. Pronzato, and M.J. Rendas. Most likely maximum entropy for population analysis with region-censored data. *Entropy*, 17(6) :3963–3988, 2015.

[C+67] I. Csiszár et al. Information-type measures of ence of probability distributions and indirect observations. *Studia Sci. Math. Hungar.*, 2 :299–318, 1967.

[CR00] S.F. Chen and R. Rosenfeld. A survey of smoothing techniques for ME models. *Speech and Audio Processing, IEEE Transactions on*, 8(1) :37–50, 2000.

[DLR77] A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 1–38, 1977.

[DPDPL97] V.J. Della Pietra, S.A. Della Pietra, and J. Lafferty. Inducing features of random fields. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 19(4) :380–393, 1997.

[DPS07] M. Dudík, S.J. Phillips, and R.E. Schapire. Maximum entropy density estimation with generalized regularization and an application to species distribution modeling. *Journal of Machine Learning Research*, 8(6), 2007.

[DPZ08] H. Dette, A. Pepelyshev, and A. Zhigljavsky. Improving updating rules in multiplicative algorithms for computing d-optimal designs. *Computational Statistics & Data Analysis*, 53(2) :312–320, 2008.

[Dud07] M. Dudík. *Maximum entropy density estimation and modeling geographic distributions of species*. PhD thesis, Princeton University, 2007.

[EB96] O. Eftedal and A.O. Brubakk. Agreement between trained and untrained observers in grading intravascular bubble signals in ultrasonic images. *Undersea & hyperbaric medicine : journal of the Undersea and Hyperbaric Medical Society, Inc*, 24(4) :293–299, 1996.

[Efr67] B. Efron. The two sample problem with censored data. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 4, pages 831–853, 1967.

[Eft07] O.S. Eftedal. *Ultrasonic detection of decompression induced vascular microbubbles*. PhD thesis, Norwegian University of Science and Technology, 2007.

[Fad57] D.K. Faddejew. Zum begriff der entropie eines endlichen wahrscheinlichkeitsschemas. *Arbeiten zur Informationstheorie, I. Berlin*, pages 86–90, 1957.

[Fed72] V.V. Fedorov. *Theory of optimal experiments*. Access Online via Elsevier, 1972.

[Fel89] J. Fellman. An empirical study of a class of iterative searches for optimal designs. *Journal of statistical planning and inference*, 21(1) :85–92, 1989.

[FN07] J.S. Ferenc and Z. Néda. On the size distribution of Poisson Voronoi cells. *Physica A : Statistical Mechanics and its Applications*, 385(2) :518–526, 2007.

[GG94] R. Gentleman and C.J. Geyer. Maximum likelihood for interval censored data : Consistency and computation. *Biometrika*, 81(3) :618–623, 1994.

[GGI92] B. Gillespie, J. Gillespie, and B. Iglewicz. A comparison of the bias in four versions of the product-limit estimator. *Biometrika*, 79(1) :149–155, 1992.

[GMZ09] B. Grechuk, A. Molyboha, and M. Zabarankin. Maximum entropy principle with general deviation measures. *Mathematics of Operations Research*, 34(2) :445–467, 2009.

[Gol04] M.C. Golumbic. *Algorithmic graph theory and perfect graphs*, volume 57. Elsevier, 2004.

[GV01] R. Gentleman and A.C. Vandal. Computational algorithms for censored-data problems using intersection graphs. *Journal of Computational and Graphical Statistics*, 10(3) :403–421, 2001.

[GW92] P. Groeneboom and J.A. Wellner. *Information Bounds and Nonparametric Maximum Likelihood Estimation*, volume 19. Springer Science & Business Media, 1992.

[HP07] R. Harman and L. Pronzato. Improvements on removing nonoptimal support points in D-optimum design algorithms. *Statistics & Probability Letters*, 77(1) :90–94, 2007.

[HT09] R. Harman and M. Trnovská. Approximate D-optimal designs of experiments on the convex hull of a finite set of information matrices. *Mathematica Slovaca*, 59(6) :693–704, 2009.

[Hug10] J. Hugon. *Vers une modélisation biophysique de la décompression.* PhD thesis, Aix Marseille 2, 2010.

[Jay57] E.T. Jaynes. Information theory and statistical mechanics. *Physical review*, 106(4) :620, 1957.

[KM58] E.L. Kaplan and P. Meier. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(282) :457–481, 1958.

[KMG78] K.E. Kisman, G. Masurel, and R. Guillerm. Bubble evaluation code for doppler ultrasonic decompression data. *Undersea Biomedical Research*, 5(1 Supplement) :5–28, 1978.

[KT03] J. Kazama and J. Tsujii. Evaluation and extension of maximum entropy models with inequality constraints. In *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing*, pages 137–144. Association for Computational Linguistics, 2003.

[Kul68] S. Kullback. *Information theory and statistics*. Courier Corporation, 1968.

[KW60] J. Kiefer and J. Wolfowitz. The equivalence of two extremum problems. *Canadian Journal of Mathematics*, 12(363-366) :233–234, 1960.

[LB93] H.D. Van Liew and M.E. Burkard. Density of decompression bubbles and competition for gas among bubbles, tissue, and blood. *Journal of Applied Physiology*, 75(5) :2293–2301, 1993.

[LBW+00] J.K. Lindsey, W.D. Byrom, J. Wang, P. Jarvis, and B. Jones. Generalized nonlinear models for pharmacokinetic data. *Biometrics*, pages 81–88, 2000.

[LCB94] H.D. Van Liew, J. Conkin, and M.E. Burkard. Probabilistic model of altitude decompression sickness based on mechanistic premises. *Journal of Applied Physiology*, 76(6) :2726–2734, 1994.

[Lin83] B.G. Lindsay. The geometry of mixture likelihoods : a general theory. *The Annals of Statistics*, 11(1) :86–94, 1983.

[Liu05] X. Liu. *Nonparametric estimation with censored data : a discrete approach*. PhD thesis, McGill University, Montreal, 2005.

[Maa03] M.H. Maathuis. Nonparametric maximum likelihood estimation for bivariate censored data. Master's thesis, Delft University of Technology, The Netherlands, 2003.

[Mal86] A. Mallet. A maximum likelihood estimation method for random coefficient regression models. *Biometrika*, 73(3) :645–656, 1986.

[Nis90] R.Y. Nishi. Doppler evaluation of decompression tables. *Man in the Sea*, 1 :297–316, 1990.

[Páz86] A. Pázman. *Foundations of optimum experimental design*, volume 14. Springer, 1986.

[Pet73] R. Peto. Experimental survival curves for interval-censored data. *Applied Statistics*, 22(1) :86–91, 1973.

[Puk93] F. Pukelsheim. *Optimal design of experiments*. Wiley, 1993.

[PX94] P.M. Pardalos and J. Xue. The maximum clique problem. *Journal of Global Optimization*, 4(3) :301–328, 1994.

[Ŕé61] A. Rényi. On measures of entropy and information. In *Fourth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 547–561, 1961.

[Rya11] D. Ryabko. On the relation between realizable and nonrealizable cases of the sequence prediction problem. *The Journal of Machine Learning Research*, 12 :2161–2180, 2011.

[Sha01] C.E. Shannon. A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review*, 5(1) :3–55, 2001.

[SJ00] Khudanpur S. and Wu J. Maximum entropy techniques for exploiting syntactic, semantic and collocational dependencies language modeling. pages 355–372, 2000.

[SN90] K.D. Sawatzky and R.Y. Nishi. Assessment of inter-rater agreement on the grading of intravascular bubble signals. *Undersea Biomedical Research*, 18(5-6) :373–396, 1990.

[Son01] S. Song. *Estimation With Bivariate Interval-Censored Data*. PhD thesis, University of Washington Seattle, 2001.

[Spe76] M.P. Spencer. Decompression limits for compressed air determined by ultrasonically detected blood bubbles. *Journal of Applied Physiology*, 40(2) :229–235, 1976.

[Ste12] M.L. Stein. *Interpolation of spatial data : some theory for kriging*. Springer Science & Business Media, 2012.

[STT78] S.D. Silvey, D.H. Titterington, and B. Torsney. An algorithm for optimal designs on a design space. *Communications in Statistics – Theory and Methods*, 7(14) :1379–1389, 1978.

[SY00] A. Schick and Q. Yu. Consistency of the GMLE with mixed case interval-censored data. *Scandinavian Journal of Statistics*, 27(1) :45–55, 2000.

[TIAS77] S. Tsukiyama, M. Ide, H. Ariyoshi, and I. Shirakawa. A new algorithm for generating all the maximal independent sets. *SIAM Journal on Computing*, 6(3) :505–517, 1977.

[Tit76] D.M. Titterington. Algorithms for computing D-optimal designs on a finite design space. In *Proc. of the 1976 Conference on Information Science and Systems, John Hopkins University*, volume 3, pages 213–216, 1976.

[Tor83] B. Torsney. A moment inequality and monotonicity of an algorithm. In *Semi-Infinite Programming and Applications*, pages 249–260. Springer, Heidelberg, 1983.

[TTT06] E. Tomita, A. Tanaka, and H. Takahashi. The worst-case time complexity for generating all maximal cliques and computational experiments. *Theoretical Computer Science*, 363(1) :28–42, 2006.

[Tur76] B.W. Turnbull. The empirical distribution function with arbitrarily grouped, censored and truncated data. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 290–295, 1976.

[vDVW96] A. van Der Vaart and J.A. Wellner. *Weak Convergence*. Springer, 1996.

[vdVW00] A. van der Vaart and J.A. Wellner. Preservation theorems for Glivenko-Cantelli and uniform Glivenko-Cantelli classes. In *High Dimensional Probability II*, pages 115–133, Birkhäuser Basel, 2000.

[WG96] Z. Wang and J.C. Gardiner. A class of estimators of the survival function from interval-censored data. *The Annals of Statistics*, 24(2) :647–658, 1996.

[WP94] É. Walter and L. Pronzato. *Identification de Modèles Paramétriques à partir de Données Expérimentales*. Masson, 1994.

[You79] D.E. Yount. Application of a bubble formation model to decompression sickness in rats and humans. *Aviation, Space, and Environmental Medicine*, 50(1) :44–50, 1979.

[YSLG98] Q. Yu, A. Schick, L. Li, and G.Wong. Asymptotic properties of the GMLE with case 2 interval-censored data. *Statistics & Probability Letters*, 37(3) :223–228, 1998.

[YSLW98] Q. Yu, A. Schick, L. Li, and G. Wong. Asymptotic properties of the GMLE in the case 1 interval-censorship model with discrete inspection times. *Canadian Journal of Statistics*, 26(4) :619–627, 1998.

[Yu10a] Y. Yu. Monotonic convergence of a general algorithm for computing optimal designs. *The Annals of Statistics*, pages 1593–1606, 2010.

[Yu10b] Y. Yu. Strict monotonicity and convergence rate of titterington's algorithm for computing d-optimal designs. *Computational Statistics & Data Analysis*, 54(6) :1419–1425, 2010.

- RÉSUMÉ -

Cette thèse propose une nouvelle méthode pour l'estimation non-paramétrique de densité à partir de données censurées par des régions de formes quelconques, éléments de partitions du domaine paramétrique. Ce travail a été motivé par le besoin d'estimer la distribution des paramètres d'un modèle biophysique de décompression afin d'être capable de prédire un risque d'accident. Dans ce contexte, les observations (grades de plongées) correspondent au comptage quantifié du nombre de bulles circulant dans le sang pour un ensemble de plongeurs ayant exploré différents profils de plongées (profondeur, durée), le modèle biophysique permettant de prédire le volume de gaz dégagé pour un profil de plongée donné et un plongeur de paramètres biophysiques connus.
Dans un premier temps, nous mettons en évidence les limitations de l'estimation classique de densité au sens du maximum de vraisemblance non-paramétrique. Nous proposons plusieurs méthodes permettant de calculer cet estimateur et montrons qu'il présente plusieurs anomalies : en particulier, il concentre la masse de probabilité dans quelques régions seulement, ce qui le rend inadapté à la description d'une population naturelle.
Nous proposons ensuite une nouvelle approche reposant à la fois sur le principe du maximum d'entropie, afin d'assurer une régularité convenable de la solution, et mettant en jeu le critère du maximum de vraisemblance, ce qui garantit une forte attache aux données. Il s'agit de rechercher la loi d'entropie maximale dont l'écart maximal aux observations (fréquences de grades observées) est fixé de façon à maximiser la vraisemblance des données. Nous illustrons par divers exemples que la solution proposée offre de meilleures performances, notamment en généralisation, que l'estimateur non-paramétrique classique du maximum de vraisemblance.

- ABSTRACT -

This thesis proposes a new method for nonparametric density estimation from censored data, where the censing regions can have arbitrary shape and are elements of partitions of the parametric domain. This study has been motivated by the need for estimating the distribution of the parameters of a biophysical model of decompression, in order to be able to predict the risk of decompression sickness. In this context, the observations correspond to quantified counts of bubbles circulating in the blood of a set of divers having explored a variety of diving profiles (depth, duration) ; the biophysical model predicts of the gaz volume produced along a given diving profile for a diver with known biophysical parameters.
In a first step, we point out the limitations of the classical nonparametric maximum-likelihood estimator. We propose several methods for its calculation and show that it suffers from several problems : in particular, it concentrates the probability mass in a few regions only, which makes it inappropriate to the description of a natural population.
We then propose a new approach relying both on the maximum-entropy principle, in order to ensure a convenient regularity of the solution, and resorting to the maximum-likelihood criterion, to guarantee a good fit to the data. It consists in searching for the probability law with maximum entropy whose maximum deviation from empirical averages is set by maximizing the data likelihood. Several examples illustrate the superiority of our solution compared to the classic nonparametric maximum-likelihood estimator, in particular concerning generalisation performance.