

Université de Montréal

**Régression de Cox avec partitions latentes issues du
modèle de Potts**

par

Danae Mirel Martínez Vargas

Département de mathématiques et de statistique
Faculté des arts et des sciences

Mémoire présenté à la Faculté des études supérieures
en vue de l'obtention du grade de
Maître ès sciences (M.Sc.)
en Statistique

2 juillet 2019

Université de Montréal

Faculté des études supérieures

Ce mémoire intitulé

**Régression de Cox avec partitions latentes issues du
modèle de Potts**

présenté par

Danae Mirel Martínez Vargas

a été évalué par un jury composé des personnes suivantes :

François Perron

(président-rapporteur)

Alejandro Murua

(directeur de recherche)

Maciej Augustyniak

(membre du jury)

Mémoire accepté le

2 juillet 2019

SOMMAIRE

Le but de ce projet de recherche est de développer un modèle de régression bayésien non paramétrique issu de partitions aléatoires dans un contexte d'analyse de survie. Notre objectif final est de construire un système de prévision qui dans un premier temps consiste à regrouper les observations ayant des caractéristiques semblables. Une fois les sous-groupes formés, la survie au sein de chaque sous-groupe est évaluée à l'aide d'un modèle bayésien non paramétrique.

Le nombre de sous-groupes dans la population est aléatoire. Nous proposons l'utilisation du modèle de classification de Potts (MURUA, STANBERRY et STUETZLE [29]) appliqué à l'espace des covariables afin de générer la formation des partitions aléatoires d'individus. Pour toute partition donnée, le modèle proposé dans ce projet suppose une régression de Cox par intervalles avec taux de risque de base Weibull au sein de chaque grappe. Cette méthodologie a été inspiré du travail de IBRAHIM [18].

Les estimations et l'inférence sont effectuées à l'aide de méthodes MCMC. Nous utilisons également la méthode d'approximation de Laplace (SHUN et MCCULLAGH [36]) pour estimer certaines constantes et proposer des mises à jour de paramètres dans l'application de l'algorithme MCMC.

Finalement, nous comparons les performances de notre modèle à celles d'une régression de Cox classique et au modèle bayésien non paramétrique de partition-produit indexé par des covariables, PPM_x tout ceci à l'aide de nombreuses simulations. En général, notre modèle a offert des résultats comparables à ceux de ses compétiteurs et s'est avéré dans certains cas le meilleur choix.

Mots-clés : analyse de survie, modèle de Cox, classification, inférence bayésienne non paramétrique, modèles de partition aléatoire, modèle de Potts.

SUMMARY

The goal of this research project is to develop a non-parametric Bayesian regression model derived from random partitions in a survival analysis context. Our final objective is to build a forecasting system that initially consists of grouping observations with similar characteristics. Once subgroups are formed, survival within each subgroup is assessed using a non-parametric Bayesian model.

The number of subgroups in the population is random. We propose the use of the Potts classification model ([29]) applied to the covariate space to generate the formation of random partitions of individuals.

For any given partition, the model proposed in this project assumes a interval-wise Weibull distribution for the baseline hazard rate inspired from the model first proposed by IBRAHIM [18]. Estimates and inference are made using MCMC methods. We also use the Laplace approximation method ([36]) to estimate some constants and propose parameter updates in the application of the MCMC algorithm.

Finally, we compare the performance of our model with that of a classical Cox regression and the non-parametric Bayesian model of product partition indexed by covariates, PPM_x , all this with the help of numerous simulations. In general, our model has delivered results comparable to those of its competitors and has in some cases proven to be the best choice.

Key words : Cox model, Clustering, non parametric Bayesian inference, Random Partition Models, Potts model, Proportional hazards models, Cox regression model, survival analysis.

TABLE DES MATIÈRES

Sommaire	v
Summary	vii
Liste des tableaux	xiii
Liste des figures	xv
Liste des sigles et des abréviations	xvii
Remerciements	xix
Introduction	1
Chapitre 1. NOTIONS PRÉLIMINAIRES EN ANALYSE DE SURVIE	3
1.1. Introduction.....	3
1.2. Fonctions de base.....	3
1.2.1. Fonction de Survie.....	3
1.2.2. Fonction de Risque.....	4
1.2.3. Relations entre fonctions.....	5
1.3. Censure.....	5
1.3.1. Censure à droite.....	5
1.3.2. Censure à gauche et censure par intervalle	6
1.3.2.1. Hypothèse de censure indépendante.....	7
1.4. Modèles de régression pour données de survie.....	7
1.5. Modèle de régression de Cox.....	8
1.5.1. Vraisemblance partielle de Cox	9
1.6. Régression de Cox avec fonction de risque par intervalles de loi exponentielle	10
Chapitre 2. MODÈLES DE PARTITIONS ALÉATOIRES.....	11

2.1.	Introduction	11
2.2.	Le paradigme bayésien	11
2.3.	Introduction à la statistique bayésienne non paramétrique	12
2.3.1.	Processus de Dirichlet	12
2.3.1.1.	Représentation Stick-breaking du Processus de Dirichlet	13
2.3.2.	Processus de Dirichlet Mixtes	13
2.4.	Quelques exemples populaires de modèles de partition aléatoire	14
2.4.1.	Modèle d'échantillonnage par espèce	14
2.4.2.	Modèles partition-produit	15
2.4.2.1.	Modèles partition-produit avec covariables	16
2.5.	Modèle de Potts	17
2.5.1.	Le modèle de Potts appliqué à un graphe	17
2.6.	Méthodes de Monte-Carlo par chaînes de Markov	19
2.6.1.	Échantillonnage de Gibbs	20
2.6.2.	Algorithme de Metropolis-Hastings	21
Chapitre 3.	PRÉSENTATION DU MODÈLE	23
3.1.	Introduction	23
3.2.	Régression de Cox avec partitions latentes issues du modèle de Potts	23
3.3.	Simulation <i>a posteriori</i> dans le cas général	24
3.4.	Inférence au sein de chaque sous-ensemble	26
3.4.1.	Vraisemblance du modèle	26
3.5.	Densité <i>a priori</i> et <i>a posteriori</i> des paramètres	27
3.5.1.	Densité <i>a priori</i> de α et λ	28
3.5.2.	Méthode d'intégration de Laplace	29
3.5.3.	Approximation lisse de la fonction $\ell(\alpha, \lambda)$	31
3.6.	Échantillonnage de la loi <i>a posteriori</i>	31
3.6.1.	Échantillonnage de la loi de β	32
3.6.2.	Échantillonnage du vecteur (α, λ)	33
3.7.	Inférence prédictive <i>a posteriori</i>	34
3.7.1.	La valeur prédictive moyenne	36

Chapitre 4. RÉSULTATS NUMÉRIQUES	39
4.1. Génération de populations artificielles issues d'un mélange gaussien	39
4.2. Génération de la variable réponse	41
4.3. Évaluation de la prévision	42
Chapitre 5. Conclusions	47
Références	49
Annexe A. Calcul des densités <i>a priori</i> marginales des paramètres alpha et lambda selon le nombre d'intervalles J	A-i
Annexe B. Calcul du gradient et de la matrice hessienne de la fonction $g(\cdot)$	B-i
B.1. Le gradient	B-i
B.2. Le matrice hessienne	B-ii
Annexe C. Calcul du gradient et la matrice hessienne de la fonction \mathcal{L}, douce	C-i
Annexe D. Obtention des intégrales dans la loi prédictive <i>a posteriori</i>	D-i
Annexe E. Détails de la simulation	E-i
E.1. Génération de données à partir d'un modèle de régression de Cox aux risques par intervalles selon la loi weibull	E-i
E.2. Paramètres de chevauchement choisis pour la génération des jeux de données issus d'un modèle de mélange gussienne	E-iii
E.3. Simulation des populations issues d'une mélange gaussienne	E-iv
E.4. Simulation de la variable réponse	E-vi
E.5. Paramètres utilisés pour les simulations de la variable réponse y	E-viii
Annexe F. Résultats supplémentaires	F-i
F.1. Approximation de Laplace : cas général	F-i
F.2. Rappel sur les lois conditionnelles de la distribution Normale multivariée	F-i

LISTE DES TABLEAUX

4. I	Hyper-paramètres choisis pour la génération de populations issues d'un modèle de mélange gaussien.	40
4. II	Carré hyper-greco-latin qui résume les résultats mesurant la performance du modèle de Cox-Potts pour des différents valeurs de I	43
4. III	Carré hyper-greco-latin qui résume les résultats mesurant la performance du modèle de Cox-Potts et ses compétiteurs : PPM_x et Cox. L'hyper-paramètre I est fixé à 1.	44
4. IV	Résultats des effets séparés selon les quatre facteurs différents considérés. Le facteur nombre d'intervalles I est fixé à 1.	46
E. I	Chevauchement appliqué à la génération des population artificiels. Pour chaque valeur de K , on a généré des populations de dimension $p = 2$, $p = 5$ et $p = 10$	E-iii
E. II	Choix de paramètres de départ pour les populations de dimension $p = 2$ et $J = 1$. Le nombre de sous-populations est $K = 3$ (haut) et $K = 6$ (bas) Le paramètre $\alpha = 1$	E-ix
E. III	Choix de paramètres de départ pour les populations de dimension $p = 2$, $J = 1$ et $K = 10$. Le paramètre $\alpha = 1$	E-x
E. IV	Choix de paramètres de départ pour les populations de dimension $p = 2$, $J = 2$, $K = 3$ (haut) et $K = 6$ (bas). Le paramètre $\alpha = 1$	E-xi
E. V	Choix de paramètres de départ pour les populations de dimension $p = 2$, $J = 2$ et $K = 10$. Le paramètre $\alpha = 1$	E-xii
E. VI	Choix de paramètres de départ pour les populations de dimension $p = 2$ et $J = 3$. Le paramètre $\alpha = 1$	E-xiii
E. VII	Choix de paramètres de départ pour les populations de dimension $p = 2$, $J = 3$ et $K = 10$. Le paramètre $\alpha = 1$	E-xiv
E. VIII	Choix de paramètres de départ pour les populations de dimension $p = 5$ et $J = 1$. Le paramètre $\alpha = 1$	E-xv

E. IX	Choix de paramètres de départ pour les populations de dimension $p = 5$ et $J = 1$. Le paramètre $\alpha = 1$	E-xvi
E. X	Choix de paramètres de départ pour les populations de dimension $p = 5$, $J = 2$, $K = 3$ (haut) et $K = 6$ (bas). Le paramètre $\alpha = 1$	E-xvii
E. XI	Choix de paramètres de départ pour les populations de dimension $p = 5$ et $J = 2$. Le paramètre $\alpha = 1$	E-xviii
E. XII	Choix de paramètres de départ pour les populations de dimension $p = 5$ et $J = 3$. Le paramètre $\alpha = 1$	E-xix
E. XIII	Choix de paramètres de départ pour les populations de dimension $p = 5$ et $J = 3$. Le paramètre $\alpha = 1$	E-xx

LISTE DES FIGURES

4.1	Échantillon de 1000 observations générées d'une mélange gaussien et une transformation de Box-Cox avec paramètre $\nu = 0,5$; $K = 3$ et $p = 2$. Pour l'échantillon ci-haut, le paramètre de chevauchement est $\bar{\omega} = 0,005$ et pour l'échantillon avec chevauchement modéré, $\bar{\omega} = 0,05$	41
4.2	Résultats de la simulation (en fixant $I = 1$) selon les facteurs dimension, nombre de <i>clusters</i> , chevauchement et nombre d'intervalles.	45

LISTE DES SIGLES ET DES ABRÉVIATIONS

v.a.	Variable aléatoire
i.i.d.	Indépendantes et identiquement distribuées (à propos de variables aléatoires)
ind.	Indépendantes (à propos de variables aléatoires ou d'observations)
p.s.	Convergence presque sûre
MCMC	Méthodes de Monte-Carlo par chaîne de Markov, de l'anglais <i>Markov Chain Monte Carlo</i>
BNP	Bayésien non paramétrique (à propos de méthodes), de l'anglais <i>Bayesian nonparametric</i>
DP	Processus de Dirichlet, de l'anglais <i>Dirichlet process</i>
DPM	Processus de Dirichlet Mixtes
RPM	Modèles de partition aléatoire, de l'anglais <i>random partition models</i>
SSMs	Modèle d'échantillonnage par espèces de l'anglais <i>Species Sampling Model</i>
PPM	Modèles de partition-produit, de l'anglais <i>Product Partition Models</i>
MSE	Erreur quadratique moyenne, de l'anglais <i>Mean Square Error</i>
RMSE	Racine carré de l'erreur quadratique moyenne, de l'anglais <i>Root mean square error</i>
MH	Algorithme de Metropolis-Hastings

REMERCIEMENTS

Je tiens à remercier mon directeur de recherche, Alejandro Murua, pour son soutien et sa grande implication tout au long de l'élaboration de ce mémoire, ainsi que pour le temps et l'énergie qu'il a consacrés à ce projet.

Je remercie également les membres du jury, François Perron et Maciej Augustyniak, pour le temps qu'ils ont consacré à la révision de ce projet de recherche et pour leurs commentaires. À mes professeurs, pour toutes les connaissances qu'ils m'ont apportées pendant ma formation.

Je remercie le département de mathématiques et statistique de l'Université de Montréal de m'avoir accueillie pendant mes études de maîtrise et de m'avoir donné l'opportunité de détenir un travail étudiant où je faisais une des choses que j'aime le plus. Je suis reconnaissante également envers le gouvernement du Québec et l'Institut des sciences mathématiques du Québec, qui m'ont accordé une bourse sans laquelle ce projet n'aurait pas pu être réalisé.

Merci à mes parents et à mon frère pour leur amour inconditionnel, pour être présents en tout moment malgré la longue distance qui nous sépare. Particulièrement, je voudrais remercier ma mère pour le courage qu'elle m'inspire et le constant soutien qu'elle m'a accordé.

Merci à Nicolas, d'avoir écouté mes idées et mes questions, merci d'être devenu le correcteur extra officiel de ce mémoire et aussi un conseiller quand j'en avais besoin. Mais surtout merci d'être la personne qui m'inspire et m'encourage constamment.

Finalement, je voudrais adresser un petit mot de remerciement à l'équipe de Thèsez-vous. Merci pour le beau travail de rassembler les étudiants des cycles supérieurs en rédaction. Apprendre à briser l'isolement, formuler des objectifs *SMART* et travailler avec la technique des *pomodoros* pendant des séances de rédaction solidaire m'a redonné le goût de travailler sur mon mémoire.

INTRODUCTION

Dans divers contextes, on s'intéresse à développer des systèmes de prévision sur la base d'une analyse rationnelle. Par exemple, en analyse de survie, on cherche à prédire le temps écoulé avant la réalisation d'un certain évènement futur.

On sait également que dans certaines situations, il serait convenable de classer les observations en sous-groupes *a priori* afin d'appliquer un système de prévision différent sur chaque sous-groupe. On s'attend à ce que ceci aide à mieux caractériser les observations et prédire les réponses. Par exemple, en oncologie, les groupes de risque sont habituellement constitués en divisant la population en blocs de patients ayant des conditions de santé et des niveaux démographiques similaires. Un modèle de prévision au sein de chaque groupe de risque avec des paramètres différents qui s'adaptent mieux à la sous-population en question est donc appliqué.

Il existe diverses façon de construire des tels sous-groupes. Beaucoup de méthodes de regroupement dont l'objectif est de former des groupes d'observations similaires tout en séparant les observations dissemblables, ont été proposées. Pour un aperçu complet des méthodes de regroupement, voir XU et WUNSCH [44]. Cependant la plupart de ces méthodes sont purement algorithmiques. On propose ici l'application d'un modèle de partition aléatoire pour la formation de grappes *a priori*, développé dans un contexte d'inférence bayésienne non paramétrique (voir ORBANZ et TEH [32]).

Une fois cette première étape franchie, on procède à l'application du système de pronostique au sein de chaque sous-ensemble. Ce projet de recherche s'attaque à développer un système de pronostique dans un contexte d'analyse de survie. Nous nous sommes inspirés en partie du travail de IBRAHIM [18], cependant notre modèle inclut des détails substantiellement différents, car on adopte une approche bayésienne pour la construction de la vraisemblance. Pour l'incorporation du modèle de classification aléatoire indexée aux covariables décrivant les caractéristiques des observations nous avons suivi une approche similaire à celle de MURUA et QUINTANA [28].

Le reste du travail est organisé comme suit. Dans le premier chapitre, les concepts principaux en analyse de survie sont révisés, incluant une brève description du modèle original de Ibrahim.

Le chapitre deux est dédié aux modèles de partition aléatoire, on met l'accent sur la présentation du modèle de Potts, qui est le modèle choisi pour le développement du présent projet de recherche. Nous y effectuons un survol du cadre de travail de l'analyse bayésienne classique et bayésienne non paramétrique (BNP). On présente également une brève description des différences entre cette dernière par rapport à la statistique paramétrique classique. Des méthodes numériques qui permettent de développer des inférences dans le contexte bayésien seront aussi discutées, le fonctionnement des MCMC, notamment par la présentation de l'algorithme fondateur de Metropolis-Hastings.

Suite à cette mise en contexte, on présente le modèle proposé dans le chapitre trois, qui sera désigné par le nom de modèle Cox-Potts. Les détails de la construction et les principaux résultats de base sont présentés.

La flexibilité du modèle Cox-Potts est évaluée et comparée à des alternatives compétitives dans un exercice de simulation présenté au chapitre quatre.

Finalement, dans le chapitre cinq nous résumons les points importants de cette recherche et formulons certaines pistes d'exploration futures. Les preuves des résultats techniques sont données dans les annexes.

Chapitre 1

NOTIONS PRÉLIMINAIRES EN ANALYSE DE SURVIE

Nous abordons dans ce chapitre quelques résultats classiques de l'analyse de survie pour le cas des variables aléatoires continues. Les fonctions de base dans un tel contexte ainsi que les liens entre elles sont mentionnés à la section 1.2. Le concept de censure ainsi que les différents types de censure sont définis à la section 1.3. Nous introduisons ensuite brièvement les modèles de régression dont le but est de mesurer la dépendance entre la survie et un ensemble de variables indépendantes. On décrit plus en détail le modèle à risques proportionnels de COX [4] à la section 1.5 et le modèle exponentiel par intervalles de IBRAHIM [18] à la section 1.6, car notre recherche est en grande partie basée sur ces deux classes de modèles.

1.1. INTRODUCTION

Soit Y , le temps écoulé jusqu'à l'occurrence d'un événement spécifique. Cet événement peut être la mort, l'apparition d'une tumeur, le développement d'une certaine maladie, la panne d'une machine, etc. Comme nous nous intéressons à un laps de temps avant la survenance d'un certain événement, on aura $Y > 0$. Nous discuterons dans cette section de trois fonctions qui caractérisent la distribution de Y , à savoir la fonction de survie, la fonction de risque, et la fonction de densité de probabilité. Si nous connaissons l'une de ces fonctions, alors les deux autres peuvent être déterminées de manière unique.

1.2. FONCTIONS DE BASE

1.2.1. Fonction de Survie

La fonction de base employée pour décrire ce type de données est la fonction de survie, $S(y)$, définie comme la probabilité pour un individu de survivre au-delà du temps y . Si Y est une variable aléatoire continue,

$$S(y) = P(Y > y) = \int_y^\infty \pi(t)dt, \quad (1.2.1)$$

où $\pi(\cdot)$ est la fonction de densité associée à Y . Notons que $\pi(y)$ est non-négative et que l'aire sous sa courbe est égale à un. La fonction de survie est le complément de la fonction de répartition, c'est-à-dire $S(y) = 1 - F(y)$, où $F(y) = P(Y \leq y)$. De (1.2.1), on peut écrire la fonction de densité comme

$$\pi(y) = -\frac{dS(y)}{dy}.$$

La fonction de survie $S(y)$ a les propriétés suivantes,

1. S est une fonction monotone, non croissante,
2. $S(0) = 1$,
3. $\lim_{y \rightarrow \infty} S(y) = 0$,
4. S est continue à droite.

1.2.2. Fonction de Risque

Une autre quantité fondamentale en analyse de survie est la fonction de risque, aussi appelée taux de panne, définie comme

$$h(y) = \lim_{\Delta y \rightarrow 0} \frac{P(y \leq Y < y + \Delta y \mid Y \geq y)}{\Delta y}. \quad (1.2.2)$$

La quantité $h(y)\Delta y$ peut être considérée comme la probabilité approximative pour un individu d'âge y de réaliser l'événement d'intérêt dans l'instant suivant. Si Y est une variable aléatoire continue, en notant que

$$P(y \leq Y < y + \Delta y \mid Y \geq y) = \frac{P(y \leq Y < y + \Delta y)}{P(Y \geq y)} = \frac{F(y + \Delta y) - F(y)}{S(y)},$$

nous pouvons réécrire (1.2.2) ainsi

$$h(y) = \frac{1}{S(y)} \lim_{\Delta y \rightarrow 0} \frac{F(y + \Delta y) - F(y)}{\Delta y} = \frac{F'(y)}{S(y)},$$

et donc,

$$h(y) = \frac{\pi(y)}{S(y)} = \frac{-S'(y)}{S(y)} = -\frac{d}{dy} \log(S(y)). \quad (1.2.3)$$

Une fonction connexe est la fonction de risque cumulé $H(y)$, définie comme

$$H(y) = \int_0^y h(t)dt = -\log(S(y)). \quad (1.2.4)$$

Ainsi, pour Y continue,

$$S(y) = \exp(-H(y)) = \exp\left(-\int_0^y h(t)dt\right).$$

La fonction de risque est généralement plus informative que la fonction de survie, car son analyse (intervalles de croissance et de décroissance de la fonction) nous renseigne sur le type

de phénomène étudié. Pour cette raison, elle est plus souvent utilisée dans la description des données. Les seules restrictions sur $h(y)$ est qu'elle doit être non négative et que $\int_0^y h(t)dt \xrightarrow{y \rightarrow \infty} \infty$. À titre d'exemple considérons la distribution de Weibull. Supposons que Y possède la fonction de densité suivante

$$\pi(y) = \begin{cases} \alpha \lambda y^{\alpha-1} \exp(-\lambda y^\alpha) & \text{si } y > 0, \alpha > 0, \lambda > 0, \\ 0 & \text{sinon.} \end{cases} \quad (1.2.5)$$

De (1.2.1) on trouve que la fonction de survie de Y est $S(y) = \exp(-\lambda y^\alpha)$. De (1.2.3), la fonction de risque est $h(y) = \alpha \lambda y^{\alpha-1}$, qui est monotone croissante quand $\alpha > 1$ et monotone décroissante si $0 < \alpha < 1$. Finalement de (1.2.4) la fonction de risque cumulé est donnée par $H(y) = \lambda y^\alpha$.

1.2.3. Relations entre fonctions

Les relations entre les diverses quantités discutées précédemment, pour une durée de vie continue Y , peuvent être résumées ainsi

$$S(y) = \int_y^\infty \pi(t)dt = \exp\left(-\int_0^y h(t)dt\right) = \exp(-H(y)), \quad (1.2.6)$$

$$\pi(y) = -\frac{d}{dy}S(y) = h(y)S(y), \quad (1.2.7)$$

$$h(y) = -\frac{d}{dy} \log(S(y)) = \frac{\pi(y)}{S(y)}. \quad (1.2.8)$$

1.3. CENSURE

Une caractéristique particulière, souvent présente dans les données de survie, est connue sous le nom de censure. Quand une donnée est censurée, on sait uniquement que l'occurrence de l'événement d'intérêt est située dans un certain intervalle. Il existe différentes catégories de censure : la censure à droite, la censure à gauche et la censure par intervalle. La vraisemblance du modèle utilisé sera différente selon le type de censure présent dans les données.

1.3.1. Censure à droite

Il y a trois types différents de censure à droite. Nous commençons par décrire la censure de *type I*, où l'événement d'intérêt est observé seulement s'il se produit avant un certain temps prédéfini à l'avance. Par exemple, une étude clinique commence avec un nombre fixe de patients auxquels un traitement est appliqué. À cause des limitations budgétaires ou temporelles, le chercheur mettra fin à l'étude ou rapportera les résultats avant que tous les patients ne subissent l'événement d'intérêt. Dans ce cas, s'il n'y a pas de pertes de sujets, toutes les observations censurées ont des temps de survie égaux à la durée de la période d'étude. Soit C_r (C_r pour *right*), le temps de fin de l'étude. La durée de vie exacte d'un

individu, dénoté par la lettre T , sera connue si, et seulement si, T est inférieure ou égale à C_r . Si T est supérieure à C_r , le temps de survie, Y est censuré à C_r . Les données censurées à droite peuvent être facilement représentées par des paires de variables aléatoires (Y, ν) , où ν indique si la durée de vie Y est observée, ($\nu = 1$) ou censurée ($\nu = 0$). Le temps de survie, Y est égal à T , si la durée de vie est observée, et à C_r si elle est censurée, c'est-à-dire $Y = \min\{T, C_r\}$. Dans ce mémoire nous nous concentrons à développer un modèle qui tient compte de ce type de censure.

La censure de *type II*, quant à elle, consiste à poursuivre l'étude jusqu'à la mort des premiers r individus, où r est un entier déterminé à l'avance et qui est inférieur au nombre total d'individus dans l'étude ($r < n$). Les expériences où l'on observe la censure de *type II* sont souvent utilisées pour tester la durée de vie d'équipements spécialisés. Une telle expérience peut aider à économiser du temps et de l'argent, car cela peut prendre longtemps avant que tous les individus sous étude éprouvent l'événement d'intérêt. Il est à noter que l'inférence statistique concernant des données censurées de *type II* est plus simple, car l'échantillon est composé des r plus petites durées de vie parmi n individus choisis aléatoirement, de sorte que la théorie des statistiques d'ordre est directement applicable à toute technique d'inférence utilisée. Il est à noter aussi que le temps de censure ici $T_{(r)}$, correspond à la durée de vie ordonnée de la r -ième observation, qui est une quantité aléatoire.

Le troisième type de censure à droite, dite censure aléatoire, survient lorsque le temps d'entrée et de sortie des individus étudiées ne sont pas fixes, mais aléatoires. Dans ce cas-là, le temps de censure peut varier d'une observation à l'autre.

1.3.2. Censure à gauche et censure par intervalle

Une durée de vie Y est censurée à gauche si l'événement d'intérêt a déjà eu lieu pour l'individu avant que cette personne ne soit observée dans l'étude au temps C_l (C_l pour *left*). Pour ces individus, nous savons qu'ils ont éprouvé l'événement avant le temps C_l , mais le temps exact est inconnu. La durée de vie exacte sera connue si, et seulement si, T est supérieure ou égale à C_l . Les données d'un schéma d'échantillonnage censuré à gauche peuvent être représentées par des paires de variables aléatoires (Y, ϵ) , où ϵ indique si la durée de vie exacte T est observée ($\epsilon = 1$) ou non ($\epsilon = 0$). Notez que, pour la censure à gauche, contrairement à la censure à droite, $Y = \max\{T, C_l\}$.

Un type de censure plus général se produit lorsque la durée de vie d'un individu n'est connue que dans un intervalle. Un tel type de censure peut survenir lorsque les patients ont un suivi périodique et le temps où le patient subit l'événement n'est alors connu que dans un intervalle $(L_i, R_i]$. Ce type de censure peut, par exemple, se produire dans des expériences industrielles où il y a une inspection périodique des équipements (T. et JOHN [40]).

1.3.2.1. Hypothèse de censure indépendante

La conception des expériences dans un contexte d'analyse de survie avec des données censurées doit être soigneusement prise en compte lors de la construction de la vraisemblance. Une hypothèse de base est que les durées de vie et les temps de censure sont indépendants.

1.4. MODÈLES DE RÉGRESSION POUR DONNÉES DE SURVIE

Il est souvent d'intérêt de connaître la relation entre le temps de survie, Y et une ou plusieurs variables explicatives. Ce serait le cas si, par exemple, l'on comparait les fonctions de survie de deux groupes recevant des traitements différents ou bien, si l'on cherchait à identifier les facteurs de risque d'une maladie particulière. Un autre problème important est de prédire la distribution du temps de l'événement à partir d'un ensemble de variables indépendantes.

Considérons le temps de survie, $Y > 0$, et un vecteur $\mathbf{x}' = (x_1, \dots, x_p)$ de variables explicatives associées à une certaine réalisation de la variable Y . Le vecteur \mathbf{x} peut inclure des variables quantitatives ou qualitatives. Deux approches de modélisation des effets des covariables sont devenues populaires dans la littérature statistique. La première est analogue à l'approche de régression linéaire classique : le logarithme naturel du temps de survie, $W = \log(Y)$, est modélisé comme

$$W = \mu + \mathbf{x}'\boldsymbol{\beta} + \sigma w,$$

où $\boldsymbol{\beta}$ est le vecteur de coefficients et w est une v.a. Des choix communs pour w sont la distribution normale standard, qui mène à un modèle de régression log-normal, la distribution des valeurs extrêmes, qui mène à un modèle de régression Weibull et la distribution logistique qui mène à une régression log-logistique.

L'estimation des coefficients de régression est effectuée à l'aide de la méthode du maximum de vraisemblance. Le taux de risque d'un individu avec une valeur de covariable \mathbf{x} pour cette classe de modèles est lié à un taux de risque de base h_0 donné par

$$h(y | \mathbf{x}) = h_0(y \exp(-\mathbf{x}'\boldsymbol{\beta})) \exp(-\mathbf{x}'\boldsymbol{\beta}).$$

La deuxième approche de modélisation consiste à exprimer le taux de risque conditionnel en fonction des variables indépendantes à travers deux classes générales de modèles : la famille des modèles multiplicatifs et la famille des modèles additifs. Pour la famille des modèles multiplicatifs, le taux de risque conditionnel d'un individu avec le vecteur des covariables \mathbf{x} est le produit d'un taux de risque de base $h_0(y)$ et une fonction non-négative des covariables, $c(\mathbf{x}'\boldsymbol{\beta})$,

$$h(y | \mathbf{x}) = h_0(y)c(\mathbf{x}'\boldsymbol{\beta}).$$

Ceci est un modèle semi-paramétrique parce qu'une forme paramétrique est supposée seulement pour l'effet des covariables. Comme $h(y | \mathbf{x})$ doit être positive, la plupart des applications utilisent le modèle de COX [4], avec $c(\mathbf{x}'\boldsymbol{\beta}) = \exp(\mathbf{x}'\boldsymbol{\beta})$. Alors, nous pouvons réécrire (1.2.6) comme

$$S(y | \mathbf{x}) = \exp\left(-\int_0^y h(t | \mathbf{x})dt\right) = \exp\left(-\int_0^y h_0(t) \exp(\mathbf{x}'\boldsymbol{\beta})dt\right). \quad (1.4.1)$$

En continuation à l'exemple introduit en (1.2.5), considérons le modèle de Cox pour la distribution Weibull, le taux de risque de base est alors donné par $h_0(y) = \alpha\lambda y^{\alpha-1}$, on obtient donc que le taux de risque est exprimé comme $h(y | \mathbf{x}) = \alpha\lambda y^{\alpha-1} \exp(\mathbf{x}'\boldsymbol{\beta})$. Ainsi, la fonction de survie devient

$$S(y | \mathbf{x}) = \exp\left(-\int_0^y \alpha\lambda t^{\alpha-1} \exp(\mathbf{x}'\boldsymbol{\beta})dt\right) = \exp(-\lambda y^\alpha \exp(\mathbf{x}'\boldsymbol{\beta})),$$

et la fonction de répartition est donnée par, $F(y | \mathbf{x}) = 1 - S(y | \mathbf{x})$. La deuxième classe des modèles est la famille des modèles à taux de risque additif. Ici, nous spécifions la fonction de risque conditionnel par

$$h(y | \mathbf{x}) = h_0(y) + \sum_{j=1}^p x_j(y)\beta_j(y). \quad (1.4.2)$$

Les coefficients de régression pour cette classe des modèles sont des fonctions du temps. Les p fonctions de régression peuvent être positives ou négatives, mais leur valeurs sont contraintes, car (1.4.2) doit être positif. L'estimation de cette classe de modèles est généralement faite par des méthodes non paramétriques des moindres carrés (KLEIN et MOESCHBERGER [20]).

Un exemple de modèle qui utilise une taux de risque additif est le modèle de régression de Cox avec fonction de risque par intervalles de loi exponentielle, qui sera détaillé plus tard dans la section 1.6. Pour cette classe de modèle on a que,

$$H_0(y) = \lambda_j(y - s_{j-1}) + \sum_{g=1}^{j-1} \lambda_g(s_g - s_{g-1}).$$

1.5. MODÈLE DE RÉGRESSION DE COX

Dans la section 1.4, nous avons introduit des modèles qui nous permettent de quantifier la relation entre le temps écoulé avant la réalisation d'un événement et un ensemble de variables explicatives. Dans cette section, nous examinerons plus en détail le modèle à risques multiplicatifs largement utilisé, dû à Cox [4].

Comme précédemment, considérons un temps de défaillance $Y > 0$ et un vecteur $\mathbf{x} \in \mathbb{R}^p$ de variables explicatives associées à Y . Nous allons considérer ici le cas des covariables fixes dans le temps. Soit $h(y | \mathbf{x})$ le taux de risque au temps y pour un individu avec vecteur de

covariables \mathbf{x} . Le modèle de Cox est défini par la relation suivante

$$h(y | \mathbf{x}) = h_0(y) \exp(\mathbf{x}'\boldsymbol{\beta}). \quad (1.5.1)$$

Le modèle de Cox est souvent appelé modèle à risques proportionnels, car si nous examinons deux individus avec des valeurs de covariables \mathbf{x}_1 et \mathbf{x}_2 respectivement, le rapport de leur taux de risque est

$$\frac{h(y | \mathbf{x}_1)}{h(y | \mathbf{x}_2)} = \frac{h_0(y) \exp(\mathbf{x}'_1\boldsymbol{\beta})}{h_0(y) \exp(\mathbf{x}'_2\boldsymbol{\beta})} = \exp((\mathbf{x}_1 - \mathbf{x}_2)'\boldsymbol{\beta}),$$

qui ne dépend pas du temps.

1.5.1. Vraisemblance partielle de Cox

D'un côté, une observation correspondant à un temps de survie exact fournit des informations sur la probabilité que l'événement se produise à ce moment précis, ce qui est approximativement égal à la fonction de densité de Y pour ce temps observé. D'un autre côté, une observation censurée à droite nous indique que le vrai temps de réalisation de l'événement d'intérêt est plus grand que le temps de censure, donc l'information fournie par cette observation-là est quantifiée par la fonction de survie évaluée au temps de censure C_r .

La vraisemblance de l'échantillon complet pour le modèle de Cox est la suivante,

$$L = \prod_{i=1}^n \pi(y_i)^{\nu_i} S(y_i)^{1-\nu_i}, \quad (1.5.2)$$

en (1.2.7) on a vu que $\pi(y_i) = h(y_i)S(y_i)$, donc, on peut écrire (1.5.2),

$$L = \prod_{i=1}^n h(y_i)^{\nu_i} \exp(-H(y_i)).$$

Définissons $D_n = (Y^n, X^n, \boldsymbol{\nu}^n)$, l'ensemble de données complets, où $X^n = (\mathbf{x}_1, \dots, \mathbf{x}_n)$, $\mathbf{x}_i \in \mathbb{R}^p$, correspond au vecteur de covariables pour la i -ième observation, $Y^n = (y_1, \dots, y_n)$ est le vecteur de temps de défaillance et finalement, $\boldsymbol{\nu}^n \in \mathbb{R}^n$ est le vecteur de variables indicatrices de censure. Si nous considérons maintenant le modèle de Cox, l'expression ci-dessus peut être réécrite comme

$$\begin{aligned} L(\boldsymbol{\beta}, h_0(y) | D_n) &\propto \prod_{i=1}^n h(y_i | \mathbf{x}_i)^{\nu_i} \exp(-H(y_i | \mathbf{x}_i)) \\ &\propto \prod_{i=1}^n h_0(y_i)^{\nu_i} \exp(\mathbf{x}'_i\boldsymbol{\beta})^{\nu_i} \exp(-H_0(y_i) \exp(\mathbf{x}'_i\boldsymbol{\beta})), \end{aligned} \quad (1.5.3)$$

qui est la vraisemblance de Cox.

1.6. RÉGRESSION DE COX AVEC FONCTION DE RISQUE PAR INTERVALLES DE LOI EXPONENTIELLE

Dans cette section nous faisons une brève description du modèle premièrement présenté par IBRAHIM, CHEN et KIM [19], duquel nous nous sommes en partie inspirés pour le développement de ce mémoire. Il s'agit d'un modèle qui permet l'utilisation de différents taux de risque. On divise l'axe du temps en J sous-intervalles, $(0, s_1], (s_2, s_3], \dots, (s_{J-1}, s_J]$, et dans chacun des ces intervalles on suppose une fonction de risque, $h(y)$ différente pour les observations qui s'y trouvent. Un choix judicieux des points de séparation devrait nous permettre d'approcher raisonnablement bien à la fonction de risque objectif. Il serait avantageux d'utiliser des intervalles étroits où le risque varie rapidement et des intervalles plus larges où le risque change plus lentement.

Plus spécifiquement, dans le j -ième intervalle, nous supposons une fonction de risque de base, $h_0(y) = \lambda_j$ constante pour $y \in I_j = (s_{j-1}, s_j]$.

Si l'on écrit $\lambda = (\lambda_1, \dots, \lambda_J)'$, de (1.2.4) la fonction de risque cumulé pour $y \in (s_{j-1}, s_j]$ est

$$H_0(y) = \int_0^{s_1} \lambda_1 dx + \int_{s_1}^{s_2} \lambda_2 dx + \dots + \int_{s_{j-1}}^y \lambda_j dx = \lambda_j(y - s_{j-1}) + \sum_{g=1}^{j-1} \lambda_g(s_g - s_{g-1}).$$

De (1.5.3) on peut écrire la vraisemblance de (β, λ) pour les n sujets comme

$$L(\beta, \lambda | D_n) = \prod_{i=1}^n \prod_{j=1}^J \left(\lambda_j \exp(\mathbf{x}'_i \beta) \right)^{\delta_{ij} \nu_i} \exp \left\{ -\delta_{ij} \left[\lambda_j (y_i - s_{j-1}) + \sum_{g=1}^{j-1} \lambda_g (s_g - s_{g-1}) \right] \exp(\mathbf{x}'_i \beta), \right\} \quad (1.6.1)$$

où $\delta_{ij} = 1$ si $y_j \in (s_{j-1}, s_j]$ et 0 sinon. Notons que la variable indicatrice δ_{ij} est indispensable pour définir la vraisemblance sur les J intervalles. Remarquons également que si $J = 1$, ce modèle est réduit au modèle paramétrique exponentielle avec taux de risque $\lambda = \lambda_1$.

Chapitre 2

MODÈLES DE PARTITIONS ALÉATOIRES

2.1. INTRODUCTION

Les algorithmes de classification visent à diviser un ensemble de données en différents sous-ensembles homogènes. Habituellement le nombre de sous-ensembles ainsi que leur taille n'est pas connu *a priori*. Nous nous concentrons ici dans les modèles de partition aléatoire, qui sont des modèles probabilistes pour générer des partitions, de sorte que les procédures algorithmiques classiques sont évités.

Dans ce chapitre, on présente une brève description des processus les plus populaires en statistique bayésienne non paramétrique. On établit les différences entre l'approche bayésienne classique et l'approche bayésienne non paramétrique, ainsi que les avantages de cette dernière sur la statistique paramétrique classique. Nous introduisons aussi le processus de Dirichlet, qui servira comme concept de base pour la construction de certains modèles de partition aléatoire. Dans la section 2.5 on détaille le modèle de Potts, qui est le modèle de partition aléatoire employé dans ce mémoire. On présente également des méthodes numériques pour l'implémentation : l'échantillonnage de Gibbs et l'algorithme classique de Metropolis-Hastings, qui sont brièvement décrits dans la section 2.6.

2.2. LE PARADIGME BAYÉSIEN

Dans le cadre de l'inférence bayésienne, on considère que l'incertitude sur le paramètre θ d'un certain modèle peut être décrite par une distribution de probabilité $\pi(\cdot)$ sur θ , appelée distribution *a priori*. La loi de θ conditionnelle au vecteur d'observations Y^n , est écrite comme $\pi(\theta | Y^n)$, et est appelée distribution *a posteriori*. Pour obtenir une telle loi, le théorème de Bayes est appliqué,

$$\pi(\theta | Y^n) = \frac{L(Y^n | \theta)\pi(\theta)}{\int_{\Theta} L(Y^n | \theta)\pi(\theta)d\theta} \quad \theta \in \Theta, \quad (2.2.1)$$

où Θ représente le support des paramètres et comme avant, $L(Y^n | \theta)$ dénote la fonction de vraisemblance. Nous supposerons par la suite que toutes ces lois sont continues.

Dans le paradigme bayésien, le paramètre inconnu θ n'est plus considéré comme inconnu et déterministe, mais comme une variable aléatoire.

Cette approche d'inférence statistique est moderne et apte à répondre aux problèmes de modélisation les plus avancés dans nombreuses disciplines. Cependant, les deux difficultés majeures liées à l'approche bayésienne sont le choix de l'*a priori* et le calcul de l'intégral au dénominateur en (2.2.1), car souvent cette quantité n'a pas une solution analytique connue. Pour répondre à cette dernière difficulté, des méthodes numériques, mieux connues sous le nom de MCMC (de l'anglais *Markov chain Monte Carlo*), ont été développées. Cette classe d'algorithmes sera présentée plus en détail dans la section 2.6.

2.3. INTRODUCTION À LA STATISTIQUE BAYÉSIENNE NON PARAMÉTRIQUE

L'approche bayésienne non paramétrique, BNP, est une méthode alternative à la statistique paramétrique classique. Les modèles traditionnels paramétriques utilisent un nombre fixe et fini de paramètres, et ceci peut causer un sur-ajustement ou un sous-ajustement quand les données et la complexité du modèle ne sont pas en accord (TEH [41]). D'un autre côté, les modèles BNP utilisent un espace de paramètres de dimension infinie et seulement un sous-ensemble des paramètres disponibles est utilisé pour décrire les données (ORBANZ et TEH [32]).

L'inférence bayésienne non paramétrique nous permet de supposer une incertitude quant au choix du modèle et en même temps de garder une méthode réalisable. Un modèle BNP est défini comme un modèle de probabilité sur un espace paramétrique infini, comme par exemple un espace fonctionnel (MÜLLER et QUINTANA [27]). Effectivement, l'inférence BNP se distingue de l'inférence bayésienne classique à cause de la loi *a posteriori* sur laquelle elle est basée. La loi *a posteriori* dans un contexte BNP est donc indexée par un ensemble infini de paramètres. Pour illustrer les affirmations précédentes, considérons l'exemple suivant : un problème typique en inférence consiste à construire une approximation de la loi ayant généré un échantillon d'observations. Dans ce cas-là, on aura les observations y_1, \dots, y_n indépendantes et identiquement distribuées, i.i.d. avec distribution F , $y_i \sim F$. Dans l'approche bayésienne classique, on assignerait une distribution *a priori* à F et après on estimerait la loi *a posteriori* avec les données fournies. D'habitude, cette loi *a priori* appartient à une famille paramétrique, qui dans certains cas est restreinte et réduit donc les types d'inférences qui peuvent être effectués. L'inférence BNP utilise une loi *a priori* provenant d'une famille de distributions avec grand support, habituellement l'espace de toutes les distributions (TEH [41]).

2.3.1. Processus de Dirichlet

Le processus de Dirichlet, DP, introduit par FERGUSON [9], est un processus stochastique utilisé dans la modélisation bayésienne non paramétrique, il est l'un des modèles les plus

populaires dans ce contexte. Grâce à sa flexibilité, il admet diverses représentations. Il s'agit d'une distribution de distributions, autrement dit, chaque tirage d'un DP est considérée une distribution. Ce processus doit son nom au fait que ses lois marginales (de dimensions finies) ont une distribution de Dirichlet. Les distributions tirées d'un DP sont discrètes, mais elles ne peuvent pas être caractérisées en utilisant un nombre fini de paramètres, d'où la classification de non paramétrique (FERGUSON [10]).

2.3.1.1. Représentation Stick-breaking du Processus de Dirichlet

Cette représentation est due à SETHURAMAN [35]. Soit $\varphi_\theta(\cdot)$ une masse ponctuelle en θ , et G une mesure de probabilité discrète définie ainsi,

$$G(\theta) = \sum_{h=1}^{\infty} \pi_h \varphi_{\theta_h}, \quad (2.3.1)$$

où $\theta_h \sim G^*$, i.i.d. et $\pi_h = v_h \prod_{\ell < h} (1 - v_\ell)$ avec $v_h \sim \text{Beta}(1, \alpha)$ i.i.d. De façon générale, on écrit $G \sim DP(\alpha, G^*)$ pour dénoter un DP sur une mesure de probabilité aléatoire. Le paramètre α est appelé paramètre de masse totale et détermine l'incertitude de G , tandis que le paramètre G^* est appelé mesure de base et il spécifie la moyenne du processus $E(G) = G^*$. Une propriété intéressante des DP est la simplicité de la loi *a posteriori* (TEH et JORDAN [42]). Soit

$$y_1, \dots, y_n \mid F \sim F \text{ et } F \sim DP(\alpha, G_0),$$

la loi *a posteriori* de F est

$$F \mid y_1, \dots, y_n \sim DP(\alpha + n, G_1),$$

avec $G_1 = \frac{\alpha}{\alpha+n} G_0 + \frac{1}{\alpha+n} \sum_{i=1}^n \varphi_{\theta_i}$. Remarquons ainsi que le processus de Dirichlet est un processus discret qui peut être représenté comme la somme de masses ponctuelles (MÜLLER et QUINTANA [27]).

2.3.2. Processus de Dirichlet Mixtes

L'une des applications importantes du processus de Dirichlet est dans l'estimation de la densité. Ici, nous sommes intéressés à la modélisation de la loi correspondant à une suite de tirages observés. Afin de ne pas nous limiter à une seule classe paramétrique, on utilise encore une loi *a priori* non paramétrique ayant comme support toutes les densités.

Dans la section précédente, on a mentionné le fait que les tirages provenant d'une DP sont discrets, de sorte que si nous sommes intéressés à estimer la densité, $y_i \sim G$, $i = 1, \dots, n$, il serait inapproprié d'assumer que $G \sim DP$, si la distribution que nous cherchons est en fait continue. Afin d'être en mesure de modéliser une loi continue à travers un processus discret, on prend $y_i \sim F$ où

$$F(y) = \int \pi(y \mid \theta) G(\theta) d\theta,$$

avec $G \sim DP(\alpha, G_0)$ et $\pi(y | \theta)$ dénote une famille de densités indexée par θ (ESCOBAR [8]). Autrement dit, la distribution inconnue est écrite comme un mélange de distributions avec un DP comme *a priori*. Ce modèle est connu sous le nom de Processus de Dirichlet Mixte, DPM (MACEachern [22]).

Une autre application populaire du DP est en classification. Dans les problèmes de classification, on modélise une suite d'observations, y_1, y_2, \dots, y_n à l'aide d'un ensemble de paramètres latents $\{\theta_1, \dots, \theta_n\}$. Chaque θ_i est tiré de façon indépendante de la loi G , tandis que chaque y_i suit la loi $F(\theta_i)$ paramétrée par θ_i

$$\begin{aligned} y_i | \theta_i &\sim F(\theta_i) \\ \theta_i | G &\sim G \\ G | \alpha, G_0 &\sim DP(\alpha, G_0). \end{aligned} \tag{2.3.2}$$

Ici, la loi G est discrète, ce qui implique que plusieurs variables θ_i peuvent prendre la même valeur, de sorte que ce modèle peut être vu comme un modèle de mélange, où les observations y_i issues de la même valeur de θ_i appartiennent au même sous-ensemble. La représentation en (2.3.2) est le modèle hiérarchique équivalent à un DPM, puisque lorsque l'on marginalise par rapport à θ_i , le modèle résultant est encore $y_i \sim \int \pi(y | \theta)G(\theta)d\theta$, i.i.d. (MÜLLER et MITRA [25]).

2.4. QUELQUES EXEMPLES POPULAIRES DE MODÈLES DE PARTITION ALÉATOIRE

Cette section a pour but d'introduire le concept de partition aléatoire, ainsi que de donner quelques exemples de modèles qui génèrent de telles partitions aléatoires. On commence d'abord par introduire quelques notations de base. Soit $U = \{1, 2, \dots, n\}$ l'ensemble des unités étudiées, une partition de U est alors une famille de sous-ensembles $\rho_n = \{U_1, \dots, U_K\}$, où $\bigcup_{k=1}^K U_k = U$, et $U_{k_1} \cap U_{k_2} = \emptyset$ si $k_1 \neq k_2$. Quand ρ_n est issue d'un processus aléatoire, on dit que $\pi(\rho_n)$ est une partition aléatoire. Ainsi, les modèles de partition aléatoire, RPM (de l'anglais *random partition models*), sont des modèles probabilistes pour générer des partitions.

Tel qu'illustré dans la section précédente, il n'est pas rare que dans plusieurs applications les chercheurs s'intéressent à la partition induite par un modèle, plutôt qu'à la mesure aléatoire, G associée. Comme exemple d'introduction, on décrit le modèle d'échantillonnage par espèce, SSMs (de l'anglais *Species Sampling Model*).

2.4.1. Modèle d'échantillonnage par espèce

Le modèle d'échantillonnage par espèces, SSMs, (PITMAN [33]) génère une mesure de probabilité aléatoire de façon indirecte à travers l'établissement d'une règle de prévision pour $\theta_{n+1} | \theta_1, \dots, \theta_n$. On considère que dans une échantillon aléatoire $\theta_i \sim G$, quand la mesure

de probabilité G est discrète, certaines observations dans l'échantillon peuvent être liées entre elles, comme nous l'avons vu précédemment. On utilise ces groupes d'observations liées pour former une partition. Pour mieux décrire l'énoncé précédent, introduisons la notation suivante :

1. Soit K le nombre de valeurs uniques parmi $(\theta_1, \dots, \theta_n)$,
2. les quantités θ_k^* , $k = 1, \dots, K$, indiquent ces valeurs uniques,
3. n_k dénote le nombre de θ_i égaux à la k -ième valeur unique θ_k^* ,
4. $\mathbf{n} = (n_1, \dots, n_K)$ est le vecteur de tailles des sous-ensembles générés par les observations liées.

Ainsi, les sous-ensembles sont définis comme $U_k = \{i : \theta_i = \theta_k^*\}$. Autrement dit, les partitions sont formées à l'aide des $K \leq n$ valeurs uniques de l'échantillon. Remarquons le fait que n'importe quel modèle de probabilité avec mesure aléatoire discrète, G , génère une partition aléatoire $\pi(\rho_n)$ en regroupant les échantillons aléatoires comme décrit ci-haut. Cette affirmation reste valide quand G est une mesure de probabilité aléatoire discrète avec *a priori*, $\pi(G)$. Des choix courants pour $\pi(G)$ sont les DP décrits dans la première section de ce chapitre.

En particulier, la version mixte du Processus de Dirichlet, DPM, révisée dans la section précédente fonctionne de façon similaire aux SSMS, car les paramètres latents θ_i en (2.3.2) introduisent des liens qui définissent de nouvelles partitions. Soit θ_k^* , $k = 1, \dots, K$ les K valeurs uniques parmi les θ_i , $i = 1, \dots, n$, alors les sous-ensembles de la partition sont définis de la même façon que dans le SSMS, c'est-à-dire, $U_k = \{i : \theta_i = \theta_k^*\}$, pour $k = 1, \dots, K$. Une autre propriété intéressante des DPM dans ce contexte est la simplicité de la loi *a priori* de ρ_n , (MÜLLER et MITRA [25]; QUINTANA et IGLESIAS [34]).

$$\pi(\rho_n) \propto \alpha^K \prod_{k=1}^K (n_k - 1)!, \quad (2.4.1)$$

donc, ceci est un modèle qui dépend du cardinal de chaque sous-ensemble généré U_k . Ici, comme avant, n_k indique la taille du k -ième grappe, U_k , et α est le paramètre de masse totale introduit dans la section 2.3.1.1. Remarquons que (2.4.1) peut être vu comme un produit de fonctions $c(U_k) = \alpha(n_k - 1)!$, où chaque terme dépend d'un seul sous-ensemble. Les modèles de partition aléatoire de la forme $\pi(\rho_n) \propto \prod_{k=1}^K c(U_k)$ pour une certaine fonction $c(U_k)$ sont connus sous le nom de modèles partition-produit, *Product Partition Models*, PPM (BARRY et HARTIGAN [1]). Dans la section suivante, on décrit plus en détail cette classe de modèles.

2.4.2. Modèles partition-produit

Une autre classe de modèles de partitions aléatoire est la classe des modèles partition-produit, PPM (HARTIGAN [15]; BARRY et HARTIGAN [1]; CROWLEY [5]). Les PPM sont définis sous une loi de probabilités $\pi(\rho_n)$ sur toutes les partitions possibles. Cette classe

de modèles utilise une fonction non-négative $c(U_k)$, mieux connue sous le nom de fonction de cohésion, afin de déterminer la probabilité associée à une certaine partition. En effet, la fonction de cohésion décrit à quel point l'utilisateur pense que les éléments de U sont regroupés *a priori*. On dit qu'une partition aléatoire $\pi(\rho_n)$, est issue d'un PPM si elle peut être écrite comme un produit de facteurs qui dépendent seulement de U_k , $k = 1, \dots, K$.

$$\pi(\rho_n) = Z \prod_{k=1}^K c(U_k), \quad (2.4.2)$$

où Z est la constante de normalisation pour cette loi. Soit y_i la réponse pour la i -ième observation, $y_k^* = \{y_i, i \in U_k\}$ les réponses ordonnés par sous-groupes et soit $Y^n = (y_1, \dots, y_n)$ les données au complet. Les PPM combinent la loi *a priori*, $\pi(\rho_n)$ avec un modèle $\pi(Y^n | \rho_n)$ qui favorise le regroupement des observations similaires et assume que les partitions sont échangeables, c'est-à-dire,

$$\pi(Y^n | \rho_n) = \prod_{k=1}^K \pi(y_k^*),$$

où $\pi(y_k^*)$ est la loi marginale de $y_i, i \in U_k$ sous la partition ρ_n . De façon plus formelle, une séquence de variables aléatoires, v.a. X_1, X_2, \dots, X_n sous un espace de probabilité est dite échangeable si la loi de distribution conjointe est invariante sous une permutation de variables, c'est-à-dire, si P est la distribution conjointe et σ est une permutation de $U = \{1, 2, \dots, n\}$, alors,

$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = P(X_1 = x_{\sigma(1)}, X_2 = x_{\sigma(2)}, \dots, X_n = x_{\sigma(n)}).$$

Pour un modèle échangeable $\pi(y_k^*)$, conditionné à une certaine partition, le PPM assume indépendance entre partitions. On a donc,

$$\pi(Y^n | \rho_n) = \prod_k \pi(y_k^* | \mu_k^*),$$

où μ_k^* dénote les paramètres de la partition. Plusieurs applications des PPM utilisent souvent l'interchangeabilité de y_i à travers $i \in U_k$ en assumant que les $y_i, i \in U_k$ sachant μ_k^* sont indépendantes. Une autre propriété intéressante des PPM est que la loi a posteriori $\pi(\rho_n | Y^n)$ est encore un PPM, avec fonction de cohésion $c(U_k)\pi(y_k^*)$ (QUINTANA et IGLESIAS [34]).

2.4.2.1. Modèles partition-produit avec covariables

MÜLLER, QUINTANA et ROSNER [31] ont introduit une version alternative du PPM dépendante des covariables associées aux observations, connue sous le nom de PPMx. Cette version modifie l'*a priori* en (2.4.2) par

$$\pi(\rho_n | X^n) \propto \prod_{k=1}^K c(U_k)g(\mathbf{x}_k^*), \quad (2.4.3)$$

avec constante de normalisation $\sum_{\rho_n} \prod_{k=1}^K c(U_k) g(\mathbf{x}_k^*)$. Ici, $g(\mathbf{x}_k^*)$ est une mesure de similarité pour \mathbf{x}_k^* . Rappelons nous que $\mathbf{x}_k^* = \{\mathbf{x}_i : i \in U_k\}$ est l'ensemble de covariables associées aux individus appartenant au k -ième sous-ensemble, $k = 1, \dots, K$ et $X^n = (\mathbf{x}_1 \dots, \mathbf{x}_n)$. Des grandes valeurs de la fonction $g(\cdot)$ indiquent des valeurs des covariables homogènes au sein du sous-ensemble correspondant. Nous clarifions que la notation « $| X^n$ » ne fait pas référence à un conditionnement, mais plutôt à une indexation de la distribution résultante due à l'effet des covariables. Sous ce modèle, deux individus i_1 et i_2 avec vecteurs de covariables similaires \mathbf{x}_{i_1} et \mathbf{x}_{i_2} devraient être plus susceptibles de se trouver dans le même sous-ensemble *a priori*.

2.5. MODÈLE DE POTTS

Le modèle présenté dans cette section est un modèle probabiliste de classification, premièrement introduit dans un contexte statistique par MURUA et collab. [29]. Il est inspiré du modèle de Potts, largement utilisé en physique pour décrire les phénomènes de ferromagnétisme. Ce modèle est appliqué à un système de particules, représentées par des observations, et leurs interactions, représentées par une mesure de similarité entre observations. La distribution d'un tel système dépend d'une température donnée, T . Pour chaque température il existe une probabilité associée à chaque partition possible. Un des grands avantages de ce modèle, est qu'il prend en considération l'information des covariables et qu'il a donc la propriété d'encourager le regroupement d'individus avec caractéristiques similaires.

2.5.1. Le modèle de Potts appliqué à un graphe

Pour faciliter le développement à venir, nous considérons que nos données, (X^n, Y^n) sont représentées au moyen du graphe $\mathcal{G} = \{(X^n, Y^n), E\}$, avec n sommets $((\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n))$ et un ensemble d'arêtes, E , qui représente en quelque sorte un lien de proximité entre les covariables. Afin de faciliter la compréhension, on fera référence au graphe $\mathcal{G}_X = \{X^n, E\}$ avec sommets $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ mais dont la structure est exactement la même que \mathcal{G} .

La loi de probabilité associée au modèle de Potts dépend des paramètres T et σ , ainsi que de X^n . Le paramètre $\sigma > 0$ est appelé paramètre de largeur de bande. On introduit également les variables $\{z_{ij}\}$ qui représentent l'ensemble des étiquettes de couleur dont la fonctionnalité est décrite comme suit. Soit $r > 1$, un entier indiquant le nombre de couleurs différentes possibles dans le graphe, et soit z_{ji} la variable indicatrice de l'étiquette de couleur,

$$z_{ji} = \begin{cases} 1 & \text{si } \mathbf{x}_j \text{ est assignée à la couleur } i \\ 0 & \text{sinon.} \end{cases}$$

Nous définissons également,

$$\alpha_{ij} = \begin{cases} 1 & \text{si } \mathbf{x}_i \text{ et } \mathbf{x}_j \text{ sont voisins dans le graphe } \mathcal{G}_X \\ 0 & \text{sinon,} \end{cases}$$

c'est-à-dire, $\alpha_{ij} = 1$ s'il y a une arête entre \mathbf{x}_i et \mathbf{x}_j , et 0 sinon. Finalement, soit γ_{ij}

$$\gamma_{ij} = \begin{cases} 1 & \text{si } \mathbf{x}_i \text{ et } \mathbf{x}_j \text{ ont la même étiquette de couleur} \\ 0 & \text{sinon.} \end{cases}$$

Le modèle de probabilité associé est donné par,

$$\pi(\{z_{ij}\} | X^n) = \frac{1}{Z} \exp \left\{ -\frac{1}{T} \sum_{\alpha_{ij}=1} \kappa_{ij}(\sigma)(1 - \gamma_{ij}) \right\}. \quad (2.5.1)$$

Ici Z est la constante de normalisation. Les similitudes entre les paires de vecteurs de covariables sont définies par $\kappa_{ij} = K(\mathbf{x}_i, \mathbf{x}_j)$, où $K(\cdot, \cdot)$ est une fonction noyau qui mesure la similarité entre observations, pour ce projet on a considéré la distance $\|\mathbf{x}_i - \mathbf{x}_j\|$, en donc la fonction κ_{ij} aurait la forme $\kappa_{ij} = \kappa_{ij}(\sigma) = K(\|\mathbf{x}_i - \mathbf{x}_j\|/\sigma)$.

Afin de tirer des échantillons provenant de la loi associée au Potts en (2.5.1), SWENDSEN et WANG [38], ont proposé un algorithme basé sur l'augmentation du modèle en introduisant un ensemble de variables binaires latentes, appelés *liens*, \mathbf{b} . Leur fonctionnement est décrit comme suit. Si \mathbf{x}_i et \mathbf{x}_j sont des points voisins et possèdent la même étiquette de couleur, c'est-à dire si $\alpha_{ij} = \gamma_{ij} = 1$, le *lien* b_{ij} pourra prendre la valeur de 1 avec probabilité $p_{ij} = 1 - \exp\{-\frac{1}{T}\kappa_{ij}(\sigma)\}$. Autrement la valeur de $b_{ij} = 0$. La fonction de distribution conjointe des étiquettes de couleur z_{ij} et des liens b_{ij} est donnée par

$$\pi(\{z_{ij}\}, \{b_{ij}\} | X^n) = \frac{1}{Z} \prod_{\alpha_{ij}=1} \left\{ (1 - p_{ij})(1 - b_{ij}) + p_{ij}b_{ij}\gamma_{ij} \right\}. \quad (2.5.2)$$

Ceci est connu comme le modèle de Fortuin-Kasteleyn-Swendsen-Wang (SOKAL [37]). En marginalisant par \mathbf{b} dans (2.5.2) on obtient l'équation (2.5.1).

Tout comme pour le paramètre de largeur de bande, la température T doit être estimée. Notre approche considère l'estimation du vecteur (σ, T) à partir du mode d'une loi *a priori* de ces deux paramètres, tel que suggéré par MURUA et WICKER [30].

Le graphe de départ est établi selon le contexte d'application et selon un certain critère de proximité, par exemple, pour assigner les valeurs des α_{ij} dans le graphe de départ on peut établir des arêtes entre \mathbf{x}_i et ses k plus proches voisin. Les valeurs de k recommandées dans la littérature oscillent entre $10 \leq k \leq 30$ (MURUA et collab. [29]; BLATT, WISEMAN et DOMANY [3]).

Le modèle de partition aléatoire associé, $\pi(\mathbf{b} | X^n)$ est obtenu en marginalisant par $\{z_{ij}\}$ dans (2.5.2). La distribution de probabilité pour les partitions est donc donnée par

$$\pi(\rho_n | X^n) = \sum_{\mathbf{b} \Rightarrow \rho_n} \pi(\mathbf{b} | X^n). \quad (2.5.3)$$

Ici, la notation $\mathbf{b} \Rightarrow \rho_n$ indique que la partition ρ_n provient des composantes connexes du graphe dont les arêtes sont induites par \mathbf{b} . SWENDSEN et WANG [39] ont utilisé la loi conjointe

en (2.5.2) pour proposer une stratégie de simulation qui s'apparente à l'échantillonnage de Gibbs. Cette stratégie a pour but de générer des observations provenant du modèle de partition aléatoire, $\pi(\mathbf{b} \mid X^n)$ et par conséquent des partitions issues du modèle de Potts associé. L'algorithme en question est le suivant,

0. Les étiquettes de couleur sont tirées de façon indépendante d'une loi uniforme discrète ayant comme support toutes les valeurs de $\{0, 1, \dots, r\}$.
1. Mettre à jour les liens. Étant donné les étiquettes de couleur, $\{z_{ij}\}$, chaque lien b_{ij} peut prendre la valeur de 1, indépendamment des autres, avec probabilité $p_{ij} = 1 - \exp\{-\frac{1}{T}\kappa_{ij}(\sigma)\}$ si $\alpha_{ij} = \gamma_{ij} = 1$, c'est à dire, si les points \mathbf{x}_i et \mathbf{x}_j sont voisins dans le graphe et possèdent la même couleur. Dans le cas contraire, b_{ij} vaut 0.
Les liens permis entre deux points sont établis avec probabilité p_{ij} . Ainsi, la formation de partitions est dictée par les liens, dont la distribution est à son tour déterminée par les similitudes entre les vecteurs de covariables \mathbf{x}_i et \mathbf{x}_j . En conséquence, le modèle privilégie *a priori* les partitions formées par des sous-ensembles d'individus ayant des valeurs de covariables similaires.
2. Mettre à jour les couleurs. Étant donné les liens résultant de l'étape précédente, $\{b_{ij}\}$ chaque composante connexe du graphe, (sous-ensemble connecté par ces liens) se voit assigner la même étiquette de couleur tirée aléatoirement selon une loi uniforme.
3. Répéter les étapes 1 et 2.

Remarquons le fait qu'à la fin de l'étape 2, tous les sommets d'une même composante connexe partagent la même étiquette de couleur. En assignant une nouvelle couleur aux sommets dans cette composante connexe, les sommets voisins avec la même couleur deviennent des candidats pour former une nouvelle composante connexe dans l'itération suivante (ils se fusionnent). Au contraire, les sommets liés durant une itération donnée, ont une probabilité $1 - p_{ij}$ de ne plus l'être, en conséquent ils ne feront possiblement plus partie de la même composante (ils se fragmentent).

Donc, les liens peuvent être vus comme un sous-ensemble des arêtes qui aident à séparer ou fusionner des composantes dans le graphe, au fil des itérations. Chaque composante connexe générée par les b_{ij} à la fin de l'étape 1 forme un sous-ensemble différent, qui tous ensembles déterminent une nouvelle partition des données. Ainsi, chaque itération génère une partition de données différente. (MURUA et QUINTANA [28]).

Les partitions aléatoires issues du modèle de Potts correspondent donc au composantes connexes générés dans le graphe au fil des itérations.

2.6. MÉTHODES DE MONTE-CARLO PAR CHAÎNES DE MARKOV

Dans la section 2.2 on a introduit le besoin de calculer une solution approximative à $m(Y^n) = \int_{\Theta} L(Y^n \mid \theta)\pi(\theta)d\theta$, car souvent cette quantité n'a pas une solution analytique.

Les méthodes de Monte-Carlo par chaînes de Markov, MCMC, sont une classe de méthodes d'échantillonnage à partir de distributions de probabilité visant à calculer des intégrales en plusieurs dimensions. On cherche à calculer une bonne approximation des intégrales de la forme

$$\int_{\Theta} h(\theta)\pi(\theta)d\theta, \quad (2.6.1)$$

où $\pi(\theta)$ est la densité de θ avec support en Θ . Notons que (2.6.1) correspond à l'espérance *a posteriori* de $h(\theta)$, $E_{\pi}(h(\theta))$. Donc, la stratégie consiste à générer n variables aléatoires, $\theta_i \in \Theta$, $i = 1, \dots, m$, selon la distribution de θ , pour ensuite estimer (2.6.1) comme

$$\frac{1}{m} \sum_{i=1}^m h(\theta_i). \quad (2.6.2)$$

Dans le cadre des MCMC, les variables θ_i ne seront pas indépendantes, mais proviendront plutôt d'une chaîne de Markov ergodique avec distribution stationnaire $\pi(\cdot)$. La convergence du processus peut être justifiée de façon formelle à l'aide du théorème ergodique pour les chaînes de Markov (FOGUEL [12]).

Théorème 2.6.1. *Soit une fonction $h : \Theta \rightarrow \mathbb{R}$ telle que $\pi(h) = E_{\pi}(h(\theta)) \in \mathbb{R}$. Alors,*

$$\frac{1}{m} \sum_{i=1}^m h(\theta_i) \xrightarrow[m \rightarrow \infty]{p.s.} E_{\pi}(h(\theta)) = \pi(h), \quad (2.6.3)$$

où *p.s.* dénote la convergence presque sûre. Ici, on suppose que la suite de variables aléatoires $\{\theta_i\}_{i \geq 0}$ est une chaîne de Markov avec distribution stationnaire $\pi(\cdot)$.

Afin de générer cette chaîne de Markov, plusieurs algorithmes ont été développés, on décrit dans les sections suivantes l'échantillonnage de Gibbs et l'algorithme de Metropolis-Hastings, MH, dû à METROPOLIS [24] et à HASTINGS [16].

2.6.1. Échantillonnage de Gibbs

L'échantillonnage de Gibbs, formellement introduit par GEMAN et GEMAN [14] est décrit comme suit. Soit $\theta = (\theta_1, \theta_2, \dots, \theta_p)$ le vecteur de paramètres de dimension p et soit $\pi(\theta | Y^p)$ la distribution *a posteriori* de θ .

0. Choisir un point de départ arbitraire $\theta_0 = (\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_p^{(0)})'$, et $i = 0$

1. Générer $\theta^{(i+1)} = (\theta_1^{(i+1)}, \theta_2^{(i+1)}, \dots, \theta_p^{(i+1)}, Y^p)$, comme suit

$$\text{Générer } \theta_1^{(i+1)} \sim \pi(\theta_1 | \theta_2^{(i)}, \dots, \theta_p^{(i)}, Y^p);$$

$$\text{Générer } \theta_2^{(i+1)} \sim \pi(\theta_2 | \theta_1^{(i+1)}, \theta_3^{(i)}, \dots, \theta_p^{(i)}, Y^p);$$

⋮

$$\text{Générer } \theta_p^{(i+1)} \sim \pi(\theta_p | \theta_1^{(i+1)}, \theta_2^{(i+1)}, \dots, \theta_{p-1}^{(i+1)}, Y^p).$$

2. $i = i + 1$ et répéter le pas 1

GELFAND et SMITH [13] ont montré que sous certaines conditions de régularité la distribution stationnaire du vecteur $\{\theta^{(i)}, i = 1, 2, \dots\}$ est $\pi(\theta | Y^p)$. Cet algorithme est un outil qui nous permet d'échantillonner de la loi $\pi(\theta | Y^p)$ sans avoir besoin de connaître la constante de normalisation.

2.6.2. Algorithme de Metropolis-Hastings

À partir d'une densité cible $\pi(\cdot)$ donnée, l'algorithme de Metropolis-Hastings, (HASTINGS [16]) ne nécessite qu'une valeur de départ $\theta^{(0)}$ pour la chaîne et le choix d'une distribution instrumentale conditionnelle $q(\theta, \check{\theta}) = q(\check{\theta} | \theta)$ à partir de laquelle on va proposer, basé sur l'état actuel de la chaîne, $\{\theta^{(i)}\}$, un candidat pour l'état suivante. La version générale de l'algorithme MH pour échantillonner de la loi *a posteriori* $\pi(\theta | Y^n)$ est décrit comme suit

1. À partir de la valeur $\theta^{(i)}$, générer une valeur candidate $\check{\theta}^{(i+1)} \sim q(\theta^{(i)}, \cdot)$
2. Calculer la probabilité d'acceptation,

$$\alpha(\theta^{(i)}, \check{\theta}^{(i+1)}) = \min \left(\frac{\pi(\check{\theta}^{(i+1)})q(\check{\theta}^{(i+1)}, \theta^{(i)})}{\pi(\theta^{(i)})q(\theta^{(i)}, \check{\theta}^{(i+1)})}, 1 \right)$$

3. Choisir

$$\theta^{(i+1)} = \begin{cases} \check{\theta}^{(i+1)} & \text{avec probabilité } \alpha(\theta^{(i)}, \check{\theta}^{(i+1)}); \\ \theta^{(i)} & \text{avec probabilité } 1 - \alpha(\theta^{(i)}, \check{\theta}^{(i+1)}) \end{cases}$$

3. $i = i + 1$ et répéter 1.

En recommençant ces étapes pour $i = 0, \dots, p$, on construit un échantillon à partir duquel il sera possible d'estimer (2.6.1) avec (2.6.2). Il faut remarquer le fait qu'on a besoin d'un grand nombre d'itérations pour que le processus converge vers la distribution stationnaire cible et il faut aussi vérifier que la chaîne est irréductible.

Chapitre 3

PRÉSENTATION DU MODÈLE

3.1. INTRODUCTION

L'objectif du modèle présenté dans ce chapitre est d'étudier la relation entre le temps de survie d'un certain individu, y_i et l'ensemble de variables indépendantes associées à celui-ci, \mathbf{x}_i . Un autre objectif primaire est de prédire la réponse y_{n+1} associée à une nouvelle observation \mathbf{x}_{n+1} . Dans les paragraphes suivants, nous introduisons le modèle que nous avons développé à partir de celui de MURUA et QUINTANA [28]. On considère que nos données sont partitionnées *a priori* à l'aide du modèle de partition aléatoire associé à Potts, qui a été introduit dans la section 2.5. Une fois la partition établie, on applique un modèle bayésien semi-paramétrique au sein de chaque grappe générée. Les détails de l'inférence associés à un tel modèle sont décrits dans la section 3.4. Dans les sections suivantes, on présente l'approche que nous avons adoptée, ainsi que les principaux résultats théoriques que nous avons développés avec la grande contribution de la direction de ce projet.

3.2. RÉGRESSION DE COX AVEC PARTITIONS LATENTES ISSUES DU MODÈLE DE POTTS

On emploie ici la même notation que celle introduite dans la section 2.4 du chapitre précédent. On dénote par U l'ensemble des n unités étudiées et $\rho_n = \{U_1, \dots, U_K\}$ dénote une partition de cet ensemble. Soit K le nombre de composantes dans la partition. On va considérer un modèle hiérarchique avec une structure similaire à celle introduite en (2.3.2). Soit $\vartheta = (\vartheta_1, \dots, \vartheta_K)$, où chacune des entrées $\vartheta_k = (\alpha, \lambda, \beta)$, $k = 1, \dots, K$, représente le vecteur de paramètres associés au modèle de prévision appliqué au sein de la k -ième composante, qui sera décrit dans les prochaines sections. On introduit également la variable indicatrice u_i qui prend la valeur de k si l'observation \mathbf{x}_i appartient au k -ième sous-ensemble, c'est-à-dire $u_i = k$ si $\mathbf{x}_i \in U_k$. Le modèle en question est représenté par les lois de probabilité

suivantes

$$\begin{aligned} y_i \mid \mathbf{x}_i, \vartheta_{u_i}, \rho_n &\stackrel{ind.}{\sim} \pi(y_i \mid \mathbf{x}_i, \vartheta_{u_i}) & \vartheta_k \mid \rho_n &\stackrel{ind.}{\sim} \pi(\vartheta \mid \rho_n) \\ \rho_n &\sim \pi(\rho_n \mid X^n), \end{aligned} \tag{3.2.1}$$

où la loi $\pi(y_i \mid \mathbf{x}_i, \vartheta_{u_i})$ dans la première ligne correspond à un modèle de régression de Cox avec fonction de risque par intervalles de loi Weibull, qui est détaillé dans la section suivante (comme nous l'avons mentionné avant, on suppose des paramètres différents sur chaque sous-intervalle de l'axe du temps). On détaille également dans les sections qui suivent l'obtention de la loi $\pi(\vartheta \mid \rho_n)$. Finalement,

$$\pi(\rho_n \mid X^n) = \sum_{\mathbf{b} \Rightarrow \rho_n} \pi(\mathbf{b} \mid X^n), \tag{3.2.2}$$

fait référence au modèle de partition aléatoire associé au modèle de Potts, (voir équation (2.5.3) dans la section 2.5 du chapitre 2). On remarque encore que la notation « $\mid X^n$ » dans cette dernière expression fait référence à une indexation de la distribution résultante due à l'effet des covariables. Ainsi, on s'attend à ce que deux individus avec vecteurs de covariables similaires soient plus susceptibles de se trouver dans le même sous-ensemble *a priori*. La vraisemblance complète associée à (3.2.1) sera donc,

$$\pi(Y^n \mid X^n, \vartheta, \rho_n) = \prod_{k=1}^K \prod_{i \in U_k} \pi(y_i \mid \mathbf{x}_i, \vartheta_k). \tag{3.2.3}$$

Ne perdons pas toutefois de vue que l'un des objectifs finals est de trouver la loi $\pi(Y^{n+1} \mid \mathbf{x}_{n+1}, X^n, \boldsymbol{\nu}^n)$ (où $\boldsymbol{\nu}^n$ est la variable indicatrice de censure) et un système de pronostique pour celle-ci.

3.3. SIMULATION *a posteriori* DANS LE CAS GÉNÉRAL

Notre approche générale de modélisation s'exprime comme (3.2.1) conjointement avec (3.2.2). La distribution *a posteriori* de ϑ est un élément crucial pour notre objectif d'inférence prédictive. Afin de déterminer les paramètres optimaux ainsi que l'espérance *a posteriori* de notre modèle nous avons choisi d'effectuer un échantillonnage des réalisations de la loi *a posteriori* à l'aide des méthodes MCMC.

Pour faciliter les développements à venir on écrira $\vartheta = (\vartheta_1, \dots, \vartheta_K)$ comme $(\rho_n, \vartheta_{\rho_n})$, où ϑ_{ρ_n} correspond à l'ensemble de paramètres optimaux pour la partition ρ_n , c'est-à-dire $\vartheta_{\rho_n} = (\hat{\vartheta}_1, \dots, \hat{\vartheta}_K)$ et K , (le nombre de composantes dans la partition) est variable d'une partition à l'autre. Pour obtenir des échantillons *a posteriori* de ces partitions nous utilisons l'algorithme proposé par SWENDSEN et WANG [39] décrit dans la section 2.5.1 dans le cadre de notre approche MCMC pour échantillonner de la distribution cible $\pi(\vartheta \mid Y^n, X^n)$.

Une autre difficulté importante à considérer est que le fait d'incorporer le modèle de Potts dans notre vraisemblance a comme conséquence l'utilisation d'une partition aléatoire. Il faudrait donc dans une étape MCMC donnée, changer la partition, disons de ρ_n à ρ'_n et ceci implique de changer la dimension du vecteur des paramètres ϑ de K à K' . Il est donc nécessaire de recourir à une méthode MCMC multidimensionnelle (MURUA et QUINTANA [28]). Nous avons choisi de suivre la méthodologie proposée par BESAG [2]. L'approche de Besag évite la spécification d'une bijection entre les différentes dimensions. Pour ce faire, on insère le problème dans un espace à très haute dimension. L'idée est la suivante, soit $\Omega = \{\rho(1), \rho(2), \dots, \rho(B_n)\}$ la collection de toutes les partitions possibles de (X^n, Y^n) , dans le modèle de Potts associé à un graphe, ceci représente toutes les partitions possibles avec n sommets dans le graphe correspondant. La notation B_n fait référence au nombre de Bell. Appelons encore $\vartheta = (\vartheta_{\rho(1)}, \vartheta_{\rho(2)}, \dots, \vartheta_{\rho(B_n)})$ la collection des paramètres associés aux éléments de Ω , et L_ρ l'espace de paramètres associé à la partition $\rho \in \Omega$. Aussi, pour chaque $\rho \in \Omega$, on denotera $\pi_0(\vartheta_\rho | \rho, Y^n, X^n)$ une distribution quelconque de support L_ρ . La distribution cible est donc exprimé comme,

$$\tilde{\pi}(\rho, \vartheta | Y^n, X^n) = \pi(\rho, \vartheta_\rho | Y^n, X^n) \prod_{R \neq \rho} \pi_0(\vartheta_R | R, Y^n, X^n)$$

La loi marginale de (ρ, ϑ_ρ) sous cette densité est exactement $\pi(\rho, \vartheta_\rho | Y^n, X^n)$. Le problème est donc réduit à échantillonner de la loi $\tilde{\pi}(\rho, \vartheta | Y^n, X^n)$. BESAG [2] a considéré deux différents types de procédures pour l'échantillonnage :

1. À partir de l'état présent (ρ, ϑ_ρ) , proposer un vecteur de paramètres mis à jour ϑ'_ρ et laisser donc les ϑ_R inchangés, pour $R \neq \rho$. Cela conduit à un rapport Metropolis-Hastings égal à

$$\frac{\tilde{\pi}(\rho, \vartheta' | Y^n, X^n) q(\vartheta'_\rho, \vartheta_\rho)}{\tilde{\pi}(\rho, \vartheta | Y^n, X^n) q(\vartheta_\rho, \vartheta'_\rho)} = \frac{\pi(\rho, \vartheta'_\rho | Y^n, X^n) q(\vartheta'_\rho, \vartheta_\rho)}{\pi(\rho, \vartheta_\rho | Y^n, X^n) q(\vartheta_\rho, \vartheta'_\rho)} = \frac{\pi(\vartheta'_\rho | \rho, Y^n, X^n) q(\vartheta'_\rho, \vartheta_\rho)}{\pi(\vartheta_\rho | \rho, Y^n, X^n) q(\vartheta_\rho, \vartheta'_\rho)}.$$

On choisira $q(\vartheta_\rho, \vartheta'_\rho) = \pi(\vartheta'_\rho | \rho, Y^n, X^n)$, de tel sorte que le rapport devient 1, et la mise à jour est toujours acceptée.

2. La deuxième procédure consiste à proposer une certaine nouvelle partition $\rho' \neq \rho$ et forcément donc, un autre vecteur de paramètres $\vartheta'_{\rho'} \neq \vartheta_\rho$. Afin de simplifier le seuil d'acceptation, on génère la mise à jour avec une proposition $\pi_0(\vartheta'_{\rho'} | \rho, Y^n, X^n) \pi(\rho' | X^n) q_{\rho\rho'}(\vartheta_\rho, \vartheta'_{\rho'})$. Ceci mène à un ratio d'acceptation de Metropolis-Hastings qui est exprimé comme,

$$\frac{\pi(\rho', \vartheta'_{\rho'} | Y^n, X^n) \pi_0(\vartheta'_{\rho'} | \rho, Y^n, X^n) \pi_0(\vartheta_\rho | \rho', Y^n, X^n) \pi(\rho | X^n) q_{\rho\rho'}(\vartheta'_{\rho'}, \vartheta_\rho)}{\pi(\rho, \vartheta_\rho | Y^n, X^n) \pi_0(\vartheta_{\rho'} | \rho', Y^n, X^n) \pi_0(\vartheta'_{\rho'} | \rho, Y^n, X^n) \pi(\rho' | X^n) q_{\rho\rho'}(\vartheta_\rho, \vartheta'_{\rho'})}$$

$$\begin{aligned}
&= \frac{\pi(\rho', \boldsymbol{\vartheta}'_{\rho'} \mid Y^n, X^n) \pi(\rho \mid X^n) q_{\rho'\rho}(\boldsymbol{\vartheta}'_{\rho'}, \boldsymbol{\vartheta}_\rho)}{\pi(\rho, \boldsymbol{\vartheta}_\rho \mid Y^n, X^n) \pi(\rho' \mid X^n) q_{\rho\rho'}(\boldsymbol{\vartheta}_\rho, \boldsymbol{\vartheta}'_{\rho'})} \\
&= \frac{\pi(\boldsymbol{\vartheta}'_{\rho'} \mid \rho', X^n, Y^n) q_{\rho'\rho}(\boldsymbol{\vartheta}'_{\rho'}, \boldsymbol{\vartheta}_\rho)}{\pi(\boldsymbol{\vartheta}_\rho \mid \rho, X^n, Y^n) q_{\rho\rho'}(\boldsymbol{\vartheta}_\rho, \boldsymbol{\vartheta}'_{\rho'})}.
\end{aligned}$$

Dans cette dernière expression, si nous fixons $q_{\rho\rho'}(\boldsymbol{\vartheta}_\rho, \boldsymbol{\vartheta}'_{\rho'}) = \pi(\boldsymbol{\vartheta}'_{\rho'} \mid \rho', Y^n, X^n)$, le ratio d'acceptation devient,

$$\frac{\pi(Y^n \mid \rho', X^n)}{\pi(Y^n \mid \rho, X^n)}.$$

Notons que l'astuce pour simplifier le seuil d'acceptation en (1.) consiste à générer un nouveau vecteur de paramètres $\boldsymbol{\vartheta}'_\rho$ pour l'ancienne partition ρ , tandis que en (2.) on considère une nouvelle partition ρ' et on génère aussi de paramètres $\boldsymbol{\vartheta}'_\rho$ et $\boldsymbol{\vartheta}'_{\rho'}$ pour l'ancienne partition et la nouvelle partition. Remarquons cependant qu'il n'est pas nécessaire de générer réellement $\boldsymbol{\vartheta}'_\rho$, ni aucun des autres $\boldsymbol{\vartheta}'_R$, $R \neq \rho'$. La proposition est réduite à la génération d'une nouvelle partition ρ' de $\pi(\rho' \mid X^n)$. Si la proposition est acceptée, alors on génère (et automatiquement accepte) une mise à jour de $\boldsymbol{\vartheta}'_{\rho'}$ de $\pi(\boldsymbol{\vartheta}'_{\rho'} \mid \rho', Y^n, X^n)$. Si la partition n'est pas acceptée, il n'est pas nécessaire de générer $\boldsymbol{\vartheta}'_{\rho'}$. La mise en oeuvre pratique du schéma de simulation *a posteriori* décrit ci-dessus nécessite donc de calculer $\pi(\boldsymbol{\vartheta}_\rho \mid \rho, Y^n, X^n)$.

3.4. INFÉRENCE AU SEIN DE CHAQUE SOUS-ENSEMBLE

3.4.1. Vraisemblance du modèle

Les résultats décrits dans cette section sont appliqués de façon indépendante sur chacun des sous-ensembles de la partition induite à l'aide du modèle de partition aléatoire associé au modèle de Potts. Dans les développements à venir, on reprend la notation introduite lors du premier chapitre. Soit $Y_i = \min\{C_i, T_i\}$, le minimum entre le temps de censure, C_i et le temps exact de survie, T_i , pour $i = 1, 2, \dots, n$ où n est le nombre total d'observations, et soit $\mathbf{x}_i \in \mathbb{R}^p$ le vecteur des covariables associées à la variable réponse y_i pour le i -ième sujet. On dénote par $\boldsymbol{\beta} \in \mathbb{R}^p$ le vecteur de coefficients de régression et par $\nu_i = \mathbb{1}_{\{T_i \leq C_i\}}$ la variable indicatrice de censure pour le i -ième sujet. Tel que décrit dans la section 1.6, nous divisons d'abord l'axe du temps en J sous-intervalles, $(s_0, s_1], (s_1, s_2], \dots, (s_{J-1}, s_J]$, où $0 = s_0 < s_1 < s_2 < \dots < s_J$. Les équations (1.2.6) et (1.5.1) impliquent que la fonction de survie associée à ce modèle est

$$S(y_i \mid \mathbf{x}_i, \boldsymbol{\vartheta}_{u_i}) = \exp\left\{-H_0(y_i \mid \alpha, \lambda) \exp(\mathbf{x}'_i \boldsymbol{\beta})\right\}. \quad (3.4.1)$$

Rappelons que $u_i = k$ si $y_i \in U_k$. Le vecteur $\boldsymbol{\vartheta}_k = (\alpha, \lambda, \boldsymbol{\beta})$, de taille $2J + p$ contient l'information complète sur les paramètres associés à la régression de Cox. Ici, $\alpha = (\alpha_1, \dots, \alpha_J)$ et $\lambda = (\lambda_1, \dots, \lambda_J)$, $\alpha_j > 0, \lambda_j > 0, j = 1, \dots, J$, dénotent les vecteurs de paramètres associés

à la fonction de risque de base, qui suit une loi de Weibull, c'est-à dire,

$$h_0(y_i | \alpha, \lambda) = \begin{cases} \alpha_j \lambda_j y_i^{\alpha_j - 1} & \text{si } y_i \in (s_{j-1}, s_j], \\ 0 & \text{si } y_i < s_0 \text{ ou } y_i > s_J. \end{cases}$$

La fonction de risque cumulé de base, $H_0(y_i | \alpha, \lambda)$, est donc donnée par

$$H_0(y_i | \alpha, \lambda) = \int_0^{y_i} h_0(t | \alpha, \lambda) dt = \lambda_j (y_i^{\alpha_j} - s_{j-1}^{\alpha_j}) + \sum_{g=1}^{j-1} \lambda_g (s_g^{\alpha_g} - s_{g-1}^{\alpha_g}), \text{ pour } s_{j-1} < y_i \leq s_j.$$

Le code avec la fonction numérique pour évaluer $H(y | \alpha, \lambda)$ peut être consulté dans la section E.4 de l'annexe E. En utilisant (1.5.3), on obtient la fonction de vraisemblance suivante

$$L(\vartheta_k | D_k) = \prod_{i=1}^n \prod_{j=1}^J \left\{ \left(\alpha_j \lambda_j y_i^{\alpha_j - 1} \exp(\mathbf{x}'_i \boldsymbol{\beta}) \right)^{\delta_{ij} \nu_i} \times \exp \left[-\delta_{ij} \exp(\mathbf{x}'_i \boldsymbol{\beta}) \left(\lambda_j (y_i^{\alpha_j} - s_{j-1}^{\alpha_j}) + \sum_{g=1}^{j-1} \lambda_g (s_g^{\alpha_g} - s_{g-1}^{\alpha_g}) \right) \right] \right\}. \quad (3.4.2)$$

Ici, on utilise la notation $D_k = (Y^n, X^n, \boldsymbol{\nu}^n)$ pour dénoter l'ensemble complet de données qui composent le k -ième sous-ensemble, où $\boldsymbol{\nu}^n \in \mathbb{R}^n$ est le vecteur de variables indicatrices de censure, $X^n = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$, $Y^n = \{y_1, \dots, y_n\}$. Rappelons aussi que δ_{ij} est une variable indicatrice d'intervalle,

$$\delta_{ij} = \begin{cases} 1 & \text{si } y_i \in I_j, \\ 0 & \text{sinon.} \end{cases}$$

La log-vraisemblance pour (3.4.2), est donc donnée par

$$\ln(L(\vartheta_k | D_k)) = \sum_{j=1}^J \sum_{i=1}^n \left\{ \delta_{ij} \nu_i \left[\ln(\alpha_j) + \ln(\lambda_j) + (\alpha_j - 1) \ln(y_i) + \mathbf{x}'_i \boldsymbol{\beta} \right] - \delta_{ij} \exp(\mathbf{x}'_i \boldsymbol{\beta}) \left[\lambda_j (y_i^{\alpha_j} - s_{j-1}^{\alpha_j}) + \sum_{g=1}^{j-1} \lambda_g (s_g^{\alpha_g} - s_{g-1}^{\alpha_g}) \right] \right\}. \quad (3.4.3)$$

À partir de cette expression, nous allons mener le développement de l'inférence de notre modèle.

3.5. DENSITÉ *a priori* ET *a posteriori* DES PARAMÈTRES

Nous spécifions tout d'abord une distribution *a priori* pour le vecteur de paramètres $\vartheta_k = (\alpha, \lambda, \boldsymbol{\beta})$. Afin d'alléger la notation de cette section, nous omettrons l'indice k désignant le sous-ensemble sur lequel l'inférence actuelle s'applique. Conditionnellement à X^n , nous proposons une loi *a priori* semi-conjuguée comme suit,

$$\pi(\vartheta | X^n, Y_0^n, a_0) \propto \pi(\boldsymbol{\beta} | \alpha, \lambda, X^n, Y_0^n, a_0) \pi(\alpha, \lambda | \theta) \pi(\theta)$$

$$\propto \left\{ \prod_{i=1}^n \prod_{j=1}^J (\alpha_j \lambda_j y_{0i}^{\alpha_j - 1} \exp(\mathbf{x}'_i \boldsymbol{\beta}))^{a_0 \delta_{ij}^0} \exp \left[-a_0 \delta_{ij}^0 \exp(\mathbf{x}'_i \boldsymbol{\beta}) \right. \right. \\ \left. \left. \left(\lambda_j (y_{0i}^{\alpha_j} - s_{j-1}^{\alpha_j}) + \sum_{g=1}^{j-1} \lambda_g (s_g^{\alpha_g} - s_{g-1}^{\alpha_g}) \right) \right] \right\} \pi(\alpha, \lambda | \theta) \pi(\theta). \quad (3.5.1)$$

Ceci est une loi augmentée, car la distribution $\pi(\theta)$ est incorporée dans l'expression afin de modéliser l'*a priori* du vecteur (α, λ) . Les variables a_0 et Y_0^n sont des paramètres auxiliaires *a priori*, plus précisément $Y_0^n = (y_{01}, y_{02}, \dots, y_{0n})$, où chaque entrée y_{0i} peut être vu comme une prévision de la moyenne marginale de y_i . De son côté, $a_0 > 0$ est un scalaire qui peut être vu comme un paramètre de précision qui quantifie la qualité du choix antérieur de Y_0^n . Ainsi, a_0 ne contrôle que la précision de la loi *a priori* de $\boldsymbol{\beta}$.

La variable $\delta_{ij}^0 = 1$ si $s_{j-1} < y_{0i} \leq s_j$ et 0 sinon. La distribution $\pi(\alpha, \lambda | \theta)$ est une distribution initiale pour (α, λ) . On supposera également que chaque composante du vecteur $\theta = (\theta_1, \theta_2) \in \mathbb{R}^2$ suit une loi gamma avec des paramètres fixes respectifs p_j et q_j . La densité de θ_j est donc donnée par,

$$\pi(\theta_j) = \frac{(q_j \theta_j)^{p_j - 1}}{\Gamma(p_j)} q_j e^{-q_j \theta_j} \quad \text{où } p_j > 0, q_j > 0 \text{ et } \theta_j > 0, \quad j = 1, 2.$$

La densité *a priori* en (3.5.1) est appelée semi-conjuguée car, en ignorant les termes $\pi(\alpha, \lambda | \theta) \pi(\theta)$, cette fonction a une forme identique à celle donnée en (3.4.2). Pour plus de détails sur l'utilisation de cette type de lois, voir FINK [11].

Pour ce modèle de Weibull par intervalles, nous considérons que toutes les y_{0i} prennent la même valeur, c'est-à-dire $y_0 = y_{01} = \dots = y_{0n}$, avec $s_0 < y_0 \leq s_1$. Sur cette spécification de Y_0^n , la densité semi-conjuguée en (3.5.1) est réduite à

$$\pi(\vartheta, \theta | X^n, Y_0^n, a_0) = c_n(\vartheta, X^n, Y_0^n, a_0) \psi(\vartheta | X^n, Y_0^n, a_0) \pi(\alpha, \lambda | \theta) \pi(\theta), \quad (3.5.2)$$

où $c_n(\vartheta, \theta, X^n, Y_0^n, a_0)$ est l'inverse de la constante de normalisation de

$$\eta(\vartheta, \theta | X^n, Y_0^n, a_0) = \psi(\vartheta | X^n, Y_0^n, a_0) \pi(\alpha, \lambda | \theta) \pi(\theta), \quad (3.5.3)$$

avec

$$\psi(\vartheta | X^n, Y_0^n, a_0) = \left[\prod_{i=1}^n (\alpha_1 \lambda_1 y_0^{\alpha_1 - 1} \exp(\mathbf{x}'_i \boldsymbol{\beta}))^{a_0} \exp(-a_0 \exp(\mathbf{x}'_i \boldsymbol{\beta}) \lambda_1 y_0^{\alpha_1}) \right]. \quad (3.5.4)$$

La loi *a priori* $\psi(\vartheta | X^n, Y_0^n, a_0)$ est une loi augmentée, car on s'en sert des paramètres auxiliaires a_0 et Y_0^n .

3.5.1. Densité *a priori* de α et λ

Dans le but d'établir des *a priori*s associés aux paramètres, on propose de prendre la valeur absolue de la différence entre λ_{j+1} et λ_j ainsi que la valeur absolue de la différence

entre α_{j+1} et α_j de loi double exponentielle, $|\lambda_{j+1} - \lambda_j| \sim \text{DE}(\theta_1)$ et $|\alpha_{j+1} - \alpha_j| \sim \text{DE}(\theta_2)$. Cette loi est associée à la pénalisation lasso de fusion, la motivation étant que les paramètres provenant d'intervalles consécutifs soient égales si leur différence est petite. TIBSHIRANI, SAUNDERS, ROSSET, ZHU et KNIGHT [43]. Les valeurs initiales de λ_1 et α_1 seront de loi exponentielle $\lambda_1 \sim \text{Exp}(\theta_1)$, $\alpha_1 \sim \text{Exp}(\theta_2)$, la densité conjointe des vecteurs α et λ peut être obtenue comme

$$\pi(\lambda \mid \theta_1) = \frac{\theta_1 e^{-\theta_1 \lambda_1} e^{-\sum_{j=1}^{J-1} |\lambda_{j+1} - \lambda_j| \theta_1}}{\int_0^\infty \dots \int_0^\infty \theta_1 e^{-\theta_1 \lambda_1} e^{-\sum_{j=1}^{J-1} |\lambda_{j+1} - \lambda_j| \theta_1} d\lambda_1 \dots d\lambda_J},$$

et de façon similaire,

$$\pi(\alpha \mid \theta_2) = \frac{\theta_2 e^{-\theta_2 \alpha_1} e^{-\sum_{j=1}^{J-1} |\alpha_{j+1} - \alpha_j| \theta_2}}{\int_0^\infty \dots \int_0^\infty \theta_2 e^{-\theta_2 \alpha_1} e^{-\sum_{j=1}^{J-1} |\alpha_{j+1} - \alpha_j| \theta_2} d\alpha_1 \dots d\alpha_J}.$$

Le calcul des densités *a priori* marginales pour différentes valeurs de J (différents nombres d'intervalles sur l'axe du temps), peut être consulté dans l'annexe A. Ainsi, en supposant l'indépendance entre α et λ , et à l'aide des équations (A.0.1) et (A.0.2), la fonction de distribution conjointe de α et λ est donné par l'expression suivante

$$\pi(\alpha, \lambda \mid \theta) = K_J^2 \theta_1^J \theta_2^J \exp \left\{ -\theta_1 \lambda_1 - \theta_2 \alpha_1 - \sum_{j=1}^{J-1} |\lambda_{j+1} - \lambda_j| \theta_1 - \sum_{j=1}^{J-1} |\alpha_{j+1} - \alpha_j| \theta_2 \right\} \quad (3.5.5)$$

où K_J est une constante qui dépend du nombre d'intervalles considérés (voir annexe A).

3.5.2. Méthode d'intégration de Laplace

Une fois la fonction de densité *a priori* établie, on cherche à déterminer la densité *a posteriori*. À cette fin, la méthode de LAPLACE [21] pour estimer des intégrales sans solution analytique a été appliquée. Nous sommes intéressés à la densité *a posteriori* des paramètres $\vartheta = (\alpha, \lambda, \beta)$. Selon la règle de Bayes décrite en (2.2.1),

$$\pi(\vartheta, \theta \mid D_k, Y_0^n, a_0) = \frac{L(\vartheta \mid D_k) \eta(\vartheta, \theta \mid X^n, Y_0^n, a_0)}{\int L(\vartheta \mid D_k) \eta(\vartheta, \theta \mid X^n, Y_0^n, a_0) d\vartheta d\theta}, \quad (3.5.6)$$

où la fonction η est celle présentée dans l'équation (3.5.3). Le dénominateur en (3.5.6) peut être exprimé de la façon suivante,

$$\begin{aligned} & \int L(\vartheta \mid D_k) \eta(\vartheta, \theta \mid X^n, Y_0^n, a_0) d\vartheta d\theta \\ &= \int L(\vartheta \mid D_k) \psi(\vartheta \mid X^n, Y_0^n, a_0) \pi(\alpha, \lambda \mid \theta) \pi(\theta) d\vartheta d\theta \\ &= \int \exp \left\{ \ln(L(\vartheta \mid D_k)) + \ln(\psi(\vartheta \mid X^n, Y_0^n, a_0)) \right\} \pi(\alpha, \lambda \mid \theta) \pi(\theta) d\vartheta d\theta. \end{aligned} \quad (3.5.7)$$

Si l'on écrit

$$\ell(\alpha, \lambda) = \int \pi(\alpha, \lambda | \theta) \pi(\theta) d\theta, \quad (3.5.8)$$

l'intégrale (3.5.7) peut être exprimée comme

$$\int e^{g(\vartheta)} \ell(\vartheta_{(\alpha, \lambda)}) d\vartheta,$$

avec $g(\vartheta) = \ln(L(\vartheta | D_k)) + \ln(\psi(\vartheta | X^n, Y_0^n, a_0))$. On utilisera la notation $\vartheta_{(a,b)}$ pour indiquer le sous-vecteur de variables en ϑ , ainsi $\vartheta_{(\alpha, \lambda)} = (\alpha, \lambda)$. En utilisant la méthode de Laplace, on obtient l'approximation suivante pour le dénominateur en (3.5.6).

$$\int L(\vartheta | D_k) \eta(\vartheta, \theta | X^n, Y_0^n, a_0) d\vartheta d\theta \approx e^{g(\hat{\vartheta})} \ell(\hat{\vartheta}_{(\alpha, \lambda)}) (-2\pi)^{(d/2)} \left| \frac{\partial^2 g(\hat{\vartheta})}{\partial \vartheta \partial \vartheta'} \right|, \quad (3.5.9)$$

où $d = 2J + p$ et $\hat{\vartheta}$ maximise $g(\vartheta)$. Ainsi, la fonction à optimiser est

$$\begin{aligned} g(\vartheta) &= \ln(L(\vartheta | D_k)) + \ln(\psi(\vartheta | X^n, Y_0^n, a_0)) \\ &= \sum_{j=1}^J \sum_{i=1}^n \left\{ \delta_{ij} \nu_i \left[\ln(\alpha_j) + \ln(\lambda_j) + (\alpha_j - 1) \ln(y_i) + \mathbf{x}'_i \boldsymbol{\beta} \right] \right. \\ &\quad \left. - \delta_{ij} \exp(\mathbf{x}'_i \boldsymbol{\beta}) \left[\lambda_j (y_i^{\alpha_j} - s_{j-1}^{\alpha_j}) + \sum_{g=1}^{j-1} \lambda_g (s_g^{\alpha_g} - s_{g-1}^{\alpha_g}) \right] \right\} \\ &\quad + na_0 [\ln(\alpha_1) + \ln(\lambda_1) + (\alpha_1 - 1) \ln(y_0)] \\ &\quad + \sum_{i=1}^n [a_0 \mathbf{x}'_i \boldsymbol{\beta} - a_0 \exp(\mathbf{x}'_i \boldsymbol{\beta}) \lambda_1 y_0^{\alpha_1}]. \end{aligned} \quad (3.5.10)$$

D'un autre côté, on voit que l'intégrale $\ell(\vartheta_{(\alpha, \lambda)})$ en (3.5.8) peut être écrite comme

$$\ell(\vartheta_{(\alpha, \lambda)}) = \int_0^\infty \int_0^\infty \pi(\alpha | \theta_2) \pi(\theta_2) \pi(\lambda | \theta_1) \pi(\theta_1) d\theta_1 d\theta_2.$$

où la solution pour $\int_0^\infty \pi(\lambda | \theta_1) \pi(\theta_1) d\theta_1$ est

$$\begin{aligned} \int_0^\infty \pi(\lambda | \theta_1) \pi(\theta_1) d\theta_1 &= \int_0^\infty K_J \theta_1^J e^{-\theta_1 \lambda_1 - \sum_{j=1}^{J-1} |\lambda_{j+1} - \lambda_j| \theta_1} \left[\frac{(q_1 \theta_1)^{p_1 - 1}}{\Gamma(p_1)} q_1 e^{-q_1 \theta_1} \right] d\theta_1 \\ &= \frac{K_J q_1^{p_1}}{\Gamma(p_1)} \int_0^\infty \theta_1^{J+p_1-1} e^{-\theta_1 [\lambda_1 + q_1 + \sum_{j=1}^{J-1} |\lambda_{j+1} - \lambda_j|]} d\theta_1 \\ &= \gamma_1 \left[\lambda_1 + q_1 + \sum_{j=1}^{J-1} |\lambda_{j+1} - \lambda_j| \right]^{-(J+p_1)} \end{aligned}$$

Ainsi, $\ell(\vartheta_{(\alpha, \lambda)})$ est égale à

$$\ell(\vartheta_{(\alpha, \lambda)}) = \gamma \left[\lambda_1 + q_1 + \sum_{j=1}^{J-1} |\lambda_{j+1} - \lambda_j| \right]^{-(J+p_1)} \left[\alpha_1 + q_2 + \sum_{j=1}^{J-1} |\alpha_{j+1} - \alpha_j| \right]^{-(J+p_2)} \quad (3.5.11)$$

où γ_i dénote la constante $\gamma_i = \frac{K_J q_i^{p_i} \Gamma(J+p_i)}{\Gamma(p_i)}$, $i = 1, 2$ et

$$\gamma = \gamma_1 \gamma_2 = \frac{K_J^2 q_1^{p_1} q_2^{p_2} \Gamma(J+p_1) \Gamma(J+p_2)}{\Gamma(p_1) \Gamma(p_2)}.$$

Le calcul détaillé du gradient et de la matrice hessienne de $g(\vartheta)$ est inclus dans l'annexe B. L'application de la méthode d'approximation de Laplace exige que la fonction $\ell(\vartheta_{(\alpha, \lambda)})$ soit dérivable, on introduit dans la section suivante une approximation lisse de cette fonction.

3.5.3. Approximation lisse de la fonction $\ell(\alpha, \lambda)$

Considérons la fonction

$$\xi(w) = \kappa^{-1} \left(\log(1 + \cosh(\kappa w)) + \log(2) \right), \quad w \in \mathbb{R}, \quad \kappa \in \mathbb{R}^+,$$

dont la dérivée est donnée par $\xi'(w) = \sinh(\kappa w) / (1 + \cosh(\kappa w))$, HERTY, KLAR, SINGH et SPELLUCCI [17] montrent que $|\xi(w) - |w|| \leq 8/3\kappa^{-1} e^{-\kappa|w|}$. Cette condition garantit que l'erreur d'approximation est petit pour des grands valeurs de la constante κ . Ainsi, on peut remplacer la fonction $\ell(\alpha, \lambda)$ par une nouvelle version lisse, $\mathcal{L}(\alpha, \lambda)$, qui est donnée par

$$\mathcal{L}(\alpha, \lambda) = \gamma \left[\lambda_1 + q_1 + \sum_{j=1}^{J-1} \xi(\lambda_{j+1} - \lambda_j) \right]^{-(J+p_1)} \left[\alpha_1 + q_2 + \sum_{j=1}^{J-1} \xi(\alpha_{j+1} - \alpha_j) \right]^{-(J+p_2)}. \quad (3.5.12)$$

Autrement dit, on remplace $|\lambda_{j+1} - \lambda_j|$ par $\xi(\lambda_{j+1} - \lambda_j)$ en 3.5.11. Les détails du calcul du gradient et de la matrice hessienne de $\mathcal{L}(\alpha, \lambda)$ sont inclus dans la section B.2 de l'annexe B.

En considérant la fonction \mathcal{L} dans l'approximation de Laplace en (3.5.9), la fonction $g(\vartheta)$ sera remplacée par $\tilde{g}(\vartheta) = \ln(L(\vartheta | D_k)) + \ln(\psi(\vartheta | X^n, Y_0^n, a_0)) + \ln(\mathcal{L}(\alpha, \lambda))$ et l'approximation de l'intégrale en considérant cette fonction est alors donnée par

$$\int L(\vartheta | D_k) \eta(\vartheta, \theta | X^n, Y_0^n, a_0) d\vartheta d\theta \approx (2\pi)^{(d/2)} e^{\tilde{g}(\vartheta^*)} \left| -\frac{\partial^2 \tilde{g}(\vartheta^*)}{\partial \vartheta \partial \vartheta'} \right| \frac{\ell(\vartheta_{(\alpha, \lambda)}^*)}{\mathcal{L}(\vartheta_{(\alpha, \lambda)}^*)}, \quad (3.5.13)$$

où ϑ^* maximise la fonction $\tilde{g}(\vartheta)$. Les dérivés qui servent à construire le gradient et la matrice hessienne de $\tilde{g}(\vartheta)$ peuvent être consultés dans l'annexe C.

3.6. ÉCHANTILLONNAGE DE LA LOI *a posteriori*

Dans le but d'estimer les paramètres du modèle, nous avons besoin de tirer un échantillon de la loi *a posteriori* $\pi(\vartheta, \theta | D_k, Y_0^n, a_0)$ établie en (3.5.6). À cette fin, on utilise les résultats de l'approximation de Laplace introduite précédemment. À partir des expressions (3.5.6) et (3.5.13) et à l'aide d'une approximation par séries de Taylor de la fonction $\tilde{g}(\vartheta)$ autour de

ϑ^* , on a

$$\begin{aligned}
\pi(\vartheta, \theta \mid D_k, Y_0^n, a_0) &\propto \int L(\vartheta \mid D_k) \eta(\vartheta, \theta \mid X^n, Y_0^n, a_0) d\vartheta d\theta \\
&\propto \exp\{\tilde{g}(\vartheta^*)\} \frac{\ell(\vartheta_{(\alpha, \lambda)}^*)}{\mathcal{L}(\vartheta_{(\alpha, \lambda)}^*)} \\
&\approx \exp\left\{\tilde{g}(\vartheta^*) + \nabla_{\tilde{g}(\vartheta^*)}(\vartheta - \vartheta^*) - \frac{1}{2}(\vartheta - \vartheta^*)' H(\vartheta - \vartheta^*)\right\} \frac{\ell(\vartheta_{(\alpha, \lambda)}^*)}{\mathcal{L}(\vartheta_{(\alpha, \lambda)}^*)} \\
&\approx \exp\left\{\tilde{g}(\vartheta^*) - \frac{1}{2}(\vartheta - \vartheta^*)' H(\vartheta - \vartheta^*)\right\} \frac{\ell(\vartheta_{(\alpha, \lambda)}^*)}{\mathcal{L}(\vartheta_{(\alpha, \lambda)}^*)},
\end{aligned} \tag{3.6.1}$$

où H est la matrice qui dénote le négatif de la matrice hessienne de $\tilde{g}(\vartheta)$ évaluée en ϑ^* . Nous allons indiquer par $H_{(a,b)}$ la sous-matrice de H composée des lignes et colonnes indicées par a et b respectivement. Par exemple, $H_{\beta,(\alpha, \lambda)}$ indique la sous-matrice composée des lignes associées à β et des colonnes associées à (α, λ) .

$$H = - \begin{pmatrix} \frac{\partial^2 \tilde{g}(\vartheta^*)}{\partial \alpha^2} & \frac{\partial^2 \tilde{g}(\vartheta^*)}{\partial \alpha \partial \lambda} & \frac{\partial^2 \tilde{g}(\vartheta^*)}{\partial \alpha \partial \beta} \\ \frac{\partial^2 \tilde{g}(\vartheta^*)}{\partial \lambda \partial \alpha} & \frac{\partial^2 \tilde{g}(\vartheta^*)}{\partial \lambda^2} & \frac{\partial^2 \tilde{g}(\vartheta^*)}{\partial \lambda \partial \beta} \\ \frac{\partial^2 \tilde{g}(\vartheta^*)}{\partial \beta \partial \alpha} & \frac{\partial^2 \tilde{g}(\vartheta^*)}{\partial \beta \partial \lambda} & \frac{\partial^2 \tilde{g}(\vartheta^*)}{\partial \beta^2} \end{pmatrix}$$

$$H_{(\alpha, \lambda), (\alpha, \lambda)} = - \begin{pmatrix} \frac{\partial^2 \tilde{g}(\vartheta^*)}{\partial \alpha^2} & \frac{\partial^2 \tilde{g}(\vartheta^*)}{\partial \alpha \partial \lambda} \\ \frac{\partial^2 \tilde{g}(\vartheta^*)}{\partial \lambda \partial \alpha} & \frac{\partial^2 \tilde{g}(\vartheta^*)}{\partial \lambda^2} \end{pmatrix} \quad H_{\beta, (\alpha, \lambda)} = - \begin{pmatrix} \frac{\partial^2 \tilde{g}(\vartheta^*)}{\partial \beta \partial \alpha} & \frac{\partial^2 \tilde{g}(\vartheta^*)}{\partial \beta \partial \lambda} \end{pmatrix}$$

Ainsi, l'approximation établie en 3.6.1 nous permet de générer des observations provenant de la loi *a posteriori* associé aux paramètres $\vartheta = (\alpha, \lambda, \beta)$.

3.6.1. Échantillonnage de la loi de β

Soit $S = H^{-1}$. De l'expression (3.6.1) ci-dessus, on remarque que le vecteur $(\beta \mid \alpha, \lambda)$ suit approximativement une loi normale. Des résultats de la loi normale multivariée conditionnelle, on déduit que la moyenne d'une telle distribution est donnée par,

$$\beta^* + S_{\beta, (\alpha, \lambda)} S_{(\alpha, \lambda), (\alpha, \lambda)}^{-1} \begin{pmatrix} \alpha - \alpha^* \\ \lambda - \lambda^* \end{pmatrix}$$

alors que la matrice de précision est donnée par

$$S_{\beta, \beta} - S_{\beta, (\alpha, \lambda)} S_{(\alpha, \lambda), (\alpha, \lambda)}^{-1} S_{(\alpha, \lambda), \beta}$$

Par conséquent, étant données des valeurs de (α, λ) , on peut proposer une stratégie d'échantillonnage à l'aide des méthodes MCMC et échantillonner une nouvelle valeur de β à l'aide de cette distribution normale. Tout comme précédemment, la notation $*$ est utilisée pour indiquer les paramètres optimaux.

3.6.2. Échantillonnage du vecteur (α, λ)

La loi de probabilité approximative du vecteur (α, λ) est plus compliquée que celle de β . Notons qu'une approximation de la distribution de ces variables peut être obtenue à partir de (3.6.1) en intégrant par rapport à β . Ceci mène à

$$\begin{aligned} \pi(\alpha, \lambda \mid Y_0^n, X^n, a_0, D_k) &\propto \int_{\beta} \exp\{\tilde{g}(\vartheta^*)\} \frac{\ell(\vartheta_{(\alpha, \lambda)}^*)}{\mathcal{L}(\vartheta_{(\alpha, \lambda)}^*)} d\beta \\ &\approx \int_{\beta} \exp\left\{\tilde{g}(\vartheta^*) - \frac{1}{2}(\vartheta - \vartheta^*)' H(\vartheta - \vartheta^*)\right\} \frac{\ell(\vartheta_{(\alpha, \lambda)}^*)}{\mathcal{L}(\vartheta_{(\alpha, \lambda)}^*)} d\beta \quad (3.6.2) \\ &\propto \exp\left\{-\frac{1}{2}(\vartheta_{(\alpha, \lambda)} - \vartheta_{(\alpha, \lambda)}^*)' H_{(\alpha, \lambda), (\alpha, \lambda)}(\vartheta_{(\alpha, \lambda)} - \vartheta_{(\alpha, \lambda)}^*)\right\} \frac{\ell(\vartheta_{(\alpha, \lambda)}^*)}{\mathcal{L}(\vartheta_{(\alpha, \lambda)}^*)}, \end{aligned}$$

car sous cette approximation le terme dans l'intégrale (3.6.2) suit approximativement une loi normale multivariée (ceci est comme si on marginalisait une vecteur normal multivariée).

La complexité de cette densité est due au terme $\ell(\vartheta_{(\alpha, \lambda)}^*)$. Une stratégie pour l'estimer consiste à tirer un échantillon aléatoire $(\vartheta_{(\alpha, \lambda)}^{(1)}, \dots, \vartheta_{(\alpha, \lambda)}^{(M)})$ provenant d'une loi normale de moyenne $\vartheta_{(\alpha, \lambda)}^*$ et de matrice de covariances $\frac{1}{2}S_{(\alpha, \lambda), (\alpha, \lambda)}$. Remarquons également que la quantité

$$\hat{\pi}(\alpha, \lambda) = \frac{1}{M} \sum_{m=1}^M \ell(\alpha, \lambda) \phi\left(\alpha, \lambda \mid \vartheta_{(\alpha, \lambda)}^{(m)}, \frac{1}{2}S_{(\alpha, \lambda), (\alpha, \lambda)}\right)$$

converge vers la valeur attendue sur $\tilde{\vartheta}_{(\alpha, \lambda)}$ de $\ell(\alpha, \lambda) \phi\left(\alpha, \lambda \mid \tilde{\vartheta}_{(\alpha, \lambda)}, \frac{1}{2}S_{(\alpha, \lambda), (\alpha, \lambda)}\right)$ sous la distribution normale supposée, en raison de la propriété reproductrice du noyau de la loi normale. Ici, on utilise la notation ϕ pour indiquer la densité d'une loi normale multivariée de paramètres spécifiés.

Ainsi, l'estimateur $\hat{\pi}(\alpha, \lambda)$ converge vers $\ell(\alpha, \lambda) \phi\left(\alpha, \lambda \mid \hat{\vartheta}_{(\alpha, \lambda)}, \frac{1}{2}S_{(\alpha, \lambda), (\alpha, \lambda)}\right)$, de sorte que nous pourrions prélever des échantillons de $\hat{\pi}(\alpha, \lambda)$. Cependant, cette expression est encore complexe en raison de l'évaluation de la fonction ℓ . L'astuce consiste à exploiter le fait qu'il n'y a presque pas de poids loin du $\vartheta_{(\alpha, \lambda)}^{(m)}$, (car c'est la moyenne de la distribution normale supposé) ainsi on peut considérer l'option de remplacer les termes

$$\ell(\alpha, \lambda) \phi\left(\alpha, \lambda \mid \vartheta_{(\alpha, \lambda)}^{(m)}, \frac{1}{2}S_{(\alpha, \lambda), (\alpha, \lambda)}\right)$$

par

$$\ell(\vartheta_{(\alpha, \lambda)}^{(m)}) \phi\left(\alpha, \lambda \mid \vartheta_{(\alpha, \lambda)}^{(m)}, \frac{1}{2}S_{(\alpha, \lambda), (\alpha, \lambda)}\right).$$

L'échantillonnage se fait en tirant un nombre m parmi $\{1, \dots, M\}$ avec poids proportionnels au $\ell(\vartheta_{(\alpha, \lambda)}^{(m)})$, et après on tire une valeur de (α, λ) à l'aide de la loi normale avec moyenne $\vartheta_{(\alpha, \lambda)}^{(m)}$ et matrice de variances $\frac{1}{2}S_{(\alpha, \lambda), (\alpha, \lambda)}$.

Notons qu'un problème potentiel avec la méthode décrite ci-dessus est que les tirages peuvent mener à des valeurs négatives, ce qui n'est pas souhaitable car les composantes de

α et λ doivent être positives. Une façon possible de contourner ce problème est d'envisager un tirage pour $(\log(\alpha), \log(\lambda))$. Soit $\boldsymbol{\theta} = (\log(\alpha), \log(\lambda), \boldsymbol{\beta})$ et $\hat{\boldsymbol{\theta}} = (\log(\hat{\vartheta}_{(\alpha,\lambda)}), \hat{\boldsymbol{\beta}})$. Dans ce cas, on aura

$$\begin{aligned}
\pi(\log(\alpha), \log(\lambda) \mid Y_0^n, X^n, a_0, D_n) &\propto \int_{\boldsymbol{\beta}} \exp \{g(\vartheta)\} \ell(\alpha, \lambda) \left[\prod_i \alpha_i \lambda_i \right] d\boldsymbol{\beta} \\
&\approx \int_{\boldsymbol{\beta}} \exp \left\{ g(\hat{\vartheta}) - \frac{1}{2} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' \tilde{H} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \right\} \ell(\alpha, \lambda) \left[\prod_i \alpha_i \lambda_i \right] d\boldsymbol{\beta} \\
&\propto \exp \left\{ -\frac{1}{2} \left((\log(\alpha), \log(\lambda)) - \log(\hat{\vartheta}_{(\alpha,\lambda)}) \right)' \right. \\
&\quad \left. \tilde{H}_{(\alpha,\lambda),(\alpha,\lambda)} \left((\log(\alpha), \log(\lambda)) - \log(\hat{\vartheta}_{(\alpha,\lambda)}) \right) \right\} \ell(\alpha, \lambda) \left[\prod_i \alpha_i \lambda_i \right].
\end{aligned} \tag{3.6.3}$$

Encore une fois on obtient le résultat (3.6.3) grâce aux propriétés de la loi normale multivariée. Remarquons que le terme $[\prod_i \alpha_i \lambda_i]$ apparaît dans l'expression ci-dessus lors de l'application du théorème de changement de variable. La notation \tilde{H} fait référence au négatif de la matrice de dérivées secondes de $g(\vartheta)$ par rapport à $\boldsymbol{\theta}$. Comme le gradient de $g(\vartheta)$ en $(\log(\alpha), \log(\lambda))$ évalué en $\log(\hat{\vartheta}_{(\alpha,\lambda)})$ s'annule, on a que

$$\begin{aligned}
\frac{\partial^2 g(\vartheta)}{\partial \log(\alpha_i) \partial \log(\alpha_j)} &= \frac{\partial^2 g(\vartheta)}{\partial \alpha_i \partial \alpha_j} \alpha_i \alpha_j \\
\frac{\partial^2 g(\vartheta)}{\partial \log(\lambda_i) \partial \log(\alpha_j)} &= \frac{\partial^2 g(\vartheta)}{\partial \lambda_i \partial \alpha_j} \lambda_i \alpha_j \\
\frac{\partial^2 g(\vartheta)}{\partial \log(\lambda_i) \partial \log(\lambda_j)} &= \frac{\partial^2 g(\vartheta)}{\partial \lambda_i \partial \lambda_j} \lambda_i \lambda_j.
\end{aligned}$$

L'échantillonnage est fait de la même façon qu'avant, sauf que cette fois-ci, $\ell(\alpha, \lambda)$ est remplacé par $\left[\prod_i \alpha_i \lambda_i \right] \ell(\alpha, \lambda)$.

3.7. INFÉRENCE PRÉDICTIONNELLE *a posteriori*

Le but de cette section est de présenter une expression explicite pour la distribution de prévision $\pi(y_{n+1} \mid D_k, \mathbf{x}_{n+1}, \nu_{n+1})$. Les développements qui suivent proviennent du travail de MURUA et QUINTANA [28]. Soit ρ_{n+1} une partition quelconque des données complètes $D_k = (X^n, Y^n, \boldsymbol{\nu})$ au sein du k -ième grappe plus une nouvelle observation $(\mathbf{x}_{n+1}, y_{n+1}, \nu_{n+1})$. Si nous conditionnons $\pi(y_{n+1} \mid D_k, \mathbf{x}_{n+1}, \nu_{n+1})$ sur cette partition, nous avons,

$$\pi(y_{n+1} \mid D_k, \mathbf{x}_{n+1}, \nu_{n+1}, \rho_{n+1}) = \frac{\pi(Y^n, y_{n+1} \mid X^n, \mathbf{x}_{n+1}, \nu_{n+1}, \rho_{n+1})}{\pi(Y^n \mid X^n, \mathbf{x}_{n+1}, \nu_{n+1}, \rho_{n+1})}. \tag{3.7.1}$$

Une façon d'exprimer ce rapport est d'utiliser la fonction de vraisemblance et la fonction η (voir équation (3.5.6)).

$$\frac{\int L(\vartheta \mid D_k, \mathbf{x}_{n+1}, \nu_{n+1}, y_{n+1}, \rho_{n+1}) \eta(\vartheta, \theta \mid X^n, \mathbf{x}_{n+1}, \nu_{n+1}, Y_0^n, a_0, \rho_{n+1}) d\vartheta d\theta}{\int_{y_{n+1}} \int L(\vartheta \mid D_k, \mathbf{x}_{n+1}, \nu_{n+1}, y_{n+1}, \rho_{n+1}) \eta(\vartheta, \theta \mid X^n, \mathbf{x}_{n+1}, \nu_{n+1}, Y_0^n, a_0, \rho_{n+1}) d\vartheta d\theta dy_{n+1}}. \quad (3.7.2)$$

Si on marginalise sur toutes les partitions possibles ρ_{n+1} , la loi de prévision *a posteriori* peut s'exprimer comme

$$\pi(y_{n+1} \mid D_k, \mathbf{x}_{n+1}, \nu_{n+1}) = \sum_{\rho_{n+1}} \frac{\pi(Y^n, y_{n+1} \mid X^n, \mathbf{x}_{n+1}, \nu_{n+1}, \rho_{n+1})}{\pi(Y^n \mid X^n, \mathbf{x}_{n+1}, \nu_{n+1}, \rho_{n+1})} \pi(\rho_{n+1} \mid D_k, \mathbf{x}_{n+1}, \nu_{n+1}). \quad (3.7.3)$$

On reprend maintenant les concepts introduits dans la section 2.5.1. Considérons une partition, ρ_n *a posteriori* des données (X^n, Y^n) et soit $\mathcal{A}(\rho_{n+1})$ la collection de toutes les partitions ρ_n donnant lieu à ρ_{n+1} , en générant des *liens* appropriés \mathbf{b}_{n+1} de \mathbf{x}_{n+1} dans le graphe associé aux données, $\mathcal{G} = \{(X^n, Y^n), E\}$. On peut aussi penser à ρ_n comme une partition des données originales qui survient en marginalisant tous les *liens* de \mathbf{x}_{n+1} . Nous avons ainsi,

$$\begin{aligned} \pi(\rho_{n+1} \mid D_k, \mathbf{x}_{n+1}, \nu_{n+1}) &= \sum_{\rho_n \in \mathcal{A}(\rho_{n+1})} \pi(\rho_n \mid D_k) \pi(\rho_{n+1} \mid X^n, \mathbf{x}_{n+1}, \nu_{n+1}, \rho_n) \\ &= \sum_{\rho_n \in \mathcal{A}(\rho_{n+1})} \pi(\rho_n \mid D_k) \sum_{\mathbf{b}_{n+1} \Rightarrow \rho_{n+1}} p(\mathbf{b}_{n+1} \mid X^n, \mathbf{x}_{n+1}, \nu_{n+1}, \rho_n). \end{aligned} \quad (3.7.4)$$

La dernière égalité survient de l'expression (3.2.2). Remarquons que le modèle de Potts qui est appliqué dans cette partie permet la formation de liens entre les points du graphe \mathcal{G}_X et le nouveau point \mathbf{x}_{n+1} avec probabilité $p_{n+1,j}$ si le nouveau point \mathbf{x}_{n+1} est voisin avec \mathbf{x}_j et s'ils partagent la même étiquette de couleur.

Considérons maintenant $[\rho_n, \mathbf{b}_{n+1}]$, une partition de $X^n \cup \{\mathbf{x}_{n+1}\}$ générée de ρ_n par les liens \mathbf{b}_{n+1} associés à \mathbf{x}_{n+1} . Les équations (3.7.3) et (3.7.4) ci-dessus nous permettent d'écrire la distribution de prévision *a posteriori* $\pi(y_{n+1} \mid D_k, \mathbf{x}_{n+1}, \nu_{n+1})$, comme

$$\sum_{\rho_n} \sum_{\mathbf{b}_{n+1}} \pi(\rho_n \mid D_k) \pi(\mathbf{b}_{n+1} \mid X^n, \mathbf{x}_{n+1}, \nu_{n+1}, \rho_n) \frac{\pi(Y^n, y_{n+1} \mid X^n, \mathbf{x}_{n+1}, \nu_{n+1}, [\rho_n, \mathbf{b}_{n+1}])}{\pi(Y^n \mid X^n, \mathbf{x}_{n+1}, \nu_{n+1}, [\rho_n, \mathbf{b}_{n+1}])}.$$

Par conséquent, étant donné un échantillon *a posteriori* de partitions $\{\rho_{n,m} : m = 1, \dots, M\}$ et liens $\{\mathbf{b}_{n+1,m} : m = 1, \dots, M\}$, on peut estimer la loi de prévision *a posteriori* à l'aide de l'expression suivante,

$$\hat{\pi}(y_{n+1} \mid D_k, \mathbf{x}_{n+1}, \nu_{n+1}) = \sum_{m=1}^M \frac{\pi(Y^n, y_{n+1} \mid X^n, \mathbf{x}_{n+1}, \nu_{n+1}, [\rho_{n,m}, \mathbf{b}_{n+1,m}])}{\pi(Y^n \mid X^n, \mathbf{x}_{n+1}, \nu_{n+1}, [\rho_{n,m}, \mathbf{b}_{n+1,m}])}. \quad (3.7.5)$$

Il est nécessaire, pour utiliser cette estimation, de calculer les intégrales du côté droit de l'expression ci-haut (celles qui correspondent à (3.7.2)). Nous utiliserons l'approximation de Laplace pour réaliser les calculs en question. Le développement détaillé peut être consulté en annexe D. Soit,

$$f_{n+1}(y \mid \vartheta, y_0, \mathbf{x}, \nu, a_0, [\rho_n, \mathbf{b}_{n+1}]) = (\alpha_j \lambda_j e^{\mathbf{x}'\beta} y^{\alpha_j-1})^\nu \exp\{-\lambda_j e^{\mathbf{x}'\beta} y^{\alpha_j}\} (\alpha_1 \lambda_1 y_0^{\alpha_1-1} e^{\mathbf{x}'\beta})^{a_0} \times \\ \exp\left\{e^{\mathbf{x}'\beta} \left[-a_0 \lambda_1 y_0^{\alpha_1} + \lambda_j s_{j-1}^{\alpha_j} - \sum_{g=1}^{j-1} \lambda_g (s_g^{\alpha_g} - s_{g-1}^{\alpha_g})\right]\right\}, \quad (3.7.6)$$

si $s_{j-1} < y \leq s_j$, $j > 1$ et

$$f_{n+1}(y \mid \vartheta, y_0, \mathbf{x}, \nu, a_0, [\rho_n, \mathbf{b}_{n+1}]) = (\alpha_1 \lambda_1 e^{\mathbf{x}'\beta} y^{\alpha_1-1})^\nu \exp\{-\lambda_1 e^{\mathbf{x}'\beta} y^{\alpha_1}\} (\alpha_1 \lambda_1 y_0^{\alpha_1-1} e^{\mathbf{x}'\beta})^{a_0} \times \\ \exp\{-a_0 \lambda_1 e^{\mathbf{x}'\beta} y_0^{\alpha_1}\}, \quad (3.7.7)$$

si $j = 1$. L'approximation résultante est donc,

$$\hat{\pi}(y_{n+1} \mid D_k, \mathbf{x}_{n+1}, \nu_{n+1}) = \sum_{m=1}^M \frac{f_{n+1}(y_{n+1} \mid \vartheta_{m,k(\mathbf{x}_{n+1})}^*, y_0, \mathbf{x}_{n+1}, \nu_{n+1}, a_0, [\rho_{n,m}, \mathbf{b}_{n+1,m}])}{\int_{y_{n+1}} f_{n+1}(y_{n+1} \mid \vartheta_{m,k(\mathbf{x}_{n+1})}^*, y_0, \mathbf{x}_{n+1}, \nu_{n+1}, a_0, [\rho_{n,m}, \mathbf{b}_{n+1,m}]) dy_{n+1}}. \quad (3.7.8)$$

La notation $k(\mathbf{x}_{n+1})$ indique le grappe dans la partition, $[\rho_{n,m}, \mathbf{b}_{n+1,m}]$ où se trouve l'observation \mathbf{x}_{n+1} . Maintenant, afin de définir l'intégral au dénominateur en (3.7.8), on introduit la fonction

$$c_j(\mathbf{x}) = (\alpha_1 \lambda_1 y_0^{\alpha_1-1} e^{\mathbf{x}'\beta})^{a_0} \exp\left\{e^{\mathbf{x}'\beta} \left[-a_0 \lambda_1 y_0^{\alpha_1-1} + \lambda_j s_{j-1}^{\alpha_j} - \sum_{g=1}^{j-1} \lambda_g (s_g^{\alpha_g} - s_{g-1}^{\alpha_g})\right]\right\},$$

pour $j \in \{1, \dots, J\}$. Notons que ces constantes dépendent de la partition. Alors, une approximation de l'intégrale dans le dénominateur de l'expression ci-dessus est,

$$\sum_{j=1}^J c_j(\mathbf{x}) \left(\exp\{-\lambda_j e^{\mathbf{x}'\beta} s_{j-1}^{\alpha_j}\} - \exp\{-\lambda_j e^{\mathbf{x}'\beta} s_j^{\alpha_j}\} \right),$$

si $\nu = 1$. En particulier, quand $J = 1$ et $\nu = 1$ la loi de prévision *a posteriori* est approximativement un mélange de lois Weibull par intervalles tronquées.

3.7.1. La valeur prédictive moyenne

Soit

$$\tilde{c}_j(\mathbf{x}) = c_j(\mathbf{x}) \left((\alpha_1 \lambda_1 y_0^{\alpha_1-1} e^{\mathbf{x}'\beta})^{a_0} \exp\{-a_0 \lambda_1 e^{\mathbf{x}'\beta} y_0^{\alpha_1-1}\} \right)^{-1}.$$

Nous notons que pour calculer la prévision moyenne à partir d'un ensemble de paramètres $\{\vartheta_k\}_j$, on aura

$$\begin{aligned}
E(y \mid \{\vartheta_k\}_{j=1}^J) &\approx \frac{\sum_{j=1}^J \tilde{c}_j(\mathbf{x}) \int_{s_{j-1}}^{s_j} \alpha_j \lambda_j e^{\mathbf{x}'\beta} y^{\alpha_j} \exp\{-\lambda_j e^{\mathbf{x}'\beta} y^{\alpha_j}\} dy}{\sum_{j=1}^J \tilde{c}_j(\mathbf{x}) \int_{s_{j-1}}^{s_j} \alpha_j \lambda_j e^{\mathbf{x}'\beta} y^{\alpha_j-1} \exp\{-\lambda_j e^{\mathbf{x}'\beta} y^{\alpha_j}\} dy} \\
&= \frac{\sum_{j=1}^J \tilde{c}_j(\mathbf{x}) \lambda_j^{-1/\alpha_j} e^{-\mathbf{x}'\beta/\alpha_j} \Gamma(1 + (1/\alpha_j)) \left[G(s_j^{\alpha_j}; 1 + (1/\alpha_j), \lambda_j e^{\mathbf{x}'\beta}) - G(s_{j-1}^{\alpha_j}; 1 + (1/\alpha_j), \lambda_j e^{\mathbf{x}'\beta}) \right]}{\sum_{j=1}^J \tilde{c}_j(\mathbf{x}) \left(\exp\{-\lambda_j e^{\mathbf{x}'\beta} s_{j-1}^{\alpha_j}\} - \exp\{-\lambda_j e^{\mathbf{x}'\beta} s_j^{\alpha_j}\} \right)},
\end{aligned}$$

où $G(w; a, b)$ est la fonction de distribution de la loi gamma avec paramètre de forme a et paramètre d'échelle b , évaluée en w .

Chapitre 4

RÉSULTATS NUMÉRIQUES

Afin de comprendre les scénarios dans lesquels notre modèle donne les meilleurs ou les pires résultats, on a effectué plusieurs simulations en considérant divers choix d'hyperparamètres. Dans ce chapitre, nous étudierons le comportement du modèle présenté dans ce mémoire en pratique et le comparerons à d'autres techniques proposées pour résoudre le même genre de problème. Plus spécifiquement, nous comparerons notre modèle au modèle PPM_x présenté dans la section 2.4.2.1 et à un modèle de régression de Cox traditionnel, tel qu'introduit dans la section 1.5.

On a généré artificiellement différentes populations de grande taille issues d'un mélange de lois gaussiennes, puis différents hyper-paramètres pour cette simulation ont été considérés. Ensuite, on a tiré 100 échantillons de taille 1000 chaque population que nous avons utilisés comme ensembles d'entraînement afin d'optimiser nos paramètres. Nous avons également créé un échantillon indépendant de taille 100 pour évaluer la qualité de l'ajustement des modèles entraînés. Les détails des choix de l'implémentation ainsi que les résultats de cette expérience sont présentés dans les prochaines sections. Le code pour l'implémentation du modèle à été réalisé en C/C++ avec la contribution de la direction de ce projet.

4.1. GÉNÉRATION DE POPULATIONS ARTIFICIELLES ISSUES D'UN MÉLANGE GAUSSIEN

On a considéré la génération de *clusters* de données, car notre objectif est de regrouper les individus avec des caractéristiques similaires pour ensuite appliquer l'inférence décrite dans la section 3.4 sur chacun des *clusters*. Cette simulation a été effectuée à l'aide de la bibliothèque statistique **MixSim** de R (MELNYKOV, CHEN et MAITRA [23]), qui permet la génération de mélanges gaussiens de la forme,

$$g(z, \Upsilon) = \sum_{k=1}^K \xi_k f(z; \mu_k, \Sigma_k),$$

où f est une densité gaussienne paramétrée par (μ_k, Σ_k) , $k = 1, \dots, K$ et ξ_1, \dots, ξ_K sont des probabilités satisfaisant $\xi_k \geq 0, k = 1, \dots, K$ et $\sum_{k=1}^K \xi_k = 1$. On emploie la notation Υ pour dénoter le vecteur de paramètres $\Upsilon = (\xi_1, \dots, \xi_K, \mu_1, \dots, \mu_K, \Sigma_1, \dots, \Sigma_K)$. On peut interpréter une variable aléatoire issue d'une distribution de mélange gaussien comme une variable générée aléatoirement d'une population de K groupes avec proportions $\xi_k, k = 1, \dots, K$.

La librairie **MixSim** de R nous permet également de construire des jeux de données à partir d'un v.a. issue du modèle de mélange tout en considérant différents niveaux de chevauchement entre les composantes de ce mélange. On a tout d'abord simulé $n = 10000$ observations du vecteur $X^n = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ pour différentes valeurs de p (dimension du vecteur \mathbf{x}_i) et pour différents valeurs de K (nombre de sous-groupes dans la population). Les détails de toutes les différentes populations simulées sont résumés dans le tableau suivant.

Tableau 4. I. Hyper-paramètres choisis pour la génération de populations issues d'un modèle de mélange gaussien.

Hyper-paramètre	Description	Valeur
N	taille de chaque population	10000
p	dimension du vecteur de covariables	2, 5, 10
K	nombre de <i>clusters</i> simulés dans chaque population	3,6,10
M	nombre d'échantillons générés de chaque population	100
n	taille de chaque échantillon	1000

Nous avons également généré séparément un ensemble de test de taille 100 pour chaque population artificielle afin de mesurer la performance de notre modèle. On a répété cette même expérience trois fois, en faisant varier le paramètre qui contrôle le degré de chevauchement entre les composantes du mélange. Ce paramètre est défini comme la somme de deux probabilités d'erreurs de classification, plus précisément le chevauchement entre la i -ième et la j -ième composante du mélange est défini comme,

$$\omega_{ij} = \omega_{i|j} + \omega_{j|i},$$

où $\omega_{j|i}$ est la probabilité de mal classer une observation provenant de la i -ième composante comme provenant de la j -ième composante, c'est-à-dire,

$$\omega_{j|i} = P(\xi_i f(z; \mu_i, \Sigma_i) < \xi_j f(z; \mu_j, \Sigma_j) \mid Z \sim N_p(\mu_i, \Sigma_i)).$$

Ainsi, ce paramètre ω_{ij} mesure le degré d'interaction entre les composantes du modèle de mélange et peut être facilement utilisé pour contrôler la complexité de regroupement des données simulées. Un exemple de deux échantillons avec chevauchements modéré et élevé est représenté dans la figure 4.1. Les détails des paramètres de chevauchement choisis pour chaque dimension, p et nombre de *clusters*, K au sein de chaque population générée peuvent

être consultés en annexe E dans le tableau E. I. La différence entre le choix de ces paramètres est due à la complexité ajoutée par le nombre de sous-ensembles.

Additionnellement, on a considéré une transformation de Box-Cox inversée (DAIMON [6]) de la forme, $x = (vx^* + 1)^{\frac{1}{v}} - 1$ afin de générer des données qui ne sont pas forcément distribués selon la loi normale. Ici x et x^* représentent l'observation originale et l'observation transformée respectivement, cette dernière suit la loi normale pour un v quelconque. Dans le cadre de cette expérience, on a choisi $v = 0,5$ pour la génération de tous nos jeux de données.

Il est important de mentionner aussi que la taille de chaque grappe peut être également contrôlée à partir des paramètres ξ_1, \dots, ξ_K . On a choisi ces paramètres de sorte que la taille de tous les sous-ensembles soient à peu près équilibrée, mais pas exactement identique.

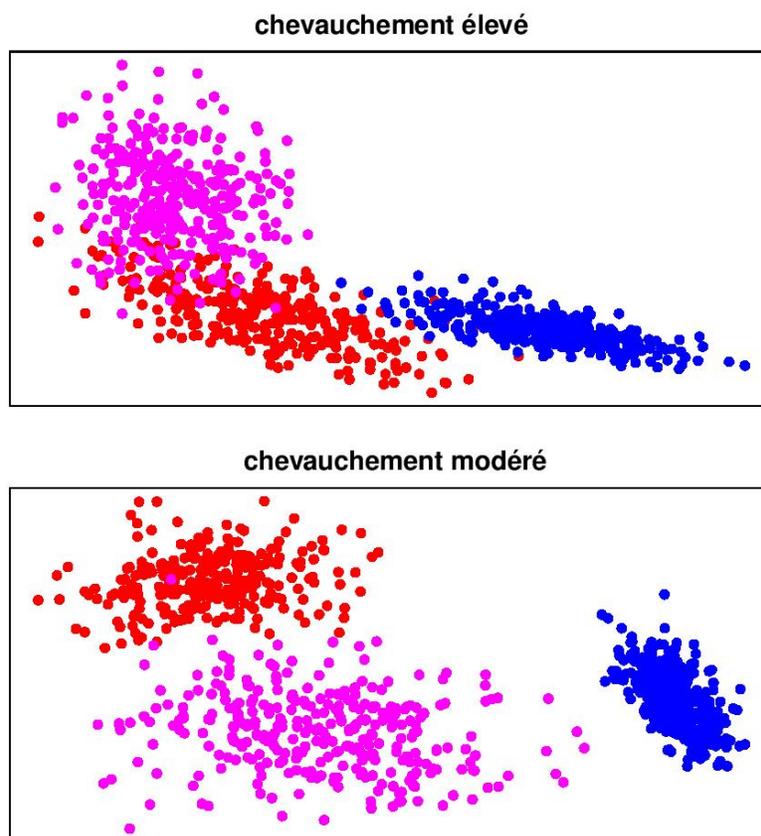


FIGURE 4.1. Échantillon de 1000 observations générées d'une mélange gaussien et une transformation de Box-Cox avec paramètre $v = 0,5$; $K = 3$ et $p = 2$. Pour l'échantillon ci-haut, le paramètre de chevauchement est $\bar{\omega} = 0,005$ et pour l'échantillon avec chevauchement modéré, $\bar{\omega} = 0,05$.

4.2. GÉNÉRATION DE LA VARIABLE RÉPONSE

La prochaine étape consiste à générer des réponses y correspondant aux covariables générées selon la description de la section précédente. Pour ce faire, nous avons appliqué la

méthode de la transformée inverse (DEVROYE [7]) sur la fonction de distribution du modèle décrit dans la section 3.4, définie comme

$$1 - S(y_i | \mathbf{x}_i, \vartheta) = 1 - \exp\{-H_0(y_i | \alpha, \lambda) \exp(\mathbf{x}'_i \boldsymbol{\beta})\}.$$

Les détails techniques de l'inversion de la vraisemblance peuvent être consultés dans l'annexe E. Pour la génération de ces variables réponse y , différents choix d'hyperparamètres s'imposent également, notamment le nombre d'intervalles J , ainsi que les valeurs de départ de α , λ et $\boldsymbol{\beta}$. De plus, étant donné le nombre d'intervalles, il faut choisir une partition *a priori* de l'axe du temps (vecteur S dans la notation du chapitre 3). Le choix de la partition *a priori*, S n'est pas évident, car celle-ci vient influencer la dispersion des y tout comme le paramètre λ . Pour cette simulation on a considéré $S = (0, 60)$ pour $J = 1$, $S = (0, 30, 60)$ pour $J = 2$ et $S = (0, 20, 40, 60)$ pour $J = 3$. On a également fixé le paramètre $\alpha = 1$. Le résumé des valeurs du reste des hyperparamètres choisis pour chacune des populations peut être consulté dans la section E.5 de l'annexe E.

Une fois les données générées, on a tiré $M = 100$ échantillons de taille $n = 1000$ de chaque population à partir d'un plan stratifié en prenant comme strates chacun des grappe présents dans la population. Ainsi, les proportions établies par les ξ_1, \dots, ξ_K sont respectées dans les échantillons. La figure 4.1 correspond à un exemple d'échantillon tiré de l'une des populations générées lors de cette expérience.

4.3. ÉVALUATION DE LA PRÉVISION

On présente dans cette section des résultats qui comparent la réponse moyenne $E(y_i | \mathbf{x}_i)$ prédite par le modèle de Cox-Potts avec la vraie réponse moyenne générée. Notre mesure de performance est la racine carrée de l'erreur quadratique moyenne (RMSE, *Root Mean Squared Error*). On considère également la comparaison du modèle Cox-Potts avec ces deux modèles compétiteurs : le modèle de régression de Cox et le modèle PPM_x . Les résultats sont présentés sous forme de carré hyper-greco-latin, car ceci nous permet d'avoir un aperçu général du comportement des modèles sous une diversité de scénarios. On a fait varier cinq différents facteurs : à savoir, le nombre de *clusters* dans la population ($K3$, $K6$, $K10$), la dimension des covariables ($p2$, $p5$, $p10$), le paramètre de chevauchement entre les composantes du mélange gaussienne dont les données sont issues (HO, MO, LO), le nombre d'intervalles théoriques, ($J1$, $J2$, $J3$) c'est-à-dire, le nombre d'intervalles choisis dans la simulation et finalement le nombre d'intervalles *a priori* qui est choisi par l'utilisateur ($I1$, $I2$, $I3$). Ce dernier facteur est propre au modèle de Cox-Potts et n'a pas donc d'effet sur ses deux compétiteurs.

Dans un premier temps on a roulé des simulations du modèle Cox-Potts sous trois scénarios différents. Le premier où on fait varier I comme étant un facteur du carré latin, le

deuxième en fixant $I = J$ et le troisième en fixant $I = 1$. Les résultats de cette première comparaison sont présentées dans le tableau suivant.

Tableau 4. II. Carré hyper-greco-latin qui résume les résultats mesurant la performance du modèle de Cox-Potts pour des différents valeurs de I .

<i>clusters</i>	modèle	$J = 1$		$J = 2$		$J = 3$	
		Moy.	É-type	Moy.	É-type	Moy.	É-type
$K = 3$		p5LOI2		p2MOI1		p10HOI3	
	I facteur	7,3791	0,4192	3,0816	0,2238	5,2427	0,4461
	$I = J$	5,5292	0,1333	5,0460	0,5162	5,2427	0,4461
	$I = 1$	5,5292	0,1333	3,0816	0,2238	3,9394	0,1489
$K = 6$		p2HOI3		p10LOI2		p5MOI1	
	I facteur	7,0570	0,6210	8,7602	0,3253	3,8504	0,2327
	$I = J$	5,4515	0,2765	8,7602	0,3253	5,1848	0,6188
	$I = 1$	5,4515	0,2765	6,6559	0,1646	3,8504	0,2327
$K = 10$		p10MOI1		p5HOI3		p2LOI2	
	I facteur	6,5046	0,1642	7,3675	0,8072	7,4848	0,7197
	$I = J$	6,5046	0,1642	7,6673	0,6570	6,7106	0,9302
	$I = 1$	6,5046	0,1642	4,1336	0,2877	4,0625	0,2126
$K = 3$		p2LOI1		p10MOI3		p5HOI2	
	I facteur	3,5747	0,1621	5,2519	0,5714	5,0141	0,4906
	$I = J$	3,5747	0,1621	6,0577	0,5594	4,1740	0,4205
	$I = 1$	3,5747	0,1621	4,3437	0,1636	3,7849	0,1860
$K = 6$		p10HOI2		p5LOI1		p2MOI3	
	I facteur	9,4553	0,4358	4,5099	0,2274	6,3624	0,9384
	$I = J$	7,5147	0,0859	7,8042	0,7836	6,3624	0,9384
	$I = 1$	7,5147	0,0859	4,5099	0,2274	4,9601	0,1921
$K = 10$		p5MOI3		p2HOI2		p10LOI1	
	I facteur	9,7067	0,6684	6,7747	0,4765	4,2943	0,3452
	$I = J$	8,2176	0,3449	6,7747	0,4765	6,3624	0,9384
	$J = 1$	8,2176	0,3449	4,2979	0,3553	4,2943	0,3452

Notre mesure de performance est la racine carré de l'erreur quadratique moyenne (RMSE, *Root Mean Squared Error*).

Le tableau 4. II suggère que quand le nombre d'intervalles *a priori* I est différent de 1, le modèle de Cox-Potts est pénalisé. Remarquons que quand le nombre d'intervalles réel $J = 1$ est combiné avec $I = 3$ ou encore avec $I = 2$, l'erreur associée est considérablement plus grande. Ainsi, le choix *a priori* de $I = 1$ semble favoriser notre modèle. On prend donc l'hyper-paramètre $I = 1$ fixe afin de réaliser la comparaison avec les deux modèles compétiteurs. On a basé cette étude comparative sur l'application des fonctions incluses

dans la librairie **ppmx** de R. Cette librairie contient les fonctions qui correspondent au modèle décrit dans MULLER, QUINTANA et ROSNER [26] et qui a été également présenté dans la section 2.4.2.1 de ce mémoire. L'ajustement du modèle de régression de Cox a été généré avec la librairie **survival** du paquet statistique R. Les résultats sont présentés dans le tableau suivant.

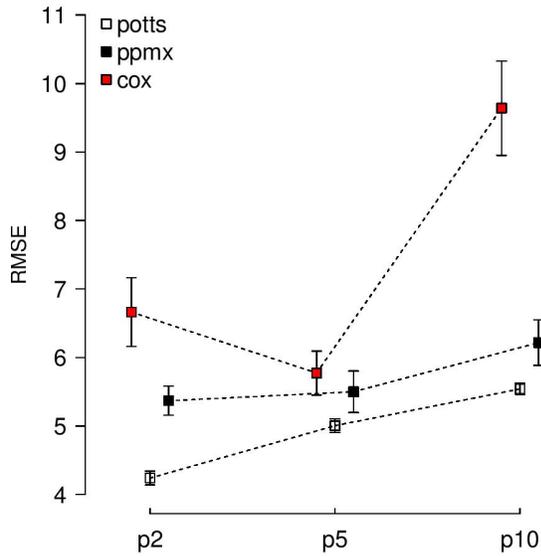
Tableau 4. III. Carré hyper-greco-latin qui résume les résultats mesurant la performance du modèle de Cox-Potts et ses compétiteurs : PPM_x et Cox. L'hyper-paramètre I est fixé à 1.

<i>clusters</i>	modèle	$J = 1$		$J = 2$		$J = 3$	
		Moy.	É-type	Moy.	É-type	Moy.	É-type
$K = 3$		p5LO		p2MO		p10HO	
	cox-potts	5,5292	0,1333	3,0816	0,2238	3,9394	0,1489
	ppmx	6,1082	0,4264	4,4406	0,5589	4,7775	0,6286
	cox	5,9554	0,8094	9,7992	2,4878	7,4582	1,6066
$K = 6$		p2HO		p10LO		p5MO	
	cox-potts	5,4515	0,2765	6,6559	0,1646	3,8504	0,2327
	ppmx	6,5765	0,3440	6,9299	0,4779	5,5737	0,6549
	cox	5,0389	0,1421	14,4463	1,7846	4,0123	0,3989
$K = 10$		p10MO		p5HO		p2LO	
	cox-potts	6,5046	0,1642	4,1336	0,2877	4,0625	0,2126
	ppmx	7,2829	0,9569	4,8549	1,0136	5,2072	0,2756
	cox	12,3686	2,0414	5,1129	1,0192	5,1491	0,7032
$K = 3$		p2LO		p10MO		p5HO	
	cox-potts	3,5747	0,1621	4,3437	0,1636	3,7849	0,1860
	ppmx	5,5543	0,7312	6,3436	1,2444	4,6826	0,7354
	cox	11,7727	1,3618	5,8641	1,3015	8,6412	0,3299
$K = 6$		p10HO		p5LO		p2MO	
	cox-potts	7,5147	0,0859	4,5099	0,2274	4,9601	0,1921
	ppmx	7,3671	0,6394	4,2474	0,7873	4,1679	0,2172
	cox	12,5032	1,7922	4,6391	1,2277	4,0671	0,2879
$K = 10$		p5MO		p2HO		p10LO	
	cox-potts	8,2176	0,3449	4,2979	0,3553	4,2943	0,3452
	ppmx	7,5294	0,7000	6,2631	0,7290	4,6012	0,6846
	cox	6,2751	0,4493	4,1466	0,6186	5,1980	1,4811

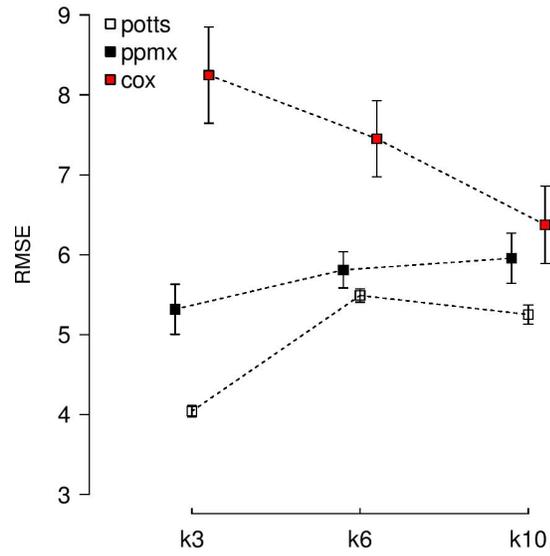
Notre mesure de performance est la racine carrée de l'erreur quadratique moyenne (RMSE, *Root Mean Squared Error*).

Notons que dans 11 des 18 configurations testées, c'est le modèle de Cox-Potts qui s'avère le plus performant et même dans les cas où il est battu par ses compétiteurs, les différences

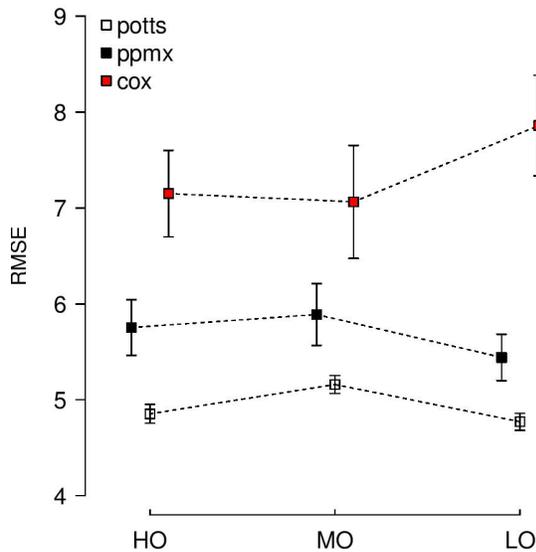
des erreurs sont relativement faibles. D'ailleurs les écart-types du RMSE associés à Cox-Potts sont toujours plus petits. Afin de mieux apprécier l'effet de chacun des facteurs, les résultats présentés dans le tableau 4. III ont été regroupés par facteurs, puis les erreurs ont été recalculées, les résultats sont illustrés par les graphiques suivants.



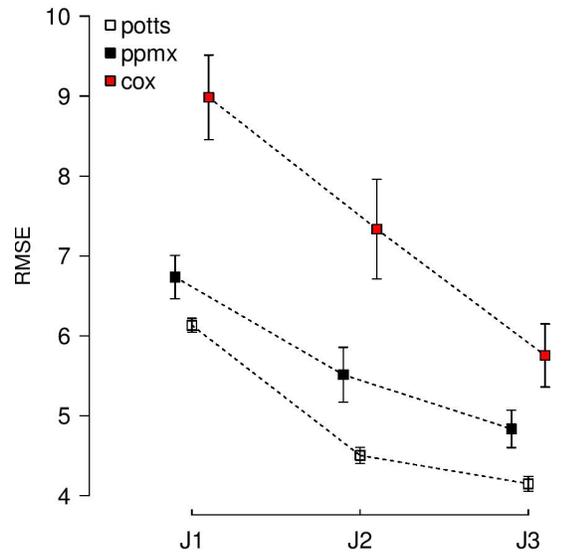
(a) résultats selon la dimension



(b) résultats selon le nombre de *clusters*



(c) résultats selon le chevauchement



(d) résultats selon le nombre d'intervalles J

FIGURE 4.2. Résultats de la simulation (en fixant $I = 1$) selon les facteurs dimension, nombre de *clusters*, chevauchement et nombre d'intervalles.

La figure 4.2 ci-haut nous permet d'observer qu'en fixant $I = 1$, c'est le modèle cox-potts qui offre les meilleurs résultats pour chacun des facteurs. Tel qu'on s'y attendait, c'est le

modèle PPM_x qui le suit en termes de performance, pour laisser loin derrière le modèle de régression de Cox traditionnel.

Quand au nombre d'intervalles J , on voit bien que les erreurs décroissent avec l'augmentation de la valeur de J . Effectivement, les plus petites erreurs sont produites pour le niveau $J = 3$ et ceci s'applique pour les trois modèles testés. Les erreurs des modèles de Cox-Potts et de PPM_x ont tendance à augmenter avec la dimension, mais le modèle de Cox a un comportement différent accompagné des erreurs les plus grandes dans les trois niveaux de ce facteur. Contrairement à ce que l'on s'attendrait à voir, un paramètre de chevauchement élevé et un nombre de *clusters* élevé ne sont pas forcément liés à une erreur plus élevée. Un autre avantage de notre modèle par rapport à ses compétiteurs est que l'on observe en général des écart-types du RMSE moins élevés.

Les moyennes et écart-types des erreurs représentés dans la figure 4.2 sont présentés en forme de tableau en 4. IV.

Tableau 4. IV. Résultats des effets séparés selon les quatre facteurs différents considérés. Le facteur nombre d'intervalles I est fixé à 1.

modèle	$p1$		$p5$		$p10$	
	Moy.	É-type	Moy.	É-type	Moy.	É-type
cox-potts	4,2381	0,1002	5,0043	0,1000	5,5421	0,0799
ppmx	5,3683	0,2121	5,4994	0,3023	6,2171	0,3319
cox	6,6623	0,5007	5,7727	0,3193	9,6397	0,6878
	$K3$		$K6$		$K10$	
cox-potts	4,0423	0,0702	5,4904	0,0840	5,2518	0,1201
ppmx	5,3178	0,3124	5,8104	0,2268	5,9565	0,3122
cox	8,2485	0,6028	7,4512	0,4763	6,3750	0,4853
	HO		MO		LO	
cox-potts	4,8537	0,0985	5,1597	0,0934	4,7711	0,0893
ppmx	5,7536	0,2897	5,8897	0,3224	5,4414	0,2421
cox	7,1502	0,4516	7,0644	0,5891	7,8601	0,5241
	$J = 1$		$J = 2$		$J = 3$	
cox-potts	6,1321	0,0872	4,5038	0,1006	4,1486	0,0931
ppmx	6,7364	0,2713	5,5133	0,3443	4,8351	0,2331
cox	8,9857	0,5300	7,3347	0,6235	5,7543	0,3951

Notre mesure de performance est la racine carré de l'erreur quadratique moyenne (RMSE, *Root Mean Squared Error*).

Chapitre 5

CONCLUSIONS

Le but principal de ce travail de recherche était le développement et l'utilisation d'un modèle bayésien semi-paramétrique avec une dépendance des covariables indexées à travers une régression de Cox par intervalles. La discussion a porté sur un exemple particulier : le modèle de Potts, introduit dans la littérature statistique par MURUA et collab. [29].

Le premier chapitre de ce mémoire résume les concepts principaux en analyse de survie. C'est dans ce chapitre que nous avons introduit les détails du modèle de régression de Cox et du modèle original de Ibrahim.

Le chapitre deux est dédié aux modèles de partition aléatoire, incluant la présentation du modèle de partition aléatoire de Potts et du modèle de partition aléatoire avec covariables, PPM_x . Un survol du cadre de travail de l'analyse bayésienne classique et non paramétrique fut également inclus.

Les développements et descriptions techniques du modèle proposé dans ce travail de recherche se trouvent dans le chapitre trois. Nous avons présenté et justifié de façon détaillée l'inférence effectuée au sein de chaque sous-ensemble ainsi que l'inférence prédictive *a posteriori*.

Finalement, la performance du modèle fut évaluée et comparée à des alternatives compétitives dans un exercice de simulation présenté au chapitre quatre. Dans nos exemples de simulation, les covariables étaient toutes de type continu, mais cela ne doit pas être considéré comme une restriction de la méthode car tout type de covariable peut être utilisé.

Globalement, on a observé que le modèle de Cox-Potts possède des écart-types du RMSE plus faibles par rapport aux deux modèles concurrents et dans la plupart des cas il semble être le modèle le plus performant. Cependant le fait de mal choisir le paramètre I pénalise le modèle et fait augmenter les erreurs.

Nous avons découvert que fixer la valeur $I = 1$ était la meilleure option pour produire des erreurs plus faibles. En effet, lors de nos exercices de simulation, on a constaté que le choix du paramètre $I = 1$ fonctionne mieux que d'essayer de deviner la vraie valeur de J . La figure 4.2 montre que le modèle de Cox-Potts est le plus performant lorsque $I = 1$ et ceci est

vrai pour tous les niveaux des quatre facteurs considérés. Ceci suggère que l'incorporation du modèle de Ibrahim dans la vraisemblance du modèle n'a pas vraiment de gain en termes de performance. Il serait donc naturel d'envisager, dans des travaux futurs, effectuer plus de simulations en faisant varier le paramètre α (propre à la fonction de risque de base de chaque sous-intervalle) et augmenter la valeur de J et de I afin de confirmer ou infirmer cette hypothèse.

Il faut aussi considérer que l'incorporation de plusieurs intervalles, J introduit une complexité considérablement plus grande par rapport aux modèles alternatifs. Même si en termes de performance, le modèle de Cox-Potts semble être supérieur au modèle PPM_x quand $I = 1$, la complexité de calcul le désavantage, car le temps d'exécution et le coût computationnel sont considérablement plus grands.

On pourrait d'autre part envisager dans des travaux futurs, de mesurer la performance du modèle de Cox-Potts sous des scénarios plus complexes. Par exemple, l'incorporation de variables de type catégoriel et discret dans nos jeux de données. Il serait également intéressant d'appliquer le modèle sur de vraies données.

Références

- [1] BARRY, D. et J. A. HARTIGAN. 1992, «Product partition models for change point problems», Ann. Statist., vol. 20, n° 1, doi :10.1214/aos/1176348521, p. 260–279, ISSN 0090-5364. URL <https://doi.org/10.1214/aos/1176348521>.
- [2] BESAG, J. 2004, «Markov chain monte carlo methods for statistical inference», .
- [3] BLATT, M., S. WISEMAN et E. DOMANY. 1996, «Superparamagnetic clustering of data», Phys. Rev. Lett., vol. 76, doi :10.1103/PhysRevLett.76.3251, p. 3251–3254. URL <https://link.aps.org/doi/10.1103/PhysRevLett.76.3251>.
- [4] COX, D. R. 1972, «Regression models and life-tables», J. Roy. Statist. Soc. Ser. B, vol. 34, p. 187–220, ISSN 0035-9246. URL [http://links.jstor.org/sici?sici=0035-9246\(1972\)34:2<187:RMAL>2.0.CO;2-6&origin=MSN](http://links.jstor.org/sici?sici=0035-9246(1972)34:2<187:RMAL>2.0.CO;2-6&origin=MSN), with discussion by F. Downton, Richard Peto, D. J. Bartholomew, D. V. Lindley, P. W. Glassborow, D. E. Barton, Susannah Howard, B. Benjamin, John J. Gart, L. D. Meshalkin, A. R. Kagan, M. Zelen, R. E. Barlow, Jack Kalbfleisch, R. L. Prentice and Norman Breslow, and a reply by D. R. Cox.
- [5] CROWLEY, E. M. 1997, «Product partition models for normal means», Journal of the American Statistical Association, vol. 92, n° 437, doi :10.1080/01621459.1997.10473616, p. 192–198. URL <https://doi.org/10.1080/01621459.1997.10473616>.
- [6] DAIMON, T. 2011, Box–Cox Transformation, Springer Berlin Heidelberg, Berlin, Heidelberg, ISBN 978-3-642-04898-2, p. 176–178, doi :10.1007/978-3-642-04898-2_152. URL https://doi.org/10.1007/978-3-642-04898-2_152.
- [7] DEVROYE, L. 1986, Non-Uniform Random Variate Generation(originally published with, Springer-Verlag. URL <http://cg.scs.carleton.ca/~luc/rnbookindex.html>.
- [8] ESCOBAR, M. D. 1994, «Estimating normal means with a Dirichlet process prior», J. Amer. Statist. Assoc., vol. 89, n° 425, p. 268–277, ISSN 0162-1459. URL [http://links.jstor.org/sici?sici=0162-1459\(199403\)89:425<268:ENMWAD>2.0.CO;2-6&origin=MSN](http://links.jstor.org/sici?sici=0162-1459(199403)89:425<268:ENMWAD>2.0.CO;2-6&origin=MSN).
- [9] FERGUSON, T. S. 1973, «A bayesian analysis of some nonparametric problems», Ann. Statist., vol. 1, n° 2, doi :10.1214/aos/1176342360, p. 209–230. URL <https://doi.org/10.1214/aos/1176342360>.
- [10] FERGUSON, T. S. 1982, «Sequential estimation with Dirichlet process priors», dans Statistical decision theory and related topics, III, Vol. 1 (West Lafayette, Ind., 1981), Academic Press, New York, p. 385–401.

- [11] FINK, D. 1997, «A compendium of conjugate priors», Technical Report.
- [12] FOGUEL, S. R. 1966, «The ergodic theorem for markov processes», Israel Journal of Mathematics, vol. 4, n° 1, doi :10.1007/BF02760066, p. 11–22, ISSN 1565-8511. URL <https://doi.org/10.1007/BF02760066>.
- [13] GELFAND, A. E. et A. F. M. SMITH. 1990, «Sampling-based approaches to calculating marginal densities», Journal of the American Statistical Association, vol. 85, n° 410, doi :10.1080/01621459.1990.10476213, p. 398–409. URL <https://www.tandfonline.com/doi/abs/10.1080/01621459.1990.10476213>.
- [14] GEMAN, S. et D. GEMAN. 1984, «Stochastic relaxation, gibbs distributions, and the bayesian restoration of images», IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-6, n° 6, doi :10.1109/TPAMI.1984.4767596, p. 721–741, ISSN 0162-8828.
- [15] HARTIGAN, J. 1990, «Partition models», Communications in Statistics - Theory and Methods, vol. 19, n° 8, doi :10.1080/03610929008830345, p. 2745–2756. URL <https://doi.org/10.1080/03610929008830345>.
- [16] HASTINGS, W. K. 1970, «Monte Carlo sampling methods using Markov chains and their applications», Biometrika, vol. 57, n° 1, doi :10.1093/biomet/57.1.97, p. 97–109, ISSN 0006-3444. URL <https://doi.org/10.1093/biomet/57.1.97>.
- [17] HERTY, M., A. KLAR, A. K. SINGH et P. SPELLUCCI. 2007, «Smoothed penalty algorithms for optimization of nonlinear models», Comput. Optim. Appl., vol. 37, n° 2, doi : 10.1007/s10589-007-9011-6, p. 157–176, ISSN 0926-6003. URL <http://dx.doi.org/10.1007/s10589-007-9011-6>.
- [18] IBRAHIM, J. G. 2001, Bayesian survival analysis, Springer series in statistics, Springer, New York, ISBN 0387952772.
- [19] IBRAHIM, J. G., M.-H. CHEN et S. KIM. 2008, «Bayesian variable selection for the Cox regression model with missing covariates», Lifetime Data Anal., vol. 14, n° 4, doi :10.1007/s10985-008-9101-5, p. 496–520, ISSN 1380-7870. URL <https://doi.org/10.1007/s10985-008-9101-5>.
- [20] KLEIN, J. P. et M. L. MOESCHBERGER. 2003, Survival Analysis Techniques for Censored and Truncated Data, 2^e éd..
- [21] LAPLACE, P. S. 1986, «Memoir on the probability of the causes of events», Statist. Sci., vol. 1, n° 3, doi :10.1214/ss/1177013621, p. 364–378. URL <https://doi.org/10.1214/ss/1177013621>.
- [22] MACEACHERN, S. N. 1994, «Estimating normal means with a conjugate style Dirichlet process prior», Comm. Statist. Simulation Comput., vol. 23, n° 3, doi :10.1080/03610919408813196, p. 727–741, ISSN 0361-0918. URL <https://doi.org/10.1080/03610919408813196>.
- [23] MELNYKOV, V., W.-C. CHEN et R. MAITRA. 2012, «Mixsim : An r package for simulating data to study performance of clustering algorithms», .
- [24] METROPOLIS, A. W. R. M. N. T. A. H. T. E., NICHOLAS ; ROSENBLUTH. 1953, «MANIAC», dans Equation of State Calculations by Fast Computing Machines, 1953, Sauls Lithograph Co. (for the Association for Computing Machinery), Washington, D. C., p. 13–17.

- [25] MÜLLER, P. et R. MITRA. 2013, «Bayesian nonparametric inference—why and how», Bayesian Anal., vol. 8, n° 2, doi :10.1214/13-BA811, p. 269–302, ISSN 1936-0975. URL <https://doi.org/10.1214/13-BA811>.
- [26] MULLER, P., F. QUINTANA et G. ROSNER. 2009, «Bayesian clustering with regression», .
- [27] MÜLLER, P. et F. A. QUINTANA. 2004, «Nonparametric Bayesian data analysis», Statist. Sci., vol. 19, n° 1, doi :10.1214/088342304000000017, p. 95–110, ISSN 0883-4237. URL <https://doi.org/10.1214/088342304000000017>.
- [28] MURUA, A. et F. A. QUINTANA. 2017, «Semiparametric Bayesian regression via Potts model», J. Comput. Graph. Statist., vol. 26, n° 2, doi :10.1080/10618600.2016.1172015, p. 265–274, ISSN 1061-8600. URL <https://doi.org/10.1080/10618600.2016.1172015>.
- [29] MURUA, A., L. STANBERRY et W. STUETZLE. 2008, «On potts model clustering, kernel k-means and density estimation», Journal of Computational and Graphical Statistics, vol. 17, n° 3, doi : 10.1198/106186008X318855, p. 629–658. URL <https://doi.org/10.1198/106186008X318855>.
- [30] MURUA, A. et N. WICKER. 2014, «The conditional-Potts clustering model», J. Comput. Graph. Statist., vol. 23, n° 3, doi :10.1080/10618600.2013.837828, p. 717–739, ISSN 1061-8600. URL <https://doi.org/10.1080/10618600.2013.837828>.
- [31] MÜLLER, P., F. QUINTANA et G. L. ROSNER. 2011, «A product partition model with regression on covariates», Journal of Computational and Graphical Statistics, vol. 20, n° 1, doi :10.1198/jcgs.2011.09066, p. 260–278. URL <https://doi.org/10.1198/jcgs.2011.09066>.
- [32] ORBANZ, P. et Y. W. TEH. 2010, «Bayesian nonparametric models», dans Encyclopedia of Machine Learning, Springer.
- [33] PITMAN, J. 1996, Some developments of the Blackwell-MacQueen urn scheme, Lecture Notes–Monograph Series, vol. Volume 30, Institute of Mathematical Statistics, Hayward, CA, p. 245–267, doi :10.1214/lnms/1215453576. URL <https://doi.org/10.1214/lnms/1215453576>.
- [34] QUINTANA, F. A. et P. L. IGLESIAS. 2003, «Bayesian clustering and product partition models», Journal of the Royal Statistical Society. Series B (Statistical Methodology), vol. 65, n° 2, p. 557–574, ISSN 13697412, 14679868. URL <http://www.jstor.org/stable/3647521>.
- [35] SETHURAMAN, J. 2012, «A short proof of the Feigin-Tweedie theorem on the existence of the mean functional of a Dirichlet process», dans Nonparametric statistical methods and related topics, World Sci. Publ., Hackensack, NJ, p. 127–136.
- [36] SHUN, Z. et P. MCCULLAGH. 1995, «Laplace approximation of high dimensional integrals», Journal of the Royal Statistical Society. Series B (Methodological), vol. 57, n° 4, p. 749–760, ISSN 00359246. URL <http://www.jstor.org/stable/2345941>.
- [37] SOKAL, A. 1997, «Monte Carlo methods in statistical mechanics : foundations and new algorithms», dans Functional integration (Cargèse, 1996), NATO Adv. Sci. Inst. Ser. B Phys., vol. 361, Plenum, New York, p. 131–192.
- [38] SWENDSEN, R. et J.-S. WANG. 1987, «Nonuniversal critical dynamics in monte carlo simulation», Physical review letters, vol. 58, doi :10.1103/PhysRevLett.58.86, p. 86–88.

- [39] SWENDSEN, R. H. et J.-S. WANG. 1987, «Nonuniversal critical dynamics in monte carlo simulations», Phys. Rev. Lett., vol. 58, doi :10.1103/PhysRevLett.58.86, p. 86–88. URL <https://link.aps.org/doi/10.1103/PhysRevLett.58.86>.
- [40] T., L. E. et W. JOHN. 2003, Statistical Methods for Survival Analysis, 2^e éd..
- [41] TEH, Y. W. 2010, «Dirichlet processes», dans Encyclopedia of Machine Learning, Springer.
- [42] TEH, Y. W. et M. I. JORDAN. 2010, «Hierarchical Bayesian nonparametric models with applications», dans Bayesian Nonparametrics : Principles and Practice, édité par N. Hjort, C. Holmes, P. Müller et S. Walker, Cambridge University Press.
- [43] TIBSHIRANI, R., M. SAUNDERS, S. ROSSET, J. ZHU et K. KNIGHT. 2005, «Sparsity and smoothness via the fused lasso», Journal of the Royal Statistical Society Series B, p. 91–108.
- [44] XU, R. et D. WUNSCH. 2009, Clustering, Wiley-IEEE Press, ISBN 9780470276808.

Annexe A

CALCUL DES DENSITÉS *A PRIORI* MARGINALES DES PARAMÈTRES ALPHA ET LAMBDA SELON LE NOMBRE D'INTERVALLES J

Remarque : on se servira des résultats suivants

$$\int_0^\infty x^n e^{-\theta x - \theta|\lambda - x|} dx = e^{-\theta\lambda} \left(\frac{\lambda^{n+1}}{n+1} + \sum_{i=0}^n \frac{n!}{i!} \frac{\lambda^i}{(2\theta)^{n+1-i}} \right) \text{ pour } n \in \mathbb{N} + \{0\}, \lambda > 0, \theta > 0.$$

$$\int_0^\infty \lambda^n e^{-\lambda\theta} d\lambda = \frac{n!}{\theta^{n+1}} \text{ pour } n \in \mathbb{N} + \{0\}, \lambda > 0, \theta > 0.$$

Ainsi, pour deux intervalles, $J = 2$, on aura

$$\int_0^\infty \int_0^\infty \theta e^{-\theta\lambda_1 - \theta|\lambda_2 - \lambda_1|} d\lambda_1 d\lambda_2 = \theta \left(\frac{1}{\theta^2} \right) + \frac{1}{2} \left(\frac{1}{\theta} \right) = \frac{3}{2\theta},$$

et donc, la densité conjointe de $\lambda = (\lambda_1, \lambda_2)$ est donnée par

$$\pi(\lambda_1, \lambda_2 | \theta) = \frac{\theta e^{-\theta\lambda_1 - \theta|\lambda_2 - \lambda_1|}}{\int_0^\infty \int_0^\infty \theta e^{-\theta\lambda_1 - \theta|\lambda_2 - \lambda_1|} d\lambda_1 d\lambda_2} = \frac{2}{3} \theta^2 e^{-\theta\lambda_1 - \theta|\lambda_2 - \lambda_1|}.$$

Si $J = 3$,

$$\int_0^\infty \int_0^\infty \int_0^\infty \theta e^{-\theta\lambda_1 - \theta|\lambda_3 - \lambda_2| - \theta|\lambda_2 - \lambda_1|} d\lambda_1 d\lambda_2 d\lambda_3 = \frac{\theta}{2} \left(\frac{2}{\theta^3} \right) + \frac{1}{\theta^2} + \frac{1}{2\theta} \left(\frac{1}{\theta} \right) = \frac{5}{2\theta^2},$$

et donc, la densité conjointe de $\lambda = (\lambda_1, \lambda_2, \lambda_3)$ est donnée par

$$\pi(\lambda_1, \lambda_2, \lambda_3 | \theta) = \frac{\theta e^{-\theta\lambda_1 - \theta|\lambda_3 - \lambda_2| - \theta|\lambda_2 - \lambda_1|}}{\int_0^\infty \int_0^\infty \int_0^\infty \theta e^{-\theta\lambda_1 - \theta|\lambda_3 - \lambda_2| - \theta|\lambda_2 - \lambda_1|} d\lambda_1 d\lambda_2 d\lambda_3} = \frac{2}{5} \theta^3 e^{-\theta\lambda_1 - \theta|\lambda_3 - \lambda_2| - \theta|\lambda_2 - \lambda_1|}.$$

Pour quatre intervalles, $J = 4$, on aura

$$\begin{aligned} & \int_0^\infty \dots \int_0^\infty \theta e^{-\theta\lambda_1 - \theta|\lambda_4 - \lambda_3| - \theta|\lambda_3 - \lambda_2| - \theta|\lambda_2 - \lambda_1|} d\lambda_1 \dots d\lambda_4 \\ &= \frac{\theta}{6} \left(\frac{6}{\theta^4} \right) + \frac{3}{4} \left(\frac{2}{\theta^3} \right) + \frac{5}{4\theta} \left(\frac{1}{\theta^2} \right) + \frac{5}{8\theta^2} \left(\frac{1}{\theta} \right) = \frac{35}{8\theta^3}, \end{aligned}$$

et la densité conjointe de $\lambda = (\lambda_1, \dots, \lambda_4)$ est donnée par

$$\begin{aligned}\pi(\lambda_1, \lambda_2, \lambda_3, \lambda_4 \mid \theta) &= \frac{\theta e^{-\theta\lambda_1 - \theta|\lambda_4 - \lambda_3| - \theta|\lambda_3 - \lambda_2| - \theta|\lambda_2 - \lambda_1|}}{\int_0^\infty \dots \int_0^\infty \theta e^{-\theta\lambda_1 - \theta|\lambda_4 - \lambda_3| - \theta|\lambda_3 - \lambda_2| - \theta|\lambda_2 - \lambda_1|} d\lambda_1 \dots d\lambda_4} \\ &= \frac{8}{35} \theta^4 e^{-\theta\lambda_1 - \theta|\lambda_4 - \lambda_3| - \theta|\lambda_3 - \lambda_2| - \theta|\lambda_2 - \lambda_1|}.\end{aligned}$$

Pour $J = 5$,

$$\begin{aligned}\int_0^\infty \dots \int_0^\infty \theta e^{-\theta\lambda_1 - \sum_{j=1}^4 |\lambda_{j+1} - \lambda_j| \theta} d\lambda_1 \dots d\lambda_5 \\ = \frac{1}{\theta^4} + \frac{2}{\theta^4} + \frac{9}{4\theta^4} + \frac{7}{4\theta^4} + \frac{7}{8\theta^4} = \frac{63}{8\theta^4},\end{aligned}$$

et la densité conjointe de $\lambda = (\lambda_1, \dots, \lambda_5)$ est donnée par

$$\pi(\lambda_1, \dots, \lambda_5 \mid \theta) = \frac{\theta e^{-\theta\lambda_1 - \sum_{j=1}^4 |\lambda_{j+1} - \lambda_j| \theta}}{\int_0^\infty \dots \int_0^\infty \theta e^{-\theta\lambda_1 - \sum_{j=1}^4 |\lambda_{j+1} - \lambda_j| \theta} d\lambda_1 \dots d\lambda_5} = \frac{8}{63} \theta^5 e^{-\theta\lambda_1 - \sum_{j=1}^4 |\lambda_{j+1} - \lambda_j| \theta}.$$

Pour $J = 6$,

$$\begin{aligned}\int_0^\infty \dots \int_0^\infty \theta e^{-\theta\lambda_1 - \sum_{j=1}^5 |\lambda_{j+1} - \lambda_j| \theta} d\lambda_1 \dots d\lambda_6 \\ = \frac{1}{\theta^5} + \frac{5}{2\theta^5} + \frac{7}{2\theta^5} + \frac{7}{2\theta^5} + \frac{21}{8\theta^5} + \frac{21}{16\theta^5} = \frac{231}{16\theta^5},\end{aligned}$$

et la densité conjointe de $\lambda = (\lambda_1, \dots, \lambda_6)$ est donnée par

$$\pi(\lambda_1, \dots, \lambda_6 \mid \theta) = \frac{\theta e^{-\theta\lambda_1 - \sum_{j=1}^5 |\lambda_{j+1} - \lambda_j| \theta}}{\int_0^\infty \dots \int_0^\infty \theta e^{-\theta\lambda_1 - \sum_{j=1}^5 |\lambda_{j+1} - \lambda_j| \theta} d\lambda_1 \dots d\lambda_6} = \frac{16}{231} \theta^6 e^{-\theta\lambda_1 - \sum_{j=1}^5 |\lambda_{j+1} - \lambda_j| \theta}.$$

Pour $J = 7$,

$$\begin{aligned}\int_0^\infty \dots \int_0^\infty \theta e^{-\theta\lambda_1 - \sum_{j=1}^6 |\lambda_{j+1} - \lambda_j| \theta} d\lambda_1 \dots d\lambda_7 \\ = \frac{1}{\theta^6} + \frac{3}{\theta^6} + \frac{5}{\theta^6} + \frac{6}{\theta^6} + \frac{45}{8\theta^6} + \frac{33}{8\theta^6} + \frac{33}{16\theta^6} = \frac{429}{16\theta^6},\end{aligned}$$

et la densité conjointe de $\lambda = (\lambda_1, \dots, \lambda_7)$ est donnée par

$$\pi(\lambda_1, \dots, \lambda_7 \mid \theta) = \frac{\theta e^{-\theta\lambda_1 - \sum_{j=1}^6 |\lambda_{j+1} - \lambda_j| \theta}}{\int_0^\infty \dots \int_0^\infty \theta e^{-\theta\lambda_1 - \sum_{j=1}^6 |\lambda_{j+1} - \lambda_j| \theta} d\lambda_1 \dots d\lambda_7} = \frac{16}{429} \theta^7 e^{-\theta\lambda_1 - \sum_{j=1}^6 |\lambda_{j+1} - \lambda_j| \theta}.$$

Pour $J = 8$,

$$\begin{aligned}\int_0^\infty \dots \int_0^\infty \theta e^{-\theta\lambda_1 - \sum_{j=1}^7 |\lambda_{j+1} - \lambda_j| \theta} d\lambda_1 \dots d\lambda_8 \\ = \frac{1}{\theta^7} + \frac{7}{2\theta^7} + \frac{27}{4\theta^7} + \frac{75}{8\theta^7} + \frac{165}{16\theta^7} + \frac{297}{32\theta^7} + \frac{429}{64\theta^7} + \frac{429}{128\theta^7} = \frac{6435}{128\theta^7},\end{aligned}$$

et la densité conjointe de $\lambda = (\lambda_1, \dots, \lambda_8)$ est donnée par

$$\pi(\lambda_1, \dots, \lambda_8 \mid \theta) = \frac{\theta e^{-\theta\lambda_1 - \sum_{j=1}^7 |\lambda_{j+1} - \lambda_j|\theta}}{\int_0^\infty \dots \int_0^\infty \theta e^{-\theta\lambda_1 - \sum_{j=1}^7 |\lambda_{j+1} - \lambda_j|\theta} d\lambda_1 \dots d\lambda_8} = \frac{128}{6435} \theta^8 e^{-\theta\lambda_1 - \sum_{j=1}^7 |\lambda_{j+1} - \lambda_j|\theta}.$$

Pour $J = 9$,

$$\begin{aligned} & \int_0^\infty \dots \int_0^\infty \theta e^{-\theta\lambda_1 - \sum_{j=1}^8 |\lambda_{j+1} - \lambda_j|\theta} d\lambda_1 \dots d\lambda_9 \\ &= \frac{1}{\theta^8} + \frac{4}{\theta^8} + \frac{35}{4\theta^8} + \frac{55}{4\theta^8} + \frac{275}{16\theta^8} + \frac{143}{8\theta^8} + \frac{1001}{64\theta^8} + \frac{715}{64\theta^8} + \frac{715}{128\theta^8} = \frac{12155}{128\theta^8}, \end{aligned}$$

et la densité conjointe de $\lambda = (\lambda_1, \dots, \lambda_9)$ est donnée par

$$\begin{aligned} \pi(\lambda_1, \dots, \lambda_9 \mid \theta) &= \frac{\theta e^{-\theta\lambda_1 - \sum_{j=1}^8 |\lambda_{j+1} - \lambda_j|\theta}}{\int_0^\infty \dots \int_0^\infty \theta e^{-\theta\lambda_1 - \sum_{j=1}^8 |\lambda_{j+1} - \lambda_j|\theta} d\lambda_1 \dots d\lambda_9} \\ &= \frac{128}{12155} \theta^9 e^{-\theta\lambda_1 - \sum_{j=1}^8 |\lambda_{j+1} - \lambda_j|\theta}. \end{aligned}$$

Finalement, pour $J = 10$,

$$\begin{aligned} & \int_0^\infty \dots \int_0^\infty \theta e^{-\theta\lambda_1 - \sum_{j=1}^9 |\lambda_{j+1} - \lambda_j|\theta} d\lambda_1 \dots d\lambda_{10} \\ &= \frac{1}{\theta^9} + \frac{9}{2\theta^9} + \frac{11}{\theta^9} + \frac{77}{4\theta^9} + \frac{429}{16\theta^9} + \frac{1001}{32\theta^9} + \frac{1001}{32\theta^9} + \frac{429}{16\theta^9} + \frac{2431}{128\theta^9} + \frac{2431}{256\theta^9} = \frac{46189}{256\theta^9}, \end{aligned}$$

et la densité conjointe de $\lambda = (\lambda_1, \dots, \lambda_{10})$ est donné par

$$\pi(\lambda_1, \dots, \lambda_{10} \mid \theta) = \frac{\theta e^{-\theta\lambda_1 - \sum_{j=1}^9 |\lambda_{j+1} - \lambda_j|\theta}}{\int_0^\infty \dots \int_0^\infty \theta e^{-\theta\lambda_1 - \sum_{j=1}^9 |\lambda_{j+1} - \lambda_j|\theta} d\lambda_1 \dots d\lambda_{10}} = \frac{256}{46189} \theta^{10} e^{-\theta\lambda_1 - \sum_{j=1}^9 |\lambda_{j+1} - \lambda_j|\theta}.$$

En général, on peut écrire la densité conjointe de $\lambda \in \mathbb{R}^J$ comme

$$\pi(\lambda \mid \theta_1) = K_J \theta_1^J \exp \left\{ -\theta_1 \lambda_1 - \sum_{j=1}^{J-1} |\lambda_{j+1} - \lambda_j| \theta_1 \right\}, \quad (\text{A.0.1})$$

où K_J est une constante qui change en fonction de J . Les mêmes résultats ci-dessus s'appliquent pour trouver les densités conjointes de α , ainsi, on a que pour $\alpha \in \mathbb{R}^J$

$$\pi(\alpha \mid \theta_2) = K_J \theta_2^J \exp \left\{ -\theta_2 \alpha_1 - \sum_{j=1}^{J-1} |\alpha_{j+1} - \alpha_j| \theta_2 \right\}. \quad (\text{A.0.2})$$

Annexe B

CALCUL DU GRADIENT ET DE LA MATRICE HESSIENNE DE LA FONCTION $G(\cdot)$

B.1. LE GRADIENT

Le gradient de la fonction $g(\vartheta)$ est formée des composantes suivantes,

$$\nabla_{g(\vartheta)} = \begin{pmatrix} \frac{\partial g(\vartheta)}{\partial \alpha} \\ \frac{\partial g(\vartheta)}{\partial \lambda} \\ \frac{\partial g(\vartheta)}{\partial \beta} \end{pmatrix},$$

dont nous spécifions le calcul séparément,

$$\begin{aligned} \frac{\partial g(\vartheta)}{\partial \alpha_1} &= \frac{\partial}{\partial \alpha_1} \ln(L(\vartheta | D_n)) + \frac{\partial}{\partial \alpha_1} \ln(\psi(\vartheta | X^n, Y_0^n, a_0)) \\ &= \sum_{i=1}^n \left\{ \delta_{i1} \nu_i \left(\frac{1}{\alpha_1} + \ln(y_i) \right) - \delta_{i1} \exp(\mathbf{x}'_i \boldsymbol{\beta}) \lambda_1 (y_i^{\alpha_1} \ln(y_i)) - \sum_{g=2}^J \delta_{ig} \exp(\mathbf{x}'_i \boldsymbol{\beta}) \lambda_1 (s_1^{\alpha_1} \ln(s_1)) \right\} \\ &\quad + \frac{na_0}{\alpha_1} + na_0 \ln(y_0) - \sum_{i=1}^n a_0 \exp(\mathbf{x}'_i \boldsymbol{\beta}) \lambda_1 y_0^{\alpha_1} \ln(y_0), \end{aligned}$$

$$\begin{aligned} \frac{\partial g(\vartheta)}{\partial \alpha_k} &= \frac{\partial}{\partial \alpha_k} \ln(L(\vartheta | D_n)) + \frac{\partial}{\partial \alpha_k} \ln(\psi(\vartheta | X^n, Y_0^n, a_0)) \\ &= \sum_{i=1}^n \left\{ \delta_{ik} \nu_i \left(\frac{1}{\alpha_k} + \ln(y_i) \right) - \delta_{ik} \exp(\mathbf{x}'_i \boldsymbol{\beta}) \lambda_k [y_i^{\alpha_k} \ln(y_i) - s_{k-1}^{\alpha_k} \ln(s_{k-1})] \right. \\ &\quad \left. - \sum_{g=k+1}^J \delta_{ig} \exp(\mathbf{x}'_i \boldsymbol{\beta}) \lambda_k [s_k^{\alpha_k} \ln(s_k) - s_{k-1}^{\alpha_k} \ln(s_{k-1})] \right\}, \quad \text{pour } k > 1. \end{aligned}$$

$$\frac{\partial g(\vartheta)}{\partial \lambda_1} = \frac{\partial}{\partial \lambda_1} \ln(L(\vartheta | D_n)) + \frac{\partial}{\partial \lambda_1} \ln(\psi(\vartheta | X^n, Y_0^n, a_0))$$

$$= \sum_{i=1}^n \left[\frac{\delta_{i1}\nu_i}{\lambda_1} - \delta_{i1} \exp(\mathbf{x}'_i \boldsymbol{\beta}) y_i^{\alpha_1} - \sum_{g=2}^J \delta_{ig} \exp(\mathbf{x}'_i \boldsymbol{\beta}) s_1^{\alpha_1} \right] + \frac{na_0}{\lambda_1} - \sum_{i=1}^n a_0 \exp(\mathbf{x}'_i \boldsymbol{\beta}) y_0^{\alpha_1}$$

$$\begin{aligned} \frac{\partial g(\vartheta)}{\partial \lambda_k} &= \frac{\partial}{\partial \lambda_k} \ln(L(\vartheta | D_n)) + \frac{\partial}{\partial \lambda_k} \ln(\psi(\vartheta | X^n, Y_0^n, a_0)) \\ &= \sum_{i=1}^n \left[\frac{\delta_{ik}\nu_i}{\lambda_k} - \delta_{ik} \exp(\mathbf{x}'_i \boldsymbol{\beta}) (y_i^{\alpha_k} - s_{k-1}^{\alpha_k}) - \sum_{g=k+1}^J \delta_{ig} \exp(\mathbf{x}'_i \boldsymbol{\beta}) (s_k^{\alpha_k} - s_{k-1}^{\alpha_k}) \right], \quad \text{pour } k > 1. \end{aligned}$$

$$\begin{aligned} \frac{\partial g(\vartheta)}{\partial \boldsymbol{\beta}} &= \frac{\partial}{\partial \boldsymbol{\beta}} \ln(L(\vartheta | D_n)) + \frac{\partial}{\partial \boldsymbol{\beta}} \ln(\psi(\vartheta | X^n, Y_0^n, a_0)) \\ &= \sum_{j=1}^J \sum_{i=1}^n \left\{ \delta_{ij}\nu_i \mathbf{x}'_i - \delta_{ij} \exp(\mathbf{x}'_i \boldsymbol{\beta}) \mathbf{x}'_i \left[\lambda_j (y_i^{\alpha_j} - s_{j-1}^{\alpha_j}) + \sum_{g=1}^{j-1} \lambda_g (s_g^{\alpha_g} - s_{g-1}^{\alpha_g}) \right] \right\} \\ &\quad + \sum_{i=1}^n a_0 \mathbf{x}'_i - a_0 \lambda_1 y_0^{\alpha_1} \exp(\mathbf{x}'_i \boldsymbol{\beta}) \mathbf{x}'_i \end{aligned}$$

B.2. LE MATRICE HESSIENNE

La matrice hessienne est donnée par,

$$\begin{pmatrix} \frac{\partial^2 g(\vartheta)}{\partial \alpha_1^2} & \frac{\partial^2 g(\vartheta)}{\partial \alpha \partial \lambda} & \frac{\partial^2 g(\vartheta)}{\partial \alpha \partial \boldsymbol{\beta}} \\ \frac{\partial^2 g(\vartheta)}{\partial \lambda \partial \alpha} & \frac{\partial^2 g(\vartheta)}{\partial \lambda^2} & \frac{\partial^2 g(\vartheta)}{\partial \lambda \partial \boldsymbol{\beta}} \\ \frac{\partial^2 g(\vartheta)}{\partial \boldsymbol{\beta} \partial \alpha} & \frac{\partial^2 g(\vartheta)}{\partial \boldsymbol{\beta} \partial \lambda} & \frac{\partial^2 g(\vartheta)}{\partial \boldsymbol{\beta}^2} \end{pmatrix}$$

le calcul des éléments de chaque bloc sont présentés dans cette section.

$$\begin{aligned} \frac{\partial^2 g(\vartheta)}{\partial \alpha_1^2} &= \frac{\partial^2}{\partial \alpha_1^2} \ln(L(\vartheta | D_n)) + \frac{\partial^2}{\partial \alpha_1^2} \ln(\psi(\vartheta | X^n, Y_0^n, a_0)) \\ &= \sum_{i=1}^n \left\{ -\frac{\delta_{i1}\nu_i}{\alpha_1^2} - \delta_{i1} \exp(\mathbf{x}'_i \boldsymbol{\beta}) \lambda_1 y_i^{\alpha_1} \ln^2(y_i) - \sum_{g=2}^J \delta_{ig} \exp(\mathbf{x}'_i \boldsymbol{\beta}) \lambda_1 s_1^{\alpha_1} \ln^2(s_1) \right\} \\ &\quad - \frac{na_0}{\alpha_1^2} - \sum_{i=1}^n a_0 \exp(\mathbf{x}'_i \boldsymbol{\beta}) \lambda_1 y_0^{\alpha_1} \ln^2(y_0) \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 g(\vartheta)}{\partial \alpha_k^2} &= \frac{\partial^2}{\partial \alpha_k^2} \ln(L(\vartheta | D_n)) + \frac{\partial^2}{\partial \alpha_k^2} \ln(\psi(\vartheta | X^n, Y_0^n, a_0)) \\ &= \sum_{i=1}^n \left\{ -\frac{\delta_{ik}\nu_i}{\alpha_k^2} - \delta_{ik} \exp(\mathbf{x}'_i \boldsymbol{\beta}) \lambda_k [y_i^{\alpha_k} \ln^2(y_i) - s_{k-1}^{\alpha_k} \ln^2(s_{k-1})] - \right. \end{aligned}$$

$$\left. \sum_{g=k+1}^J \delta_{ig} \exp(\mathbf{x}'_i \boldsymbol{\beta}) \lambda_k [s_k^{\alpha_k} \ln^2(s_k) - s_{k-1}^{\alpha_k} \ln^2(s_{k-1})] \right\} \text{ pour } k > 1$$

$$\begin{aligned} \frac{\partial g(\vartheta)}{\partial \alpha_1 \partial \lambda_1} &= \frac{\partial}{\partial \alpha_1 \partial \lambda_1} (\ln(L(\vartheta | D_n))) + \frac{\partial}{\partial \alpha_1 \partial \lambda_1} (\ln(\psi(\vartheta | X^n, Y_0^n, a_0))) \\ &= \sum_{i=1}^n \left\{ -\delta_{i1} \exp(\mathbf{x}'_i \boldsymbol{\beta}) (y_i^{\alpha_1} \ln(y_i)) - \sum_{g=2}^J \delta_{ig} \exp(\mathbf{x}'_i \boldsymbol{\beta}) (s_1^{\alpha_1} \ln(s_1)) \right\} \\ &\quad - \sum_{i=1}^n a_0 \exp(\mathbf{x}'_i \boldsymbol{\beta}) y_0^{\alpha_1} \ln(y_0) \end{aligned}$$

$$\begin{aligned} \frac{\partial g(\vartheta)}{\partial \alpha_k \partial \lambda_k} &= \frac{\partial^2}{\partial \alpha_k \partial \lambda_k} \ln(L(\vartheta | D_n)) + \frac{\partial^2}{\partial \alpha_k \partial \lambda_k} (\ln(\psi(\vartheta | X^n, Y_0^n, a_0))) \\ &= \sum_{i=1}^n \left\{ -\delta_{ik} \exp(\mathbf{x}'_i \boldsymbol{\beta}) [y_i^{\alpha_k} \ln(y_i) - s_{k-1}^{\alpha_k} \ln(s_{k-1})] \right. \\ &\quad \left. - \sum_{g=k+1}^J \delta_{ig} \exp(\mathbf{x}'_i \boldsymbol{\beta}) [s_k^{\alpha_k} \ln(s_k) - s_{k-1}^{\alpha_k} \ln(s_{k-1})] \right\} \text{ pour } k > 1 \end{aligned}$$

$$\begin{aligned} \frac{\partial g(\vartheta)}{\partial \alpha_1 \partial \boldsymbol{\beta}} &= \frac{\partial}{\partial \alpha_1 \partial \boldsymbol{\beta}} \ln(L(\vartheta | D_n)) + \frac{\partial}{\partial \alpha_1 \partial \boldsymbol{\beta}} (\ln(\psi(\vartheta | X^n, Y_0^n, a_0))) \\ &= \sum_{i=1}^n \left\{ -\delta_{i1} \exp(\mathbf{x}'_i \boldsymbol{\beta}) \mathbf{x}'_i \lambda_1 (y_i^{\alpha_1} \ln(y_i)) - \sum_{g=2}^J \delta_{ig} \exp(\mathbf{x}'_i \boldsymbol{\beta}) \mathbf{x}'_i \lambda_1 (s_1^{\alpha_1} \ln(s_1)) \right\} \\ &\quad - \sum_{i=1}^n a_0 \exp(\mathbf{x}'_i \boldsymbol{\beta}) \mathbf{x}'_i \lambda_1 y_0^{\alpha_1} \ln(y_0) \end{aligned}$$

$$\begin{aligned} \frac{\partial g(\vartheta)}{\partial \alpha_k \partial \boldsymbol{\beta}} &= \frac{\partial}{\partial \alpha_k \partial \boldsymbol{\beta}} \ln(L(\vartheta | D_n)) + \frac{\partial}{\partial \alpha_k \partial \boldsymbol{\beta}} (\ln(\psi(\vartheta | X^n, Y_0^n, a_0))) \\ &= \sum_{i=1}^n \left\{ -\delta_{ik} \exp(\mathbf{x}'_i \boldsymbol{\beta}) \mathbf{x}'_i \lambda_k [y_i^{\alpha_k} \ln(y_i) - s_{k-1}^{\alpha_k} \ln(s_{k-1})] \right. \\ &\quad \left. - \sum_{g=k+1}^J \delta_{ig} \exp(\mathbf{x}'_i \boldsymbol{\beta}) \mathbf{x}'_i \lambda_k [s_k^{\alpha_k} \ln(s_k) - s_{k-1}^{\alpha_k} \ln(s_{k-1})] \right\} \end{aligned}$$

$$\frac{\partial^2 g(\vartheta)}{\partial \lambda_1^2} = \frac{\partial^2}{\partial \lambda_1^2} \ln(L(\vartheta | D_n)) + \frac{\partial^2}{\partial \lambda_1^2} \ln(\psi(\vartheta | X^n, Y_0^n, a_0)) = \sum_{i=1}^n -\frac{\delta_{i1} \nu_i}{\lambda_1^2} - \frac{na_0}{\lambda_1^2}$$

$$\frac{\partial^2 g(\vartheta)}{\partial \lambda_k^2} = \frac{\partial^2}{\partial \lambda_k^2} \ln(L(\vartheta | D_n)) + \frac{\partial^2}{\partial \lambda_k^2} \ln(\psi(\vartheta | X^n, Y_0^n, a_0)) = \sum_{i=1}^n -\frac{\delta_{ik} \nu_i}{\lambda_k^2}$$

$$\begin{aligned} \frac{\partial g(\vartheta)}{\partial \lambda_1 \partial \beta} &= \frac{\partial}{\partial \lambda_1 \partial \beta} \ln(L(\vartheta | D_n)) + \frac{\partial}{\partial \lambda_1 \partial \beta} \ln(\psi(\vartheta | X^n, Y_0^n, a_0)) \\ &= \sum_{i=1}^n \left[-\delta_{i1} \exp(\mathbf{x}'_i \beta) \mathbf{x}'_i y_i^{\alpha_1} - \sum_{g=2}^J \delta_{ig} \exp(\mathbf{x}'_i \beta) \mathbf{x}'_i s_1^{\alpha_1} \right] - \sum_{i=1}^n a_0 \exp(\mathbf{x}'_i \beta) \mathbf{x}'_i y_0^{\alpha_1} \end{aligned}$$

$$\begin{aligned} \frac{\partial g(\vartheta)}{\partial \lambda_k \partial \beta} &= \frac{\partial}{\partial \lambda_k \partial \beta} \ln(L(\vartheta | D_n)) + \frac{\partial}{\partial \lambda_k \partial \beta} (\ln(\psi(\vartheta | X^n, Y_0^n, a_0))) \\ &= \sum_{i=1}^n \left[-\delta_{ik} \exp(\mathbf{x}'_i \beta) \mathbf{x}'_i (y_i^{\alpha_k} - s_{k-1}^{\alpha_k}) - \sum_{g=k+1}^J \delta_{ig} \exp(\mathbf{x}'_i \beta) \mathbf{x}'_i (s_k^{\alpha_k} - s_{k-1}^{\alpha_k}) \right] \quad \text{pour } k > 1 \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 g(\vartheta)}{\partial \beta^2} &= \frac{\partial^2}{\partial \beta^2} \ln(L(\vartheta | D_n)) + \frac{\partial^2}{\partial \beta^2} \ln(\psi(\vartheta | X^n, Y_0^n, a_0)) \\ &= \sum_{j=1}^J \sum_{i=1}^n \left\{ -\delta_{ij} \exp(\mathbf{x}'_i \beta) \mathbf{x}'_i \mathbf{x}_i \left[\lambda_j (y_i^{\alpha_j} - s_{j-1}^{\alpha_j}) + \sum_{g=1}^{j-1} \lambda_g (s_g^{\alpha_g} - s_{g-1}^{\alpha_g}) \right] \right\} \\ &\quad + \sum_{i=1}^n -a_0 \lambda_1 y_0^{\alpha_1} \exp(\mathbf{x}'_i \beta) \mathbf{x}'_i \mathbf{x}_i \end{aligned}$$

Annexe C

CALCUL DU GRADIENT ET LA MATRICE HESSIENNE DE LA FONCTION \mathcal{L} , DOUCE

Dans la section 3.5.3 on a introduit une approximation douce de la fonction $\ell(\alpha, \lambda)$. Pour simplifier la notation, considérons

$$\varphi(\theta) = \theta_1 + q + \sum_{j=1}^{J-1} \xi(\theta_{j+1} - \theta_j),$$

où q peut être remplacé par q_1 ou q_2 et θ_j par α_j ou encore λ_j . Rappelons que la fonction $\xi(\cdot)$ est définie comme

$$\xi(w) = \kappa^{-1} \{ \log(1 + \cosh(\kappa w)) + \log(2) \}, \quad w \in \mathbb{R}, \quad \kappa \in \mathbb{R}^+,$$

Soit

$$\gamma = \frac{K_J^2 q_1^{p_1} q_2^{p_2} \Gamma(J + p_1) \Gamma(J + p_2)}{\Gamma(p_1) \Gamma(p_2)}.$$

Ainsi, on peut réécrire la fonction $\mathcal{L}(\alpha, \lambda)$ comme

$$\mathcal{L}(\alpha, \lambda) = \gamma \left[\varphi(\lambda)^{-(J+p_1)} \varphi(\alpha)^{-(J+p_2)} \right],$$

remarquons que

$$\begin{aligned} \frac{\partial \varphi(\alpha)}{\partial \alpha_1} &= 1 - \xi'(\alpha_2 - \alpha_1) \\ \frac{\partial \varphi(\alpha)}{\partial \alpha_k} &= \xi'(\alpha_k - \alpha_{k-1}) - \mathbb{1}_{\{k < J\}} \xi'(\alpha_{k+1} - \alpha_k), \quad k \neq 1. \end{aligned}$$

$$\begin{aligned} \frac{\partial \varphi(\lambda)}{\partial \lambda_1} &= 1 - \xi'(\lambda_2 - \lambda_1) \\ \frac{\partial \varphi(\lambda)}{\partial \lambda_k} &= \xi'(\lambda_k - \lambda_{k-1}) - \mathbb{1}_{\{k < J\}} \xi'(\lambda_{k+1} - \lambda_k), \quad k \neq 1. \end{aligned}$$

Ainsi, on a que

$$\begin{aligned}\frac{\partial \mathcal{L}(\alpha, \lambda)}{\partial \alpha_1} &= -\gamma \left[\varphi(\lambda)^{-(J+p_1)} (J+p_2) \varphi(\alpha)^{-(J+p_2+1)} \right] [1 - \xi'(\alpha_2 - \alpha_1)] \\ &= -\mathcal{L}(\alpha, \lambda) (J+p_2) \varphi(\alpha)^{-1} [1 - \xi'(\alpha_2 - \alpha_1)]\end{aligned}$$

$$\begin{aligned}\frac{\partial \mathcal{L}(\alpha, \lambda)}{\partial \lambda_1} &= -\gamma \left[\varphi(\alpha)^{-(J+p_2)} (J+p_1) \varphi(\lambda)^{-(J+p_1+1)} \right] [1 - \xi'(\lambda_2 - \lambda_1)] \\ &= -\mathcal{L}(\alpha, \lambda) (J+p_1) \varphi(\lambda)^{-1} [1 - \xi'(\lambda_2 - \lambda_1)]\end{aligned}$$

Pour $k > 1$,

$$\begin{aligned}\frac{\partial \mathcal{L}(\alpha, \lambda)}{\partial \alpha_k} &= -\gamma \left[\varphi(\lambda)^{-(J+p_1)} (J+p_2) \varphi(\alpha)^{-(J+p_2+1)} \right] \left[\xi'(\alpha_k - \alpha_{k-1}) - \mathbb{1}_{\{k < J\}} \xi'(\alpha_{k+1} - \alpha_k) \right] \\ &= -\mathcal{L}(\alpha, \lambda) (J+p_2) \varphi(\alpha)^{-1} \left[\xi'(\alpha_k - \alpha_{k-1}) - \mathbb{1}_{\{k < J\}} \xi'(\alpha_{k+1} - \alpha_k) \right]\end{aligned}$$

$$\begin{aligned}\frac{\partial \mathcal{L}(\alpha, \lambda)}{\partial \lambda_k} &= -\gamma \left[\varphi(\alpha)^{-(J+p_2)} (J+p_1) \varphi(\lambda)^{-(J+p_1+1)} \right] \left[\xi'(\lambda_k - \lambda_{k-1}) - \mathbb{1}_{\{k < J\}} \xi'(\lambda_{k+1} - \lambda_k) \right] \\ &= -\mathcal{L}(\alpha, \lambda) (J+p_1) \varphi(\lambda)^{-1} \left[\xi'(\lambda_k - \lambda_{k-1}) - \mathbb{1}_{\{k < J\}} \xi'(\lambda_{k+1} - \lambda_k) \right].\end{aligned}$$

Maintenant, pour la construction de la matrice hessienne et afin de simplifier la notation et les calculs, nous écrivons

$$\Delta \xi(\lambda)_k = \begin{cases} \xi'(\lambda_k - \lambda_{k-1}) - \mathbb{1}_{\{k < J\}} \xi'(\lambda_{k+1} - \lambda_k), & \text{si } k > 1 \\ 1 - \xi'(\lambda_2 - \lambda_1), & \text{si } k = 1, \end{cases}$$

et

$$\Delta \xi(\alpha)_k = \begin{cases} \xi'(\alpha_k - \alpha_{k-1}) - \mathbb{1}_{\{k < J\}} \xi'(\alpha_{k+1} - \alpha_k), & \text{si } k > 1 \\ 1 - \xi'(\alpha_2 - \alpha_1), & \text{si } k = 1, \end{cases}$$

pour dénoter les premières dérivés de $\varphi(\lambda)$ et $\varphi(\alpha)$ par rapport à λ_k et α_k respectivement.

Remarquons également que

$$\begin{aligned}\frac{\partial^2 \varphi(\alpha)}{\partial \alpha_1^2} &= \xi''(\alpha_2 - \alpha_1) \\ \frac{\partial^2 \varphi(\alpha)}{\partial \alpha_k^2} &= \xi''(\alpha_k - \alpha_{k-1}) + \mathbb{1}_{\{k < J\}} \xi''(\alpha_{k+1} - \alpha_k) \\ \frac{\partial^2 \varphi(\alpha)}{\partial \alpha_k \partial \alpha_j} &= - \left[\mathbb{1}_{\{j=k-1\}} \xi''(\alpha_k - \alpha_{k-1}) + \mathbb{1}_{\{j=k+1, k < J\}} \xi''(\alpha_{k+1} - \alpha_k) \right]\end{aligned}$$

$$\frac{\partial^2 \varphi(\lambda)}{\partial \lambda_1^2} = \xi''(\lambda_2 - \lambda_1)$$

$$\frac{\partial^2 \varphi(\lambda)}{\partial \lambda_k^2} = \xi''(\lambda_k - \lambda_{k-1}) + \mathbb{1}_{\{k < J\}} \xi''(\lambda_{k+1} - \lambda_k)$$

$$\frac{\partial^2 \varphi(\lambda)}{\partial \lambda_k \partial \lambda_j} = - \left[\mathbb{1}_{\{j=k-1\}} \xi''(\lambda_k - \lambda_{k-1}) + \mathbb{1}_{\{j=k+1, k < J\}} \xi''(\lambda_{k+1} - \lambda_k) \right].$$

Ainsi, pour les deuxièmes dérivés, on a

$$\begin{aligned} \frac{\partial^2 \mathcal{L}(\alpha, \lambda)}{\partial \alpha_1^2} &= - \left[\frac{\partial}{\partial \alpha_1} \mathcal{L}(\alpha, \lambda) \right] \Delta \xi(\alpha)_1 (J + p_2) \varphi(\alpha)^{-1} + \mathcal{L}(\alpha, \lambda) (J + p_2) \varphi(\alpha)^{-2} [\Delta \xi(\alpha)_1]^2 \\ &\quad - \mathcal{L}(\alpha, \lambda) (J + p_2) \varphi(\alpha)^{-1} \xi''(\alpha_2 - \alpha_1) \\ &= \mathcal{L}(\alpha, \lambda) (J + p_2) \varphi(\alpha)^{-1} \left[(J + p_2 + 1) \varphi(\alpha)^{-1} [\Delta \xi(\alpha)_1]^2 - \xi''(\alpha_2 - \alpha_1) \right] \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 \mathcal{L}(\alpha, \lambda)}{\partial \alpha_k \partial \alpha_1} &= - \left[\frac{\partial}{\partial \alpha_k} \mathcal{L}(\alpha, \lambda) \right] \Delta \xi(\alpha)_1 (J + p_2) \varphi(\alpha)^{-1} + \mathcal{L}(\alpha, \lambda) (J + p_2) \varphi(\alpha)^{-2} \Delta \xi(\alpha)_1 \Delta \xi(\alpha)_k \\ &\quad + \mathcal{L}(\alpha, \lambda) (J + p_2) \varphi(\alpha)^{-1} \mathbb{1}_{\{k=2\}} \xi''(\alpha_2 - \alpha_1) \\ &= \mathcal{L}(\alpha, \lambda) (J + p_2) \varphi(\alpha)^{-1} \left[(J + p_2 + 1) \varphi(\alpha)^{-1} \Delta \xi(\alpha)_1 \Delta \xi(\alpha)_k + \mathbb{1}_{\{k=2\}} \xi''(\alpha_2 - \alpha_1) \right] \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 \mathcal{L}(\alpha, \lambda)}{\partial \alpha_k^2} &= - \left[\frac{\partial}{\partial \alpha_k} \mathcal{L}(\alpha, \lambda) \right] \Delta \xi(\alpha)_k (J + p_2) \varphi(\alpha)^{-1} + \mathcal{L}(\alpha, \lambda) (J + p_2) \varphi(\alpha)^{-2} [\Delta \xi(\alpha)_k]^2 \\ &\quad - \mathcal{L}(\alpha, \lambda) (J + p_2) \varphi(\alpha)^{-1} \left[\xi''(\alpha_k - \alpha_{k-1}) + \mathbb{1}_{\{k < J\}} \xi''(\alpha_{k+1} - \alpha_k) \right] \\ &= \mathcal{L}(\alpha, \lambda) (J + p_2) \varphi(\alpha)^{-1} \left[(J + p_2 + 1) \varphi(\alpha)^{-1} [\Delta \xi(\alpha)_k]^2 - \xi''(\alpha_k - \alpha_{k-1}) \right. \\ &\quad \left. - \mathbb{1}_{\{k < J\}} \xi''(\alpha_{k+1} - \alpha_k) \right] \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 \mathcal{L}(\alpha, \lambda)}{\partial \alpha_j \partial \alpha_k} &= \mathcal{L}(\alpha, \lambda) (J + p_2) \varphi(\alpha)^{-1} \left[(J + p_2 + 1) \varphi(\alpha)^{-1} [\Delta \xi(\alpha)_k] [\Delta \xi(\alpha)_j] \right. \\ &\quad \left. + \mathbb{1}_{\{j=k-1, k > 1\}} \xi''(\alpha_k - \alpha_{k-1}) + \mathbb{1}_{\{j=k+1, k < J\}} \xi''(\alpha_{k+1} - \alpha_k) - \mathbb{1}_{\{j=k, k > 1\}} \xi''(\alpha_k - \alpha_{k-1}) \right. \\ &\quad \left. - \mathbb{1}_{\{j=k, k < J\}} \xi''(\alpha_{k+1} - \alpha_k) \right]. \end{aligned}$$

Les dérivés par rapport à λ son similaires

$$\begin{aligned} \frac{\partial^2 \mathcal{L}(\alpha, \lambda)}{\partial \lambda_1^2} &= - \left[\frac{\partial}{\partial \lambda_1} \mathcal{L}(\alpha, \lambda) \right] \Delta \xi(\lambda)_1 (J + p_1) \varphi(\lambda)^{-1} + \mathcal{L}(\alpha, \lambda) (J + p_1) \varphi(\lambda)^{-2} [\Delta \xi(\lambda)_1]^2 \\ &\quad - \mathcal{L}(\alpha, \lambda) (J + p_1) \varphi(\lambda)^{-1} \xi''(\lambda_2 - \lambda_1) \\ &= \mathcal{L}(\alpha, \lambda) (J + p_1) \varphi(\lambda)^{-1} \left[(J + p_1 + 1) \varphi(\lambda)^{-1} [\Delta \xi(\lambda)_1]^2 - \xi''(\lambda_2 - \lambda_1) \right] \end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 \mathcal{L}(\alpha, \lambda)}{\partial \lambda_k \partial \lambda_1} &= - \left[\frac{\partial}{\partial \lambda_k} \mathcal{L}(\alpha, \lambda) \right] \Delta \xi(\lambda)_1 (J + p_1) \varphi(\lambda)^{-1} + \mathcal{L}(\alpha, \lambda) (J + p_1) \varphi(\lambda)^{-2} \Delta \xi(\lambda)_1 \Delta \xi(\lambda)_k \\
&\quad + \mathcal{L}(\alpha, \lambda) (J + p_1) \varphi(\lambda)^{-1} \mathbb{1}_{\{k=2\}} \xi''(\lambda_2 - \lambda_1) \\
&= \mathcal{L}(\alpha, \lambda) (J + p_1) \varphi(\lambda)^{-1} \left[(J + p_1 + 1) \varphi(\lambda)^{-1} \Delta \xi(\lambda)_1 \Delta \xi(\lambda)_k + \mathbb{1}_{\{k=2\}} \xi''(\lambda_2 - \lambda_1) \right]
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 \mathcal{L}(\alpha, \lambda)}{\partial \lambda_k^2} &= - \left[\frac{\partial}{\partial \lambda_k} \mathcal{L}(\alpha, \lambda) \right] \Delta \xi(\lambda)_k (J + p_1) \varphi(\lambda)^{-1} + \mathcal{L}(\alpha, \lambda) (J + p_1) \varphi(\lambda)^{-2} [\Delta \xi(\lambda)_k]^2 \\
&\quad - \mathcal{L}(\alpha, \lambda) (J + p_1) \varphi(\lambda)^{-1} \left[\xi''(\lambda_k - \lambda_{k-1}) + \mathbb{1}_{\{k < J\}} \xi''(\lambda_{k+1} - \lambda_k) \right] \\
&= \mathcal{L}(\alpha, \lambda) (J + p_1) \varphi(\lambda)^{-1} \left[(J + p_1 + 1) \varphi(\lambda)^{-1} [\Delta \xi(\lambda)_k]^2 - \xi''(\lambda_k - \lambda_{k-1}) \right. \\
&\quad \left. - \mathbb{1}_{\{k < J\}} \xi''(\lambda_{k+1} - \lambda_k) \right]
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 \mathcal{L}(\alpha, \lambda)}{\partial \lambda_j \partial \lambda_k} &= \mathcal{L}(\alpha, \lambda) (J + p_1) \varphi(\lambda)^{-1} \left[(J + p_1 + 1) \varphi(\lambda)^{-1} [\Delta \xi(\lambda)_k] [\Delta \xi(\lambda)_j] \right. \\
&\quad + \mathbb{1}_{\{j=k-1, k > 1\}} \xi''(\lambda_k - \lambda_{k-1}) + \mathbb{1}_{\{j=k+1, k < J\}} \xi''(\lambda_{k+1} - \lambda_k) - \mathbb{1}_{\{j=k, k > 1\}} \xi''(\lambda_k - \lambda_{k-1}) \\
&\quad \left. - \mathbb{1}_{\{j=k, k < J\}} \xi''(\lambda_{k+1} - \lambda_k) \right].
\end{aligned}$$

Finalement, on a que

$$\frac{\partial^2 \mathcal{L}(\alpha, \lambda)}{\partial \alpha_j \partial \lambda_k} = \mathcal{L}(\alpha, \lambda) (J + p_1) (J + p_2) \varphi(\lambda)^{-1} \varphi(\alpha)^{-1} [\Delta \xi(\lambda)_k] [\Delta \xi(\alpha)_j]$$

Annexe D

OBTENTION DES INTÉGRALES DANS LA LOI PRÉDICTIVE A *POSTERIORI*

L'équation (D.0.1) ci-dessus nous permet d'approximer la distribution de prévision *a posteriori* pour une observation future y_{n+1} .

$$\hat{\pi}(y_{n+1} | Y^n, X^n, \mathbf{x}_{n+1}) = \sum_{m=1}^M \frac{\pi(Y^n, y_{n+1} | X^n, \mathbf{x}_{n+1}, [\rho_{n,m}, \mathbf{b}_{n+1,m}])}{\pi(Y^n | X^n, \mathbf{x}_{n+1}, [\rho_{n,m}, \mathbf{b}_{n+1,m}])}. \quad (\text{D.0.1})$$

Afin de donner une expression explicite pour le résultat des intégrales contenues dans cette équation, on introduit les expressions suivantes. Pour une partition donnée, $[\rho_{n,m}, \mathbf{b}_{n+1,m}]$, on va dénoter par k le grappe dans la partition qui contient le point \mathbf{x}_{n+1} , et j est tel que $s_{j-1} < y_{n+1} \leq s_j$ (j est donc, l'intervalle dans lequel la nouvelle observation présente l'évènement d'intérêt ou est censuré). Par simplicité, nous n'utiliserons pas l'indice k lorsque la liaison est sous-entendu. Des expressions (3.4.2) et (3.5.4), nous avons

$$L(\vartheta | X^n, \mathbf{x}_{n+1}, Y^n, y_{n+1}, [\rho_n, \mathbf{b}_{n+1}]) = L(\vartheta | D_n, \rho_n) \left(\alpha_j \lambda_j y_{n+1}^{\alpha_j - 1} \exp(\mathbf{x}'_{n+1} \boldsymbol{\beta}) \right)^{\nu_{n+1}} \times \exp \left[- \exp(\mathbf{x}'_{n+1} \boldsymbol{\beta}) \left(\lambda_j (y_{n+1}^{\alpha_j} - S_{j-1}^{\alpha_j}) + \sum_{g=1}^{j-1} \lambda_g (S_g^{\alpha_g} - S_{g-1}^{\alpha_g}) \right) \right].$$

De façon similaire,

$$\psi(\vartheta | X^n, \mathbf{x}_{n+1}, Y_0^n, a_0, [\rho_n, \mathbf{b}_{n+1}]) = \psi(\vartheta | X^n, Y_0^n, a_0, \rho_n) \left(\alpha_1 \lambda_1 y_0^{\alpha_1 - 1} \exp(\mathbf{x}'_{n+1} \boldsymbol{\beta}) \right)^{a_0} \times \exp(-a_0 \exp(\mathbf{x}'_{n+1} \boldsymbol{\beta}) \lambda_1 y_0^{\alpha_1}).$$

On définit également la fonction f_{n+1} comme le produit des parties ajoutés due à l'observation \mathbf{x}_{n+1} et aux *liens*, \mathbf{b}_{n+1} des deux expressions ci-haut, c'est-à-dire

$$f_{n+1}(y | \vartheta, y_0, \mathbf{x}, \nu, a_0, [\rho_n, \mathbf{b}_{n+1}]) = \left(\alpha_j \lambda_j e^{\mathbf{x}' \boldsymbol{\beta}} y^{\alpha_j - 1} \right)^\nu \exp \left\{ - \lambda_j e^{\mathbf{x}' \boldsymbol{\beta}} y^{\alpha_j} \right\} \left(\alpha_1 \lambda_1 y_0^{\alpha_1 - 1} e^{\mathbf{x}' \boldsymbol{\beta}} \right)^{a_0} \times \quad (\text{D.0.2})$$

$$\exp \left\{ e^{x'\beta} \left[-a_0 \lambda_1 y_0^{\alpha_1} + \lambda_j s_{j-1}^{\alpha_j} - \sum_{g=1}^{j-1} \lambda_g (s_g^{\alpha_g} - s_{g-1}^{\alpha_g}) \right] \right\},$$

si $s_{j-1} < y \leq s_j$, $j > 1$ et

$$f_{n+1}(y \mid \vartheta, y_0, \mathbf{x}, \nu, a_0, [\rho_n, \mathbf{b}_{n+1}]) = \left(\alpha_1 \lambda_1 e^{x'\beta} y^{\alpha_1-1} \right)^\nu \exp \left\{ -\lambda_1 e^{x'\beta} y^{\alpha_1} \right\} \left(\alpha_1 \lambda_1 y_0^{\alpha_1-1} e^{x'\beta} \right)^{a_0} \times \exp \left\{ -a_0 \lambda_1 e^{x'\beta} y_0^{\alpha_1} \right\}, \quad (\text{D.0.3})$$

si $j = 1$. On aura ainsi que le numérateur dans l'expression désignant la loi de prévision *a posteriori*, donné par $\pi(Y^n, y_{n+1} \mid X^n, \mathbf{x}_{n+1}, \rho_{n+1})$ est égale à

$$\begin{aligned} & \int_{\vartheta} L(\vartheta \mid X^n, Y^n, \rho_n) \psi(\vartheta \mid X^n, \mathbf{x}_{n+1}, Y_0^n, a_0, \rho_n) \pi(\alpha, \lambda \mid \theta) \pi(\theta) \times \\ & f_{n+1}(y_{n+1} \mid \vartheta, y_0, \mathbf{x}_{n+1}, \nu_{n+1}, a_0, [\rho_n, \mathbf{b}_{n+1}]) d\vartheta d\theta \\ &= \int_{\vartheta} L(\vartheta \mid X^n, Y^n, \rho_n) \psi(\vartheta \mid X^n, \mathbf{x}_{n+1}, Y_0^n, a_0, \rho_n) \ell(\alpha, \lambda) \\ & f_{n+1}(y_{n+1} \mid \vartheta, y_0, \mathbf{x}_{n+1}, \nu_{n+1}, a_0, [\rho_n, \mathbf{b}_{n+1}]) d\vartheta, \end{aligned}$$

En utilisant l'approximation de Laplace, nous pouvons estimer cette dernière integrale. Nous obtenons ainsi

$$\begin{aligned} \pi(Y^n, y_{n+1} \mid X^n, \mathbf{x}_{n+1}, \rho_{n+1}) &\approx \hat{\pi}(Y^n, y_{n+1} \mid X^n, \mathbf{x}_{n+1}, \rho_{n+1}) \\ &= e^{\tilde{g}(\vartheta^*)} \frac{\ell(\vartheta_{(\alpha, \lambda)}^*)}{\mathcal{L}(\vartheta_{(\alpha, \lambda)}^*)} (2\pi)^{(d/2)} \left| -\frac{\partial^2 \tilde{g}(\vartheta^*)}{\partial \vartheta \partial \vartheta'} \right| f_{n+1}(y_{n+1} \mid \vartheta^*, y_0, \mathbf{x}_{n+1}, \nu_{n+1}, a_0, [\rho_n, \mathbf{b}_{n+1}]). \end{aligned}$$

En procédant de la même manière, nous obtenons l'approximation suivante pour le dénominateur

$$\begin{aligned} \pi(Y^n \mid X^n, \mathbf{x}_{n+1}, \rho_{n+1}) &\approx \hat{\pi}(Y^n \mid X^n, \mathbf{x}_{n+1}, \rho_{n+1}) \\ &= e^{\tilde{g}(\hat{\vartheta})} \frac{\ell(\vartheta_{(\alpha, \lambda)}^*)}{\mathcal{L}(\vartheta_{(\alpha, \lambda)}^*)} (2\pi)^{(d/2)} \left| -\frac{\partial^2 \tilde{g}(\vartheta^*)}{\partial \vartheta \partial \vartheta'} \right| \int_{y_{n+1}} f_{n+1}(y_{n+1} \mid \vartheta^*, y_0, \mathbf{x}_{n+1}, \nu_{n+1}, a_0, [\rho_n, \mathbf{b}_{n+1}]) dy_{n+1}. \end{aligned}$$

Par conséquent, la loi de prévision *a posteriori* peut être estimée comme

$$\hat{\pi}(y_{n+1} \mid Y^n, X^n, \mathbf{x}_{n+1}) = \sum_{m=1}^M \frac{f_{n+1}(y_{n+1} \mid \vartheta_{m, k(\mathbf{x}_{n+1})}^*, y_0, \mathbf{x}_{n+1}, \nu_{n+1}, a_0, [\rho_{n, m}, \mathbf{b}_{n+1, m}])}{\int_{y_{n+1}} f_{n+1}(y_{n+1} \mid \vartheta_{m, k(\mathbf{x}_{n+1})}^*, y_0, \mathbf{x}_{n+1}, \nu_{n+1}, a_0, [\rho_{n, m}, \mathbf{b}_{n+1, m}]) dy_{n+1}}, \quad (\text{D.0.4})$$

où $k(\mathbf{x}_{n+1})$ indique le grappe dans la partition, $[\rho_{n, m}, \mathbf{b}_{n+1, m}]$ où se trouve l'observation en question.

Annexe E

DÉTAILS DE LA SIMULATION

E.1. GÉNÉRATION DE DONNÉES À PARTIR D'UN MODÈLE DE RÉGRESSION DE COX AUX RISQUES PAR INTERVALLES SELON LA LOI WEIBULL

Dans le chapitre 4, nous avons présenté divers résultats produits à partir de scénarios générés artificiellement à l'aide de simulation de données. Nous détaillons ici la démarche suivie pour générer de telles données. Plus particulièrement, on décrit la démarche suivie pour générer des variables réponses y_i à partir un ensemble de covariables \mathbf{x}_i et un ensemble d'hyperparamètres. On a appliqué la méthode de la transformée inverse (DEVROYE [7]). Pour un rappel de la notation, veuillez vous référer au chapitre 3. Les détails théoriques sont les suivantes.

Dans la section 3.4, nous avons étudié les principales fonctions qui caractérisent le modèle de régression de Cox par intervalles proposé dans ce projet de recherche. La fonction de survie associé à un tel modèle est donnée par

$$S(y_i | \mathbf{x}_i, \vartheta) = \exp\left\{-H_0(y_i | \alpha, \lambda) \exp(\mathbf{x}'_i \boldsymbol{\beta})\right\},$$

Ceci implique que la fonction de répartition est

$$F(y_i | \mathbf{x}_i, \vartheta) = \frac{1}{Z} \left[1 - \exp\left\{-H_0(y_i | \alpha, \lambda) \exp(\mathbf{x}'_i \boldsymbol{\beta})\right\}\right],$$

Nous avons aussi introduit la fonction de risque cumulé de base, $H_0(y_i | \alpha, \lambda)$,

$$H_0(y_i | \alpha, \lambda) = \lambda_j (y_i^{\alpha_j} - s_{j-1}^{\alpha_j}) + \sum_{g=1}^{j-1} \lambda_g (s_g^{\alpha_g} - s_{g-1}^{\alpha_g}), \text{ pour } s_{j-1} < y_i \leq s_j.$$

Ainsi, on aura par exemple pour $0 < y < s_1$, une fonction de répartition et une fonction de répartition inverse comme suit

$$F(y | \mathbf{x}_i, \vartheta) = \frac{1}{Z} \left[1 - \exp\left\{-\lambda_1 y^{\alpha_1} \exp(\mathbf{x}'_i \boldsymbol{\beta})\right\}\right]$$

$$F^{-1}(y | \mathbf{x}_i, \vartheta) = \left[-\frac{\ln(1 - yZ)}{\lambda_1 \exp(\mathbf{x}'_i \boldsymbol{\beta})} \right]^{1/\alpha_1},$$

où Z est la constante de normalisation. Si $s_1 < y < s_2$

$$F(y | \mathbf{x}_i, \vartheta) = \frac{1}{Z} \left[1 - \exp\left\{ [\lambda_2 s_1^{\alpha_2} - \lambda_2 y^{\alpha_2} - \lambda_1 s_1^{\alpha_1}] \exp(\mathbf{x}'_i \boldsymbol{\beta}) \right\} \right]$$

$$F^{-1}(y | \mathbf{x}_i, \vartheta) = \left[\frac{(\lambda_2 s_1^{\alpha_2} - \lambda_1 s_1^{\alpha_1}) \exp(\mathbf{x}'_i \boldsymbol{\beta}) - \ln(1 - yZ)}{\lambda_2 \exp(\mathbf{x}'_i \boldsymbol{\beta})} \right]^{1/\alpha_2}.$$

D'ailleurs, si $s_2 < y < s_3$

$$F(y | \mathbf{x}_i, \vartheta) = \frac{1}{Z} \left[1 - \exp\left\{ -[\lambda_3 (y^{\alpha_3} - s_2^{\alpha_3}) + \sum_{g=1}^2 \lambda_g (s_g^{\alpha_g} - s_{g-1}^{\alpha_g})] \exp(\mathbf{x}'_i \boldsymbol{\beta}) \right\} \right]$$

$$F^{-1}(y | \mathbf{x}_i, \vartheta) = \left[\frac{[\lambda_3 s_2^{\alpha_3} - \lambda_2 (s_2^{\alpha_2} - s_1^{\alpha_2}) - \lambda_1 s_1^{\alpha_1}] \exp(\mathbf{x}'_i \boldsymbol{\beta}) - \ln(1 - yZ)}{\lambda_3 \exp(\mathbf{x}'_i \boldsymbol{\beta})} \right]^{1/\alpha_3}$$

Comme dernier exemple, si $s_3 < y < s_4$ on aura

$$F(y | \mathbf{x}_i, \vartheta) = \frac{1}{Z} \left[1 - \exp\left\{ -[\lambda_4 (y^{\alpha_4} - s_3^{\alpha_4}) + \sum_{g=1}^3 \lambda_g (s_g^{\alpha_g} - s_{g-1}^{\alpha_g})] \exp(\mathbf{x}'_i \boldsymbol{\beta}) \right\} \right]$$

$$F^{-1}(y | \mathbf{x}_i, \vartheta) = \left[\frac{[\lambda_4 s_3^{\alpha_4} - \sum_{g=1}^3 \lambda_g (s_g^{\alpha_g} - s_{g-1}^{\alpha_g})] \exp(\mathbf{x}'_i \boldsymbol{\beta}) - \ln(1 - yZ)}{\lambda_4 \exp(\mathbf{x}'_i \boldsymbol{\beta})} \right]^{1/\alpha_4}$$

En général, si $s_{j-1} < y < s_j$

$$F(y | \mathbf{x}_i, \vartheta) = \frac{1}{Z} \left[1 - \exp\left\{ -[\lambda_j (y^{\alpha_j} - s_{j-1}^{\alpha_j}) + \sum_{g=1}^{j-1} \lambda_g (s_g^{\alpha_g} - s_{g-1}^{\alpha_g})] \exp(\mathbf{x}'_i \boldsymbol{\beta}) \right\} \right]$$

$$F^{-1}(y | \mathbf{x}_i, \vartheta) = \left[\frac{[\lambda_j s_{j-1}^{\alpha_j} - \sum_{g=1}^{j-1} \lambda_g (s_g^{\alpha_g} - s_{g-1}^{\alpha_g})] \exp(\mathbf{x}'_i \boldsymbol{\beta}) - \ln(1 - yZ)}{\lambda_j \exp(\mathbf{x}'_i \boldsymbol{\beta})} \right]^{1/\alpha_j}$$

La constante de normalisation, Z est donné par

$$Z = 1 - \exp\left\{ -[\lambda_J (s_J^{\alpha_J} - s_{J-1}^{\alpha_J}) + \sum_{g=1}^{J-1} \lambda_g (s_g^{\alpha_g} - s_{g-1}^{\alpha_g})] \exp(\mathbf{x}'_i \boldsymbol{\beta}) \right\}$$

Rappelons-nous que $\lambda_1 \sim \text{Exp}(\theta_1)$ et $\alpha_1 \sim \text{Exp}(\theta_2)$. Maintenant on montre le code source contenant toutes les fonctions R utilisées dans cette partie, suivi d'un exemple de code utilisant ces fonctions pour exécuter l'une des simulations rapportées dans ce mémoire.

E.2. PARAMÈTRES DE CHEVAUCHEMENT CHOISIS POUR LA GÉNÉRATION DES JEUX DE DONNÉES ISSUS D'UN MODÈLE DE MÉLANGE GUSSIENNE

Tableau E. I. Chevauchement appliqué à la génération des population artificiels. Pour chaque valeur de K , on a généré des populations de dimension $p = 2$, $p = 5$ et $p = 10$.

Dimension de \mathbf{x}	nombre de <i>clusters</i>	chevauchement	$\bar{\omega}$	$\check{\omega}$
$p = 2$	$K3$	HO	0,0500	0,1250
		MO	0,0050	0,0125
		LO	0,0005	0,0012
	$K6$	HO	0,0100	0,0500
		MO	0,0500	0,0250
		LO	0,0025	0,0125
	$K10$	HO	0,0062	0,0625
		MO	0,0025	0,0500
		LO	0,0010	0,0400
$p = 5$	$K3$	HO	0,0250	0,0625
		MO	0,0025	0,0062
		LO	0,0002	0,0006
	$K6$	HO	0,0100	0,0500
		MO	0,0050	0,0250
		LO	0,0025	0,0125
	$K10$	HO	0,0125	0,1250
		MO	0,0050	0,1000
		LO	0,0020	0,0800
$p = 10$	$K3$	HO	0,0250	0,0625
		MO	0,0025	0,0062
		LO	0,0002	0,0006
	$K6$	HO	0,0100	0,0500
		MO	0,0050	0,0250
		LO	0,0025	0,0125
	$K10$	HO	0,0125	0,1250
		MO	0,0050	0,1000
		LO	0,0020	0,0800

$\bar{\omega}$ - chevauchement moyen,

$\check{\omega}$ - chevauchement maximal

HO - niveau de chevauchement élevé

MO - niveau de chevauchement modéré

LO - niveau de chevauchement bas

E.3. SIMULATION DES POPULATIONS ISSUES D'UNE MÉLANGE GAUSSIENNE

```
1
2 #-----#
3 # fonctions requises pour effectuer la simulation #
4 #-----#
5
6 library("MixSim")
7 library("broman")
8 library("graphics")
9 library("xtable")
10
11 #-----#
12 # génération des populations artificielles #
13 #-----#
14 conditions <- function(p_dim, K_clust, lam){
15   x <- 0
16   if( -5 > lam || lam > 5 ){
17     print("lambda must be between -5 and 5")
18     x <- x + 1
19   }
20   if( p_dim!=2 && p_dim!=5 && p_dim!=10){
21     print("choose other value of p")
22     x <- x+1
23   }
24   if( K_clust!=3 && K_clust!=6 && K_clust!=10 ){
25     print("choose other value of K")
26     x <- x+1
27   }
28   if(x!=0){
29     return(FALSE)
30   } else{
31     return(TRUE)
32   }
33 }
34
35 not_isolated_clusters <- function(M, K_clust){
36   M_ind <- matrix(0, nrow=K_clust, ncol=K_clust)
37   for(i in 1:K_clust){
38     for( j in 1:K_clust ){
39       if( M[i,j]==0 ){
40         M_ind[i,j] <- 1
41       }
42     }
43   }
44   return(sum(M_ind))
45 }
46
47
48 MixData <- function(p_dim, K_clust, mean_overlap, N_size, lam){
49
50   if(conditions(p_dim, K_clust, lam)==FALSE){
```

```

51   stop("conditions in the parameters not satisfied")
52 } else {
53   pi_min <- 1/K_clust- 0.05
54
55   if(K_clust==3){
56     max_ovelp <- mean_overlap*(K_clust*(K_clust-1)/2) - (mean_overlap/2)
57     Mix_HO <- MixSim(BarOmega = mean_overlap*10,
58     MaxOmega = max_ovelp*10,      K = K_clust, p = p_dim, PiLow = pi_min, resN=1000)
59     Mix_MO <- MixSim(BarOmega = mean_overlap,
60     MaxOmega = max_ovelp,        K = K_clust, p = p_dim, PiLow = pi_min)
61     Mix_LO <- MixSim(BarOmega = mean_overlap*(1/10),
62     MaxOmega = max_ovelp*(1/10), K = K_clust, p = p_dim, PiLow = pi_min)
63   }
64
65   if(K_clust==6){
66     max_ovelp <- mean_overlap*(K_clust*(K_clust-1)/2) - 10*mean_overlap
67     Mix_HO <- MixSim(BarOmega = mean_overlap*2,
68     MaxOmega = max_ovelp*2,      K = K_clust, p = p_dim, PiLow = pi_min, resN=1000)
69     Mix_MO <- MixSim(BarOmega = mean_overlap,
70     MaxOmega = max_ovelp,        K = K_clust, p = p_dim, PiLow = pi_min, resN=1000)
71     Mix_LO <- MixSim(BarOmega = mean_overlap*(1/2),
72     MaxOmega = max_ovelp*(1/2), K = K_clust, p = p_dim, PiLow = pi_min, resN=1000)
73   }
74
75   if(K_clust==10){
76     repeat{
77       max_ovelp <- mean_overlap*(K_clust*(K_clust-1)/2) - 25*mean_overlap
78       Mix_HO <- MixSim(BarOmega = mean_overlap*(5/2),
79       MaxOmega = max_ovelp*(5/4), K = K_clust, p = p_dim, PiLow = pi_min, resN=1000)
80       Mix_MO <- MixSim(BarOmega = mean_overlap,
81       MaxOmega = max_ovelp,      K = K_clust, p = p_dim, PiLow = pi_min, resN=1000)
82       Mix_LO <- MixSim(BarOmega = mean_overlap*(2/5),
83       MaxOmega = max_ovelp*(4/5), K = K_clust, p = p_dim, PiLow = pi_min, resN=1000)
84
85       if(not_isolated_clusters(M = Mix_MO$OmegaMap, K_clust) > (K_clust^2-K_clust)/2){
86         mean_overlap <- mean_overlap + 0.001
87       } else {
88         break
89       }
90     }
91   }
92
93   dataset_HO <- simdataset(n = N_size, Pi = Mix_HO$Pi, Mu = Mix_HO$Mu, S = Mix_HO$S,
94   = rep(lam, p_dim))
95   dataset_MO <- simdataset(n = N_size, Pi = Mix_MO$Pi, Mu = Mix_MO$Mu, S = Mix_MO$S,
96   lambda = rep(lam, p_dim))
97   dataset_LO <- simdataset(n = N_size, Pi = Mix_LO$Pi, Mu = Mix_LO$Mu, S = Mix_LO$S,
98   lambda = rep(lam, p_dim))
99
100  if(p_dim==2){
101    colors <- c("tomato", "lightpink", "lightskyblue", "darkolivegreen3",
102              "lightsteelblue1",
103              "red", "magenta", "blue", "brown", "green")

```

```

104     x11()
105     par(mfrow=c(3,1), mar= c(1, 2, 2, 2))
106     plot(dataset_HO$X, col = colors[dataset_HO$id], pch=19, cex =0.8, xlab="", ylab
107     main="High overlap", xaxt='n', yaxt='n')
108     plot(dataset_MO$X, col = colors[dataset_MO$id], pch=19, cex =0.8, xlab="", ylab="",
109     main="Moderated overlap", xaxt='n', yaxt='n')
110     plot(dataset_LO$X, col = colors[dataset_LO$id], pch=19, cex =0.8, xlab="", ylab="",
111     main="Low overlap", xaxt='n', yaxt='n')
112 }
113
114 dataset_H <- data.frame(dataset_HO$X, dataset_HO$id)
115 dataset_M <- data.frame(dataset_MO$X, dataset_MO$id)
116 dataset_L <- data.frame(dataset_LO$X, dataset_LO$id)
117
118 # create the labels
119 x_p <- rep(0, p_dim)
120 for(i in 1:p_dim){
121     x_p[i] <- paste("x", i, sep = "_")
122 }
123
124 colnames(dataset_H) <- colnames(dataset_M) <- colnames(dataset_L) <- c(x_p, "label")
125 mix_data <- vector("list", 3)
126 mix_data[[1]] <- dataset_H
127 mix_data[[2]] <- dataset_M
128 mix_data[[3]] <- dataset_L
129 return(mix_data)
130 }
131 }

```

E.4. SIMULATION DE LA VARIABLE RÉPONSE

```

1 # fonction de risque cumulé de base #
2 H <- function(j, y, S, alpha, lambda){
3     if(j ==1){
4         H_out <- lambda[j]*(y^(alpha[j]) - S[j]^(alpha[j]))
5     } else {
6         H_sum <- 0
7         for(g in seq(1,(j-1))){
8             H_sum <- H_sum + lambda[g]*(S[g+1]^(alpha[g]) - S[g]^(alpha[g]))
9         }
10        H_out <- lambda[j]*(y^(alpha[j]) - S[j]^(alpha[j])) + H_sum
11    }
12    return(H_out)
13 }
14
15 # constant de normalisation #
16 normalizedF <- function(S, alpha, lambda, u){
17     GI <- length(S)
18     w <- 1 - exp(-H(GI - 1, S[GI], S, alpha, lambda) * u)
19     return(w)
20 }

```

```

21 |
22 | # fonction de répartition #
23 | F_dist <- fonction(j, y, S, alpha, lambda, x, beta){
24 |   w <- (1-exp(-H(j, y, S, alpha, lambda)*exp(x**beta)))/normalizedF(S, alpha,
25 |     lambda, exp(x**beta) )
26 |   return(w)
27 | }
28 |
29 | H_inv <- fonction(j, alpha, lambda, S){
30 |   if(j == 1){
31 |     if(S[1] > 0){
32 |       H_i <- lambda[1]*exp(alpha[1]*log(S[1]))
33 |     } else {
34 |       H_i <- 0
35 |     }
36 |   } else {
37 |     H_sum <- 0
38 |     for(g in seq(1, (j-1))){
39 |       H_sum <- H_sum + lambda[g]*(S[g+1]^alpha[g] - S[g]^alpha[g])
40 |     }
41 |     H_i <- lambda[j]*(S[j]^alpha[j]) - H_sum
42 |   }
43 |   return(H_i)
44 | }
45 |
46 | # fonction de répartition inverse #
47 | F_inv <- fonction(j, alpha, lambda, S, beta, x, U){
48 |   uN <- normalizedF( S, alpha, lambda, exp( x**beta) )
49 |   logU <- log(1 - U *uN )
50 |   Hj <- H_inv(j, alpha, lambda, S)*exp(x**beta)
51 |   #cat("uN=", uN, " log=", log(uN), " j=", j, " Hj=", Hj, "\n")
52 |   F_i <- ((Hj-logU)/(lambda[j]*exp(x**beta)))^(1/alpha[j])
53 |
54 |   return(F_i)
55 | }
56 |
57 | simulation_one <- fonction(alpha, lambda, beta, S, x){
58 |   U <- runif(1)
59 |   J <- length(alpha)
60 |   F_d <- rep(0,J)
61 |   for(j in seq(1,J)){
62 |     F_d[j] <- F_dist(j, S[j+1], S, alpha, lambda, x, beta)
63 |   }
64 |   #normalized, make sure last term is 1
65 |   F_d <- c(0, F_d) /F_d[J]
66 |   #print(F_d)
67 |   y <- 0
68 |   if(J > 1){
69 |     for(i in seq(1,J)){
70 |       #cat("U=", U, " F[" , i, "]=", F_d[i], " , ", F_d[i+1], "\n")
71 |       if(F_d[i] <= U & U < F_d[i+1]){
72 |         #cat("Found between Fd[" , i, "]=", F_d[i], " and Fd[" , i+1, "]=", F_d[i+1], "\n")
73 |         y <- F_inv(i, alpha, lambda, S, beta, x, U)

```

```
74     }  
75   }  
76 } else {  
77   y <- F_inv(1, alpha, lambda, S, beta, x, U)  
78 }  
79 return(y)  
80 }
```

E.5. PARAMÈTRES UTILISÉS POUR LES SIMULATIONS DE LA VARIABLE RÉ- PONSE Y

Tableau E. II. Choix de paramètres de départ pour les populations de dimension $p = 2$ et $J = 1$. Le nombre de sous-populations est $K = 3$ (haut) et $K = 6$ (bas) Le paramètre $\alpha = 1$.

Chevauchement	grappe	β_1	β_2	$E(\mathbf{x}'\boldsymbol{\beta})$	λ
HO	C1	1,00	2,00	19,697	0,003
	C2	-1,00	1,00	0,562	0,045
	C3	0,00	1,00	1,367	0,012
MO	C1	1,00	2,00	10,750	0,005
	C2	-1,00	1,00	0,410	0,061
	C3	0,00	1,00	3,263	0,005
LO	C1	1,00	2,00	21,923	0,002
	C2	-1,00	1,00	2,977	0,008
	C3	0,00	1,00	2,057	0,008
HO	C1	1,00	2,00	10,787	0,009
	C2	-1,00	1,00	0,371	0,135
	C3	0,00	1,00	3,169	0,011
	C4	1,00	0,00	1,226	0,020
	C5	1,00	-1,00	0,905	0,022
	C6	2,00	1,00	2,343	0,007
MO	C1	1,00	2,00	4,254	0,024
	C2	-1,00	1,00	1,193	0,042
	C3	0,00	1,00	1,316	0,025
	C4	1,00	0,00	1,308	0,019
	C5	1,00	-1,00	1,712	0,012
	C6	2,00	1,00	1,767	0,009
LO	C1	1,00	2,00	4,577	0,022
	C2	-1,00	1,00	1,389	0,036
	C3	0,00	1,00	1,356	0,025
	C4	1,00	0,00	2,523	0,010
	C5	1,00	-1,00	0,955	0,021
	C6	2,00	1,00	1,556	0,011

Tableau E. III. Choix de paramètres de départ pour les populations de dimension $p = 2$, $J = 1$ et $K = 10$. Le paramètre $\alpha = 1$.

Chevauchement	grappe	β_1	β_2	$E(\mathbf{x}'\boldsymbol{\beta})$	λ
HO	C1	1,00	2,00	15,718	0,013
	C2	-1,00	1,00	2,737	0,037
	C3	0,00	1,00	1,218	0,055
	C4	1,00	0,00	2,138	0,023
	C5	1,00	-1,00	2,876	0,014
	C6	2,00	1,00	20,645	0,002
	C7	-1,00	0,00	0,450	0,063
	C8	0,00	-1,00	0,958	0,026
	C9	2,00	-1,00	7,342	0,003
	C10	-2,00	1,00	1,113	0,018
MO	C1	1,00	2,00	5,689	0,035
	C2	-1,00	1,00	1,547	0,065
	C3	0,00	1,00	1,453	0,046
	C4	1,00	0,00	1,447	0,035
	C5	1,00	-1,00	0,951	0,042
	C6	2,00	1,00	4,057	0,008
	C7	-1,00	0,00	0,667	0,043
	C8	0,00	-1,00	0,672	0,037
	C9	2,00	-1,00	3,661	0,006
	C10	-2,00	1,00	0,651	0,031
LO	C1	1,00	2,00	1,375	0,145
	C2	-1,00	1,00	1,047	0,096
	C3	0,00	1,00	2,677	0,025
	C4	1,00	0,00	1,537	0,033
	C5	1,00	-1,00	1,501	0,027
	C6	2,00	1,00	24,061	0,001
	C7	-1,00	0,00	0,586	0,049
	C8	0,00	-1,00	0,776	0,032
	C9	2,00	-1,00	1,195	0,019
	C10	-2,00	1,00	1,471	0,014

Tableau E. IV. Choix de paramètres de départ pour les populations de dimension $p = 2$, $J = 2$, $K = 3$ (haut) et $K = 6$ (bas). Le paramètre $\alpha = 1$.

Chevauchement	grappe	β			λ	
		β_1	β_2	$E(\mathbf{x}'\beta)$	λ_1	λ_2
HO	C1	1,00	2,00	19,70	0,003	0,001
	C2	-1,00	1,00	0,56	0,059	0,036
	C3	0,00	1,00	1,37	0,018	0,012
MO	C1	1,00	2,00	10,750	0,005	0,002
	C2	-1,00	1,00	0,410	0,081	0,049
	C3	0,00	1,00	3,263	0,008	0,005
LO	C1	1,00	2,00	21,923	0,002	0,001
	C2	-1,00	1,00	2,977	0,011	0,007
	C3	0,00	1,00	2,057	0,012	0,008
HO	C1	1,00	2,00	10,787	0,009	0,003
	C2	-1,00	1,00	0,371	0,135	0,067
	C3	0,00	1,00	3,169	0,011	0,006
	C4	1,00	0,00	1,226	0,020	0,014
	C5	1,00	-1,00	0,905	0,022	0,016
	C6	2,00	1,00	2,343	0,007	0,005
MO	C1	1,00	2,00	4,254	0,024	0,008
	C2	-1,00	1,00	1,193	0,042	0,021
	C3	0,00	1,00	1,316	0,025	0,015
	C4	1,00	0,00	1,308	0,019	0,013
	C5	1,00	-1,00	1,712	0,012	0,008
	C6	2,00	1,00	1,767	0,009	0,007
LO	C1	1,00	2,00	4,577	0,022	0,007
	C2	-1,00	1,00	1,389	0,036	0,018
	C3	0,00	1,00	1,356	0,025	0,015
	C4	1,00	0,00	2,523	0,010	0,007
	C5	1,00	-1,00	0,955	0,021	0,015
	C6	2,00	1,00	1,556	0,011	0,008

Tableau E. V. Choix de paramètres de départ pour les populations de dimension $p = 2$, $J = 2$ et $K = 10$. Le paramètre $\alpha = 1$.

Chevauchement	grappe	β		$E(\mathbf{x}'\beta)$	λ	
		β_1	β_2		λ_1	λ_2
HO	C1	1,00	2,00	15,718	0,006	0,002
	C2	-1,00	1,00	2,737	0,018	0,009
	C3	0,00	1,00	1,218	0,027	0,016
	C4	1,00	0,00	0,797	0,012	0,008
	C5	1,00	-1,00	1,001	0,007	0,005
	C6	2,00	1,00	2,607	0,001	0,001
	C7	-1,00	0,00	3,127	0,032	0,025
	C8	0,00	-1,00	0,691	0,013	0,010
	C9	2,00	-1,00	1,452	0,002	0,001
	C10	-2,00	1,00	2,779	0,009	0,007
MO	C1	1,00	2,00	5,689	0,018	0,006
	C2	-1,00	1,00	1,547	0,032	0,016
	C3	0,00	1,00	1,453	0,023	0,014
	C4	1,00	0,00	0,812	0,017	0,012
	C5	1,00	-1,00	2,976	0,021	0,015
	C6	2,00	1,00	1,387	0,004	0,003
	C7	-1,00	0,00	3,091	0,021	0,017
	C8	0,00	-1,00	1,168	0,019	0,015
	C9	2,00	-1,00	1,438	0,003	0,002
	C10	-2,00	1,00	3,007	0,015	0,013
LO	C1	1,00	2,00	1,375	0,073	0,024
	C2	-1,00	1,00	1,047	0,048	0,024
	C3	0,00	1,00	2,677	0,012	0,007
	C4	1,00	0,00	1,564	0,016	0,011
	C5	1,00	-1,00	1,618	0,013	0,010
	C6	2,00	1,00	2,993	0,001	0,001
	C7	-1,00	0,00	1,214	0,024	0,019
	C8	0,00	-1,00	0,461	0,016	0,013
	C9	2,00	-1,00	1,299	0,009	0,008
	C10	-2,00	1,00	1,807	0,007	0,006

Tableau E. VI. Choix de paramètres de départ pour les populations de dimension $p = 2$ et $J = 3$. Le paramètre $\alpha = 1$.

Chevauchement	grappe	β			λ		
		β_1	β_2	$E(\mathbf{x}'\beta)$	λ_1	λ_2	λ_3
HO	C1	1,00	2,00	19,697	0,003	0,001	0,001
	C2	-1,00	1,00	0,562	0,059	0,036	0,025
	C3	0,00	1,00	1,367	0,018	0,012	0,009
MO	C1	1,00	2,00	10,750	0,005	0,002	0,002
	C2	-1,00	1,00	0,410	0,081	0,049	0,035
	C3	0,00	1,00	3,263	0,008	0,005	0,004
LO	C1	1,00	2,000	21,923	0,002	0,001	0,001
	C2	-1,00	1,000	2,977	0,011	0,007	0,005
	C3	0,00	1,000	2,057	0,012	0,008	0,006
HO	C1	1,00	2,00	10,787	0,009	0,003	0,002
	C2	-1,00	1,00	0,371	0,135	0,067	0,045
	C3	0,00	1,00	3,169	0,011	0,006	0,005
	C4	1,00	0,00	1,226	0,020	0,014	0,010
	C5	1,00	-1,00	0,905	0,022	0,016	0,012
	C6	2,00	1,00	2,343	0,007	0,005	0,004
MO	C1	1,00	2,00	4,254	0,024	0,008	0,005
	C2	-1,00	1,00	1,193	0,042	0,021	0,014
	C3	0,00	1,00	1,316	0,025	0,015	0,011
	C4	1,00	0,00	1,308	0,019	0,013	0,010
	C5	1,00	-1,00	1,712	0,012	0,008	0,006
	C6	2,00	1,00	1,767	0,009	0,007	0,006
LO	C1	1,00	2,00	4,577	0,022	0,007	0,004
	C2	-1,00	1,00	1,389	0,036	0,018	0,012
	C3	0,00	1,00	1,356	0,025	0,015	0,011
	C4	1,00	0,00	2,523	0,010	0,007	0,005
	C5	1,00	-1,00	0,955	0,021	0,015	0,012
	C6	2,00	1,00	1,556	0,011	0,008	0,006

Tableau E. VII. Choix de paramètres de départ pour les populations de dimension $p = 2$, $J = 3$ et $K = 10$. Le paramètre $\alpha = 1$.

Chevauchement	grappe	β			λ		
		β_1	β_2	$E(\mathbf{x}'\beta)$	λ_1	λ_2	λ_3
HO	C1	1,00	2,00	15,718	0,006	0,002	0,001
	C2	-1,00	1,00	2,737	0,018	0,009	0,006
	C3	0,00	1,00	1,218	0,027	0,016	0,012
	C4	1,00	0,00	0,797	0,012	0,008	0,006
	C5	1,00	-1,00	1,001	0,007	0,005	0,004
	C6	2,00	1,00	2,607	0,001	0,001	0,000
	C7	-1,00	0,00	3,127	0,032	0,025	0,020
	C8	0,00	-1,00	0,691	0,013	0,010	0,009
	C9	2,00	-1,00	1,452	0,002	0,001	0,001
	C10	-2,00	1,00	2,779	0,009	0,007	0,006
MO	C1	1,00	2,00	5,689	0,018	0,006	0,004
	C2	-1,00	1,00	1,547	0,032	0,016	0,011
	C3	0,00	1,00	1,453	0,023	0,014	0,010
	C4	1,00	0,00	0,812	0,017	0,012	0,009
	C5	1,00	-1,00	2,976	0,021	0,015	0,012
	C6	2,00	1,00	1,387	0,004	0,003	0,002
	C7	-1,00	0,00	3,091	0,021	0,017	0,014
	C8	0,00	-1,00	1,168	0,019	0,015	0,012
	C9	2,00	-1,00	1,438	0,003	0,002	0,002
	C10	-2,00	1,00	3,007	0,015	0,013	0,010
LO	C1	1,00	2,00	1,375	0,073	0,024	0,015
	C2	-1,00	1,00	1,047	0,048	0,024	0,016
	C3	0,00	1,00	2,677	0,012	0,007	0,005
	C4	1,00	0,00	1,564	0,016	0,011	0,008
	C5	1,00	-1,00	1,618	0,013	0,010	0,007
	C6	2,00	1,00	2,993	0,001	0,001	0,000
	C7	-1,00	0,00	1,214	0,024	0,019	0,016
	C8	0,00	-1,00	0,461	0,016	0,013	0,011
	C9	2,00	-1,00	1,299	0,009	0,008	0,006
	C10	-2,00	1,00	1,807	0,007	0,006	0,005

Tableau E. VIII. Choix de paramètres de départ pour les populations de dimension $p = 5$ et $J = 1$. Le paramètre $\alpha = 1$.

Chevauchement	grappe	β_1	β_2	β_3	β_4	β_5	$E(\mathbf{x}'\boldsymbol{\beta})$	λ_1
HO	C1	1,00	2,00	0,00	-2,00	1,00	1,641	0,0300
	C2	-1,00	1,00	2,00	0,00	1,00	5,515	0,0050
	C3	0,00	-1,00	-1,00	2,00	1,00	1,208	0,0140
MO	C1	1,00	2,00	0,00	-2,00	1,00	25,755	0,0020
	C2	-1,00	1,00	2,00	0,00	1,00	16,327	0,0020
	C3	0,00	-1,00	-1,00	2,00	1,00	9,556	0,0020
LO	C1	1,00	2,00	0,00	-2,00	1,00	0,252	0,1980
	C2	-1,00	1,00	2,00	0,00	1,00	10,583	0,0020
	C3	0,00	-1,00	-1,00	2,00	1,00	0,277	0,0600
HO	C1	1,00	2,00	0,00	-2,00	1,00	58,745	0,0020
	C2	-1,00	1,00	2,00	0,00	1,00	27,584	0,0020
	C3	0,00	1,00	-1,00	2,00	1,00	19,696	0,0020
	C4	1,00	0,00	0,00	-2,00	1,00	11,294	0,0010
	C5	1,00	-1,00	2,00	0,00	1,00	248,954	0,0010
	C6	2,00	1,00	-1,00	2,00	1,00	1789,898	0,0010
MO	C1	1,00	2,00	0,00	-2,00	1,00	6,2649	0,0160
	C2	-1,00	1,00	2,00	0,00	1,00	77,7953	0,0006
	C3	0,00	1,00	-1,00	2,00	1,00	256,1611	0,0001
	C4	1,00	0,00	0,00	-2,00	1,00	27,1663	0,0009
	C5	1,00	-1,00	2,00	0,00	1,00	690,2736	0,0000
	C6	2,00	1,00	-1,00	2,00	1,00	1061,6418	0,0000
LO	C1	1,00	2,00	0,00	-2,00	1,00	2,3784	0,0420
	C2	-1,00	1,00	2,00	0,00	1,00	16,0673	0,0031
	C3	0,00	1,00	-1,00	2,00	1,00	45,2778	0,0007
	C4	1,00	0,00	0,00	-2,00	1,00	14,0929	0,0018
	C5	1,00	-1,00	2,00	0,00	1,00	327,9262	0,0001
	C6	2,00	1,00	-1,00	2,00	1,00	2415,7653	0,0000

Tableau E. IX. Choix de paramètres de départ pour les populations de dimension $p = 5$ et $J = 1$. Le paramètre $\alpha = 1$.

Chevauchement	grappe	β_1	β_2	β_3	β_4	β_5	$E(\mathbf{x}'\boldsymbol{\beta})$	λ_1
HO	C1	1,00	2,00	0,00	-2,00	1,00	15,4957	0,0129
	C2	-1,00	1,00	2,00	0,00	1,00	2,7163	0,0368
	C3	0,00	1,00	-1,00	2,00	1,00	1,4542	0,0458
	C4	1,00	0,00	0,00	-2,00	1,00	0,1864	0,2682
	C5	1,00	-1,00	2,00	0,00	1,00	1,1091	0,0361
	C6	2,00	1,00	-1,00	2,00	1,00	16,9033	0,0020
	C7	-1,00	0,00	1,00	-2,00	1,00	0,2021	0,1414
	C8	0,00	-1,00	2,00	0,00	1,00	4,4483	0,0056
	C9	2,00	-1,00	3,00	1,00	1,00	17,4026	0,0013
	C10	-2,00	1,00	-1,00	2,00	1,00	0,8264	0,0242
MO	C1	1,00	2,00	0,00	-2,00	1,00	8,6765	0,0231
	C2	-1,00	1,00	2,00	0,00	1,00	1,9880	0,0503
	C3	0,00	1,00	-1,00	2,00	1,00	2,1735	0,0307
	C4	1,00	0,00	0,00	-2,00	1,00	0,4741	0,1055
	C5	1,00	-1,00	2,00	0,00	1,00	22,0231	0,0018
	C6	2,00	1,00	-1,00	2,00	1,00	19,4945	0,0017
	C7	-1,00	0,00	1,00	-2,00	1,00	0,2980	0,0959
	C8	0,00	-1,00	2,00	0,00	1,00	1,4902	0,0168
	C9	2,00	-1,00	3,00	1,00	1,00	238,8376	0,0001
	C10	-2,00	1,00	-1,00	2,00	1,00	8,3831	0,0024
LO	C1	1,00	2,00	0,00	-2,00	1,00	34,7089	0,0058
	C2	-1,00	1,00	2,00	0,00	1,00	5,8469	0,0171
	C3	0,00	1,00	-1,00	2,00	1,00	4,0488	0,0165
	C4	1,00	0,00	0,00	-2,00	1,00	0,8001	0,0625
	C5	1,00	-1,00	2,00	0,00	1,00	13,0957	0,0031
	C6	2,00	1,00	-1,00	2,00	1,00	48,5122	0,0007
	C7	-1,00	0,00	1,00	-2,00	1,00	2,6819	0,0107
	C8	0,00	-1,00	2,00	0,00	1,00	3,4990	0,0071
	C9	2,00	-1,00	3,00	1,00	1,00	6,4226	0,0035
	C10	-2,00	1,00	-1,00	2,00	1,00	0,2010	0,0995

Tableau E. X. Choix de paramètres de départ pour les populations de dimension $p = 5$, $J = 2$, $K = 3$ (haut) et $K = 6$ (bas). Le paramètre $\alpha = 1$.

Chevauchement	grappe	β					$E(\mathbf{x}'\beta)$	λ	
		β_1	β_2	β_3	β_4	β_5		λ_1	λ_2
HO	C1	1,00	2,00	0,00	-2,00	1,00	1,641	0,030	0,015
	C2	-1,00	1,00	2,00	0,00	1,00	5,515	0,006	0,004
	C3	0,00	-1,00	-1,00	2,00	1,00	1,208	0,021	0,014
MO	C1	1,00	2,00	0,00	-2,00	1,00	25,755	0,002	0,001
	C2	-1,00	1,00	2,00	0,00	1,00	16,327	0,002	0,001
	C3	0,00	-1,00	-1,00	2,00	1,00	9,556	0,003	0,002
LO	C1	1,00	2,00	0,00	-2,00	1,00	0,252	0,198	0,099
	C2	-1,00	1,00	2,00	0,00	1,00	10,583	0,003	0,002
	C3	0,00	-1,00	-1,00	2,00	1,00	0,277	0,090	0,060
HO	C1	1,00	2,00	0,00	-2,00	1,00	58,745	0,002	0,001
	C2	-1,00	1,00	2,00	0,00	1,00	27,584	0,002	0,001
	C3	0,00	1,00	-1,00	2,00	1,00	19,696	0,002	0,001
	C4	-1,00	1,00	2,00	0,00	1,00	11,294	0,002	0,001
	C5	0,00	1,00	-1,00	2,00	1,00	248,954	0,001	0,001
	C6	0,00	1,00	-1,00	2,00	1,00	1789,898	0,001	0,001
MO	C1	1,00	2,00	0,00	-2,00	1,00	6,2649	0,016	0,005
	C2	-1,00	1,00	2,00	0,00	1,00	77,7953	0,001	0,000
	C3	0,00	1,00	-1,00	2,00	1,00	256,1611	0,000	0,000
	C4	-1,00	1,00	2,00	0,00	1,00	27,1663	0,001	0,000
	C5	0,00	1,00	-1,00	2,00	1,00	690,2736	0,000	0,000
	C6	0,00	1,00	-1,00	2,00	1,00	1061,6418	0,000	0,000
LO	C1	1,00	2,00	0,00	-2,00	1,00	2,3784	0,042	0,014
	C2	-1,00	1,00	2,00	0,00	1,00	16,0673	0,003	0,002
	C3	0,00	1,00	-1,00	2,00	1,00	45,2778	0,001	0,000
	C4	-1,00	1,00	2,00	0,00	1,00	14,0929	0,002	0,001
	C5	0,00	1,00	-1,00	2,00	1,00	327,9262	0,000	0,000
	C6	0,00	1,00	-1,00	2,00	1,00	2415,7653	0,000	0,000

Tableau E. XI. Choix de paramètres de départ pour les populations de dimension $p = 5$ et $J = 2$. Le paramètre $\alpha = 1$.

Chevauchement	grappe	β					$E(\mathbf{x}'\beta)$	λ	
		β_1	β_2	β_3	β_4	β_5		λ_1	λ_2
HO	C1	1,00	2,00	0,00	-2,00	1,00	15,4957	0,0065	0,0022
	C2	-1,00	1,00	2,00	0,00	1,00	2,7163	0,0184	0,0092
	C3	0,00	1,00	-1,00	2,00	1,00	1,4542	0,0229	0,0138
	C4	1,00	0,00	0,00	-2,00	1,00	0,1864	0,1341	0,0894
	C5	1,00	-1,00	2,00	0,00	1,00	1,1091	0,0180	0,0129
	C6	2,00	1,00	-1,00	2,00	1,00	16,9033	0,0010	0,0007
	C7	-1,00	0,00	1,00	-2,00	1,00	0,2021	0,0707	0,0550
	C8	0,00	-1,00	2,00	0,00	1,00	4,4483	0,0028	0,0022
	C9	2,00	-1,00	3,00	1,00	1,00	17,4026	0,0006	0,0005
	C10	-2,00	1,00	-1,00	2,00	1,00	0,8264	0,0121	0,0101
MO	C1	1,00	2,00	0,00	-2,00	1,00	8,6765	0,0115	0,0038
	C2	-1,00	1,00	2,00	0,00	1,00	1,9880	0,0252	0,0126
	C3	0,00	1,00	-1,00	2,00	1,00	2,1735	0,0153	0,0092
	C4	1,00	0,00	0,00	-2,00	1,00	0,4741	0,0527	0,0352
	C5	1,00	-1,00	2,00	0,00	1,00	22,0231	0,0009	0,0006
	C6	2,00	1,00	-1,00	2,00	1,00	19,4945	0,0009	0,0006
	C7	-1,00	0,00	1,00	-2,00	1,00	0,2980	0,0479	0,0373
	C8	0,00	-1,00	2,00	0,00	1,00	1,4902	0,0084	0,0067
	C9	2,00	-1,00	3,00	1,00	1,00	238,8376	0,0000	0,0000
	C10	-2,00	1,00	-1,00	2,00	1,00	8,3831	0,0012	0,0010
LO	C1	1,00	2,00	0,00	-2,00	1,00	34,7089	0,0029	0,0010
	C2	-1,00	1,00	2,00	0,00	1,00	5,8469	0,0086	0,0043
	C3	0,00	1,00	-1,00	2,00	1,00	4,0488	0,0082	0,0049
	C4	1,00	0,00	0,00	-2,00	1,00	0,8001	0,0312	0,0208
	C5	1,00	-1,00	2,00	0,00	1,00	13,0957	0,0015	0,0011
	C6	2,00	1,00	-1,00	2,00	1,00	48,5122	0,0003	0,0003
	C7	-1,00	0,00	1,00	-2,00	1,00	2,6819	0,0053	0,0041
	C8	0,00	-1,00	2,00	0,00	1,00	3,4990	0,0036	0,0029
	C9	2,00	-1,00	3,00	1,00	1,00	6,4226	0,0017	0,0014
	C10	-2,00	1,00	-1,00	2,00	1,00	0,2010	0,0498	0,0415

Tableau E. XII. Choix de paramètres de départ pour les populations de dimension $p = 5$ et $J = 3$. Le paramètre $\alpha = 1$.

Chevauchement	grappe	β					$E(\mathbf{x}'\beta)$	λ		
		β_1	β_2	β_3	β_4	β_5		λ_1	λ_2	λ_3
HO	C1	1,00	2,00	0,00	-2,00	1,00	1,64	0,030	0,015	0,010
	C2	-1,00	1,00	2,00	0,00	1,00	5,52	0,006	0,004	0,003
	C3	0,00	-1,00	-1,00	2,00	1,00	1,21	0,021	0,014	0,010
MO	C1	1,00	2,00	0,00	-2,00	1,00	25,755	0,002	0,001	0,001
	C2	-1,00	1,00	2,00	0,00	1,00	16,327	0,002	0,001	0,001
	C3	0,00	-1,00	-1,00	2,00	1,00	9,556	0,003	0,002	0,001
LO	C1	1,00	2,00	0,00	-2,00	1,00	0,252	0,198	0,099	0,066
	C2	-1,00	1,00	2,00	0,00	1,00	10,583	0,003	0,002	0,001
	C3	0,00	-1,00	-1,00	2,00	1,00	0,277	0,090	0,060	0,045
HO	C1	1,00	2,00	0,00	-2,00	1,00	58,745	0,002	0,001	0,001
	C2	-1,00	1,00	2,00	0,00	1,00	27,584	0,002	0,001	0,001
	C3	0,00	1,00	-1,00	2,00	1,00	19,696	0,002	0,001	0,001
	C4	-1,00	1,00	2,00	0,00	1,00	11,294	0,002	0,001	0,001
	C5	0,00	1,00	-1,00	2,00	1,00	248,954	0,001	0,001	0,001
	C6	0,00	1,00	-1,00	2,00	1,00	1789,898	0,001	0,001	0,001
MO	C1	1,00	2,00	0,00	-2,00	1,00	6,2649	0,0160	0,0053	0,0032
	C2	-1,00	1,00	2,00	0,00	1,00	77,7953	0,0006	0,0003	0,0002
	C3	0,00	1,00	-1,00	2,00	1,00	256,1611	0,0001	0,0001	0,0001
	C4	-1,00	1,00	2,00	0,00	1,00	27,1663	0,0009	0,0006	0,0005
	C5	0,00	1,00	-1,00	2,00	1,00	690,2736	0,0000	0,0000	0,0000
	C6	0,00	1,00	-1,00	2,00	1,00	1061,6418	0,0000	0,0000	0,0000
LO	C1	1,00	2,00	0,00	-2,00	1,00	2,3784	0,0420	0,0140	0,0084
	C2	-1,00	1,00	2,00	0,00	1,00	16,0673	0,0031	0,0016	0,0010
	C3	0,00	1,00	-1,00	2,00	1,00	45,2778	0,0007	0,0004	0,0003
	C4	-1,00	1,00	2,00	0,00	1,00	14,0929	0,0018	0,0012	0,0009
	C5	0,00	1,00	-1,00	2,00	1,00	327,9262	0,0001	0,0000	0,0000
	C6	0,00	1,00	-1,00	2,00	1,00	2415,7653	0,0000	0,0000	0,0000

Tableau E. XIII. Choix de paramètres de départ pour les populations de dimension $p = 5$ et $J = 3$. Le paramètre $\alpha = 1$.

Chevauchement	grappe	β					$E(\mathbf{x}'\beta)$	λ		
		β_1	β_2	β_3	β_4	β_5		λ_1	λ_2	λ_3
HO	C1	1,00	2,00	0,00	-2,00	1,00	15,4957	0,0065	0,0022	0,0013
	C2	-1,00	1,00	2,00	0,00	1,00	2,7163	0,0184	0,0092	0,0061
	C3	0,00	1,00	-1,00	2,00	1,00	1,4542	0,0229	0,0138	0,0098
	C4	1,00	0,00	0,00	-2,00	1,00	0,1864	0,1341	0,0894	0,0671
	C5	1,00	-1,00	2,00	0,00	1,00	1,1091	0,0180	0,0129	0,0100
	C6	2,00	1,00	-1,00	2,00	1,00	16,9033	0,0010	0,0007	0,0006
	C7	-1,00	0,00	1,00	-2,00	1,00	0,2021	0,0707	0,0550	0,0450
	C8	0,00	-1,00	2,00	0,00	1,00	4,4483	0,0028	0,0022	0,0019
	C9	2,00	-1,00	3,00	1,00	1,00	17,4026	0,0006	0,0005	0,0004
	C10	-2,00	1,00	-1,00	2,00	1,00	0,8264	0,0121	0,0101	0,0086
MO	C1	1,00	2,00	0,00	-2,00	1,00	8,6765	0,0115	0,0038	0,0023
	C2	-1,00	1,00	2,00	0,00	1,00	1,9880	0,0252	0,0126	0,0084
	C3	0,00	1,00	-1,00	2,00	1,00	2,1735	0,0153	0,0092	0,0066
	C4	1,00	0,00	0,00	-2,00	1,00	0,4741	0,0527	0,0352	0,0264
	C5	1,00	-1,00	2,00	0,00	1,00	22,0231	0,0009	0,0006	0,0005
	C6	2,00	1,00	-1,00	2,00	1,00	19,4945	0,0009	0,0006	0,0005
	C7	-1,00	0,00	1,00	-2,00	1,00	0,2980	0,0479	0,0373	0,0305
	C8	0,00	-1,00	2,00	0,00	1,00	1,4902	0,0084	0,0067	0,0056
	C9	2,00	-1,00	3,00	1,00	1,00	238,8376	0,0000	0,0000	0,0000
	C10	-2,00	1,00	-1,00	2,00	1,00	8,3831	0,0012	0,0010	0,0009
LO	C1	1,00	2,00	0,00	-2,00	1,00	34,7089	0,0029	0,0010	0,0006
	C2	-1,00	1,00	2,00	0,00	1,00	5,8469	0,0086	0,0043	0,0029
	C3	0,00	1,00	-1,00	2,00	1,00	4,0488	0,0082	0,0049	0,0035
	C4	1,00	0,00	0,00	-2,00	1,00	0,8001	0,0312	0,0208	0,0156
	C5	1,00	-1,00	2,00	0,00	1,00	13,0957	0,0015	0,0011	0,0008
	C6	2,00	1,00	-1,00	2,00	1,00	48,5122	0,0003	0,0003	0,0002
	C7	-1,00	0,00	1,00	-2,00	1,00	2,6819	0,0053	0,0041	0,0034
	C8	0,00	-1,00	2,00	0,00	1,00	3,4990	0,0036	0,0029	0,0024
	C9	2,00	-1,00	3,00	1,00	1,00	6,4226	0,0017	0,0014	0,0012
	C10	-2,00	1,00	-1,00	2,00	1,00	0,2010	0,0498	0,0415	0,0355

Annexe F

RÉSULTATS SUPPLÉMENTAIRES

F.1. APPROXIMATION DE LAPLACE : CAS GÉNÉRAL

La méthode de Laplace, due à Pierre-Simon de Laplace, est une méthode pour l'évaluation numérique d'intégrales de la forme

$$\int_B e^{-\lambda g(\boldsymbol{\vartheta})} h(\boldsymbol{\vartheta}) d\boldsymbol{\vartheta}$$

où g est une fonction deux fois dérivable, λ est un grand nombre réel et les bornes peuvent éventuellement être infinies. Soit $\boldsymbol{\vartheta}^*$ le point qui minimise la fonction g , considérons l'expansion de Taylor de la fonction autour de $\boldsymbol{\vartheta}^*$

$$g(\boldsymbol{\vartheta}) \approx g(\boldsymbol{\vartheta}^*) + \frac{\partial g(\boldsymbol{\vartheta}^*)}{\partial \boldsymbol{\vartheta}} (\boldsymbol{\vartheta} - \boldsymbol{\vartheta}^*) + \frac{1}{2} (\boldsymbol{\vartheta} - \boldsymbol{\vartheta}^*)^T \frac{\partial^2 g(\boldsymbol{\vartheta}^*)}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}^T} (\boldsymbol{\vartheta} - \boldsymbol{\vartheta}^*),$$

le vecteur de premiers dérivés s'annule, car $\boldsymbol{\vartheta}^*$ minimise g , ainsi, on obtient que l'intégrale peut être estimée de la façon suivante

$$\begin{aligned} \int_B e^{-\lambda g(\boldsymbol{\vartheta})} h(\boldsymbol{\vartheta}) d\boldsymbol{\vartheta} &= e^{-\lambda g(\boldsymbol{\vartheta}^*)} h(\boldsymbol{\vartheta}^*) \int_B e^{\frac{-\lambda}{2} (\boldsymbol{\vartheta} - \boldsymbol{\vartheta}^*)^T \frac{\partial^2 g(\boldsymbol{\vartheta}^*)}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}^T} (\boldsymbol{\vartheta} - \boldsymbol{\vartheta}^*)} d\boldsymbol{\vartheta} \\ &= e^{-\lambda g(\boldsymbol{\vartheta}^*)} h(\boldsymbol{\vartheta}^*) \left(\frac{2\pi}{\lambda} \right)^{d/2} \left| \frac{\partial^2 g(\boldsymbol{\vartheta}^*)}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}^T} \right|^{-1/2} \end{aligned}$$

La deuxième intégrale peut être estimée à l'aide d'une intégrale de Gauss en remplaçant les bornes par $-\infty$ et ∞

F.2. RAPPEL SUR LES LOIS CONDITIONNELLES DE LA DISTRIBUTION NORMALE MULTIVARIÉE

Proposition : Soit $X = (X_1, X_2) \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ où X_i représente un vecteur aléatoire de taille $p_i < p$ pour $i = 1, 2$, $\boldsymbol{\Sigma} > 0$ et où

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$$

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

Alors,

$$X_2|X_1 \sim N_{p_2} \left(\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(X_1 - \mu_1), \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \right).$$