

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

L'ÉCHANTILLONNAGE DE GIBBS POUR L'ESTIMATION
BAYÉSIENNE DANS L'ANALYSE DE SURVIE

MÉMOIRE

PRÉSENTÉ

COMME EXIGENCE PARTIELLE

DE LA MAÎTRISE EN MATHÉMATIQUES

PAR

LOTFI KHRIBI

FÉVRIER 2007

UNIVERSITÉ DU QUÉBEC À MONTRÉAL
Service des bibliothèques

Avertissement

La diffusion de ce mémoire se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.01-2006). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

REMERCIEMENTS

Je tiens à remercier ma directrice Brenda McGibbon et mon codirecteur Glenn Shorrock pour leurs efforts exceptionnels et pour m'avoir soutenu d'une façon remarquable pendant la réalisation de ce projet de mémoire. Leur intérêt pour l'approche bayésienne en statistique et, en particulier, celui de Brenda McGibbon en biostatistique, ont sans aucun doute contribué à mener à terme ce projet de mémoire.

J'aimerais aussi remercier Sorana Froda et mon ami Severien Nkurunziza pour leurs encouragements tout au long de ces années que j'ai passées au Département de mathématiques. J'aimerais aussi souligner le travail remarquable de Gisèle Legault, analyste en informatique, pour sa disponibilité et son aide.

Finalement, j'aimerais remercier ma femme, Imen, mes deux enfants, Yasmine et Youssef, et ma grande famille, en particulier ma mère, et tous ceux qui m'ont encouragé pendant ces années d'études.

TABLE DES MATIÈRES

LISTE DES TABLEAUX	vi
LISTE DES FIGURES	viii
RÉSUMÉ	ix
INTRODUCTION	1
CHAPITRE I	
INTRODUCTION À L'INFÉRENCE BAYÉSIENNE	4
1.1 Approche statistique classique	5
1.2 Approche bayésienne	5
1.2.1 Modèle bayésien	6
1.2.2 Modèle bayésien hiérarchique	8
1.3 Lien entre les deux approches	9
1.4 Inférence bayésienne	10
1.4.1 Coût quadratique	11
1.4.2 Estimateur de Bayes	11
1.4.3 Intervalle de crédibilité bayésien	11
1.4.4 Méthodes d'estimation bayésiennes	12
1.5 Méthodes d'estimation Monte-Carlo	12
1.6 Méthodes de simulation par chaînes de Markov	13
1.6.1 Pratique des méthodes MCMC	14
1.6.2 Algorithme de Metropolis-Hastings	14
1.6.3 L'échantillonnage de Gibbs	16
1.7 Efficacité asymptotique des estimateurs bayésiens	19
1.7.1 Efficacité asymptotique de l'estimateur du maximum de vraisemblance	19
1.7.2 Efficacité asymptotique des estimateurs bayésiens	21

CHAPITRE II	
ANALYSE BAYÉSIENNE ET FRÉQUENTISTE DES MODÈLES DE DURÉE AVEC CENSURE	
	26
2.1	Introduction à l'analyse de survie
	26
2.1.1	Fonction de survie et de risque instantané
	27
2.1.2	Données censurées
	28
2.2	Méthodes fréquentistes d'analyse de survie
	31
2.2.1	Méthodes paramétriques
	31
2.2.2	Méthodes non paramétriques
	33
2.2.3	Méthodes semi paramétriques
	35
2.3	Analyse bayésienne des modèles paramétriques de durée avec censure
	39
2.3.1	Fonctions de vraisemblance des modèles paramétriques censurés
	40
CHAPITRE III	
MÉTHODES BAYÉSIENNES EN ANALYSE DE SURVIE	
	42
3.1	Introduction
	42
3.1.1	Logiciel de simulation MCMC : BUGS
	42
3.2	Méthode semi paramétrique : la méthode de Kalbfleisch
	43
3.2.1	Processus de comptage
	44
3.2.2	Présentation de la méthode de Kalbfleisch
	45
3.3	Exemples d'application de la méthode Kalbfleisch
	47
3.3.1	Étude de simulation
	47
3.3.2	Cas de leucémie chez les enfants
	52
3.3.3	Estimation des paramètres avec WinBUGS
	53
3.4	Résultats avec WinBUGS pour la base de données pour le cancer du sein
	56
3.5	Analyse et résultats
	57
3.6	Méthodes non paramétriques : la méthode de Florens et Rolin
	61
3.6.1	La distribution de Dirichlet
	61
3.6.2	Processus de Dirichlet
	61
3.6.3	Modèle non paramétrique de durée de vie avec censure
	62

3.7	Méthode bayésienne paramétrique pour des modèles avec points de rupture : la méthode de Carlin, Gelfand et Smith	66
3.7.1	Formulation bayésienne	67
3.7.2	Choix de modèle et des densités <i>a priori</i> pour les paramètres	69
3.7.3	Lois <i>a posteriori</i> des paramètres	70
3.8	Exemple d'application de la méthode de Carlin, Gelfand et Smith	71
3.8.1	Cas des accidents des mines	71
3.8.2	Estimation avec WinBUGS	72
	CONCLUSION	76
	APPENDICE A	
	PROGRAMMES WINBUGS POUR LES DEUX MÉTHODES BAYÉSIENNES UTILISÉES DANS LE CHAPITRE 3	78
	BIBLIOGRAPHIE	84

LISTE DES TABLEAUX

1.1	Exemple de lois conjuguées naturelles pour des lois appartenant à la famille exponentielle	8
3.1	Erreur quadratique moyenne pour les estimations de la fonction de survie avec les méthodes de Kaplan-Meier et de Kalbfleisch du groupe placebo pour une simulation.	50
3.2	Erreur quadratique moyenne pour les estimations de la fonction de survie avec les méthodes de Kaplan-Meier et de Kalbfleisch du groupe traité pour une simulation.	51
3.3	Données tirées de l'ouvrage de Klein et Moeschberger (1997) pour le cas de leucémie chez les enfants.	52
3.4	Moyennes <i>a posteriori</i> de survie pour le groupe traité et les intervalles de crédibilité $IC_{95\%}$ dans l'exemple de leucémie.	53
3.5	Estimation par la méthode de K.M. de la fonction de survie et des intervalles de confiance (95%) pour le groupe 6-MP.	55
3.6	Données tirées de l'ouvrage de Klein et Moeschberger (1997) pour le cas de cancer du sein (Exemple (1.5) Chapitre 1).	57
3.7	Valeurs <i>a posteriori</i> des moyennes de survie pour les femmes avec une immunoperoxydase négative et leurs intervalles de crédibilité $IC_{95\%}$ dans l'exemple de cancer du sein.	58

3.8	Estimation par la méthode de K.M. de la fonction de survie et des intervalles de confiance (95%) pour le groupe immunoperoxydase négatif.	59
3.9	Base de données sur les accidents des mines en Grande-Bretagne entre 1851 et 1962 citée dans l'article de Carlin, Gelfand et Smith.	72
3.10	Estimation des densités <i>a posteriori</i> $k Y$, θY et λY pour les données sur les accidents des mines en Grande-Bretagne entre 1851 et 1962 ($a_1 = a_2 = 0.5$; $c_1 = c_2 = 0$; $d_1 = d_2 = 1$).	73

LISTE DES FIGURES

2.1	Exemple de censure à droite.	30
3.1	Estimation de la moyenne de survie par la méthode de Kalbfleisch pour les groupes de traitement et de placebo dans l'exemple de leucémie.	54
3.2	Valeurs <i>a posteriori</i> de quelques durées de vie et de coefficient de régression en fonction des itérations de la chaîne dans l'étude de l'exemple de cancer du sein.	60
3.3	Estimation des densités <i>a posteriori</i> $k Y$, θY et λY pour les données sur les accidents des mines en Grande-Bretagne entre 1851 et 1962 ($a_1 = a_2 = 0.5$; $c_1 = c_2 = 0$; $d_1 = d_2 = 1$).	74
3.4	Valeurs <i>a posteriori</i> des paramètres en fonction des itérations de la chaîne dans l'étude de l'exemple des données sur les accidents des mines en Grande-Bretagne entre 1851 et 1962.	75

RÉSUMÉ

Dans ce mémoire, nous étudions l'approche bayésienne dans l'estimation des durées de vie pour des problèmes avec des points de rupture et avec des données censurées.

Dans le chapitre 1, nous présentons les deux approches statistiques, l'approche fréquentiste et l'approche bayésienne. Nous montrons les points de ressemblance et de différence entre ces deux approches. Aussi, nous présentons le modèle bayésien hiérarchique avec les méthodes d'estimation Monte-Carlo avec un bref aperçu des méthodes de simulation par chaînes de Markov (MCMC). nous nous intéressons en particulier aux deux algorithmes qui sont utiles pour ces méthodes : il s'agit de l'algorithme Metropolis-Hastings et la méthode d'échantillonnage de **Gibbs**. Cette dernière a été utilisée par Geman et Geman (84) pour générer des observations à partir d'une distribution de Gibbs (distribution de Boltzmann). Il s'agit d'une forme particulière de méthode de Monte-Carlo par chaîne de Markov qui, du fait de son efficacité, est largement utilisée dans de nombreux domaines d'analyse statistique bayésienne. Nous parlons aussi dans ce chapitre du logiciel **BUGS** (*Bayesian Inference Using Gibbs Sampling*). Ce logiciel de programmation a été développé à l'unité **MRC** de Biostatistique de Cambridge. Il vise à tirer avantage des probabilités de l'échantillonnage de Gibbs dans l'inférence bayésienne sur des systèmes complexes.

Dans le chapitre 2, nous abordons quelques concepts nécessaires à l'étude de l'analyse de survie, tels les fonctions de survie et de risque, les différents types de données censurées. Nous parlons aussi des méthodes fréquentistes d'analyse de survie, notamment la méthode de Kaplan-Meier. Nous rappelons aussi comment est déterminée la fonction de vraisemblance d'un modèle bayésien paramétrique de durée avec censure.

Le chapitre 3 présente deux méthodes bayésiennes qui seront analysées et implémentées dans le logiciel **BUGS**. Une méthode qui est semi paramétrique, il s'agit de la méthode de Kalbfleisch. L'autre méthode paramétrique est celle de Carlin, Gelfand et Smith. Nous vérifions, grâce à des simulations, l'efficacité de ces deux méthodes bayésiennes. Deux exemples de simulations seront traités, un avec données censurées et l'autre avec points de rupture. Nous démontrons principalement que les estimations par simulation et l'inférence bayésienne paramétrique donnent de bons résultats par rapport aux méthodes classiques.

Mots-clés : fonction de survie, méthodes paramétrique et semi paramétrique bayésiennes, méthodes de Monte Carlo par chaîne de Markov, échantillonnage de Gibbs.

INTRODUCTION

Étudier des durées de vie, c'est s'intéresser à l'apparition d'un événement au cours du temps, quel qu'en soit son type (décès, apparition d'une maladie, récurrence d'une maladie...). Il est nécessaire pour cela de disposer du temps de suivi de chaque sujet ainsi que du moment auquel l'événement éventuel s'est produit. La particularité de ces études est l'existence de données incomplètes, appelées données censurées, pour les sujets n'ayant pas présenté l'événement en question au moment du bilan de l'étude ; ce la nécessite une méthodologie adaptée pour leur analyse. L'analyse des durées de vie trouve donc des applications dans de nombreux domaines, biomédicaux ou non, tels que l'industrie (pour des études de fiabilité d'appareils), la physique (avec l'étude de la durée de vie des particules)...

Les premiers modèles d'analyse de durées de survie ont été développés afin de modéliser, d'une manière univariée, la survie observée. Il s'agit de modèles non paramétriques tels que ceux utilisés dans la méthode de Kaplan et Meier (1958). Devant les besoins cliniques et épidémiologiques de prendre en compte simultanément plusieurs facteurs, des modèles paramétriques ont été proposés imposant une distribution *a priori* pour les données de survie (par exemple, distribution exponentielle, distribution de Weibull...) afin de pouvoir décrire l'effet des facteurs pronostiques d'une manière similaire à celle utilisée dans la régression multiple (Feigl et Zelen, 1965).

Dans un premier temps, les méthodes d'étude de la survie ont été élaborées en se basant sur la statistique inférentielle classique, c'est-à-dire selon une approche fréquentiste. Une toute autre approche de la statistique a vu le jour suite aux travaux du révérend Thomas Bayes (Barnard, 1958) et de son texte « Essai en vue de résoudre un problème de la doctrine des chances ». Thomas Bayes pose dans ce texte les bases des probabilités conditionnelles. En effet, face à un raisonnement permettant de déduire des probabilités à partir d'une cause déterminée, Thomas Bayes s'est penché sur le problème inverse consistant à analyser les causes, qui sont supposées inconnues, à partir d'événements observés. Les méthodes bayésiennes en statistique tirent donc

leurs fondements de ces travaux anciens ; il faut cependant noter qu'elles n'ont été que peu utilisées dans le domaine biomédical jusque vers le début des années 1980. Des développements relativement récents, tant sur le plan théorique, avec la théorie sur les chaînes de Markov et la mise en oeuvre de méthodes d'échantillonnage performantes, que technique, avec des moyens de calcul de plus en plus puissants, sont à l'origine de l'accroissement du développement et de l'utilisation de ces méthodes. C'est donc tout naturellement que plusieurs méthodes bayésiennes d'analyse de la survie ont été développées.

Les utilisateurs des méthodes d'analyse de survie ont donc généralement à choisir une méthode parmi toute la panoplie qui leur est proposée. Ce choix va dépendre non seulement du type de survie à estimer mais également de l'approche statistique utilisée pour l'inférence.

Dans le chapitre 1, nous présentons les deux approches en statistique : l'approche fréquentiste et l'approche bayésienne et nous montrons les points de ressemblance et de différence entre ces deux approches. Aussi, on présente le modèle bayésien hiérarchique avec les méthodes d'estimation Monte-Carlo avec un bref aperçu sur les méthodes de simulation par chaînes de Markov (MCMC) en particulier deux algorithmes qui utilisent ces méthodes : l'algorithme Metropolis-Hastings et la méthode d'échantillonnage de Gibbs. Cette dernière méthode a été utilisée par Geman et Geman (1984) pour générer des observations à partir d'une distribution de Gibbs (distribution de Boltzmann), raison pour laquelle on l'appelle la méthode de Gibbs. Il s'agit d'une forme particulière de la méthode de Monte-Carlo par chaîne de Markov (MCMC) qui, du fait de son efficacité, est largement utilisée dans de nombreux domaines d'analyse statistique bayésienne. On parle aussi dans ce chapitre du logiciel **BUGS** (*Bayesian Inference Using Gibbs Sampling*). Ce logiciel de programmation a été développé à l'unité **MRC** de Biostatistique de Cambridge. Il vise à tirer avantage des probabilités de l'échantillonnage de Gibbs dans l'inférence bayésienne sur des systèmes complexes.

Dans le chapitre 2, nous abordons quelques concepts nécessaires à l'étude de l'analyse de survie, tels les fonctions de survie et de risque, les différents types de données censurées. On parle aussi des méthodes fréquentistes d'analyse de survie, notamment la méthode de Kaplan-Meier. Nous rappelons aussi comment est déterminée la fonction de vraisemblance d'un modèle bayésien paramétrique de durée avec censure.

Le chapitre 3 présente deux méthodes bayésiennes qui seront analysées et implémentées dans

le logiciel **BUGS** ; d'abord une méthode semi paramétrique, qui est la méthode de Kalbfleisch et ensuite une autre méthode paramétrique, qui est celle de Carlin, Gelfand et Smith. Grâce à des simulations, nous vérifions l'efficacité de ces deux méthodes bayésiennes. Deux exemples de simulations seront traités, un avec données censurées et l'autre avec points de rupture. Nous démontrons principalement que les estimations par simulation et l'inférence bayésienne paramétrique donnent de meilleurs résultats par rapport aux méthodes classiques.

CHAPITRE I

INTRODUCTION À L'INFÉRENCE BAYÉSIENNE

L'objet principal de la statistique est de faire, à partir d'observations d'un phénomène aléatoire, une inférence au sujet de la loi générant ces observations en vue d'analyser le phénomène ou de prévoir un événement futur.

Pour réduire la complexité du phénomène étudié, deux approches statistiques sont utilisées dans la littérature : modélisation non paramétrique et modélisation paramétrique. Dans la modélisation nonparamétrique, on considère que l'inférence statistique doit prendre en considération la complexité autant que possible et donc on cherche à estimer la distribution de l'ensemble du phénomène, mettant en oeuvre l'estimation des fonctionnelles (densité, régression, etc.). Par opposition, le modèle paramétrique consiste d'observations d'une variable $X = \{X_1, \dots, X_n\}$, d'une fonction de densité $f(x | \theta)$, où seul le paramètre θ , appartenant à un sous-ensemble Θ d'un espace vectoriel de dimension finie, n'est pas connu.

D'une manière plus générale, une fois le modèle construit, on cherche à établir une inférence sur θ , c'est-à-dire à utiliser les observations x afin d'évaluer θ , en vue d'une décision liée à ces paramètres.

Notons que la démarche statistique est fondamentalement une démarche d'inversion qui est de remonter des « effets »(observations) aux « causes »(paramètres).

Pour certains auteurs, ces deux approches (fréquentiste ou probabiliste, bayésienne) s'opposent alors que d'autres voient en elles une certaine complémentarité. Cette opposition réside en fait dans la démarche même de ces deux approches.

1.1 Approche statistique classique

Dans l'approche fréquentiste, les données observées sont considérées comme des observations de variables aléatoires. Elles servent alors à faire porter l'inférence sur les paramètres Θ ayant dirigé leur mécanisme de génération. Autrement dit, l'information provenant des données observées est l'unique source d'information.

Dans cette approche, l'inversion est flagrante dans la notion de vraisemblance. En effet, on écrit

$$l(\theta | x) = f(\mathbf{x} | \theta).$$

en considérant $l(\theta | x)$ comme une fonction de θ , on la normalise (quand cela est possible) pour en faire une fonction de densité sur θ et on l'utilise dans l'estimation de θ . Par exemple, en estimation, on cherche la valeur $\hat{\theta}_{MV}$ qui maximise $l(\theta | x)$, c'est l'estimation au sens du maximum de vraisemblance :

$$\hat{\theta}_{MV} = \arg_{\theta \in \Theta} \max\{l(\theta | x)\}.$$

On utilise $l(\theta | x)$ comme si elle était une fonction de densité de probabilité conditionnelle aux observations \mathbf{x} . Cette inversion est purement formelle alors que, dans l'approche bayésienne, comme nous le verrons un peu plus tard, cette inversion se fait d'une manière plus satisfaisante par la règle de Bayes.

1.2 Approche bayésienne

Dans la section précédente, nous avons considéré le cas où le paramètre inconnu θ est **non aléatoire**. Dans l'approche bayésienne, l'idée de base consiste à traiter le paramètre inconnu θ comme une variable aléatoire admettant une densité de probabilité $\pi(\theta)$ qui s'appelle **densité a priori**.

L'objectif est donc d'utiliser cette information supplémentaire. Sachant que l'information contenue dans les observations \mathbf{x} est contenue dans $\pi(\mathbf{x} | \theta)$ et l'information *a priori* sur θ dans $\pi(\theta)$, on peut utiliser la règle de Bayes pour combiner ces deux types d'informations en définissant la **densité a posteriori** par :

$$\pi(\theta | \mathbf{x}) = \frac{\pi(\mathbf{x} | \theta)\pi(\theta)}{\int \pi(\mathbf{x} | \theta)\pi(\theta)d\theta}.$$

qui contiendra donc toutes les informations sur θ .

On remarque que l'inversion de cause à effet est ici beaucoup plus naturelle. Elle se fait d'une manière cohérente, car l'état de connaissance *a priori* sur θ traduite par la densité *a priori* $\pi(\theta)$ est transformé, après les observations \mathbf{x} , en état de connaissance *a posteriori* par la densité *a posteriori* $\pi(\theta | \mathbf{x})$.

Remarquons que la densité *a posteriori* peut s'écrire

$$\pi(\theta | \mathbf{x}) \propto l(\theta | \mathbf{x})\pi(\theta).$$

1.2.1 Modèle bayésien

On appelle **modèle bayésien** la donnée d'un modèle paramétrique, $\pi(\mathbf{x} | \theta)$, et d'une densité *a priori* $\pi(\theta)$ sur les paramètres.

Étant donné la densité des observations $\pi(\mathbf{x} | \theta)$ et la densité *a priori* $\pi(\theta)$ on peut construire :

i) la densité jointe de (θ, \mathbf{x}) ,

$$\phi(\theta, \mathbf{x}) = \pi(\mathbf{x} | \theta)\pi(\theta);$$

ii) la densité marginale de \mathbf{x} ,

$$m(\mathbf{x}) = \int \phi(\theta, \mathbf{x})d\theta = \int \pi(\mathbf{x} | \theta)\pi(\theta)d\theta;$$

et

iii) la densité *a posteriori* de θ , obtenue par formulation de Bayes,

$$\begin{aligned} \pi(\theta | \mathbf{x}) &= \frac{\pi(\mathbf{x} | \theta)\pi(\theta)}{\int \pi(\mathbf{x} | \theta)\pi(\theta)d\theta} \\ &= \frac{\pi(\mathbf{x} | \theta)\pi(\theta)}{m(\mathbf{x})} \\ &\propto \pi(\mathbf{x} | \theta)\pi(\theta). \end{aligned}$$

Autrement dit, la densité *a posteriori* représente une actualisation de l'information *a priori* au vu de l'information apportée par les observations.

Les méthodes bayésiennes se sont fortement développées ces dernières années (Richardson, 1996) notamment dans le domaine biomédical (Gelfand et Smith, 1990). Un des faits moteurs

à cela vient de l'essor des méthodes de simulations permettant de générer la distribution *a posteriori* lorsque celle-ci est d'écriture complexe et donc difficilement calculable analytiquement. En effet, il n'est pas toujours possible d'obtenir une écriture analytique simple de la distribution *a posteriori*. Cependant, lorsque les lois utilisées appartiennent à la famille exponentielle, on peut alors avoir recours aux lois dites conjuguées. Lorsque la loi des observations $\pi(\mathbf{x} | \theta)$ est une loi usuelle, une loi *a priori* conjuguée permet d'expliciter la loi *a posteriori* comme une loi du même type que la loi *a priori*. si par exemple $\pi(\mathbf{x} | \theta)$ appartient à la famille exponentielle, alors il existe une loi conjuguée naturelle pour $\pi(\theta)$ et $\pi(\theta | \mathbf{x})$. Quelques exemples de lois conjuguées naturelles sont donnés dans le tableau 1.1 (Le lecteur trouvera les démonstrations et propriétés pour ces lois dans des ouvrages de référence tels que : ((Berger, 1985) et (Robert, 1992)).

Lorsqu'il n'y a pas d'écriture analytique simple de $\pi(\theta | \mathbf{x})$ ou que la distribution *a posteriori* est multidimensionnelle, on pourra avoir recours à l'utilisation de méthodes de simulation pour générer des échantillons de variables aléatoires suivant la distribution de $\pi(\theta | \mathbf{x})$. Ces méthodes de simulations reposent sur des algorithmes stochastiques faisant intervenir des procédures itératives d'échantillonnage. On va ainsi échantillonner successivement les composantes du vecteur de paramètre θ et créer une suite dépendante de valeurs de ces composantes, mais possédant des propriétés markoviennes. Les algorithmes les plus couramment utilisés appartiennent à la famille des algorithmes de Monte Carlo par chaîne de Markov (MCMC). Il s'agit de l'échantillonneur de Gibbs ((1984) et (Gelfand et Smith, 1990)) ainsi que de l'algorithme de Metropolis (Metropolis, Rosenbluth et Teller, 1953), ou de son extension par Hastings (Hastings, 1970). Le principe de l'échantillonneur de Gibbs est décrit dans la section 1.6.

$\pi(D \theta)$	$\pi(\theta)$	$\pi(\theta D)$
$Normale(\theta, \sigma^2)$	$Normale(\mu, \tau^2)$	$Normale\left\{\rho(\sigma^2\mu + \tau^2 D), \rho\sigma^2\tau^2\right\}$ où $\rho^{-1} = \sigma^2 + \tau^2$
$Poisson(\theta)$	$Gamma(\alpha, \beta)$	$Gamma(\alpha + D, \beta + 1)$
$Binomiale(n, \theta)$	$Beta(\alpha, \beta)$	$Beta(\alpha + D, \beta + n - D)$
$Gamma(\nu, \theta)$	$Gamma(\alpha, \beta)$	$Gamma(\alpha + \nu, \beta - D)$

Tableau 1.1 Exemple de lois conjuguées naturelles pour des lois appartenant à la famille exponentielle

1.2.2 Modèle bayésien hiérarchique

L'analyse bayésienne hiérarchique modélise l'information en niveaux successifs et conditionnels, en distinguant entre apports structurels et apports proprement subjectifs et, fidèle à la logique bayésienne, en modélisant l'incertitude sur la loi *a priori* par une **loi sur les paramètres de cette loi**. Dans la plupart des cas, la loi de premier niveau sera conjuguée par souci de simplification et aussi parce que la modélisation sur les niveaux supérieurs permet de corriger éventuellement cette erreur de spécification de l'information *a priori*.

On appelle modèle bayésien hiérarchique la donnée d'un modèle statistique bayésien, $(\pi(\mathbf{x} | \theta), \pi(\theta))$, où la loi *a priori* $\pi(\theta)$ est décomposée en distributions conditionnelles

$$\pi_1(\theta | \theta_1), \pi_2(\theta_1 | \theta_2), \dots, \pi_d(\theta_{d-1} | \theta_d),$$

et en une distribution marginale $\pi_{d+1}(\theta_d)$. Les paramètres θ_i sont dits *hyperparamètres*.

Remarquons qu'un modèle bayésien hiérarchique

$$x \sim \pi(\mathbf{x} | \theta), \quad \theta \sim \pi_1(\theta | \theta_1), \quad \theta_1 \sim \pi_2(\theta_1 | \theta_2), \quad \dots, \quad \theta_{d-1} \sim \pi_d(\theta_{d-1} | \theta_d), \quad \theta_d \sim \pi_{d+1}(\theta_d),$$

est un cas particulier de modèle bayésien usuel

$$x \sim \pi(\mathbf{x} | \theta), \quad \theta \sim \pi(\theta),$$

avec

$$\pi(\theta) = \int_{\Theta_1 \times \dots \times \Theta_d} \pi_1(\theta | \theta_1) \pi_2(\theta_1 | \theta_2) \dots \pi_d(\theta_{d-1} | \theta_d) \pi_{d+1}(\theta_d) d\theta_1 \dots d\theta_d.$$

Cette réduction illustre l'aspect « classique » des modèles hiérarchiques et montre donc que cette approche particulière bénéficiera des propriétés générales de l'approche bayésienne avec, en plus, des avantages propres liés à la décomposition de la loi *a priori*.

L'approche bayésienne hiérarchique tend à modéliser le manque d'information sur les paramètres de la loi *a priori* par une nouvelle distribution sur ces paramètres.

Bien que conceptuellement plus abstraite, cette méthode est souvent préférée à une approche bayésienne empirique qui utilise les observations (et la loi marginale) pour estimer les paramètres de la loi *a priori*.

La technique bayésienne d'échantillonnage qui utilise les structures hiérarchiques du modèle bayésien est fondée sur les travaux de (1984) et (Gelfand et Smith, 1990). On évalue $\pi(\theta | x)$ de la manière suivante :

$$\pi(\theta | x) = \int_{\Theta_1} \pi_1(\theta | x, \theta_1) \pi_2(\theta_1) d\theta_1.$$

Si les lois $\pi_1(\theta | x, \theta_1)$, $\pi_2(\theta_1)$ sont connues, générer θ suivant $\pi(\theta | x)$ revient en fait à générer θ_1 suivant $\pi_2(\theta_1)$ puis θ suivant $\pi_1(\theta | x, \theta_1)$.

1.3 Lien entre les deux approches

En pensant à l'aspect d'inversion de la statistique, il est tentant de considérer, sous réserve de l'intégrabilité, la fonction de vraisemblance $l(\theta | \mathbf{x})$ comme une loi de probabilité sur θ dont l'estimateur du maximum de vraisemblance serait la mode. En effet, celle-ci est équivalente à la densité *a posteriori* $\pi(\theta | \mathbf{x})$ lorsqu'on choisit une loi *a priori* uniforme définie sur $[0, 1]$ pour $\pi(\theta)$. Laplace considérait que l'absence d'information *a priori* justifiait le choix de la loi uniforme. Cette position, défendable lorsque θ est un paramètre de position, conduit cependant à des paradoxes et à des contradictions qui montrent clairement la nécessité de la théorie bayésienne plus élaborée incluant les notions des lois *a priori non informatives*, des lois *conjugées* et des lois *a priori de références* (Robert, 1992).

L'approche à privilégier pour l'inférence bayésienne est celle passant par la loi *a posteriori*. En

effet, travaillant conditionnellement aux observations, cette approche suit d'une manière cohérente l'idée d'inversion des causes aux effets, tout en restant fidèle au principe de vraisemblance. En fait, la loi *a posteriori* représente l'actualisation de l'information *a priori*, $\pi(\theta)$, au vu de l'information contenue dans les observations \mathbf{x} , au travers de la vraisemblance $l(\theta | \mathbf{x})$.

Disposant ainsi d'une distribution de probabilité sur θ , le champ de l'inférence est beaucoup plus vaste que dans le cadre classique qui se contentait de $l(\theta | \mathbf{x})$. Nous avons calculer par exemple la moyenne, la médiane ou le mode.

À ce stade, on peut dire que l'approche statistique fréquentiste où toute l'inférence est basée sur la vraisemblance est un cas particulier de l'approche bayésienne avec une densité *a priori* uniforme définie sur $[0, 1]$. En effet, si $\pi(\theta) = c$, la densité *a posteriori* $\pi(\theta | \mathbf{x}) \propto l(\theta | \mathbf{x})$ et, d'après le principe de vraisemblance, toute inférence tirée de ces deux approches sera équivalente. Cependant, seule l'approche bayésienne permet d'introduire des informations complémentaires sur θ sous forme d'une densité *a priori* non uniforme.

Une fois la densité *a posteriori* calculée, il reste encore à savoir l'utiliser correctement.

1.4 Inférence bayésienne

Une inférence bayésienne est fondée sur la détermination rigoureuse de trois facteurs :

1. la loi des observations $\pi(\mathbf{x} | \theta)$
2. la distribution *a priori* des paramètres $\pi(\theta)$
3. le coût associé aux décisions $C(\theta, \delta(\mathbf{x}))$

Notons que les critiques fréquentes de l'approche bayésienne se font souvent sur le point 2, alors que, conceptuellement, les points 1 et 3 se trouvent sur le même rang et que le point 3, qui est un point commun entre l'approche classique et l'approche bayésienne, est encore plus subtil.

Le choix d'une fonction de coût dépend de l'application concernée. Lorsque le contexte ne conduit pas à la détermination de la fonction d'utilité qui permet d'en déduire la fonction de coût, on peut avoir recours à des fonctions de coûts classiques, qui sont à la fois simples et bien étudiées.

1.4.1 Coût quadratique

Parmi les fonctions de coûts les plus utilisées on peut noter le *coût quadratique*

$$C(\theta, d) = (\theta - d)^2.$$

qui pénalise trop fortement les grandes erreurs, du fait de sa convexité stricte.

1.4.2 Estimateur de Bayes

L'estimateur de Bayes δ de $\theta \in \Theta$, associé à la loi *a posteriori* $\pi(\theta | \mathbf{x})$ et au coût quadratique $C(\theta, d) = (\theta - d)^2$, est la **moyenne a posteriori**

$$\theta_m = E[\theta | \mathbf{x}] = \int_{\Theta} \theta \pi(\theta | \mathbf{x}) d\theta. \quad (1.1)$$

En effet, on a

$$E[C(\theta, d)] = d^2 - 2dE[\theta] + E[\theta^2] = (d - E[\theta])^2 + Var[\theta],$$

qui est minimisé lorsque $d = E[\theta]$.

Un tel estimateur prend aussi la forme de l'extremum de la distribution *a posteriori* :

$$\theta_{MAX} = \arg \max_{\theta \in \Theta} \{\pi(\theta | x)\}. \quad (1.2)$$

De tels calculs nécessitent alors de mettre en oeuvre des techniques d'intégration et/ou d'optimisation. Malheureusement, les expressions (1.1) ou (1.2) des estimateurs sont souvent compliquées et les calculs ne peuvent généralement être menés analytiquement que dans des cas simples, notamment pour des modèles linéaires et des distributions de probabilité standards.

1.4.3 Intervalle de crédibilité bayésien

L'intervalle de crédibilité est un reflet réel de la confiance que l'on peut avoir sur la valeur du paramètre concerné. Il met en évidence le point de vue original de l'approche bayésienne qui ne considère pas que le paramètre est une quantité fixe inconnue mais une variable aléatoire dont nous avons une appréciation plus ou moins exacte.

Par définition, un **intervalle de crédibilité** au niveau α est un intervalle tel que la probabilité du paramètre de lui appartenir selon la distribution *a posteriori* est de $(1 - \alpha)$. Il y a plusieurs manières de le construire ; par exemple, on peut prendre celui donné par les $\alpha/2$ et $(1 - \alpha/2)$ quantiles.

Si on s'intéresse à plusieurs paramètres simultanément, on parlera de **région de crédibilité**.

1.4.4 Méthodes d'estimation bayésiennes

L'approche bayésienne nécessite le calcul d'estimateurs suivant une distribution *a posteriori* d'intérêt. Les estimateurs de type *maximum a posteriori* et *moyenne a posteriori* sont donc calculables par des techniques d'intégration et d'optimisation (voir les expressions (1.1) et (1.2)). La mise en pratique de l'approche bayésienne est toutefois problématique car il est souvent impossible de trouver une expression analytique directement exploitable de ces estimateurs. De plus, les dimensions des variables considérées rendent souvent très difficiles les calculs d'intégration et d'optimisation, même dans le cas où une forme analytique des estimateurs est disponible explicitement.

Des solutions classiques pour résoudre ces problèmes d'implantation sont basées sur des méthodes d'approximations asymptotiques et de calculs numériques. Dans ce qui suit, des méthodes de simulation de type Monte-Carlo sont utilisées afin de calculer les estimateurs d'intérêt. Ces méthodes sont présentées dans la section suivante.

1.5 Méthodes d'estimation Monte-Carlo

Les méthodes dites « Monte-Carlo » (Robert, 1992) permettent entre autres d'estimer les modes et les espérances d'une distribution. Ces méthodes utilisent une approximation de la distribution d'intérêt construite à partir de séries d'échantillons simulés.

Le principe fondateur de ces méthodes est d'utiliser une série d'échantillons $(\theta_1, \dots, \theta_n)$ simulés suivant la loi de probabilité d'intérêt $\pi(\theta | x)$ pour mener à bien les calculs. L'estimateur de la moyenne *a posteriori* donné par (1.1) est approché d'après une série d'échantillons

$(\theta_1, \dots, \theta_n) \sim \pi(\theta | x)$ par l'expression

$$\hat{\theta}_m = \sum_{i=1}^n \theta_i. \quad (1.3)$$

Malheureusement, une telle méthode d'estimation est encore limitée par la modélisation car les lois de probabilité facilement simulables ne sont pas très nombreuses. Néanmoins, il est souvent nécessaire de mettre en oeuvre des techniques de simulation sophistiquées pour échantillonner suivant les distributions d'intérêt. Des méthodes de simulation par chaîne de Markov sont utilisées. Ces méthodes sont ainsi présentées dans la section suivante.

1.6 Méthodes de simulation par chaînes de Markov

Les techniques d'estimation Monte-Carlo permettent d'implanter le calcul d'estimateurs issus de procédures d'estimation bayésienne. Toutefois, ces approches nécessitent de pouvoir simuler des séquences d'échantillons suivant les distributions de probabilité d'intérêt. Or, ces distributions sont souvent multivariées, de type non standard. Il est donc nécessaire de considérer des méthodes de simulation statistique élaborées. Les schémas de simulation par chaînes de Markov sont ainsi considérés.

Un moyen de produire des échantillons suivant une loi d'intérêt donnée est d'utiliser une méthode de simulation par chaînes de Markov (Gilks, Richardson et Spiegelhalter, 1996) et (Robert, 1992). Le principe d'une telle méthode de simulation est de générer une suite d'échantillons $(\theta_0, \theta_1, \dots, \theta_n, \dots)$ de manière récursive : $\theta_n \sim q(\theta_n | \theta_{n-1})$ avec une loi $q(\cdot | \cdot)$ aisément simulable, de telle manière que les échantillons θ_n soient asymptotiquement distribués suivant la distribution d'intérêt : $\theta_n \sim \pi(\theta | x)$ pour n assez grand. La loi de transition $q(\cdot | \cdot)$ doit ainsi être choisie en fonction de la distribution d'intérêt $\pi(\theta | x)$ et certaines conditions sont nécessaires pour qu'un tel schéma de simulation, produisant par définition des échantillons corrélés, puisse converger et générer des variables aléatoires suivant une loi invariante. Ce dernier point est important pour assurer que le calcul des estimateurs (1.1) et (1.2) d'après les expressions (1.3) est possible à partir d'une séquence d'échantillons $(\theta_0, \theta_1, \dots, \theta_n)$. De telles approches, utilisant ces techniques d'estimation et de simulation, sont appelées méthodes de simulation Monte-Carlo par chaîne de Markov (MCMC) et permettent de traiter de nombreux problèmes

d'estimation (Gilks, Richardson et Spiegelhalter, 1996) même pour des modèles non linéaires et/ou non-Gaussiens.

1.6.1 Pratique des méthodes MCMC

L'utilisation intensive des méthodes de simulation MCMC est due à leur adaptabilité pour une vaste classe de problèmes et de modèles mais également à l'amélioration des technologies informatiques. Les algorithmes sont faciles à implanter et un simple ordinateur de bureau est suffisant pour mettre en oeuvre la réalisation des simulations. La part la plus importante du travail, et sans doute une des plus intéressantes, est l'élaboration du schéma de simulation pour un modèle donné : distribution *a priori* des paramètres du modèle (Robert, 1992), distribution candidate pour l'algorithme de Metropolis-Hastings et distributions conditionnelles pour l'échantillonnage de Gibbs (Robert, 1992) par exemple. Toutefois, un aspect important de l'implantation de ce type de méthode est qu'il faut utiliser plusieurs réalisations ou simulations sur un ensemble de données pour obtenir des résultats statistiquement significatifs, étant donné la nature aléatoire des algorithmes de simulation. Les chaînes de Markov sont généralement très riches en informations statistiques, plus qu'il n'en faut pour calculer seulement les estimateurs de type (1.1) et (1.2).

Deux algorithmes MCMC sont présentés dans la section suivante : l'algorithme de Metropolis-Hastings et l'échantillonnage de Gibbs. Ces méthodes permettent de construire un noyau de transition en fonction d'une distribution d'intérêt donnée telle que les séquences d'échantillons simulés qui forment une chaîne de Markov convergente vers la distribution stationnaire voulue.

1.6.2 Algorithme de Metropolis-Hastings

L'algorithme de Metropolis-Hastings est un schéma de simulation permettant de générer des échantillons suivant une distribution de $\pi(x)$ dont une forme analytique est disponible. Cette méthode a d'abord été utilisée pour la résolution de problèmes en mécanique statistique avant d'être élargie à un cadre plus général de simulation statistique. Bien que les premiers résultats sur cette méthode de simulation soient connus depuis les années 1950, son utilisation pratique

n'est devenue possible que depuis une trentaine d'années car elle nécessite de puissants moyens de simulation informatique.

Schéma de simulation : l'algorithme de Metropolis-Hastings est étudié et présenté dans de nombreux ouvrages consacrés aux méthodes de simulation MCMC (voir (Robert, 1992)). Le schéma de simulation est décrit ci-dessous :

Algorithme de Metropolis-Hastings

1. initialiser $\theta_0 \sim \pi_0(x)$, $k = 0$
2. itération k :
 - proposer un candidat θ^* pour θ_{k+1} : simuler :

$$\theta^* \sim q(\theta^*|\theta_k)$$

- accepter le candidat θ^* avec la probabilité

$$\alpha(\theta^*, \theta_k) = \min\left\{1, \frac{\pi(\theta^*|\mathbf{x})}{\pi(\theta_k|\mathbf{x})}\right\}$$

simuler $u \sim U_{[0,1]}$

- si $u \leq \alpha(\theta^*, \theta_k)$, alors prendre $\theta_{k+1} = \theta^*$, sinon poser $\theta_{k+1} = \theta_k$

3. $k \leftarrow k + 1$ et aller en (2)

Il est important de noter que l'algorithme de Metropolis-Hastings permet de simuler des échantillons suivant la distribution π à partir de distributions candidates q générales dont le support est inclus dans celui de π . Toutefois, le schéma de simulation ne fonctionnera pas bien en pratique si les zones de fortes probabilités de q sont localisées surtout dans les queues de distribution de π .

Notons aussi que l'algorithme de Metropolis-Hastings ne génère pas d'échantillons i.i.d. . Toutefois, la corrélation des échantillons générés peut parfois s'avérer utile pour améliorer le comportement des estimateurs et permettre de faciliter la visualisation de la distribution des échantillons comme une courbe ou une surface évoluant au cours des itérations de l'algorithme.

L'algorithme de Metropolis-Hastings peut présenter des difficultés à explorer convenablement

le support de la distribution d'intérêt. Cette situation s'aggrave d'autant plus dans le cas de distributions multidimensionnelles, le nombre d'échantillons nécessaires pour obtenir une couverture suffisante du support de la distribution d'intérêt devenant très important.

Des algorithmes MCMC sont alors spécifiquement adaptés à la simulation de distributions multidimensionnelles.

1.6.3 L'échantillonnage de Gibbs

Cette méthode a été utilisée par Geman et Geman (1984) pour générer des observations à partir d'une distribution de Gibbs (distribution de Boltzmann). Il s'agit d'une forme particulière de la méthode de Monte-Carlo par chaîne de Markov qui, du fait de son efficacité, est largement utilisée dans de nombreux domaines d'analyse statistique bayésienne.

Dans la méthode de Gibbs, après avoir choisi un point départ, les d composantes du vecteur de covariables (θ) sont générées les unes après les autres conditionnellement à toutes les autres composantes. Si $\pi(\theta | x)$ est la densité des d composantes du vecteur θ , conditionnellement aux données observées (\mathbf{x}), nous utilisons alors les densités conditionnelles $\pi(\theta_1 | \theta_2, \theta_3, \dots, \theta_d, \mathbf{x})$, $\pi(\theta_2 | \theta_1, \theta_3, \dots, \theta_d, \mathbf{x})$, et ainsi de suite. À chaque k^e étape, la distribution conditionnelle utilise les valeurs générées les plus récentes parmi toutes les autres composantes. Par la théorie des chaînes de Markov, on a que lorsque $k \rightarrow \infty$ la densité des réalisations obtenues converge vers $\pi(\theta | x)$. Le schéma de simulation est décrit ci-dessous :

Algorithme d'échantillonnage de Gibbs

1. Fixer $k=0$.
2. – Générer $\theta_1^{(k+1)}$ à partir de $\pi(\theta_1 | \theta_2^{(k)}, \theta_3^{(k)}, \dots, \theta_d^{(k)}, \mathbf{x})$
 - Générer $\theta_2^{(k+1)}$ à partir de $\pi(\theta_2 | \theta_1^{(k+1)}, \theta_3^{(k)}, \dots, \theta_d^{(k)}, \mathbf{x})$
 - ...

- Générer $\theta_{d-1}^{(k+1)}$ à partir de $\pi(\theta_{d-1} | \theta_1^{(k+1)}, \theta_2^{(k+1)}, \dots, \theta_d^{(k)}, \mathbf{x})$
 - Générer $\theta_d^{(k+1)}$ à partir de $\pi(\theta_d | \theta_1^{(k+1)}, \theta_2^{(k+1)}, \dots, \theta_{d-1}^{(k+1)}, \mathbf{x})$.
3. Si la convergence est obtenue, alors
- 2.a) retenir $\theta = \theta^{(k+1)}$
sinon
 - 2.b) fixer $k = k + 1$, et retourner à 1

Le schéma de simulation est d'autant plus efficace lorsque les distributions conditionnelles

$$\pi_i(\theta_i | \theta_1, \dots, \theta_{(i-1)}, \theta_{(i+1)}, \dots, \theta_d)$$

sont simulables rapidement.

Un des points remarquables de l'algorithme de Gibbs est qu'il rend possible la simulation de la distribution jointe $\pi(\theta) = \pi(\theta_1, \dots, \theta_d)$ à partir des distributions conditionnelles $\pi_i(\theta_i | \theta_{j \neq i})$.

Notons que l'échantillonnage de Gibbs s'applique particulièrement bien aux modèles hiérarchiques, c'est-à-dire à un modèle bayésien $(\pi(\mathbf{x} | \theta), \pi(\theta))$, où la loi *a priori* $\pi(\theta)$ est décomposée en distributions conditionnelles

$$\pi_1(\theta | \theta_1), \pi_2(\theta_1 | \theta_2), \dots, \pi_d(\theta_{d-1} | \theta_d),$$

pour des raisons structurelles ou calculatoires. De tels modèles apparaissent naturellement dans l'analyse bayésienne de structures complexes, où la diversité des informations *a priori* et la variabilité des observations nécessitent l'introduction de plusieurs niveaux de lois *a priori* (Robert, 1992).

Dans le cas particulier à deux composantes par exemple, la distribution jointe s'écrit d'après la formule de Bayes suivant les expressions

$$\pi(\theta_1, \theta_2) = \pi_1(\theta_1 | \theta_2) \times \pi^1(\theta_2) = \pi_2(\theta_2 | \theta_1) \times \pi^2(\theta_1), \quad (1.4).$$

où $\pi^1(\theta_2)$ et $\pi^2(\theta_1)$ désignent les distributions marginales. Il s'ensuit alors

$$\pi^1(\theta_2) = \frac{\pi_2(\theta_2 | \theta_1) \times \pi^2(\theta_1)}{\pi_1(\theta_1 | \theta_2)} \propto \frac{\pi_2(\theta_2 | \theta_1)}{\pi_1(\theta_1 | \theta_2)}.$$

Ainsi, il est possible d'exprimer les densités des probabilités des distributions marginales à partir de celles des distributions conditionnelles, par exemple :

$$\pi^1(\theta_2) = \left(\int \frac{\pi_2(\theta_2 | x)}{\pi_1(x | \theta_2)} dx \right)^{-1} \frac{\pi_2(\theta_2 | \theta_1)}{\pi_1(\theta_1 | \theta_2)}.$$

La densité de la distribution jointe (1.4) s'écrit donc

$$\pi(\theta_1, \theta_2) = \left(\int \frac{\pi_2(\theta_2 | x)}{\pi_1(x | \theta_2)} dx \right)^{-1} \pi_2(\theta_2 | \theta_1).$$

Le cas général pour d composantes nécessite certaines conditions sur les lois conditionnelles, (voir (Robert, 1992)).

1.7 Efficacité asymptotique des estimateurs bayésiens

1.7.1 Efficacité asymptotique de l'estimateur du maximum de vraisemblance

Soient X_1, \dots, X_n n variables aléatoires indépendantes issues de la même loi $f(\mathbf{x}, \theta)$ (par rapport à une mesure μ); supposons que le paramètre θ inconnu se représente par la relation suivante :

$$\theta = g(\theta)$$

; alors, une suite $\delta_n(X_1, \dots, X_n)$ d'estimateurs de $g(\theta)$ qui satisfait :

$$\sqrt{n} [\delta_n - g(\theta)] \xrightarrow{\mathcal{L}} \mathbf{N} [0, v(\theta)] \quad v(\theta) > 0.$$

avec

$$v(\theta) = \frac{[g'(\theta)]^2}{\mathbf{I}(\theta)}, \quad \mathbf{I}(\theta) = E_\theta \left[\left(\frac{\partial}{\partial \theta} \log \prod_{i=1}^n f(X_i | \theta) \right)^2 \right] \text{ (est l'information de Fisher)}$$

est **asymptotiquement efficace**.

Théorème 1.7.1 (Bahadur, 1964 (on trouve la preuve du théorème)) Soit X_1, \dots, X_n n variables aléatoires indépendantes issues de la même loi $f(x, \theta)$, avec $\theta \in \Theta \subseteq \mathbb{R}^k$, et supposons qu'on a les conditions de régularité suivantes :

- (i) Θ intervalle ouvert
- (ii) Les lois P_θ des X_i ont un support commun, l'ensemble $A = \{x \mid f(x, \theta) > 0\}$ est indépendant de θ .
- (iii) $\forall x \in A$, $f(x, \theta)$ est deux fois dérivable par rapport à θ et $\frac{\partial^2 f(x, \theta)}{\partial \theta^2}$ est continue en θ .
- (iv) On peut dériver 2 fois $\int f(\mathbf{x}, \theta) d\mu(\mathbf{x})$ sous le signe d'intégration par rapport à θ .
- (v) $\mathbf{I}(\theta)$ satisfait $0 < \mathbf{I}(\theta) < \infty$.

(vi) $\forall \theta_0 \in \Theta, \exists c > 0, \exists M(x)$ telle que

$$\left| \frac{\partial^2 \log f(x, \theta)}{\partial \theta^2} \right| \leq M(x),$$

$\forall x \in A, \theta_0 - c < \theta < \theta_0 + c$ et $E_{\theta_0} [M(X)] < \infty$.

sous ces conditions, si $\delta_n = \delta_n(X_1, \dots, X_n)$ est un **estimateur asymptotiquement normal**, alors on a

$$v(\theta) \geq \frac{[g'(\theta)]^2}{\mathbf{I}(\theta)},$$

$\forall \theta$ sauf pour les θ appartenant à un ensemble de mesure nulle.

On a vu qu'un estimateur du maximum de vraisemblance (E.M.V.) de θ est défini comme toute solution $\hat{\theta}_n$ de l'équation :

$$p(X, \hat{\theta}_n) = \sup_{\theta \in \Theta} p_n(X, \theta).$$

On définit la log-vraisemblance de l'observation X par :

$$L(\theta) = \log p_n(X, \theta)$$

avec

$$p_n(X, \theta) = \prod_{i=1}^n f(x_i | \theta).$$

On suppose que $L(\theta)$ est une fonction continue, de classe C^2 sur Θ . Ainsi, pour calculer et étudier $\hat{\theta}_n$, on résoudra les équations :

$$\frac{\partial L}{\partial \theta_j}(\theta) = 0 \quad j = 1, \dots, k.$$

De plus, les $\hat{\theta}_n$ vérifient les deux propriétés suivantes :

(a) **Convergence faible**

$$\forall h > 0 \quad P_{\theta} \left(|\hat{\theta}_n - \theta| > h \right) \xrightarrow{n \rightarrow +\infty} 0 \quad \text{pour tout } \theta.$$

Pour montrer la consistance de $\hat{\theta}_n$, on montre en général que, pour tout θ_0 ,

$$P_{\theta_0} \left(\sup_{|\theta - \theta_0| > 0} p_n(X, \theta) \geq p_n(X, \theta_0) \right) \xrightarrow{n \rightarrow +\infty} 0.$$

(b) **Loi asymptotique et vitesse de convergence.** Afin d'étudier la limite en loi d'un E.M.V., la méthode classique consiste à faire un développement de Taylor de $L'(\theta)$ autour de $\hat{\theta}_n$. On supposera :

(H0) Θ ouvert, $L(\theta)$ de classe C^2 sur Θ et $\hat{\theta}_n$ consistant.

Dans le cas unidimensionnel, on obtient l'équation :

$$0 = L'(\hat{\theta}_n) = L'(\theta) + (\hat{\theta}_n - \theta) \left[L''(\theta) + R_n(\hat{\theta}_n - \theta) \right]$$

avec

$$R_n(h) = \int_0^1 (L''(\theta + sh) - L''(\theta)) ds.$$

D'où :

$$\hat{\theta}_n - \theta = - \frac{L'(\theta)}{L''(\theta) + R_n(\hat{\theta}_n - \theta)}$$

La démarche à suivre est alors la suivante :

On suppose qu'il existe une fonction déterministe $c(n) = \sqrt{n}$ et telle que

$$\lim_{n \rightarrow +\infty} c(n) = +\infty$$

et :

(H1) $-\frac{1}{c^2(n)} L''(\theta)$ converge en probabilité sous P_θ vers la constante positive déterministe $\mathbf{I}(\theta)$.

(H2) $\frac{1}{c(n)} L'(\theta)$ converge en loi sous P_θ vers la loi gaussienne $\mathbf{N}(0, \mathbf{I}(\theta))$.

(H3) $\sup_{|\alpha| \leq r_n} \left| \frac{1}{c^2(n)} (R_n(\theta - \alpha) - R_n(\theta)) \right| \xrightarrow{n \rightarrow +\infty} 0$ en probabilité sous P_θ pour toute v.a. r_n telle que $r_n \xrightarrow{P_\theta} 0$.

Alors sous (H0)-(H3), $\sqrt{n}(\hat{\theta}_n - \theta)$ converge en loi sous P_θ^n vers la loi gaussienne $\mathbf{N}(0, \mathbf{I}^{-1}(\theta))$.

1.7.2 Efficacité asymptotique des estimateurs bayésiens

Un des résultats les plus importants dans cette section est le **théorème 1.7.3** (Bickel, 1969) qui montre que, sous des conditions appropriées, la distribution des estimateurs bayésiens basés sur n v.a. i.i.d. tend à devenir indépendante de la distribution *a priori* quand n tend vers l'infini et on déduit que les estimateurs bayésiens sont asymptotiquement efficaces.

Soit X_1, \dots, X_n n variables aléatoires indépendantes issues de la même loi $f(x_i, \theta)$, avec $\theta \in \Theta \subseteq R^k$. On note par θ_0 la valeur du paramètre à estimer.

(B1) La fonction $L(\theta) = \sum_{i=1}^n \log f(x_i | \theta)$ vérifie les conditions de régularité du théorème 1.7.1.

Sous les conditions du théorème 1.7.1, si $\tilde{\theta} = \tilde{\theta}_n$ est une suite d'estimateurs bayésiens telle que $\tilde{\theta} \xrightarrow{P} \theta$, alors

$$L(\theta) = L(\theta_0) + (\theta - \theta_0)L'(\theta_0) - \frac{1}{2}(\theta - \theta_0)^2 [n\mathbf{I}(\theta_0) + R_n(\theta)]$$

on a

$$\frac{1}{n}R_n(\theta) \xrightarrow{n \rightarrow \infty} 0.$$

(B2) $\forall \epsilon > 0, \exists \delta > 0$ telle que dans l'expression (1), on a

$$P \left(\sup \left\{ \left| \frac{1}{n}R_n(\theta) \right| : |\theta - \theta_0| \leq \delta \right\} \geq \epsilon \right) \xrightarrow{n \rightarrow \infty} 0$$

(B3) $\forall \delta > 0, \exists \epsilon > 0$ telle que

$$P \left(\sup \left\{ \frac{1}{n} [L(\theta) - L(\theta_0)] : |\theta - \theta_0| \geq \delta \right\} \leq -\epsilon \right) \xrightarrow{n \rightarrow \infty} 1$$

(B4) La loi *a priori* $\pi(\theta)$ est continue et positive, $\forall \theta \in \Theta \subseteq R^k$

(B5) $\int |\theta| \pi(\theta) d\theta < \infty$.

Pour établir l'efficacité asymptotique des estimateurs bayésiens sous les conditions (B1)-(B5), on montre dans le théorème 1.7.2 (Bickel, 1969) que lorsque le nombre d'observations n tend vers l'infini, on a

$$\pi(\theta | \mathbf{x}) \sim \mathbf{N} \left(\theta_0 + \frac{1}{n\mathbf{I}(\theta_0)} L'(\theta_0), \frac{1}{n\mathbf{I}(\theta_0)} \right)$$

On pose

$$T_n = \theta_0 + \frac{1}{n\mathbf{I}(\theta_0)} L'(\theta_0)$$

Théorème 1.7.2 (Bickel, 1969) Si $\pi^*(t | x)$ est la densité a posteriori de $\sqrt{n}(\theta - T_n)$ alors

(i) si les conditions (B1)-(B4) sont satisfaites,

$$\int \left| \pi^*(t | x) - \sqrt{I(\theta_0)} \Phi[t\sqrt{I(\theta_0)}] \right| dt \xrightarrow{P} 0.$$

(ii) si, de plus, la condition (B5) est satisfaite, alors

$$\int (1 + |t|) \left| \pi^*(t | \mathbf{x}) - \sqrt{I(\theta_0)} \Phi[t\sqrt{I(\theta_0)}] \right| dt \xrightarrow{P} 0.$$

Preuve :

$$\sqrt{n}(\theta - T_n) \sim \pi^*(t | \mathbf{x}) \iff \theta \sim \pi^*(T_n + \frac{t}{\sqrt{n}} | \mathbf{x}).$$

Donc

$$\pi^*(t | \mathbf{x}) = \frac{\pi(T_n + \frac{t}{\sqrt{n}}) \exp[L(T_n + \frac{t}{\sqrt{n}})]}{\int \pi(T_n + \frac{u}{\sqrt{n}}) \exp[L(T_n + \frac{u}{\sqrt{n}})] du};$$

$$\pi^*(t | \mathbf{x}) = \exp(\omega(t)) \pi(T_n + \frac{t}{\sqrt{n}}) / C_n$$

avec

$$\omega(t) = L(T_n + \frac{t}{\sqrt{n}}) - L(\theta_0) - \frac{1}{2n\mathbf{I}(\theta_0)} [L'(\theta_0)]^2$$

et

$$C_n = \int \pi(T_n + \frac{u}{\sqrt{n}}) \exp(\omega(u)) du.$$

Si on pose

$$J = \int \left| \pi(T_n + \frac{t}{\sqrt{n}}) \exp(\omega(t)) - C_n \sqrt{\mathbf{I}(\theta_0)} \Phi[t\sqrt{\mathbf{I}(\theta_0)}] \right| dt,$$

alors

$$\int \left| \pi^*(t | \mathbf{x}) - \sqrt{\mathbf{I}(\theta_0)} \Phi[t\sqrt{\mathbf{I}(\theta_0)}] \right| dt = J / C_n.$$

donc (i) \iff

$$J / C_n \xrightarrow{P} 0.$$

Si on montre que

$$(a) \quad J_1 = \int \left| \pi(T_n + \frac{t}{\sqrt{n}}) \exp(\omega(t)) - \pi(\theta_0) \exp(-\frac{1}{2} t^2 \mathbf{I}(\theta_0)) \right| dt \xrightarrow{P} 0,$$

donc

$$C_n \xrightarrow{P} \int \pi(\theta_0) \exp(-\frac{1}{2} t^2 \mathbf{I}(\theta_0)) dt = \pi(\theta_0) \sqrt{\frac{2\pi}{\mathbf{I}(\theta_0)}}.$$

$$\begin{aligned}
\text{(b) } J_2 &= \int \left| C_n \sqrt{\mathbf{I}(\theta_0)} \Phi[t\sqrt{\mathbf{I}(\theta_0)}] - \pi(\theta_0) \exp\left(-\frac{1}{2}t^2 \mathbf{I}(\theta_0)\right) \right| dt \\
&= \left| \frac{C_n \sqrt{\mathbf{I}(\theta_0)}}{\sqrt{2\pi}} - \pi(\theta_0) \right| \int \exp\left(-\frac{1}{2}t^2 \mathbf{I}(\theta_0)\right) dt \xrightarrow{P} 0
\end{aligned}$$

alors

$$J \leq J_1 + J_2 \xrightarrow{P} 0,$$

d'où le résultat (i)

Pour montrer (ii), on remarque que

$$\int (1 + |t|) \left| \pi^*(\mathbf{t} | \mathbf{x}) - \sqrt{\mathbf{I}(\theta_0)} \Phi[t\sqrt{\mathbf{I}(\theta_0)}] \right| dt = \frac{J'}{C_n} \leq \frac{1}{C_n} (J'_1 + J'_2) \xrightarrow{P} 0,$$

avec

$$\begin{aligned}
J' &= \int (1 + |t|) \left| \pi\left(T_n + \frac{t}{\sqrt{n}}\right) \exp(\omega(t)) - C_n \sqrt{\mathbf{I}(\theta_0)} \Phi[t\sqrt{\mathbf{I}(\theta_0)}] \right| dt, \\
J'_1 &= \int (1 + |t|) \left| \pi\left(T_n + \frac{t}{\sqrt{n}}\right) \exp(\omega(t)) - \pi(\theta_0) \exp\left(-\frac{1}{2}t^2 \mathbf{I}(\theta_0)\right) \right| dt, \\
J'_2 &= \int (1 + |t|) \left| C_n \sqrt{\mathbf{I}(\theta_0)} \Phi[t\sqrt{\mathbf{I}(\theta_0)}] - \pi(\theta_0) \exp\left(-\frac{1}{2}t^2 \mathbf{I}(\theta_0)\right) \right| dt.
\end{aligned}$$

Avec les résultats du théorème 1.7.2, on est en mesure de montrer le théorème 1.7.3.

Théorème 1.7.3 (Bickel, 1969) Si les conditions (B1)-(B5) sont satisfaites, et si $\hat{\theta}_n$ est l'estimateur de Bayes pour la densité *a priori* π et la fonction perte est quadratique, alors

$$\sqrt{n}(\tilde{\theta}_n - \theta_0) \xrightarrow{\mathcal{L}} N[0, 1/I(\theta_0)]$$

et $\hat{\theta}_n$ est consistante et asymptotiquement efficace.

Preuve : On a

$$\sqrt{n}(\tilde{\theta}_n - \theta_0) = \sqrt{n}(\tilde{\theta}_n - T_n) + \sqrt{n}(T_n - \theta_0),$$

et par le théorème de la limite centrale

$$\sqrt{n}(T_n - \theta_0) \xrightarrow{\mathcal{L}} N[0, 1/I(\theta_0)].$$

Donc il reste à montrer que

$$\sqrt{n}(\tilde{\theta}_n - T_n) \xrightarrow{P} 0.$$

Or l'estimateur de Bayes par l'erreur quadratique est défini par

$$\tilde{\theta}_n = E[\theta | \mathbf{x}] = \int \theta \pi(\theta | \mathbf{x}) d\theta = \int (T_n + \frac{t}{\sqrt{n}}) \pi^*(\mathbf{t} | \mathbf{x}) dt,$$

donc

$$\sqrt{n}(\tilde{\theta}_n - T_n) = \int t \pi^*(\mathbf{t} | \mathbf{x}) dt.$$

Ainsi

$$\begin{aligned} \sqrt{n}|\tilde{\theta}_n - T_n| &= \left| \int t \pi^*(\mathbf{t} | \mathbf{x}) dt - \int t \sqrt{\mathbf{I}(\theta_0)} \Phi[t \sqrt{\mathbf{I}(\theta_0)}] dt \right| \\ &= \int |t| \left| \pi^*(\mathbf{t} | \mathbf{x}) - \sqrt{\mathbf{I}(\theta_0)} \Phi[t \sqrt{\mathbf{I}(\theta_0)}] \right| dt \xrightarrow{P} 0 \quad \text{par (ii)} \end{aligned}$$

CHAPITRE II

ANALYSE BAYÉSIENNE ET FRÉQUENTISTE DES MODÈLES DE DURÉE AVEC CENSURE

2.1 Introduction à l'analyse de survie

L'analyse de la survie a connu une importante croissance dans la seconde moitié du XX^e siècle. Dans ce domaine, les développements qui ont eu le plus profond impact sur les essais cliniques sont la méthode de Kaplan-Meier (1958) pour l'estimation de la fonction de survie et le modèle des risques proportionnels (Cox, 1972) pour quantifier les effets de covariables sur le temps de survie. La théorie des martingales pour le processus de comptage, développée par Aalen (Aalen, 1978), offre un cadre unifié pour l'étude des propriétés des statistiques d'analyse de survie, aussi bien pour des petits que pour des grands échantillons. Des progrès significatifs ont été réalisés et on peut espérer de nouveaux développements dans plusieurs domaines, comme les données censurées par intervalle, la modélisation jointe de données longitudinales et de données de durée, et les méthodes bayésiennes.

Dans ce chapitre, nous exposons les principales méthodes d'estimation des durées de vie. Nous y définissons les différents concepts de survie et décrivons la dualité qui existe entre une approche fréquentiste et une approche bayésienne. On parle aussi des méthodes fréquentistes d'analyse de survie notamment la méthode de Kaplan-Meier (1958) et on la compare avec les méthodes bayésiennes de survie.

2.1.1 Fonction de survie et de risque instantané

Dans cette section, on définit quelques concepts de l'analyse de survie qui seront utiles dans le chapitre suivant. Les définitions données ici peuvent être trouvées dans les ouvrages qui traitent de l'analyse de survie, comme par exemple Klein et Moeschberger (Klein et Moeschberger, 1997).

Considérons une variable aléatoire T , non négative. La variable T représente le temps auquel un certain événement se produit. Des exemples d'événements sont : la mort après une maladie grave, la guérison après le traitement adéquat d'une maladie, la reprise du travail professionnel après une période de chômage, la panne d'une composante d'une machine, etc.

Notons par F la fonction de répartition de T , c'est-à-dire $F(t) = P\{T \leq t\}$.

La fonction de survie S au temps t est définie par :

$$S(t) = P\{T > t\} = 1 - F(t).$$

Définition 2.1 : Risque instantané de décès

On appelle *risque instantané de décès* ou *fonction de risque au temps t* la fonction :

$$h(t) = \lim_{(\Delta \rightarrow 0)_{\Delta > 0}} \frac{P\{t \leq T < t + \Delta \mid T \geq t\}}{\Delta}$$

Cette fonction décrit la manière dont la probabilité instantanée de mort d'un individu change avec le temps.

Supposons que T soit une variable aléatoire continue qui admet une densité de probabilité f .

Alors, nous avons

$$h(t) = \frac{f(t)}{1 - F(t)} = \frac{f(t)}{S(t)}, \quad t \geq 0.$$

Comme $f(t) = -\frac{d[1-F(t)]}{dt}$, on a aussi

$$h(t) = -\frac{d \log(S(t))}{dt}.$$

La fonction cumulative de risque est défini par

$$H(t) = \int_0^t h(u) du,$$

ce qui peut être réécrit, en utilisant l'expression ci-dessus de $h(t)$, comme

$$H(t) = -\log(S(t)) = -\log(1 - F(t)),$$

ou de manière équivalente

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

2.1.2 Données censurées

On cherche à évaluer l'efficacité d'un nouveau traitement chez des personnes atteintes de maladies graves, qui peuvent donner lieu à des complications, à des rechutes, ou même aboutir au décès. Dans tous les cas, le critère d'intérêt peut s'exprimer comme la durée entre l'instauration du traitement et l'apparition de l'événement témoignant de l'échec du traitement. Pour simplifier l'écriture, on supposera par la suite que l'événement considéré est le décès et on s'intéressera à la durée de survie des individus. Au moment où on effectue l'analyse des données, on dispose en général d'un échantillon composé de deux types d'information :

- Certains individus sont décédés et on dispose alors d'une information complète de la forme $(T = t)$, où T est la variable aléatoire « durée de survie ».
- D'autres individus, au contraire, sont vivants au moment de l'analyse. Tout ce qu'on connaît de leur durée de survie est que celle-ci est supérieure au temps qui s'est écoulé entre l'instauration du traitement et le moment de l'analyse. On a donc une information de la forme $(T < t)$. Une donnée dont on ne connaît qu'une borne est dite *censurée*.

Définition : Censure à droite. Une observation est dite *censurée à droite* au temps T^* si la valeur exacte de celle-ci n'est pas connue après le temps T^* .

On peut définir la censure à gauche de la même façon en substituant au mot « après » le mot « avant ».

On verra dans cette section deux types de censure : de type I et la censure de type III, appelée censure aléatoire (« random »censoring).

Données censurées de type I

Ce type de censure, rencontré souvent dans un contexte de laboratoire, est utilisé lorsque tous les individus sont observés à partir du même instant et que l'étude est soumise à une date limite. Plus formellement, soit n individus numérotés de 1 à n . Tous les individus débutent l'étude à la même date. Ils termineront l'étude à leur décès ou à la fin de celle-ci. Dénotons par C le temps entre le début et la fin de l'étude. Soit T_1, T_2, \dots, T_n le temps entre le début de l'étude et le décès du i ème individu, $i = 1, \dots, n$. L'échantillon sera donc composé de n couples (Y_i, δ_i) , $i = 1, \dots, n$ où

$$Y_i = \min(T_i, C),$$

$$\delta_i = \begin{cases} 1 & \text{si } T_i \leq C, \\ 0 & \text{si } T_i > C. \end{cases}$$

Données censurées de type III

La censure aléatoire (« random censoring ») est utilisée dans un contexte clinique. Lorsqu'une étude médicale est amorcée, l'arrivée des individus dans l'étude dépend du moment où le diagnostic médical est prononcé.

Considérons la situation suivante : une étude a été mise en oeuvre pour évaluer l'efficacité d'un traitement médical. L'étude a démarré en 1980, et s'est arrêtée en 1996, et un certain nombre d'individus ont participé à cette étude : ils ont reçu le traitement et ils ont été suivis jusqu'au moment où l'événement particulier « la mort » se produit. Notons :

- E_i = la date d'entrée de l'individu i dans l'étude ;
- D_i = la date « de la mort » de l'individu i

Le temps de « survie » de l'individu est alors

$$T_i = D_i - E_i.$$

Malheureusement, on est souvent confronté à des situations qui ne permettent pas d'observer les temps de survie de chaque individu. Des exemples :

- **La fin de l'étude** : pour certains individus, l'événement d'intérêt ne s'est pas encore produit à la fin de l'étude ;
- « **Lost to follow up** » : l'individu a décidé de quitter l'étude et on a perdu sa trace ;
- « **Drop out** » : l'individu ne veut plus participer à l'étude (effets secondaires du médicament...) et quitte l'étude.

Dans chacune de ces situations, on ne dispose pas de l'observation exacte (complète) sur la variable T . Considérons, par exemple, la première situation décrite ci-dessus, et notons

$$C_i = \text{date de la fin de l'étude} - E_i.$$

Alors, il est clair qu'il peut arriver que, pour certains individus, l'événement d'intérêt (« la mort ») ne se soit pas encore produit à ce moment-là, et donc tout ce qu'on sait est que le temps de survie T_i est supérieur au temps C_i ($T_i > C_i$). Donc, pour cet individu, on dispose d'une observation partielle sur la variable T . L'observation a été censurée par la fin de l'étude. Voir Figure 2.1 ci-dessous.

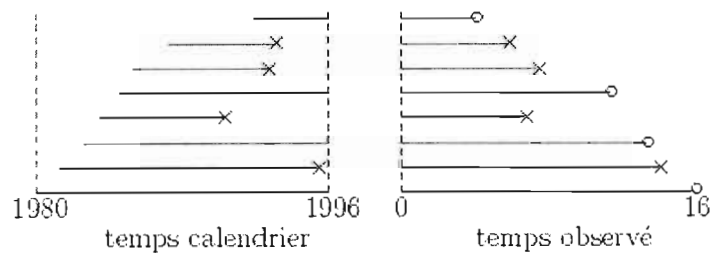


Figure 2.1 Exemple de censure à droite.

En général, dans une telle situation, on observe le minimum entre le temps de survie T_i et le temps de censure C_i , et pour chaque individu on sait si le temps observé est le temps de l'événement d'intérêt T_i , ou le temps de censure C_i .

Les observations consistent en

$$(Y_1, \delta_1), (Y_2, \delta_2), \dots, (Y_n, \delta_n),$$

avec

$$Y_i = \min(T_i, C_i) \quad \text{et} \quad \delta_i = I(T_i \leq C_i),$$

où $\delta = I(T \leq C)$ est une variable aléatoire indicatrice qui vaut 1 si on a observé le temps d'intérêt T , et qui vaut 0 si on observé plutôt le temps de censure C .

On suppose que les variables aléatoires T_i et C_i sont indépendantes pour tout i .

Ceci décrit la situation d'un modèle de *censure aléatoire à droite*.

2.2 Méthodes fréquentistes d'analyse de survie

Il existe différentes méthodes pour estimer la fonction de survie avec données censurées ; certaines font appel à des estimateurs paramétriques, d'autres à des estimateurs semi paramétriques ou encore non paramétriques. (Voir Lawless 1992).

2.2.1 Méthodes paramétriques

L'approche paramétrique stipule l'appartenance de la loi de probabilité réelle des observations à une classe particulière de lois qui dépendent d'un certain nombre (fini) de paramètres. L'avantage de cette approche est la facilitation attendue de la phase d'estimation des paramètres ainsi que l'obtention d'intervalles de confiance et la construction de tests. L'inconvénient de la méthode paramétrique est l'inadéquation pouvant exister entre le phénomène étudié et le modèle retenu.

Donc, ces méthodes reposent sur l'utilisation de modèles dans lesquels la distribution des durées de survie est une fonction paramétrique du temps. Certains modèles les plus couramment utilisés comme les modèles exponentiels de Weibull et gamma seront abordés dans les sections suivantes.

Modèle exponentiel

Cette loi, bien connue dans l'étude classique de la statistique, est décrite en analyse de survie par une fonction de risque instantané constante. En effet, si T est distribuée selon une loi exponentielle de paramètre λ où $\lambda > 0$, alors sa fonction de densité est donnée par

$$f(t) = \lambda \exp\{-\lambda t\}, \quad \lambda > 0, t > 0$$

sa fonction de répartition par

$$F(t) = 1 - \exp\{-\lambda t\},$$

sa fonction de survie par

$$S(t) = \exp\{-\lambda t\}$$

et sa fonction de risque instantané par

$$\lambda(t) = \lambda.$$

Historiquement, la loi exponentielle a été une des premières distributions utilisées, particulièrement grâce à la grande disponibilité de méthodes statistiques simples et parce qu'elle apparaît convenable pour présenter le temps de survie de plusieurs sujets. Cependant, le fait que la fonction de risque instantané soit constante se veut fort restrictif. Elle implique, entre autres, que la probabilité de décès dans l'intervalle $[t, t + dt)$, sachant que l'individu a survécu jusqu'à t , ne s'accroît pas avec l'âge, ce qui, pour de nombreuses situations, est très irréaliste.

Modèle de Weibull

Initialement utilisée pour représenter les effets de l'usure de pièces usinées (Weibull, 1951), la loi de Weibull est maintenant utilisée pour maintes applications en biologie, particulièrement pour la durée de vie d'organismes constitués de plusieurs composantes où la défaillance d'une de ces composantes peut entraîner la mort.

La fonction de densité d'une loi de Weibull de paramètres λ et β est donnée par

$$f(t) = \lambda\beta(\lambda t)^{\beta-1}\exp\{(-\lambda t)^\beta\}, \quad t > 0, \lambda > 0, \beta > 0,$$

sa fonction de survie par

$$S(t) = \exp\{(-\lambda t)^\beta\}$$

et sa fonction de risque instantané par

$$\lambda(t) = \lambda\beta(\lambda t)^{\beta-1}.$$

On remarque que $\lambda(t)$ est croissante si $\beta > 1$, décroissante pour $\beta < 1$ et constante pour $\beta = 1$. Dans ce dernier cas, la loi de Weibull devient une exponentielle de paramètre λ . Le paramètre β est souvent appelé le paramètre de forme, la valeur de λ ne modifiant en rien celle-ci.

Modèle de gamma

Bien que moins utilisée que la loi de Weibull, la loi gamma, bien connue en statistiques classiques, est largement employée en analyse de survie. Une des principales raisons qui explique la moins grande popularité du modèle gamma, comparativement à la loi de Weibull, est la moins grande facilité d'expression des fonctions de survie et de risque instantané. En effet, la fonction de survie dans le modèle gamma est donnée par

$$S(t) = 1 - \frac{1}{\Gamma(k)} \int_0^{\lambda t} u^{k-1} e^{-u} du \quad t > 0, k > 0, \lambda > 0,$$

et sa fonction de risque instantané au temps t par

$$\lambda(t) = \frac{\lambda(\lambda t)^{k-1} e^{-\lambda t}}{\Gamma(k) - \int_0^{\lambda t} u^{k-1} e^{-u} du} \quad t > 0, k > 0, \lambda > 0.$$

La fonction $\lambda(t)$ sera monotone croissante si $k > 1$ et monotone décroissante si $k < 1$.

Cette fois encore, la loi gamma est équivalente à l'exponentielle λ si $k = 1$.

La loi gamma est souvent employée quand il est nécessaire d'utiliser une variable aléatoire qui est la somme de variables aléatoires de loi exponentielle indépendantes et identiquement distribuées. En effet, si T_1, T_2, \dots, T_n sont des variables aléatoires indépendantes et toutes distribuées selon une exponentielle λ , alors $T = \sum_{i=1}^n T_i$ est distribuée selon une loi gamma de paramètre λ et avec $k = n$.

2.2.2 Méthodes non paramétriques

L'approche non paramétrique ne nécessite aucune hypothèse quant à la loi de probabilité réelle des observations – et c'est là son principal avantage. Il s'agit dès lors d'un problème d'estimation fonctionnelle, avec les ambiguïtés que cela implique, par exemple la fonction de survie, qui est continue, sera estimée par une fonction discontinue. L'inconvénient d'une telle approche est la nécessité de disposer d'un nombre important d'observations, le problème de l'estimation

d'un paramètre fonctionnel étant délicat puisqu'il appartient à un espace de dimension infinie. Deux méthodes de ce type seront présentées dans cette section. Il s'agit de la méthode de Kaplan-Meier et la méthode actuarielle. (voir (Lawless, 1992) et (Klein et Moeschberger, 1997))

Méthode de Kaplan-Meier

L'estimateur de Kaplan-Meier (1958) est une méthode non paramétrique d'estimation des fonctions de survie. Cet estimateur permet d'intégrer l'information provenant de toutes les observations disponibles, tant censurées que non censurées, parce que la survie jusqu'à tout récemment est considérée comme une série d'étapes définies par les durées de survie et les durées censurées observées.

Dans la méthode de Kaplan et Meier, le temps de participation des individus du groupe étudié est divisé en intervalles de temps de survie débutant à l'instant t_i où survient un décès et prenant fin juste avant le décès suivant. Soient d_i , le nombre de décès survenant au début de l'intervalle $[t_i, t_{i+1}[$ et c_{i-1} , le nombre de censures pendant $[t_i, t_{i+1}[$. On calcule n_i , le nombre de survivants exposés au risque de décéder juste avant l'instant t_i , à partir du nombre d'individus effectivement exposés :

$$n_i = n_{i-1} - d_{i-1} - c_{i-1}$$

La probabilité de survie en t_{i+1} sachant qu'on était en vie en t_i est estimée par :

$$\hat{S}_{t_{i+1}|t_i} = \frac{n_i - d_i}{n_i}$$

L'estimation de la probabilité de survivre juste après la date t_i s'obtient en réalisant le produit des probabilités conditionnelles de survie. Pour tout $t_i \leq t < t_{i+1}$ on a :

$$\hat{S}_{t_i} = \prod_{j=1}^i \hat{S}_{t_{j+1}|t_j}. \quad (2.1)$$

Cet estimateur est l'estimateur de Kaplan-Meier.

Un estimateur de la variance de l'estimateur de Kaplan-Meier à un temps t fixé est donné par la formule de Greenwood (Greenwood, 1926) :

$$\widehat{Var}[\hat{S}(t)] = [\hat{S}(t)]^2 \sum_{t_i \leq t} \frac{d_i}{n_i(n_i - d_i)}. \quad (2.2)$$

Méthode actuarielle

Le principe de la méthode actuarielle (Bohmer, 1912) est de découper le temps de participation maximal observé en intervalles de temps (mois, année...). Cette division doit être réalisée de manière à ce qu'il y ait un nombre minimal de décès dans chacun des intervalles de temps. Les intervalles de temps peuvent être choisis comme étant égaux ou inégaux.

Le suivi d'un certain nombre d'individus peut évidemment prendre fin pendant l'intervalle $[t_i, t_{i+1}[$ (censures autres que par le décès). Sous l'hypothèse que ces individus censurés sont soumis au même taux de mortalité que ceux décédés, on va admettre que les individus censurés sont exposés, en moyenne, au risque de décès sur la moitié de l'intervalle en question.

Notons n_i le nombre d'individus vivants et observables juste avant l'instant t_i , d_i le nombre de décès survenant pendant l'intervalle $[t_i, t_{i+1}[$ et c_i le nombre de censures pendant $[t_i, t_{i+1}[$. Ainsi, le nombre de survivants exposés au risque de décéder pendant $[t_i, t_{i+1}[$ est $N_i = n_i - c_i/2$ et la probabilité conditionnelle de survie juste avant t_{i+1} sachant qu'on était en vie en t_i est estimée par :

$$\hat{S}_{t_{i+1}|t_i} = \frac{N_i - d_i}{N_i}.$$

Pour obtenir la probabilité de survie en t_{i+1} , on va effectuer le produit des probabilités conditionnelles de survie :

$$\hat{S}_{t_i} = \prod_{j=1}^i \hat{S}_{t_{j+1}|t_j}.$$

Enfin, pour tout temps t compris entre $[t_i, t_{i+1}[$, S_t est estimé par interpolation linéaire entre \hat{S}_{t_i} et $\hat{S}_{t_{i+1}}$.

2.2.3 Méthodes semi paramétriques

L'approche semi paramétrique est une sorte de compromis entre les deux approches précédentes. La loi de probabilité réelle des observations est censée appartenir à une classe de lois dépendant de paramètres, et s'écrivant sous forme de fonction(s) non paramétrique(s). Relativement récente, elle est apparue au cours des années 70, cette approche est très répandue en analyse de la survie, notamment dans le cas du modèle de régression de Cox (Cox, 1972).

Modèle de Cox

Le modèle multivarié régressif semi paramétrique de Cox ((Cox, 1972) et (Cox, 1975)), ou modèle à taux proportionnels, va permettre d'estimer la fonction de survie en tenant compte de plusieurs variables explicatives (facteurs de pronostics, voir (Klein et Moeschberger, 1997)).

Supposons que nous observons n individus indépendants et que pour le i ème individu, $i = 1, \dots, n$, nous avons des données de la forme $(t_i, \delta_i, \{Z_i(t), 0 \leq t \leq t_i\})$, où $Z_i(t)$ est le vecteur contenant les valeurs des variables exogènes au temps t (Comme ces valeurs changent parfois dans le temps, elles sont des fonctions de t).

Le modèle semi paramétrique des risques proportionnels (ou modèle de Cox) est donné par

$$h(t|Z) = \frac{\lim_{\Delta t \rightarrow 0} P\{T \in [t, t + \Delta t) | T \geq t, Z\}}{\Delta t} = h_0(t)g(\beta'Z), \quad t \geq 0. \quad (2.3)$$

La fonction $h_0(t)$ est le risque de base qui est la même pour tous les individus. Aucune supposition paramétrique n'est faite sur cette fonction, qui constitue la partie non paramétrique du modèle. La fonction g est fixée (on prend $g(\beta'Z) = \exp(\beta'Z)$ dans la majorité des cas) et constitue la partie paramétrique du modèle. Dans ce modèle, $h(t|Z)$ qui est la fonction de risque conditionnelle sur les variables exogènes Z exprime la mortalité au temps t , pour un individu présentant les caractéristiques Z , comme étant le produit de deux fonctions :

1. la fonction $h_0(t)$ qui modélise le taux de base de mortalité pour les individus pour qui $Z = 0$ i.e pour les individus qui ne sont pas exposés aux facteurs de pronostics. Cette fonction est laissée indéterminée et dépend uniquement du temps ;
2. l'autre fonction $\exp(\beta Z)$ qui dépend uniquement des paramètres, et donc des facteurs pronostiques. Cette fonction est supposée être indépendante du temps.

Ce modèle semi paramétrique s'écrit donc :

$$h(t, Z) = h_0(t)\exp(\beta Z),$$

où β est le vecteur des paramètres associés au vecteur Z .

La fonction de vraisemblance et la fonction de vraisemblance partielle

Nous voulons estimer le paramètre β ainsi que la fonction cumulative de risque $H(t)$. Nous essayons de le faire via la méthode de maximum de vraisemblance.

La contribution à la vraisemblance conditionnelle pour le point d'observation (Z_i, δ_i, X_i) est

$$\lim_{\epsilon \rightarrow 0} \frac{P\{Z \in Z_i \pm \epsilon, \delta = \delta_i | X = X_i\}}{2\epsilon}.$$

Quand $\delta_i = 1$, cette expression est égale à

$$\lim_{\epsilon \rightarrow 0} \frac{P\{T \in Z_i \pm \epsilon, C \geq Z_i | X = X_i\}}{2\epsilon}.$$

Supposons que le temps de survie T et le temps de censure C sont indépendants, sachant X . Alors, l'expression précédente se réduit à

$$f(Z_i | X_i)(1 - G(Z_i | X_i)),$$

où $G(\cdot | x)$ dénote la fonction de répartition de C sachant $X = x$, c'est-à-dire, la distribution conditionnelle de C sachant $X = x$, et où nous avons supposé que $G(\cdot | x)$ est une fonction continue.

De façon similaire, quand $\delta_i = 0$, la contribution d'une observation (Z_i, δ_i, X_i) à la vraisemblance conditionnelle est

$$\lim_{\epsilon \rightarrow 0} \frac{P\{C \in Z_i \pm \epsilon, T \geq Z_i | X = X_i\}}{2\epsilon}.$$

ou encore

$$g(Z_i | X_i)(1 - F(Z_i | X_i)),$$

où $g(\cdot | x)$ est la densité de probabilité conditionnelle de C sachant $X = x$, et où $F(\cdot | x)$ est la fonction de répartition conditionnelle de T sachant $X = x$, supposée continue.

Nous mettons les deux types de contributions ensemble, et nous trouvons que la contribution d'une observation (Z_i, δ_i, X_i) à la vraisemblance conditionnelle est donnée par

$$f(Z_i | X_i)^{\delta_i} (1 - F(Z_i | X_i))^{1-\delta_i} g(Z_i | X_i)^{1-\delta_i} (1 - G(Z_i | X_i))^{\delta_i}.$$

Supposons maintenant que la censure est **non-informative**, c'est-à-dire que la distribution de la censure ne contient pas des informations sur la distribution de T , sachant $X = x$. Ceci implique

que les derniers deux facteurs, $g(Z_i|X_i)^{1-\delta_i}(1 - G(Z_i|X_i))^{\delta_i}$, ne dépendent pas des paramètres d'intérêt. Nous pouvons dès lors ignorer ces deux facteurs et travailler avec la contribution suivante à la vraisemblance conditionnelle

$$f(Z_i|X_i)^{\delta_i}(1 - F(Z_i|X_i))^{1-\delta_i}.$$

Notons $Z_{(1)} \leq Z_{(2)} \leq \dots \leq Z_{(n_u)}$ les observations non-censurées (n_u telles observations, avec des ex-aequo possibles), et $X_{(i)}$ la covariable X associée à l'observation $Z_{(i)}$. Sous le modèle de taux de risque proportionnels, nous trouvons donc la vraisemblance conditionnelle :

$$\begin{aligned} V(\beta) &= \prod_u f(Z_i|X_i) \prod_c (1 - F(Z_i|X_i)) \\ &= \prod_u h(Z_i|X_i) \prod_{i=1}^n (1 - F(Z_i|X_i)) \\ &= \prod_{i=1}^{n_u} h_0(Z_{(i)}) \exp\{\beta' X_{(i)}\} \prod_{i=1}^n \exp[H_0(Z_i) \exp\{\beta' X_i\}], \end{aligned}$$

où u dénote le produit par rapport aux données non-censurées, et c dénote le produit par rapport aux données censurées. La quantité $H_0(\cdot)$ est la fonction cumulative de risque, c'est-à-dire $H_0(t) = \int_0^t h_0(u) du$.

Pour estimer le paramètre β , nous maximisons $\max\{\log(V(\beta))\}$ par rapport à β . Ceci est équivalent à maximiser

$$\sum_{j=1}^N \left\{ \beta X_{(j)} - \log \left[\sum_{i \in R_j} \exp\{\beta X_i\} \right] \right\},$$

ou encore à maximiser

$$\sum_{j=1}^N \log \left(\frac{\exp(\beta X_{(j)})}{\sum_{i \in R_j} \exp\{\beta X_i\}} \right),$$

ou encore à maximiser

$$\prod_{j=1}^N \frac{\exp(\beta X_{(j)})}{\sum_{i \in R_j} \exp\{\beta X_i\}}.$$

Cette dernière expression est connue sous le nom de *vraisemblance partielle de Cox* ((Cox, 1972) et (Cox, 1975)).

En estimant le vecteur β , cela permet d'en connaître davantage sur l'effet des covariables sur la survie. On voit aussi que la vraisemblance $V(\beta)$ contient la fonction $h_0(t)$, laquelle n'attire pas trop notre attention. La solution proposée par Cox conduit à éliminer $h_0(t)$, avec une perte

d'un peu d'informations dans l'estimation de β . Notons que le modèle défini par (2.3) peut être représenté à l'aide de la fonction de survie sous la forme suivante :

$$P(T \geq t|Z) = \exp\{-H(t)\exp(\beta'Z)\}, \quad (2.4)$$

où $\exp\{-H(t)\} = P\{T \geq t|Z = 0\} = S(t)$ est la fonction de survie de base. Aucune supposition n'est faite sur cette fonction.

Le modèle défini par (2.3) est approprié pour une variable aléatoire continue T , et c'est pour cette raison qu'on l'appelle « modèle de risque proportionnel ». L'expression (2.4) a l'avantage d'incorporer les distributions de survie pour le cas continu comme pour le cas discret.

Hypothèse de proportionnalité des risques

Dans le modèle de Cox ((Cox, 1972) et (Cox, 1975)), l'effet d'une covariable sur le risque de décès est supposé être constant au cours du temps. Autrement dit, on fait l'hypothèse que, pour un facteur pronostique donné ayant n niveaux, $i = 2, \dots, n$, l'évolution de la probabilité de décéder au cours du temps reste proportionnelle entre les n niveaux. Le rapport $g(\beta'Z_1)/g(\beta'Z_2)$ mesure le risque relatif à tout moment t de subir l'événement lorsque la valeur de la variable exogène est Z_1 , par rapport au risque lorsque la valeur de la variable exogène est Z_2 .

Cette hypothèse de proportionnalité peut s'illustrer facilement à l'aide d'une variable indicatrice pouvant prendre les valeurs 0 ou 1. La fonction de risque instantané correspondante pour un individu appartenant à la classe $z = 0$ est $h(t|Z = 0) = h_0(t)$, et celle d'un individu appartenant à la classe $z = 1$ est $h(t|Z = 1) = h_0(t)e^\beta$. Par conséquent le rapport des fonctions de hasards

$$\frac{h(t|Z = 1)}{h(t|Z = 0)} = e^\beta$$

est bien constant quelque soit t .

2.3 Analyse bayésienne des modèles paramétriques de durée avec censure

Dans cette section on rappelle comment est déterminée la fonction de vraisemblance d'un modèle censuré (voir (Klein et Moeschberger, 1997)).

2.3.1 Fonctions de vraisemblance des modèles paramétriques censurés

On considère un couple de v.a. positive (T, C) , où T est la durée de survie et C un temps de censure. L'échantillon sera donc composé de n couples (Y_i, δ_i) , $i = 1, \dots, n$ où

$$Y_i = \min(T_i, C_i),$$

$$\delta_i = \begin{cases} 1 & \text{si } T_i \leq C_i, \\ 0 & \text{si } T_i > C_i. \end{cases}$$

Afin de simplifier la présentation, on suppose que T et C sont indépendants, de fonctions de répartition F et G . Pour simplifier, on ne considérera là encore que des lois à densité par rapport à la mesure de Lebesgue et soient f et g des dérivées de F et G . On a donc la densité de (Y, Δ) :

$$\begin{aligned} p(t, \delta) &= [(1 - F(t))g(t)]^{1-\delta} [(1 - G(t))f(t)]^\delta \\ &= [f(t)^\delta (1 - F(t))^{1-\delta}] [g(t)^{1-\delta} (1 - G(t))^\delta] \end{aligned}$$

Supposons alors que $F \sim \pi_1(\theta)$ et $G \sim \pi_2(\lambda)$ et que θ est le vecteur des paramètres d'intérêt.

Si $x = (y_i, \delta_i)_{i=1, \dots, n}$ est un échantillon i.i.d. de Y et Δ , on pourra écrire :

$$\pi(\theta | x) \propto \pi_1(\theta) \prod_{i=1}^n f(y_i | \theta)^{\delta_i} S(y_i | \theta)^{1-\delta_i}$$

avec $S = 1 - F$.

Fonction de vraisemblance dans un modèle exponentiel censuré

Dans un modèle exponentiel, on a :

$$\lambda \in \mathbb{R}^+, \quad f(t | \lambda) = \lambda \exp(-\lambda t) \quad \text{et} \quad S(t | \lambda) = \exp(-\lambda t).$$

Ce qui donne une fonction de vraisemblance pour un échantillon de n individus décrits par $(y_i, \delta_i)_{(i=1, \dots, n)}$

$$\begin{aligned} p(x | \lambda) &= \prod_{i=1}^n \left(\lambda \exp(-\lambda y_i) \right)^{\delta_i} \left(\exp(-\lambda y_i) \right)^{1-\delta_i} \\ &= \lambda^{\left\{ \sum_{i=1}^n \delta_i \right\}} \exp \left\{ -\lambda \sum_{i=1}^n y_i \right\} \end{aligned}$$

En prenant comme loi *a priori* une loi gamma on a :

$$\begin{aligned}\pi(\lambda) &= \frac{\beta_0^\alpha}{\Gamma(\alpha_0)} \lambda^{\alpha_0-1} \exp(-\beta_0 \lambda), \quad (\alpha_0 > 0, \beta_0 > 0) \\ &\propto \lambda^{\alpha_0-1} \exp(-\beta_0 \lambda).\end{aligned}$$

et d'après le théorème de Bayes on obtient comme fonction *a posteriori* :

$$\begin{aligned}\pi(\lambda | x) &\propto \lambda^{\alpha_0-1} \exp(-\beta_0 \lambda) \lambda^{\left(\sum_{i=1}^n \delta_i\right)} \exp\left\{-\lambda \sum_{i=1}^n y_i\right\} \\ &\propto \lambda^{\left(\alpha_0 + \sum_{i=1}^n \delta_i - 1\right)} \exp\left\{-\lambda \left(\sum_{i=1}^n y_i + \beta_0\right)\right\} \\ &\propto \Gamma\left(\alpha_0 + \sum_{i=1}^n \delta_i, \beta_0 + \sum_{i=1}^n y_i\right).\end{aligned}$$

CHAPITRE III

MÉTHODES BAYÉSIENNES EN ANALYSE DE SURVIE

3.1 Introduction

Plusieurs auteurs ont discuté de l'utilisation de l'inférence bayésienne dans le cas de modèles semi paramétriques ((Ferguson, 1973) ; (Kalbfleisch, 1978)) ; (Florens et Rolin, 2001) et paramétriques. On peut citer ici (Feigl et Zelen, 1965) ; (Carlin, Gelfand et Smith, 1992), etc. Ces différents travaux ont été développés dans des contextes méthodologiques différents, en utilisant des lois *a priori* différentes et/ou en modélisant la fonction cumulative de risque ou directement la fonction de risque instantané. Nous décrirons deux méthodes utilisant les principes des techniques bayésiennes : la première est une méthode semi paramétrique de Kalbfleisch ((Kalbfleisch, 1978), (Clayton, 1991)) qui est l'une des premières décrites dans la littérature ; la deuxième méthode est paramétrique, celle de Carlin, Gelfand et Smith (Carlin, Gelfand et Smith, 1992), il s'agit d'une méthode d'estimation par simulation du processus de Poisson dont l'inférence bayésienne paramétrique donne de bons résultats par rapport aux méthodes classiques. Ces deux méthodes peuvent être facilement mises en application grâce au code du logiciel **BUGS** (**B**ayesian inference **U**sing **G**ibbs **S**ampling) (Spiegelhalter, Thomas, Best et Gilks, 1996a).

3.1.1 Logiciel de simulation MCMC : BUGS

Les initiales **BUGS** signifient *Bayesian Inference Using Gibbs Sampling*. Ce logiciel de programmation a été développé par Spiegelhalter, Thomas, Best et Gilks

(Spiegelhalter, Thomas, Best et Gilks, 1996a) à l'unité *MRC* de Biostatistique de Cambridge. Il vise à tirer avantage des probabilités de l'échantillonnage de Gibbs dans l'inférence bayésienne sur des systèmes complexes.

BUGS repose sur un langage proche de C ou de S-Plus et comprend des déclarations sur le modèle, les données et les spécifications *a priori*, en autorisant plusieurs niveaux dans la modélisation *a priori*.

La plupart des lois standards sont reconnues par BUGS, qui autorise de plus un grand nombre de transformations. BUGS comprend également des commandes comme *compile*, *data*, *update*, *stat*. La sortie de BUGS est une table des valeurs simulées pour l'ensemble des paramètres.

Une restriction importante de ce logiciel est d'imposer l'utilisation de lois *a priori* conjuguées ou log-concaves pour permettre l'emploi de l'échantillonnage de Gibbs et de méthodes de simulation automatiques. Les autres lois peuvent être utilisées par une discrétisation de leur support. Un module de contrôle de convergence des algorithmes MCMC, nommé **CODA**, a été construit par Best, Cowles et Vines (Best, Cowles et Vines, 1995). Initialement prévu pour traiter les sorties de BUGS, ce logiciel, fonctionnant sous S-Plus, permet d'analyser les résultats d'algorithmes de **Gibbs** et de **Metropolis-Hastings**. Les techniques de contrôle sont principalement celles décrites dans Cowles et Carlin (Cowles et Carlin, 1996).

3.2 Méthode semi paramétrique : la méthode de Kalbfleisch

Comme on l'a vu à la section 2.2, le modèle de Cox à taux proportionnels (Cox, 1972) est largement utilisé dans les études de survie. Dans ce modèle, les coefficients de régression des covariables, mesurant les effets des facteurs pronostiques étudiés, sont modélisés de manière paramétrique et il n'est pas nécessaire de donner une spécification *a priori* pour le taux de base de mortalité. L'estimation fréquentiste de ces coefficients de régression est réalisée en utilisant la vraisemblance partielle de Cox ((Cox, 1972); (Cox, 1975)), qui est une approximation de la vraisemblance totale. Grâce à cette notion de vraisemblance partielle, ces coefficients et le taux de base de mortalité peuvent être estimés séparément, ce qui simplifie considérablement les calculs. L'utilisation de la vraisemblance totale aurait pour conséquence d'avoir à estimer simultanément les coefficients de régression et le taux de base de mortalité.

Plusieurs analyses bayésiennes ont été réalisées dans le cadre du modèle de Cox ((Ferguson, 1973); (Hjort, 1990)) dans lesquelles les lois *a priori* pour le taux de base de mortalité et pour les coefficients de régression des covariables sont spécifiées de manières différentes. Certaines de ces méthodes bayésiennes ont utilisé la vraisemblance totale plutôt que la vraisemblance partielle. En effet, un des avantages à utiliser les méthodes bayésiennes pour modéliser conjointement les coefficients de régression des covariables et le taux de base de mortalité est que l'on peut calculer de manière précise les distributions *a posteriori* du modèle ainsi que leurs écarts-type en utilisant les techniques MCMC. Néanmoins, le problème de la spécification rationnelle des fonctions de distributions *a priori* ainsi que la réalisation de calculs intensifs restent toujours posés. Les résultats de ces études bayésiennes de survie ont montré la précision des estimations et les avantages potentiels à utiliser ces méthodes pour analyser des données de survie.

3.2.1 Processus de comptage

Dans cette section, on fait un rappel sur la notion de processus de comptage. Les définitions que nous présentons se trouvent dans Kein (Klein et Moeschberger, 1997) et Fleming et Harrington (Fleming, 1991).

Définition 3.2.1 Un processus de comptage est une fonction $N(t); t \geq 0$ telle que

1. $N(0) = 0$;
2. $N(t) < \infty$;
3. $N(t)$ a des trajectoires constantes par paliers et continues à droite faisant uniquement des sauts de valeur unitaire.

Un processus de comptage $N(t)$ est donc une fonction de t qui compte le nombre total d'événements qui se sont produits dans l'intervalle $[0; t]$.

En analyse des durées de vie avec un échantillon ayant une censure à droite, on considère

$$N_i(t) = I[T_i \leq t; \delta_i = 1], \quad i = 1, \dots, n.$$

c.-à-d. le processus de comptage qui fait un saut d'une unité au temps t si le décès de l'individu i est observé à cet instant. Le processus

$$N(t) = \sum_{i=1}^n N_i(t) = \sum_{t_i \leq t} \delta_i$$

est le processus de comptage qui compte le nombre total de décès jusqu'à l'instant t . Il sera aussi utile de noter que

$$Y_i(t) = I[T_i \geq t]$$

et

$$Y(t) = \sum_{i=1}^n Y_i(t)$$

où le processus $Y(t)$ compte le nombre total d'individus qui sont toujours dans l'étude au temps t . On appelle parfois $Y_i(t)$ la fonction indicatrice.

Définition 3.2.2 L'historique ou la filtration d'un processus au temps t , notée \mathcal{F}_t , est toute l'information accumulée sur le processus dans l'intervalle $[0; t]$.

En analyse des durées de vie, \mathcal{F}_t contient toute l'information que l'on a sur nos n sujets au temps t .

Clairement, $\mathcal{F}_s \subseteq \mathcal{F}_t$ pour $s < t$, car plus le temps passe, plus on accumule de l'information.

Rappelons que pour toute variable aléatoire \mathbf{X} , on a que :

1. $E[\mathbf{X}] = E[E[\mathbf{X}|\mathcal{F}_t]], \forall t \geq 0$;
2. $E[E[\mathbf{X}|\mathcal{F}_t]|\mathcal{F}_s] = E[\mathbf{X}|\mathcal{F}_s]$ lorsque $\mathcal{F}_s \subseteq \mathcal{F}_t$;
3. Si \mathbf{Y} est mesurable par rapport à \mathcal{F}_t (c.-à-d. que si on a l'information sur \mathcal{F}_t , on connaît la valeur de \mathbf{Y}), alors $E[g(\mathbf{Y})\mathbf{X}|\mathcal{F}_t] = g(\mathbf{Y})E[\mathbf{X}|\mathcal{F}_t]$.

Définition 3.2.3 La fonction (ou le processus) $I(t) = E[dN(t)|\mathcal{F}_{t-}]$ est le processus d'intensité de $N(t)$. La fonction $\Lambda(t) = \int_0^t I(s)ds$ est le processus d'intensité cumulée.

3.2.2 Présentation de la méthode de Kalbfleisch

Soit le processus de comptage $N_i(t)$ défini à la section précédente comptant le nombre de sujets ($i = 1, \dots, n$) décédés au temps t . Le processus d'intensité correspondant est donné par

$$I_i(t)dt = E(dN_i(t)|\mathcal{F}_{t-}),$$

où $dN_i(t)$ correspond à l'accroissement de N_i sur l'intervalle de temps $[t, t + dt[$ et \mathcal{F}_{t-} représente les données disponibles juste avant le temps t . Donc, $dN_i(t) = 1$ si le i ème sujet décède pendant cet intervalle de temps et $dN_i(t) = 0$ sinon. En supposant que le temps soit continu ($dt \rightarrow 0$), $E(dN_i(t)|\mathcal{F}_{t-})$ correspond alors au risque instantané de décès au temps t du i ème sujet. On a alors le modèle à taux proportionnels suivant :

$$I_i(t) = Y_i(t)h_0(t)\exp(\beta\mathbf{z}_i),$$

où

$$Y_i(t) = \begin{cases} 1 & \text{si le sujet } i \text{ est observé au temps } t \\ 0 & \text{sinon} \end{cases}$$

et $h_0(t)\exp(\beta\mathbf{z}_i)$ correspond à la formulation classique du modèle de Cox.

La fonction de vraisemblance du modèle est proportionnelle à

$$\prod_{i=1}^n \left[\prod_{t \geq 0} I_i(t)^{dN_i(t)} \right] \exp\left(- \int_{t \geq 0} I_i(t) dt \right)$$

On peut considérer les incréments $dN_i(t)$ sur $[t, t + dt[$ comme des variables aléatoires indépendantes suivant une loi de Poisson. Donc ;

$$dN_i(t) \sim P(I_i(t)dt);$$

cela donne :

$$I_i(t)dt = Y_i(t)dH_0(t)\exp(\beta\mathbf{z}_i)$$

où $dH_0(t) = h_0(t)dt$ représente l'accroissement du risque instantané du décès pendant $[t, t + dt[$.

Comme la conjuguée d'une loi de Poisson est une loi gamma, Kalbfleisch (Kalbfleisch, 1978) considère que $H_0(t)$ est un processus gamma avec $E[H_0(t)] = H_0^*(t)$, qui est une fonction connue, et $Var[H_0(t)] = H_0^*(t)/c$ où c est un nombre réel positif. Les incréments $dH_0(t)$ sur $[t, t + dt[$ sont censés être indépendants et suivent des lois $Gamma(cH_0^*(t), c)$ indépendantes, où les paramètres $H_0^*(t)$ et c représentent respectivement l'estimation *a priori* de la fonction de risque instantané inconnue et le degré de confiance que l'on attache à cet loi *a priori*. Quand c est grand la croyance *a priori* sur $H_0^*(t)$ est forte, et inversement. Kalbfleisch (Kalbfleisch,

1978) choisit comme loi *a priori*, $H_0^*(t) = rdt$ où r estime le taux d'événements par unité de temps, et dt est l'intervalle de temps considéré.

Donc pour un temps $t_j (j = 1 \dots n)$, l'estimation de la fonction de survie par la méthode de Kalbfleisch est de la forme suivante :

$$SKALB[t_j] = \exp \left[- \sum_{k=1}^j H_0(t_k) \exp(\beta Z_k) \right] \quad (3.1)$$

La loi *a priori* pour les composantes de β est considérée comme étant peu informative (β suit une loi normale multivariée de moyenne 0 et de matrice de variance covariance σ^2).

Ce modèle peut être implanté avec le logiciel **WinBUGS** et l'estimation des paramètres est réalisée en utilisant l'échantillonneur de Gibbs.

3.3 Exemples d'application de la méthode Kalbfleisch

La méthode d'analyse bayésienne de la survie a été évaluée à partir d'une étude basée sur des simulations. Après cette étude et dans le but d'illustrer notre approche, nous avons réutilisé deux bases de données de survie qui ont déjà été analysées par plusieurs auteurs et avec des méthodes différentes.

3.3.1 Étude de simulation

Nous avons fait une étude de simulation qui montre la performance de la méthode de Kalbfleisch comparativement à la méthode de Kaplan-Meier. On considère un échantillon X_i de taille $n = 10$, $1 \leq i \leq n$, de distribution lognormale de moyenne $\mu = 1$ qui représente le groupe traité et un autre échantillon Y_i de taille $n = 10$, $1 \leq i \leq n$, de distribution exponentielle de moyenne $\lambda = 0.8$ pour le groupe placebo. On considère aussi un troisième échantillon indépendant C_i , $1 \leq i \leq 2n$, avec distribution uniforme $U(0, 3)$ de moyenne 1.5 représentant la censure dans les deux groupes.

Pour les observations X_i du groupe traité, on observe le minimum entre le temps de survie X_i et le temps de censure C_i , et pour chaque individu on sait si le temps observé est le temps de l'événement d'intérêt X_i , ou le temps de censure C_i .

Les observations consistent en

$$(T_1, \delta_1), (T_2, \delta_2), \dots, (T_n, \delta_n),$$

avec

$$T_i = \min(X_i, C_i) \quad \text{et} \quad \delta_i = I(X_i \leq C_i),$$

où $\delta = I(X \leq C)$ est une variable aléatoire indicatrice qui vaut 1 si on a observé le temps d'intérêt X , et qui vaut 0 si on a plutôt observé le temps de censure C .

On suppose que les variables aléatoires X_i et C_i sont indépendantes pour tout i , $1 \leq i \leq n$.

De la même manière, pour les observations Y_i du groupe placebo, on prend

$$N_i = \min(Y_i, C_{i+10}) \quad \text{et} \quad \delta_i = I(Y_i \leq C_{i+10}), \quad i = 1, \dots, n$$

Les échantillons ainsi générés ont ensuite été analysés, indépendamment les uns des autres, en utilisant :

- une méthode fréquentiste, celle de Kaplan-Meier ;
- une méthode bayésienne, celle de Kalbfleisch.

On commence par le calcul des vraies valeurs de survie pour le groupe traité $SV(T)$ avec les probabilités suivantes :

$$Pr(X > T(i))$$

où X est la variable aléatoire de distribution lognormale de moyenne $\mu = 1$. Aussi on calcule les vraies valeurs de survie $SV(N)$ pour le groupe placebo avec les probabilités suivantes :

$$Pr(Y > N(i))$$

où Y_i est de distribution exponentielle de moyenne $\lambda = 0.8$.

De l'autre côté, les temps de survie pour les deux groupes vont être estimés avec la méthode fréquentiste de Kaplan-Meier (SKM(N) et SKM(T)) définie au chapitre 2 (formule (2.1)) et la méthode bayésienne de Kalbfleisch (SKALB(N) et SKALB(T)) définie à la section (3.2.2)(formule (3.1)).

La mesure de la performance qu'on utilise est la moyenne des erreurs quadratiques moyennes où l'erreur quadratique moyenne est donnée par :

$$EQM(\hat{S}) = \frac{\sum_{i=1}^n [S(t_i) - \hat{S}(t_i)]^2}{n}$$

où \hat{S} est l'estimation du temps de survie avec l'une des deux méthodes utilisées, n est la taille de l'échantillon ($n=10$).

On définit donc la moyenne des erreurs quadratiques moyennes par :

$$MEQM(\hat{S}) = \frac{\sum_{k=1}^m EQM(\hat{S}_k)}{m}$$

où $EQM(\hat{S}_k)$ est l'erreur quadratique moyenne de la k^e simulation et m est le nombre de simulations ($m=25$).

Les résultats d'une seule simulation pour les deux groupes sont présentés dans les deux tableaux suivants :

Temps(N):Placebo	SV(N)	SKALB(N)	SKM(N)	EQM_KALB(N)	EQM_KM(N)	EQM_KALB - EQM_KM
0,00611	0,992390	0,93840	0,91667	0,0019847	0,0129039	-0,0109192
0,03347	0,959024	0,87780	0,83333			
0,21551	0,763848	0,81010	0,74074			
0,24181	0,739141	0,74300	0,64815			
0,24744	0,733963	0,67610	0,55556			
0,43253	0,582361	0,54580	0,46296			
1,17687	0,229676	0,26630	0,37037			
1,18240	0,228095	0,21810	0,27778			
2,34811	0,053123	0,04475	0,18519			

SV(N) est la vraie survie pour le groupe placebo.

SKALB est la survie estimée avec la méthode de Kalbfleisch.

SKM est la survie estimée par la méthode de Kaplan-Meier.

EQM_KALB représente l'erreur quadratique moyenne par rapport à la méthode de Kalbfleisch.

EQM_KM est l'erreur quadratique moyenne par rapport à la méthode de Kaplan-Meier.

Tableau 3.1 Erreur quadratique moyenne pour les estimations de la fonction de survie avec les méthodes de Kaplan-Meier et de Kalbfleisch du groupe placebo pour une simulation.

Pour les 25 simulations, on a une moyenne des erreurs quadratiques moyennes du groupe placebo égale à -0.031704797 .

Temps(T):Traité	SV(T)	SKALB(T)	SKM(T)	EQM_KALB(T)	EQM_KM(T)	EQM_KALB(T) - EQM_KM(T)
0,43055	0,800302	0,7602	0,90909	0,0018080	0,0082425	-0,0064344
0,69652	0,641196	0,6681	0,81818			
0,77848	0,598867	0,6203	0,72727			
0,82396	0,576768	0,5729	0,63636			
0,87923	0,551205	0,5249	0,54545			
1,22817	0,418581	0,3657	0,45455			
1,38420	0,372544	0,3063	0,36364			
1,64519	0,309293	0,2478	0,27273			

SV(T) est la vraie survie pour le groupe traité.

SKALB est la survie estimée avec la méthode de Kalbfleisch.

SKM est la survie estimée par la méthode de Kaplan-Meier.

EQM_KALB représente l'erreur quadratique moyenne par rapport à la méthode de Kalbfleisch.

EQM_KM est l'erreur quadratique moyenne par rapport à la méthode de Kaplan-Meier.

Tableau 3.2 Erreur quadratique moyenne pour les estimations de la fonction de survie avec les méthodes de Kaplan-Meier et de Kalbfleisch du groupe traité pour une simulation.

Donc une moyenne des erreurs quadratiques moyennes pour les 25 simulations du groupe traité égale à $-0,00977191$

Si nous comparons les moyennes des EQM des deux méthodes de calcul, nous pouvons dire que les résultats des simulations par la méthode de Kalbfleisch sont nettement meilleurs que ceux obtenus par la méthode de Kaplan-Meier. Cette performance est vraie pour les deux groupes (placebo et traité).

3.3.2 Cas de leucémie chez les enfants

Pour illustrer la méthode de Kalbfleisch, nous avons utilisé des données tirées de l'ouvrage de Klein et Moeschberger (Klein et Moeschberger, 1997). Freireich et al. (Freireich et al., 1963) rapportent les résultats d'une épreuve clinique d'un médicament appelé 6-mercaptopurine (médicament anticancérigène de synthèse appartenant à la famille des antimétabolites (6-MP)), comparativement à un placebo pour 42 enfants qui sont atteints d'une leucémie aiguë. L'expérience a été réalisée dans 11 hôpitaux américains. Des patients en phase de rémission complète ou partielle de leur leucémie avec un traitement de la prednisone ont été choisis. La rémission complète ou partielle signifie que la majeure partie ou tous les signes de la maladie avaient disparu de la moelle osseuse. Des patients ont été suivis jusqu'au retour de la leucémie (rechute), ou jusqu'à la fin de l'étude (en mois). Les données sont rapportées dans le tableau suivant : Dans cet exemple, le modèle ne contient qu'un seul paramètre β qui permet de mesurer l'écart

Traitement	Temps de survie						
Placebo	1	1	2	2	3	4	4
	5	5	8	8	8	8	11
6-MP	11	12	12	15	17	22	23
	6*	6	6	6	7	9*	10*
	10	11*	13	16	17*	19*	20*
	22	23	25*	32*	32*	34*	35*

Tableau 3.3 Données tirées de l'ouvrage de Klein et Moeschberger (1997) pour le cas de leucémie chez les enfants.

de mortalité du groupe « Placebo » par rapport au groupe « 6-MP ». La seule caractéristique d'un individu que l'on considère ici est son appartenance à un groupe. C'est pour cette raison que les deux groupes de traitement sont codés respectivement $Z = 0$ pour 6-MP et $Z = 1$ pour le placebo. Nous avons utilisé une distribution normale multivariée comme densité *a priori* pour le coefficient de régression β . En effet, cette distribution a montré sa flexibilité et son utilité comme distribution *a priori* dans de nombreux problèmes de régression (Geisser, 1993).

3.3.3 Estimation des paramètres avec WinBUGS

Une modification sur le programme se trouvant dans les exemples de WinBUGS (Leuk : survival analysis using Cox regression (Spiegelhalter, Thomas, Best et Gilks, 1996b)) et qui présente la méthode bayésienne de Kalbfleisch nous a permis de l'adapter à notre base de donnée.

Comme cela se fait habituellement avec les algorithmes MCMC, les premières itérations de la chaîne (les 1000 premières dans notre cas) n'ont pas été retenues et les estimations ont été calculées à partir des valeurs de la distribution *a posteriori* obtenues de ce qui reste de la chaîne (les 10 000 itérations suivantes). **WinBUGS** donné une moyenne *a posteriori* (écart-type type) du coefficient de régression $\beta = 1.539$ (0.417) et un intervalle de crédibilité de niveau 95% comme on l'a défini à la section 1.4 (chapitre 1) qui est égal à (0.747; 2.376). Ceci rivalise avec l'évaluation partielle standard de probabilité (obtenue en utilisant le procédé de SAS PHREG) de 1.59 (écart-type de 0.43) et un intervalle de confiance de niveau 95% = (0.7472; 2, 4328).

Les probabilités de survie ont été estimées pour le groupe 6-MP et les résultats sont présentés dans le tableau suivant :

Temps(t)	moyenne	écart-type	2.5%	médiane	97.5%
Survie.traite[t=6]	0.8768	0.04892	0.7655	0.8833	0.9534
Survie.traite[t=7]	0.8646	0.0522	0.7458	0.8708	0.9479
Survie.traite[t=10]	0.802	0.06827	0.6513	0.8096	0.9156
Survie.traite[t=13]	0.7102	0.08911	0.5211	0.7165	0.8659
Survie.traite[t=16]	0.6614	0.09716	0.4586	0.666	0.8367
Survie.traite[t=22]	0.5662	0.1116	0.3416	0.5681	0.776
Survie.traite[t=23]	0.4755	0.1192	0.2439	0.4755	0.7091
beta	1.538	0.4176	0.7501	1.527	2.376

Survie.traite : Représente la fonction de survie du groupe traité

Tableau 3.4 Moyennes *a posteriori* de survie pour le groupe traité et les intervalles de crédibilité $IC_{95\%}$ dans l'exemple de leucémie.

Le graphique suivant montre que la probabilité de survie calculée avec la méthode de Kalbfleisch pour le groupe traité est nettement plus élevée que celle du groupe placebo.

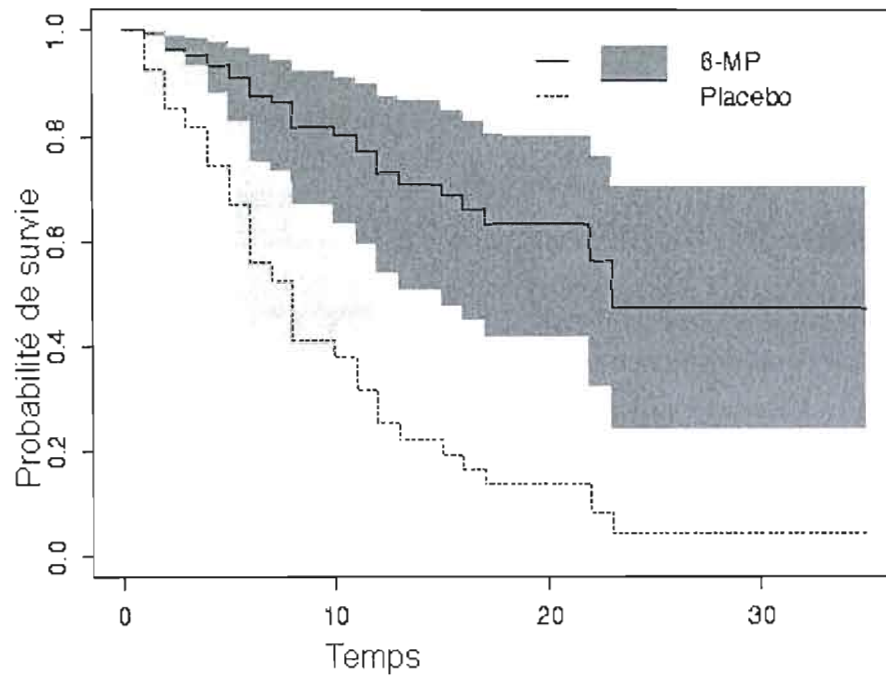


Figure 3.1 Estimation de la moyenne de survie par la méthode de Kalbfleisch pour les groupes de traitement et de placebo dans l'exemple de leucémie.

L'estimation de la fonction de survie pour le groupe 6-MP avec la méthode de Kaplan-Meier (1958), nous donne le résultat suivant :

temps_mp	survie	variance	ecartype	IC_Binf(95%)	IC_Bsup(95%)
6	0.85714	0.005831	0.07636	0.70748	1.00681
7	0.80672	0.007558	0.08694	0.63633	0.97712
10	0.75294	0.009283	0.09635	0.56410	0.94179
13	0.69020	0.011409	0.10681	0.48084	0.89955
16	0.62745	0.013008	0.11405	0.40391	0.85100
22	0.53782	0.016444	0.12823	0.28648	0.78915
23	0.44818	0.018115	0.13459	0.18438	0.71198

Tableau 3.5 Estimation par la méthode de K.M. de la fonction de survie et des intervalles de confiance (95%) pour le groupe 6-MP.

Si nous faisons une comparaison entre ces deux méthodes, nous constatons que la méthode bayésienne de Kalbfleisch est plus précise dans l'estimation de coefficient de régression (et les durées de vie du groupe 6-MP) avec un écart-type proche de 0.417 alors que l'on obtient une valeur de 0.43 pour la méthode fréquentiste de Kaplan-Meier

On obtient également des moyennes *a posteriori* et des écarts-type des estimations des durées de vie pour le groupe 6-MP qui sont meilleurs avec la méthode bayésienne de Kalbfleisch qu'avec méthode de Kaplan-Meier.

Enfin, la longueur de l'intervalle de crédibilité à 95% de coefficient de régression varie entre 0.747 et 2.376 pour la méthode bayésienne de Kalbfleisch. L'estimateur est compris entre 0.7472 et 2.4328 pour la méthode fréquentiste de Kaplan-Meier. On peut dire aussi que les intervalles de crédibilité des durées de vie pour le groupe 6-MP sont meilleurs avec la méthode bayésienne.

Quoi qu'il en soit, l'étude par simulation a montré que la méthode bayésienne fournit des estimations précises. D'une manière générale, on peut donc constater que les résultats obtenus sont sensiblement meilleurs avec la méthode bayésienne de Kalbfleisch qu'ils ne le sont lorsque nous utilisons la méthode fréquentiste de Kaplan-Meier.

3.4 Résultats avec WinBUGS pour la base de données pour le cancer du sein

Dans une étude conçue pour déterminer si la patiente atteinte de cancer du sein déjà classifiée avec des ganglions lymphatiques négatifs par la méthode usuelle « MOS », peut être mieux classifiée par un examen immunohistochimique(IH) de leurs ganglions lymphatiques. La patiente est considérée comme immunoperoxydase positive ou immunoperoxydase négative par ce test. L'intérêt de cette étude est que 16% des patientes présentant des ganglions lymphatiques axillaires négatifs par l'examen usuel ont une récurrence dans un délai de 10 ans. Nous avons choisi 45 patientes qui avaient le cancer du sein avec des ganglions lymphatiques axillaires négatifs avec un minimum de suivi de 10 ans parmi les patientes atteintes de cancer du sein dans les hôpitaux universitaires de l'Ohio. Neuf des 45 patientes sont déclarées immunoperoxydase positives, et les 36 autres sont restées négatives. Les temps de survie (en mois) pour les deux groupes de patientes sont indiqués dans le tableau 3.6 (+ dénote une observation censurée).

Ces données ont déjà été analysées par plusieurs auteurs, tels que Sedmak (Sedmak, 1989) Klein et Maeschberger (Klein et Moeschberger, 1997) et d'autres. La base de données que nous utiliserons ici provient du livre de Klein et Maeschberger (Klein et Moeschberger, 1997).

Dans cette base de données, le modèle ne contient qu'un seul paramètre β qui permet de mesurer l'écart de mortalité du groupe « immunoperoxydase positif » par rapport au groupe « immunoperoxydase négatif ». La seule caractéristique d'un individu que l'on considère ici est son appartenance à un groupe ; c'est pour cette raison que les deux groupes sont codés respectivement $Z = 1$ pour le groupe « immunoperoxydase négatif » et $Z = 0$ pour le groupe « immunoperoxydase positif ». Les durées de survie et le coefficient de régression β seront estimés avec la méthode de Kalbfleisch en utilisant le logiciel **WinBUGS**.

Immunoperoxydase négatifs : 19, 25, 30, 34, 37, 46, 47, 51, 56, 57, 61, 66, 67, 74, 78, 86, 122⁻, 123⁻, 130⁻, 130⁻, 133⁻, 134⁻, 136⁻, 141⁻, 143⁻, 148⁻, 151⁻, 152⁻, 153⁻, 154⁻, 156⁻, 162⁻, 164⁻, 165⁻, 182⁻, 189⁻,

Immunoperoxydase positifs : 22, 23, 38, 42, 73, 77, 89, 115, 144⁻

⁻Observation censurée

Tableau 3.6 Données tirées de l'ouvrage de Klein et Moeschberger (1997) pour le cas de cancer du sein (Exemple (1.5) Chapitre 1).

3.5 Analyse et résultats

Les estimations bayésiennes du coefficient de régression qui permet de mesurer l'écart de mortalité du groupe « immunoperoxydase positif » par rapport au groupe « immunoperoxydase négatif » et des fonctions de survie ont été obtenues en utilisant **WinBUGS** (tableau 3.7).

Dans cette application, nous avons également écarté les 1000 premières itérations de la chaîne MCMC et l'estimation a été calculée à partir de valeur de la distribution *a posteriori* obtenue sur les 6000 dernières itérations.

La moyenne *a posteriori* (écart-type type) du coefficient de régression β est égal à **0.9801** (un écart-type de **0.4333**) avec un intervalle de crédibilité de niveau 95% = **(0.1076; 1.794)**.

Ceci rivalise avec l'estimation obtenue dans l'ouvrage de Klein et Maeschberger (Klein et Moeschberger, 1997) (chapitre 8 exemple 8.1) qui utilise l'algorithme de Newton-Raphson. Leurs résultats étaient moins performants puisque l'estimateur ponctuel du paramètre β était de 0.9802 (un écart-type de 0.4349) avec un intervalle de confiance de niveau 95% égal à (0.1278; 1.8326).

L'estimation de coefficient de régression β par la méthode bayésienne de Kalbfleisch nous révèle qu'une patiente qui fait partie du groupe avec une immunoperoxydase positive a un risque de mourir de $e^{0.9801} = 2.66$ fois plus élevé qu'une patiente du groupe immunoperoxydase négatifs. Klein et Maeschberger (Klein et Moeschberger, 1997) donnent un intervalle de confiance

pour le risque relatif égal à $\exp(0.1278; 1.8326) = (1.14; 6.25)$. Cet intervalle est moins performant comparativement à l'intervalle $\exp(0.1076; 1.794) = (1.11; 6.01)$ défini par la méthode de Kalbfleisch.

Temps: groupe traité	moyenne	écart-type	2.5%	médiane	97.5%
Survie.trt[19]	0.9835	0.01643	0.9392	0.9886	0.9996
Survie.trt[25]	0.9327	0.03378	0.8526	0.9376	0.9816
Survie.trt[30]	0.9155	0.03793	0.8264	0.9211	0.9729
Survie.trt[34]	0.898	0.04196	0.7996	0.9035	0.9639
Survie.trt[37]	0.881	0.0456	0.7762	0.886	0.9527
Survie.trt[46]	0.8276	0.05452	0.7052	0.8326	0.92
Survie.trt[47]	0.8096	0.05722	0.6822	0.8144	0.9072
Survie.trt[51]	0.7909	0.05967	0.6599	0.7956	0.8943
Survie.trt[56]	0.773	0.06228	0.638	0.7771	0.8813
Survie.trt[57]	0.7546	0.06462	0.6143	0.7597	0.8672
Survie.trt[61]	0.736	0.06709	0.5903	0.7412	0.8541
Survie.trt[66]	0.7183	0.06866	0.5711	0.723	0.8405
Survie.trt[67]	0.6997	0.07075	0.5492	0.7042	0.8266
Survie.trt[74]	0.6632	0.07425	0.5081	0.6664	0.8004
Survie.trt[78]	0.6235	0.07703	0.464	0.6256	0.7697
Survie.trt[86]	0.604	0.07821	0.4429	0.6062	0.7517
beta	0.9801	0.4333	0.1076	0.991	1.794

Survie.trt(t) est l'estimation de la fonction de survie pour une patiente du groupe Immunoperoxydase négatif

Tableau 3.7 Valeurs *a posteriori* des moyennes de survie pour les femmes avec une immunoperoxydase négative et leurs intervalles de crédibilité $IC_{95\%}$ dans l'exemple de cancer du sein.

L'estimation de la fonction de survie pour le groupe immunoperoxydase négative avec la méthode de Kaplan-Meier (1958) nous donne le résultat suivant :

Temps	Survie_KM	écart-type	IC_BInf	IC_BSup
19	0.97368	0.025967	0.92279	1.02458
25	0.94737	0.036224	0.87637	1.01837
30	0.92105	0.043744	0.83531	1.00679
34	0.89474	0.049784	0.79716	0.99231
37	0.86842	0.054836	0.76094	0.97590
46	0.84211	0.059153	0.72617	0.95804
47	0.81579	0.062886	0.69253	0.93905
51	0.78947	0.066135	0.65985	0.91910
56	0.76316	0.068968	0.62798	0.89833
57	0.73684	0.071434	0.59683	0.87685
61	0.71053	0.073570	0.56633	0.85472
66	0.68421	0.075405	0.53642	0.83200
67	0.65789	0.076960	0.50705	0.80874
74	0.63158	0.078252	0.47821	0.78495
78	0.60526	0.079293	0.44985	0.76068
86	0.57895	0.080093	0.42196	0.73593
beta	0.9801	0.4333	0.1076	1.794

Survie_KM est l'estimation de la fonction de survie avec la méthode de Kaplan-Meier

Tableau 3.8 Estimation par la méthode de K.M. de la fonction de survie et des intervalles de confiance (95%) pour le groupe immunoperoxydase négatif.

Quoi qu'il en soit, l'étude par simulation a montré que la méthode bayésienne de Kalbfleisch fournit des estimations précises comparativement à la méthode fréquentiste de Kaplan-Meier.

La figure (3.2) représente les 6000 itérations des valeurs *a posteriori* de quelques durées de vie et de coefficient de régression qui permet de mesurer l'écart de mortalité du groupe « immunoperoxydase positif » par rapport au groupe « immunoperoxydase négatif ». Sur cette figure, on peut constater graphiquement une certaine stationnarité des valeurs *a posteriori* tout au long des 6000 itérations de la chaîne. Ceci indique que l'échantillon obtenu provient bien de la distribution souhaitée.

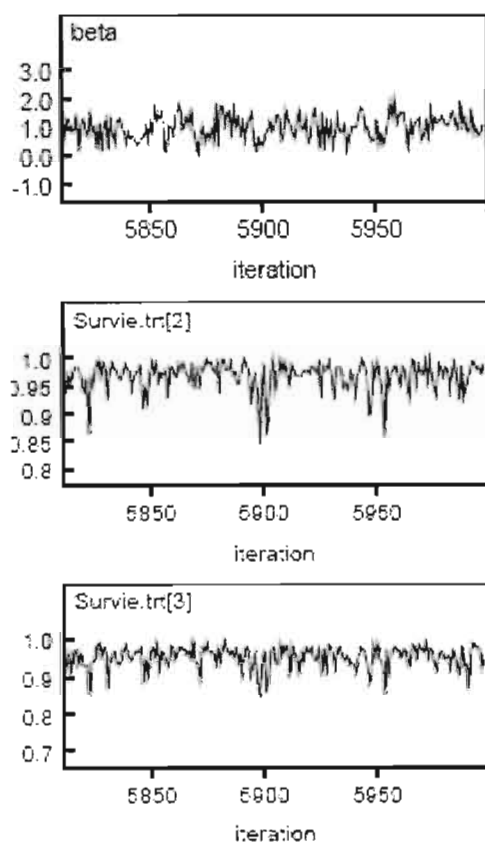


Figure 3.2 Valeurs *a posteriori* de quelques durées de vie et de coefficient de régression en fonction des itérations de la chaîne dans l'étude de l'exemple de cancer du sein.

3.6 Méthodes non paramétriques : la méthode de Florens et Rolin

Les premiers travaux publiés sur la statistique bayésienne non paramétrique sont ceux de Ferguson (Ferguson, 1973). Ils introduisent la notion de processus de Dirichlet comme probabilité sur une probabilité. Ce point de départ a suscité beaucoup de travaux étendant ce modèle à la prise en compte par exemple de données censurées (Sursarla, Van Ryzin, 1976)

Florens et Rolin (Florens et Rolin, 2001) ont démontré principalement que les estimations par simulation du processus de Dirichlet et l'inférence bayésienne non paramétrique, donnent de bons résultats par rapport aux méthodes classiques.

3.6.1 La distribution de Dirichlet

La distribution de Dirichlet de paramètre α est une distribution définie sur le simplexe S_{K-1} . Elle est caractérisée par la densité p_1, \dots, p_{K-1} par rapport à la mesure de Lebesgue dans \mathbb{R}^{K-1} et cette densité vérifie :

$$f(p_1, \dots, p_{K-1} | \alpha) = \frac{\Gamma(\alpha_1 + \dots + \alpha_K)}{\Gamma[\alpha_1] \dots \Gamma[\alpha_K]} \left[\prod_{i=1}^{K-1} p_i^{\alpha_i - 1} \right] \left[1 - \sum_{i=1}^{K-1} p_i \right]^{\alpha_K - 1}.$$

La moyenne de p_i est égale à

$$\frac{\alpha_i}{\sum_{i=1}^K \alpha_i},$$

et sa variance est

$$\frac{((\sum_{i=1}^K \alpha_i) - \alpha_i) \alpha_i}{(\sum_{i=1}^K \alpha_i)^2 + (\sum_{i=1}^K \alpha_i)^3}.$$

La distribution de Dirichlet généralise la distribution bêta : ces deux probabilités sont identiques si $K = 2$.

3.6.2 Processus de Dirichlet

Soient une probabilité Q et un réel positif ν . On dira que l'élément aléatoire P définit sur l'ensemble \mathcal{P} des probabilités suit un processus de Dirichlet $\mathcal{D}(Q, \nu)$ si $\forall (B_l)_{l=1, \dots, L}$ partition mesurable de \mathbb{R}^m , $(P(B_l))_{l=1, \dots, L}$ suit une loi de Dirichlet de paramètres $(\nu Q(B_l))_{l=1, \dots, L}$.

Le principal résultat justifiant le choix des processus de Dirichlet dans un modèle bayésien non

paramétrique est le suivant :

Théorème 3.4.1 Si P est *a priori* distribuée suivant un processus de Dirichlet $Dir(P_0, n_0)$ et si $x = (x_1, \dots, x_n)$ est un échantillon i.i.d. de loi P , on a les résultats suivants :

1. La loi *a posteriori* de P est un processus de Dirichlet $\mathcal{D}(P_*, n_*)$ avec :

$$n_* = n_0 + n, \quad (3.2)$$

$$P_* = \frac{n_0}{n + n_0} P_0 + \frac{n}{n + n_0} P_n, \quad (3.3)$$

où $P_n = (1/n) \sum \delta_{x_i}$ est la loi empirique de l'échantillon (δ_a représentant la mesure de Dirac en a).

2. La loi marginale de (x_1, \dots, x_n) peut être décomposée de la manière suivante :

–

$$x_1 \sim P_0$$

–

$$x_{i+1} | x_1, \dots, x_i \sim \frac{n_0}{i + n} P_0 + \frac{i}{i + n} P_i \quad (i = 1, \dots, n - 1)$$

où P_i est la loi empirique associée à l'échantillon (x_1, \dots, x_n) .

Le théorème 3.4.1 conduit aux remarques suivantes :

1. La distribution de Dirichlet constitue une famille fermée pour l'inférence dans un échantillon i.i.d. non paramétrique et les formules de transformation (3.1) et (3.2) des hyperparamètres (n_0, P_0) de la loi *a priori* en les hyperparamètres (n_*, P_*) de la loi *a posteriori* sont particulièrement simples.
2. Si n_0 est très faible, la distribution *a posteriori* devient égale au processus de Dirichlet $Dir(P_n, n)$ qui ne dépend que des caractéristiques de l'échantillon et qui est, en particulier, centrée sur P_n .

3.6.3 Modèle non paramétrique de durée de vie avec censure

On considère un couple aléatoire (T, C) de deux variables positives représentant respectivement la durée du phénomène d'intérêt et la durée de censure. Ces variables ne sont pas directement

observables et le statisticien ne dispose que d'un d'échantillon i.i.d. composé de n couples (Y_i, δ_i) , $i = 1, \dots, n$ où

$$Y_i = \min(T_i, C_i),$$

$$\delta_i = \begin{cases} 1 & \text{si } T_i \leq C_i, \\ 0 & \text{si } T_i > C_i. \end{cases}$$

On suppose que T et C sont indépendantes et que F et G sont des fonctions de distribution. On notera H la fonction de distribution de (Y, Δ) .

Supposons aussi que F a une distribution *a priori* de Dirichlet avec paramètres $n_0 \in \mathbb{R}^+$ et F_0 une mesure de probabilité,

$$F \sim \text{Dir}(n_0 F_0)$$

De cet échantillon $(Y_i, \delta_i)_{i=1, \dots, n}$, on extrait les temps de censure distincts que l'on ordonne en $a_1 < a_2 < \dots < a_m$ et on définit la partition :

$$B_1 = [0, a_1], B_2 =]a_1, a_2], \dots, B_m =]a_{m-1}, a_m], B_{m+1} =]a_m, \infty[.$$

Sous cette partition, F (et F_0) peut être décomposée sous la forme d'une distribution marginale et une autre conditionnelle qui sont définies par :

$$F_j = F(B_j) \quad 1 \leq j \leq m+1$$

et

$$F^j(*) = F(*|B_j) = \frac{F(* \cap B_j)}{F(B_j)} \quad 1 \leq j \leq m+1.$$

Il est clair que :

$$F = \sum_{1 \leq j \leq m+1} F_j F^j.$$

De la même manière, on exprime F_0 sous la forme suivante :

$$F_0 = \sum_{1 \leq j \leq m+1} F_{0j} F_0^j.$$

où

$$F_{0j} = F_0(B_j) \quad 1 \leq j \leq m+1$$

et

$$F_0^j(*) = F_0(*|B_j) = \frac{F_0(* \cap B_j)}{F_0(B_j)} \quad 1 \leq j \leq m+1.$$

Maintenant, les probabilités marginaux $\{F_j : 1 \leq j \leq m+1\}$ peuvent être remplacés en termes de probabilités de survie et de risque, qui seront particulièrement utiles dans le contexte de modèle non paramétrique de durée de vie avec censure.

On définit les probabilités de survie par :

$$\begin{aligned} S_0 &= 1, \\ S_j &= F((a_j, \infty]) = \sum_{j+1 \leq l \leq m+1} F_l \quad 1 \leq j \leq m, \\ S_{m+1} &= 0. \end{aligned}$$

de sorte que

$$F_j = S_{j-1} - S_j \quad 1 \leq j \leq m+1,$$

et les probabilités de risque par :

$$\begin{aligned} H_1 &= F_1 = 1 - S_1, \\ H_j &= \frac{F((a_{j-1}, a_j])}{F((a_{j-1}, \infty])} = \frac{F_j}{S_{j-1}} = 1 - \frac{S_j}{S_{j-1}} \quad 2 \leq j \leq m, \\ H_{m+1} &= 1. \end{aligned}$$

Ce qui nous permet d'exprimer en termes de probabilités de risque le produit des probabilités marginales de la manière suivante :

$$F_j = H_j \prod_{1 \leq l < j} (1 - H_l) \quad 1 \leq j \leq m+1,$$

et en terme de probabilités de survie :

$$S_j = \prod_{1 \leq l \leq j} (1 - H_l) \quad 1 \leq j \leq m+1.$$

Nous employons les mêmes notations pour les probabilités de survie et de risque définies par F_0

$$F_{0j} = H_{0j} \prod_{1 \leq l < j} (1 - H_{0l}) \quad 1 \leq j \leq m+1.$$

et en terme de probabilités de survie :

$$S_{0j} = \prod_{1 \leq l \leq j} (1 - H_{0l}) \quad 1 \leq j \leq m + 1.$$

Maintenant par une propriété connue de la distribution de Dirichlet (voir, par exemple, (Rolin, 1983)), nous avons la proposition suivante :

Proposition 3.4.1 Si $F \sim \text{Dir}(n_0 F_0)$, alors

1. Les H_j sont a priori indépendants \Leftrightarrow

$$\prod_{1 \leq j \leq m} H_j$$

2. Les H_j admettent des lois Beta de paramètres $n_0 F_{0j}$ et $n_0 S_{0j}$ \Leftrightarrow

$$H_j \sim \text{Beta}(n_0 F_{0j}, n_0 S_{0j}) \quad 1 \leq j \leq m$$

On montre alors en généralisant l'argument présenté précédemment pour une observation censurée que :

1. Les H_j et les F_j demeurent *a posteriori* indépendants.
2. H_j *a posteriori* suit une loi :

$$\text{Beta}(n_0 F_{0j} + D_j, n_0 S_{0j} + N_j + L_j)$$

où D_j est le nombre d'observations non censurées appartenant à $B_j = (a_{j-1}, a_j]$:

$$D_j = \sum_{i=1}^n \mathbf{1}(y_i \in B_j) d_i,$$

N_j est le nombre d'individus à risque juste après a_j :

$$N_j = \sum_{i=1}^n \mathbf{1}(y_i > a_j),$$

et L_j le nombre de données censurées en a_j :

$$L_j = \sum_{i=1}^n \mathbf{1}(y_i = a_j)(1 - d_i) \quad 1 \leq j \leq m.$$

Notons que

$$N_{j-1} = N_j + D_j + L_j \quad 1 \leq j \leq m.$$

3. F^j suit *a posteriori* un processus de Dirichlet de paramètres :

$$\begin{aligned} n_{*j} &= n_0 F_{0j} + n F_{unj} \\ F_*^j &= \frac{n_0 F_{0j}}{n_0 F_{0j} + n F_{unj}} F_0^j + \frac{n F_{unj}}{n_0 F_{0j} + n F_{unj}} F_{unj}^j, \end{aligned}$$

où F_{un} est la distribution empirique des observations non censurées :

$$F_{un} = \frac{1}{n} \sum_{i=1}^n \delta_{y_i} d_i,$$

où δ_{y_i} est la mesure de Dirac en y_i , à laquelle on associe

$$F_{unj} = F_{un}(B_j) = \frac{1}{n} D_j \quad 1 \leq j \leq m+1,$$

$$F_{un}^j(*) = F_{un}(**|B_j) \quad 1 \leq j \leq m+1,$$

loi conditionnelle à B_j

et

$$S_{nj} = F_n((a_j, \infty]) = \frac{1}{n} N_j \quad 1 \leq j \leq m.$$

3.7 Méthode bayésienne paramétrique pour des modèles avec points de rupture : la méthode de Carlin, Gelfand et Smith

Plusieurs articles traitent du problème de point de rupture où les observations Y_1, \dots, Y_k sont *iid* de densité $f(y|\theta)$ et Y_{k+1}, \dots, Y_n sont *iid* de densité $g(y|\lambda)$ avec un point de rupture tel que défini à la section (3.7.1) et ayant un indice appartenant à l'ensemble $\{1, \dots, n\}$. À cet effet, on peut mentionner Hinkley (Hinkley, 1970).

Chernoff et Zacks (Chernoff et Zacks, 1964) étaient les premiers à considérer le problème de rupture selon l'approche bayésienne paramétrique. Smith (Smith, 1975) a présenté une formulation bayésienne pour une séquence finie d'observations *iid*. En particulier, il considère trois cas :

1. $f(y|\theta)$ et $g(y|\lambda)$ sont connues ;
2. seulement $f(y|\theta)$ est connue ;
3. $f(y|\theta)$ et $g(y|\lambda)$ sont inconnues.

3.7.1 Formulation bayésienne

Selon Carlin, Gelfand et Smith (Carlin, Gelfand et Smith, 1992), la formulation la plus simple du problème de point de rupture est de considérer les observations $Y_i, i = 1, \dots, n$ comme étant *iid* avec

$$Y_i \sim f(y|\theta), \quad i = 1, \dots, k$$

et

$$Y_i \sim g(y), \quad i = k + 1, \dots, n,$$

où k inconnu prend sa valeur dans l'ensemble $\{1, \dots, n\}$. Ainsi, $k = n$ s'interprète comme étant l'absence de point de rupture dans le problème.

La vraisemblance pour ce modèle est donc définie par :

$$L(Y; k) = \prod_{i=1}^k f(Y_i) \prod_{i=k+1}^n g(Y_i). \quad (3.4)$$

On peut obtenir directement de l'équation précédente l'estimateur de vraisemblance maximale (EVM) du paramètre k .

Dans la formulation bayésienne de la vraisemblance $L(Y; k)$, on ajoute à k (inconnu) une densité *a priori* $\tau(k)$ définie sur $\{1, \dots, n\}$, ainsi la densité *a posteriori* de $k|Y$ est définie par :

$$k|Y \sim \frac{L(Y; k)\tau(k)}{\sum_{k=1}^n L(Y; k)\tau(k)}. \quad (3.5)$$

Avec cette densité *a posteriori* qui est facile à échantillonner ($k \in \{1, \dots, n\}$: cas discret), on peut définir toutes sortes d'indicateurs statistiques comme par exemple le mode *a posteriori*, moyenne *a posteriori* ou même l'intervalle de crédibilité de k .

Si on suppose que les densités $f(y)$ et $g(y)$ sont inconnues mais on suppose que $f(Y|\theta)$ appartient à une famille paramétrique, et que $g(Y|\lambda)$ possède une autre famille paramétrique. La vraisemblance est donnée par :

$$L(Y; k, \theta, \lambda) = \prod_{i=1}^k f(Y_i|\theta) \prod_{i=k+1}^n g(Y_i|\lambda).$$

Supposant aussi que les paramètres k, θ et λ ont une densité *a priori* égale à $\pi(\theta, \lambda, k)$; la distribution conjointe des $Y_i, i = 1, \dots, n$ et les paramètres seraient donc

$$L(Y; k, \theta, \lambda)\pi(\theta, \lambda, k). \quad (3.6)$$

Notre intérêt principal porte sur les distributions marginales *a posteriori* des paramètres k , θ et λ .

Avec l'échantillonnage de Gibbs discuté au premier chapitre de la section 1.6, on peut facilement obtenir ces distributions marginales *a posteriori* $k|Y$, $\theta|Y$ et $\lambda|Y$. La programmation de cette méthode nécessite l'échantillonnage à partir des distributions conditionnelles, $k|(Y, \theta, \lambda)$, $\theta|(Y, k, \lambda)$ et $\lambda|(Y, k, \theta)$.

Noton que chacune de ces distributions conditionnelles est en fonction de la distribution conjointe. Supposons maintenant que θ , λ et k sont indépendants, c.à.d

$$\pi(\theta, \lambda, k) = \pi_1(\theta)\pi_2(\lambda)\tau(k).$$

Soit π_1 la densité conjuguée de f et π_2 celle de g , alors la distribution conditionnelle $\theta|(Y, \lambda, k)$ ne dépendra pas de λ , c'est seulement la densité *a priori* π_1 qui a été actualisée avec les données Y_1, \dots, Y_k et de la même manière pour la distribution conditionnelle $\lambda|(Y, \theta, k)$ c'est la densité *a priori* π_2 mise à jour par les données Y_{k+1}, \dots, Y_n .

Avec

$$\theta \sim \pi_1(\theta|\alpha), \quad \lambda \sim \pi_2(\lambda|\beta),$$

où $\alpha \sim \rho_1(\alpha)$ et $\beta \sim \rho_2(\beta)$.

La distribution conjointe des données et des paramètres serait donc :

$$L(Y; k, \theta, \lambda)\tau(k)\pi_1(\theta|\alpha)\rho_1(\alpha)\pi_2(\lambda|\beta)\rho_2(\beta). \quad (3.7)$$

Ce qui nous permet, avec cette densité conjointe, de définir les densités conditionnelles. En particulier :

1. $k|(Y, \theta, \lambda, \alpha, \beta)$ indépendante des paramètres α et β ;
2. $\theta|(Y, k, \lambda, \alpha, \beta)$ indépendante des paramètres λ et β , c'est la densité *a priori* π_1 actualisée par les données Y_1, \dots, Y_k ;
3. $\lambda|(Y, k, \theta, \alpha, \beta)$ indépendante des paramètres θ et α , c'est la densité *a priori* π_2 actualisée par les données Y_{k+1}, \dots, Y_n ;
4. $\alpha|(Y, k, \theta, \lambda, \beta)$ indépendante de Y et des paramètres k , λ et β , c'est la densité *a priori* ρ_1 actualisée par θ ;

5. $\beta|(Y, k, \theta, \alpha, \lambda)$ indépendante de Y et des paramètres k, θ et α , c'est la densité *a priori* ρ_2 actualisée par λ .

3.7.2 Choix de modèle et des densités *a priori* pour les paramètres

Afin d'appliquer cette méthode, Carlin, Gelfand et Smith (Carlin, Gelfand et Smith, 1992) utilisent trois étapes hiérarchiques pour leur modèles en supposant que le point de rupture serait entre les intervalles. En premier lieu, ils choisissent le processus de Poisson pour les données c-à-d

$$Y_i \sim P(\theta t_i), \quad i = 1, \dots, k,$$

et

$$Y_i \sim P(\lambda t_i), \quad i = k + 1, \dots, n$$

Si on utilise le tableau 1.1 des lois conjuguées (section 1.2), les densités *a priori* pour les paramètres θ et λ seront de la forme suivante :

$$\theta \sim G(a_1, b_1)$$

et

$$\lambda \sim G(a_2, b_2)$$

La troisième étape consiste à prendre des lois gamma-inverse pour b_1 et b_2

$$b_1 \sim IG(c_1, d_1)$$

qui sera indépendante de

$$b_2 \sim IG(c_2, d_2)$$

Rappelons qu'une gamma-inverse est la distribution de probabilité de l'inverse d'une variable aléatoire $Y \sim G(\lambda, \alpha)$ et sa fonction de densité est donnée par

$$p(x = 1/y) = \frac{\lambda^\alpha (1/x)^{\alpha+1} \exp\{-\lambda/x\}}{\Gamma(\alpha)}.$$

On suppose que a_1, a_2, c_1, c_2, d_1 et d_2 sont connus.

Pour le paramètre k (point de rupture à estimer), on suppose que sa densité *a priori* est la loi

uniforme sur l'intervalle $[0, n]$ qui est une loi non informative. Le choix de la loi uniforme pour k est motivée par le fait que lorsqu'on ne dispose pas d'information pertinente pour choisir une loi *a priori*, et si on veut cependant utiliser l'approche bayésienne, on peut faire appel à des lois *a priori* non informatives ; si pour un paramètre qui peut prendre un nombre fini de valeurs ceci semble raisonnable et ne pose aucun problème.

3.7.3 Lois *a posteriori* des paramètres

Notre intérêt principal est de définir les densités *a posteriori* marginales des paramètres k , θ et λ en actualisant les lois *a priori* par les données $Y_i, i = 1, \dots, n$. Pour cela, on utilise l'algorithme de Gibbs. Après avoir choisi un point de départ, les 3 composantes du vecteur de covariables ($\Theta = (k, \theta, \lambda)$) sont générées les unes après les autres conditionnellement à toutes les autres composantes. Si $\pi(\Theta | y)$ est la densité des 3 composantes du vecteur Θ , conditionnellement aux données observées (y), nous utilisons alors les densités conditionnelles $\pi(k | \theta, \lambda, y)$, $\pi(\theta | k, \lambda, y)$, et ainsi de suite. À chaque j^e étape, la distribution conditionnelle utilise les valeurs générées les plus récentes parmi toutes les autres composantes. Par la théorie des chaînes de Markov, lorsque $j \rightarrow \infty$ la densité des réalisations obtenues converge vers $\pi(\Theta | y)$. Ce qui nous donne les densités *a posteriori* marginales suivantes :

$$\begin{aligned}\theta | (Y, \lambda, b_1, b_2, k) &\sim G \left\{ a_1 + \sum_{i=1}^k Y_i, \left(\sum_{i=1}^k t_i + \frac{1}{b_1} \right)^{-1} \right\}, \\ \lambda | (Y, \theta, b_1, b_2, k) &\sim G \left\{ a_2 + \sum_{i=k+1}^n Y_i, \left(\sum_{i=k+1}^n t_i + \frac{1}{b_2} \right)^{-1} \right\}, \\ b_1 | (Y, \theta, \lambda, b_2, k) &\sim IG \left\{ a_1 + c_1, \left(\theta + \frac{1}{d_1} \right)^{-1} \right\}, \\ b_2 | (Y, \theta, \lambda, b_1, k) &\sim IG \left\{ a_2 + c_2, \left(\lambda + \frac{1}{d_2} \right)^{-1} \right\}\end{aligned}$$

et

$$p(k | (Y, \theta, \lambda, b_1, b_2)) = \frac{\exp \left\{ (\lambda - \theta) \sum_{i=1}^k t_i \right\} \left(\frac{\theta}{\lambda} \right)^{\sum_{i=1}^k Y_i}}{\sum_{k=1}^n \exp \left\{ (\lambda - \theta) \sum_{i=1}^k t_i \right\} \left(\frac{\theta}{\lambda} \right)^{\sum_{i=1}^k Y_i}}$$

Une autre variable d'intérêt est $R = \frac{\theta}{\lambda}$ dont la densité *a posteriori* est définie par :

$$R | (Y, \lambda, b_1, b_2, k) \sim G \left\{ a_1 + \sum_{i=1}^k Y_i, \frac{1}{\lambda} \left(\sum_{i=1}^k t_i + \frac{1}{b_1} \right)^{-1} \right\},$$

3.8 Exemple d'application de la méthode de Carlin, Gelfand et Smith

3.8.1 Cas des accidents des mines

Les données sont relatives aux accidents des mines en Grande-Bretagne entre 1851 et 1962 (les données sont rapportées dans le tableau 3.9). Il s'agit d'un comptage du nombre d'accidents dans les mines par année dont l'objectif est l'estimation de point de changement (l'année) qui a fait que ce nombre d'accidents a diminué de manière très significative.

Cette base de données a été utilisée pour la première fois par Marguire, Pearson et Wynn (Marguire, Pearson et Wynn, 1952). L'approche fréquentiste a été analysée par plusieurs auteurs dont Worsley (Worsley, 1986) et Siegmund (Siegmund, 1986). Raftery et Akman (Raftery, 1986) ont appliqué pour la première fois l'approche bayésienne à cette base de donnée dont les valeurs initiales pour leurs modèle ont été $a_1 = a_2 = 0$ et $b_1 = b_2 = 0$.

Carlin, Gelfand et Smith (Carlin, Gelfand et Smith, 1992) ont utilisé les mêmes valeurs sauf pour les paramètres b_1 et b_2 , ont donné des densités *a priori* non informatives de type Gamma.

$$b_1 \sim \text{Gamma}(c_1 = 0, d_1 = 1) \quad \text{et} \quad b_2 \sim \text{Gamma}(c_2 = 0, d_2 = 1).$$

<i>Year</i>	<i>Count</i>	<i>Year</i>	<i>Count</i>	<i>Year</i>	<i>Count</i>	<i>Year</i>	<i>Count</i>
1851	4	1881	2	1911	0	1941	4
1852	5	1882	5	1912	1	1942	2
1853	4	1883	2	1913	1	1943	0
1854	1	1884	2	1914	1	1944	0
1855	0	1885	3	1915	0	1945	0
1856	4	1886	4	1916	1	1946	1
1857	3	1887	2	1917	0	1947	4
1858	4	1888	1	1918	1	1948	0
1859	0	1889	3	1919	0	1949	0
1860	6	1890	2	1920	0	1950	0
1861	3	1891	2	1921	0	1951	1
1862	3	1892	1	1922	2	1952	0
1863	4	1893	1	1923	1	1953	0
1864	0	1894	1	1924	0	1954	0
1865	2	1895	1	1925	0	1955	0
1866	6	1896	3	1926	0	1956	0
1867	3	1897	0	1927	1	1957	1
1868	3	1898	0	1928	1	1958	0
1869	5	1899	1	1929	0	1959	0
1870	4	1900	0	1930	2	1960	1
1871	5	1901	1	1931	3	1961	0
1872	3	1902	1	1932	3	1962	1
1873	1	1903	0	1933	1		
1874	4	1904	0	1934	1		
1875	4	1905	3	1935	2		
1876	1	1906	1	1936	1		
1877	5	1907	0	1937	1		
1878	5	1908	3	1938	1		
1879	3	1909	2	1939	1		
1880	4	1910	2	1940	2		

Tableau 3.9 Base de données sur les accidents des mines en Grande-Bretagne entre 1851 et 1962 citée dans l'article de Carlin, Gelfand et Smith.

3.8.2 Estimation avec WinBUGS

Comme cela se fait habituellement avec les algorithmes MCMC, les premières itérations de la chaîne (les 1000 premières dans notre cas) n'ont pas été retenues et les estimations ont été calculées à partir des valeurs de la distribution *a posteriori* obtenues de ce qui reste de la chaîne (les 10 000 itérations suivantes). Les moyennes *a posteriori* (écart-type) des paramètres k , θ , λ et $R = \frac{\theta}{\lambda}$ étaient respectivement $k = 40.15(2.45)$, $\beta = \text{beta}[1] = 3.07(0.2851)$, $\lambda = \text{beta}[2] = 0.9067(0.1159)$ et $R = 3.44(0.5408)$. Si on regarde le graphique de la densité *a posteriori* du paramètre k (figure 3.3), on peut remarquer la présence de trois sommets (en $k = 39$, $k = 40$ et $k = 41$); ce qui indique que le changement a eu lieu au plus tôt en 1889 ou au plus tard

node sample	mean	sd	MC error	2.5%	median	97.5%	start	
R	3.44	0.5408	0.005337	2.514	3.398	4.628	1000	11000
beta[1]	3.07	0.2851	0.002757	2.539	3.061	3.659	1000	11000
beta[2]	0.9067	0.1159	0.001313	0.6923	0.9015	1.146	1000	11000
k	40.15	2.45	0.02668	36.0	40.0	46.0	1000	11000

Tableau 3.10 Estimation des densités *a posteriori* $k|Y$, $\theta|Y$ et $\lambda|Y$ pour les données sur les accidents des mines en Grande-Bretagne entre 1851 et 1962 ($a_1 = a_2 = 0.5$; $c_1 = c_2 = 0$; $d_1 = d_2 = 1$).

en 1892. Ces résultats rivalisent avec ceux de Raftery et Akman (Raftery, 1986) qui ont obtenu comme valeur modale *a posteriori* le 10 mars 1890 et une médiane *a posteriori* le 27 août 1890 et celles de Worsley (Worsley, 1986) et sa région critique $C_\alpha = \{36, \dots, 53\}$ dans laquelle on trouve l'année ou le nombre d'accident a diminué d'une manière significatif.

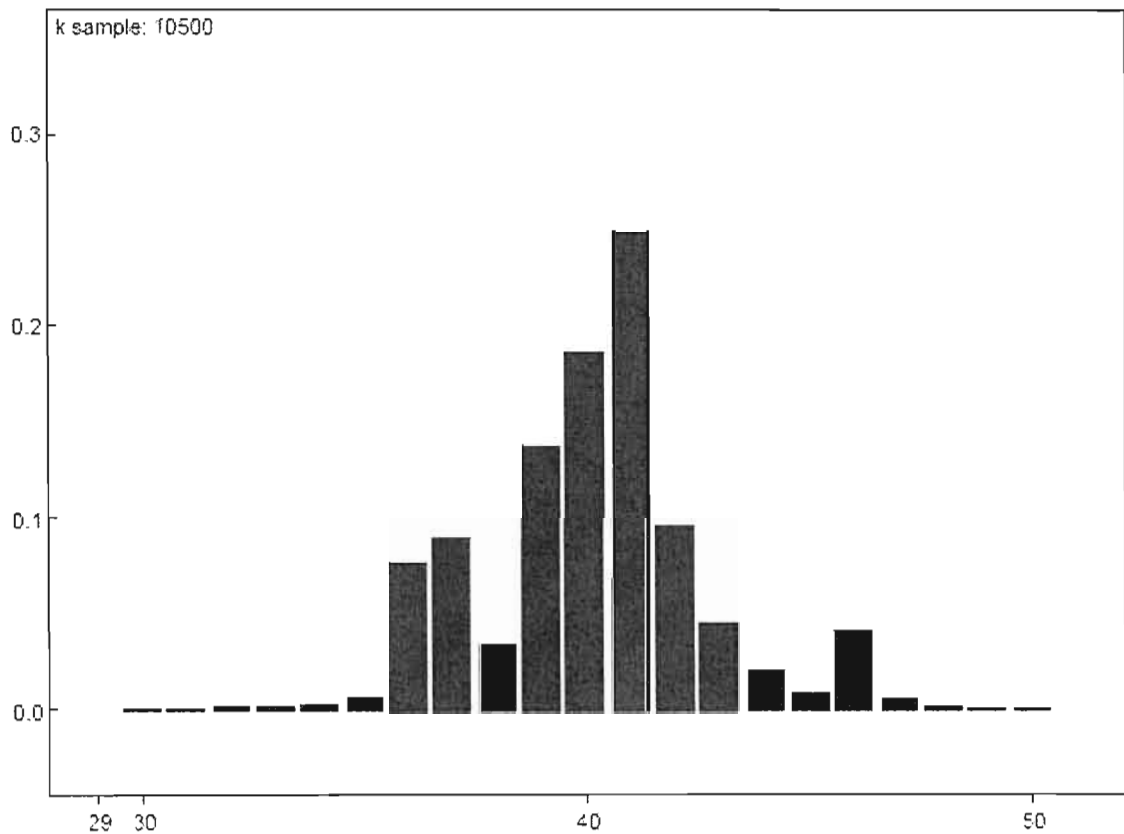


Figure 3.3 Estimation des densités *a posteriori* $k|Y$, $\theta|Y$ et $\lambda|Y$ pour les données sur les accidents des mines en Grande-Bretagne entre 1851 et 1962 ($a_1 = a_2 = 0.5$; $c_1 = c_2 = 0$; $d_1 = d_2 = 1$).

La figure(3.4) représente les 10000 itérations des valeurs *a posteriori* des paramètres R , $\beta = \text{beta}[1]$, $\lambda = \text{beta}[2]$ et k . Sur cette figure, nous pouvons constater graphiquement une certaine stationnarité des valeurs *a posteriori* tout au long des 10000 itérations de la chaîne ; ceci indique que les échantillons obtenus proviennent bien des distributions souhaitées.

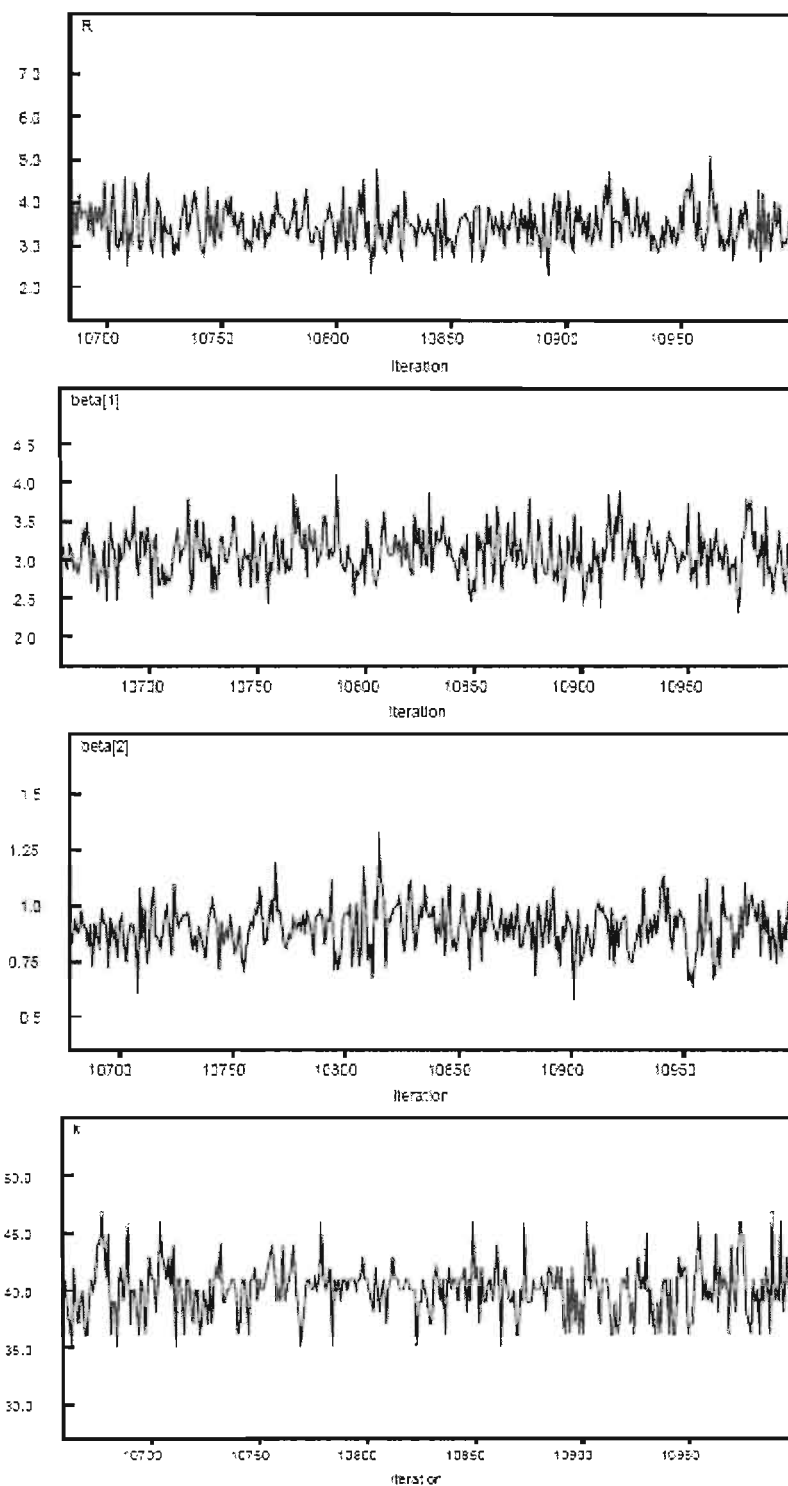


Figure 3.4 Valeurs *a posteriori* des paramètres en fonction des itérations de la chaîne dans l'étude de l'exemple des données sur les accidents des mines en Grande-Bretagne entre 1851 et 1962.

CONCLUSION

L'étude des durées de vie s'est considérablement développée dans le domaine biomédical afin d'étudier, pour des études de mortalité, l'effet de certains facteurs thérapeutiques ou pronostiques sur l'apparition d'un évènement au cours du temps (par exemple : le décès, la récurrence d'une pathologie...). Le modèle développé au début des années 1970 par David Cox (Cox, 1972) constitue l'une des contributions majeures à cette évolution. En effet, ce modèle a été le premier à permettre d'étudier simultanément l'effet de plusieurs facteurs pronostiques en établissant une relation entre eux et la distribution de survie, cette dernière n'ayant pas à être précisée.

L'analyse bayésienne constitue un autre développement important de la biostatistique de ces 20 dernières années. L'utilisation de méthodes bayésiennes peut présenter certains avantages. En effet, ces méthodes utilisent des connaissances antérieures, exprimées sous forme de distribution de probabilité, afin de modifier une information nouvelle. L'utilisation de lois de distributions, et surtout de leurs propriétés, offre une certaine souplesse pour modéliser des problèmes biomédicaux complexes (Richardson, 1996). Jusqu'à récemment, un obstacle à l'utilisation de ces méthodes venait du fait que, dans des problèmes biomédicaux complexes, la distribution a posteriori des paramètres étudiés est généralement multidimensionnelle et possède rarement une écriture analytique simple. Le développement d'algorithmes itératifs possédant des propriétés markoviennes a permis de surmonter cet obstacle dans la mesure où ceux-ci permettent d'échantillonner à partir de la distribution a posteriori, sans qu'il soit nécessaire de la spécifier explicitement, et assurent la convergence en distribution vers la distribution a posteriori. L'échantillonneur de Gibbs (1984) et l'algorithme de Metropolis (Metropolis, Rosenbluth et Teller, 1953) sont les algorithmes stochastiques de Monte-Carlo par Chaîne de Markov les plus utilisés dans les méthodes d'analyse bayésienne. Dans le domaine de l'analyse de survie, les méthodes bayésiennes permettent de modéliser conjointement le taux de base de décès et les coefficients de régression des covariables. Les analyses bayésiennes développées à partir du modèle de Cox ((Kalbfleisch, 1978), (Chen, Shao et Ibrahim, 2000)) ont montré la précision des

estimations et l'intérêt d'utiliser ces méthodes. Nous avons exploité ces avantages pour analyser et implanter avec le logiciel **BUGS** deux méthodes : une première méthode semi-paramétrique, il s'agit de la méthode de Kalbfleisch, et une seconde méthode paramétrique, celle de Carlon, Gelfand et Smith.

Des études réalisées sur des données de patients atteints de la leucémie, où la transplantation de moelle représentant le seul espoir réaliste pour la survie à long terme, ont montré l'efficacité des méthodes bayésiennes. On démontre principalement que les estimations par MCMC et l'inférence bayésienne paramétrique donnent de bons résultats par rapport aux méthodes classiques.

APPENDICE A

PROGRAMMES WINBUGS POUR LES DEUX MÉTHODES BAYÉSIENNES UTILISÉES DANS LE CHAPITRE 3

```
\#\# programme utilise dans la methode semi parametrique de  
Kalbfleisch pour le probleme avec censure (voir Spiegelhalter 1996b)\#\#\#\  
  
model Kalbfleisch;  
  
var  
beta,          # coefficient de regression  
failtime[N],  # temps de remission ou temps de censure de chaque patient  
t[T+1],       # temps de rechute avec un temps de censure maximum  
dN[N,T],      # increment dans le processus de comptage  
Y[N,T],       # 1 = si le sujet est observe; 0 = sinon  
Idt[N,T],     # processus d'intensite  
Z.trt[N],     # covariable dans le modele  
H[T],         # l'increment du risque instantane  
H.star[T],    # loi a priori de la fonction cumulative de risque  
c,            # degre de confiance que l'on attache a la loi a priori de dL0  
mu[T],        # parametre pour la loi gamma a priori ( = c0 * dL0.star)  
r,            # valeur a priori pour le taux de rechute  
failcens[N],  # rechute = 1, censure = 0  
Survie.trt[N], # fonction de survie pour le groupe traite  
Survie.placebo[N], # fonction de survie pour le groupe placebo  
  
{  
# loi a priori non informative pour le coefficient de regression
```



```

beta ~ dnorm(0.0,1.0E-7);

# Etape 1: actualisation des donnees

for(i in 1:N) {
  for(j in 1:T) {

    # risk set = 1 if failtime >= t
    Y[i,j] <- step(failtime[i] - t[j] + eps);

    # Processus de comptage jump = 1 if failtime in [ t[j], t[j+1] )
    # c.a.d, si t[j] <= failtime < t[j+1]

    dN[i,j] <- Y[i,j]*step(t[j+1] - failtime[i] - eps)*failcens[i];
  }
}

# Etape 2:Le model
for(j in 1:T){
  for (i in 1:N) {

    dN[i,j] ~ dpois(Idt[i,j]); # la vraisemblance
    Idt[i,j] <- Y[i,j]*exp(beta*Z.trt[i])*H[j]; # Le Processus d'intensite

  }
}
H[j]~dgamma(mu[j],c) # Loi a priori gamma pour le taux de hasard cumule
mu[j] <- H.star[j]*c;# moyenne a priori pour le risque instantane de deces

# Etape 3:Estimation de la fonction de survie =
#exp(- Integral{10(u) du})^{exp(beta*z)}

Survie.trt[j] <- pow(exp(-sum(H[1:j])), exp(beta* -0.5));
#Estimation de la fonction de survie pour le groupe traite

```

```

Survie.placebo[j] <- pow(exp(-sum(H[1:j])), exp(beta*0.5))
#Estimation de la fonction de survie pour le groupe placebo

    }
for(j in 1:T) {
  H.star[j] <- r * (t[j+1] -t[j])
}

Data list(N = 42, T = 17, eps = 1.0E-6,r=0.1,c=1

failtime = c(1, 1, 2, 2, 3, 4, 4, 5, 5, 8, 8, 8, 8, 11, 11, 12,
12, 15, 17, 22, 23, 6, 6, 6, 6, 7, 9, 10, 10, 11, 13, 16, 17,
19, 20, 22, 23, 25, 32,32, 34, 35),

failcens= c(1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1,1, 1,1, 0, 1, 0, 1, 0, 0, 1, 1, 0, 0, 0, 1, 1, 0, 0, 0,
0,0),

Z.trt = c(0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5,
0.5,0.5, 0.5,0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5,-0.5, -0.5, -0.5,
-0.5, -0.5, -0.5, -0.5, -0.5, -0.5, -0.5, -0.5, -0.5, -0.5, -0.5,
-0.5, -0.5, -0.5, -0.5,-0.5, -0.5, -0.5),

t = c(1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 15, 16, 17, 22, 23,
35) )

Inits list( beta = 0.0,
H=c(1.0,1.0,1.0,1.0,1.0,1.0,1.0,1.0,1.0,1.0,1.0,1.0,1.0,1.0,1.0,
1.0,1.0))

Inits list( beta = 1.0,
H=c(0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,
0.5,0.5))

\#\# programme utilise dans la methode parametrique de Carlin,

```

```

Gelfand et Smith pour
le probleme avec point de rupture\#\#\

model point_de_rupture

{
  # Etape 1: actualisation des donnees avec les lois a priori

  # lois a priori non informative pour les parametres k, b1 et b2

  b1 ~ dgamma(0.0,d1);
  b2 ~ dgamma(0.0,d2);

  blinv <- 1.0/b1;
  b2inv <- 1.0/b2;

  k ~ dcat(puni[ ]); # loi uniforme sur l'intervalle [0,N]

  beta[1] ~ dgamma(a1, blinv); # loi a priori gamma qui est le
                               conjuguee naturelle de loi de Poisson
  beta[2] ~ dgamma(a2, b2inv);

  R <- beta[1]/beta[2]; # Rapport entre les deux parametres beta[1]
  et beta[2]

  # Etape 2:Le model

  for(i in 1:N) {
    j[i] <- 1 + step(t[i] - (k + 0.5))
    Y[i] ~ dpois(beta[j[i]]) # la vraisemblance
  }
}

Data list(N = 112, a1=0.5, a2=0.5, d1=1.0, d2=1.0, t=
c(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,
22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,

```



```
0.008928571,0.008928571,0.008928571,0.008928571,0.008928571,  
0.008928571,0.008928571,  
0.008928571,0.008928571,0.008928571,0.008928571,0.008928571,  
0.008928571,0.008928571,  
0.008928571,0.008928571,0.008928571,0.008928571,0.008928571,  
0.008928571,0.008928571,  
0.008928571,0.008928571,0.008928571,0.008928571,0.008928571,  
0.008928571,0.008928571,  
0.008928571  
))  
Inits list(beta =c(3,1), b1=1.0, b2=1.0, k=112))
```

BIBLIOGRAPHIE

- Aalen OO. (1978) Nonparametric inference for a family of counting processes. *Annals of Statistics* 6 : 701-726.
- Abraham C, Daures JP. (2000) Global robustness with respect to the loss function and the prior. *Theory and decision* 48 : 359-81.
- Allison, Paul D. (1995). *Survival Analysis Using the SAS System : A Practical Guide*. Cary, NC : SAS Institute Inc.
- Anderson, P.K. (1991). Survival analysis 1982-1991 : the second decade of the proportional hazards regression model. *Statistics in medicine* 10 :1931-41.
- Barnard GA. (1958) Thomas Bayes - a biographical note. *Biometrika* 45 : 293-315.
- Berger JO. (1985) *Statistical decision theory and bayesian analysis* (2nd edition). *Springer-Verlag*, New York.
- Berger JO. (1994) An overview of robust Bayesian analysis. *Test* 3 : 5-124.
- Best, N.G., Cowles, M.K. et Vines, S.K. 1995, **CODA** :Convergence Diagnosis and Output Analysis software for Gibbs sampling output. *Medical Research Council Biostatistics Unit*, Cambridge University. [http://www.mrc-bsu.cam.ac.uk/bugs/\(2005\)](http://www.mrc-bsu.cam.ac.uk/bugs/(2005))
- Bohmer P.E.. (1912). Theorie der unabhängigen Wahrscheinlichkeiten Rapports. *Mémoires et procès verbaux du septième congrès international d'actuares*. Amsterdam, 2 :327-43.
- Carlin, B.P., Gelfand, A.E., Smith, A.F.M. (1992). Hierarchical Bayesian Analysis of Changepoint Problems. *Appl. Statist.* 41, No.2, 389-405.
- Chen, M.H., Shao, Q.M. et Ibrahim, J.G. (2000). *Monte Carlo Methods in Bayesian Computation*. New York : Springer.
- Chernoff, H et Zacks, S. (1964). Estimating the current mean of a normal distribution which is subjected to changes in time. *Ann. Math. Statist.*, 35, 999-1018.
- Clayton, D.G. (1991). A Monte Carlo Method for Bayesian Inference in Frailty Models. *Biometrics*. 47, 467-485.
- Cowles, M.K. et Carlin, B.P. (1996), Markov Chain Monte-Carlo convergence diagnostics : A comparative study. *Journal of the American Statistical Association*, 91, 883-904.
- Cox, D.R. (1972). Regression models and life tables (with discussion). *J.R. Stat. Soc. B.* 34 :187-202.
- Cox DR. (1975) Partial likelihood. *Biometrika* 62 : 269-76.
- Cox, D.R. et Oakes, D. (1984). *Analysis of survival data*. London : *Chapman and Hall*.

- Feigl, P. et Zelen, M. (1965). Estimation of exponential survival probabilities with concomitant information *Biometrics* 21 : 826-838.
- Ferguson, T.S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics* 1 :209-230.
- Ferguson TS. (1973) A Bayesian analysis of some nonparametric problems. *The Annals of Statistics* 1 : 209-230.
- Fleming, T.R., et Harrington, D.P. (1991). Counting Processes and Survival Analysis. *New York : John Wiley and Sons*
- Florens, J.P. et Rolin, J.M. (2001). Simulation of Posterior Distributions in Nonparametric Censored Analysis. *International Statistical Review* 67(2), 187-210.
- Freireich, E.J., Gehan, E., Frei, E., Schroeder, L.R., Wolman, I.J., Anbari, R., Burgert, E.O., Mills, S.D., Pinkel, D., Selawry, O.S., Moon, J.H., Gendel, B.R., Spurr, C.L., Storrs, R., Haurani, F., Hoogstraten, B. et Lee, S. (1963). The effect of 6-Mercaptopurine on the duration of Steroid-Induced remissions in Acute Leukemia : A Model for Evaluation of Other Potentially Useful Therapy. *Blood*, 21, 1963, 699-716.
- Geisser S. (1993) Predictive inference : an introduction. *New York : Chapman and Hall*.
- Gelfand AE, Smith AFM. (1990) Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association* 85 : 398-409.
- Geman, S. et Geman, D. 1984. Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *IEEE Transaction on Pattern Analysis and Machine Intelligence* 6 : 721-741.
- Gilks, W.R., Richardson, S. et Spiegelhalter, D.J. 1996. Markov Chain Monte Carlo in Practice. *Interdisciplinary Statistics. Chapman et Hall*.
- Greenwood, M. (1926) The natural duration of cancer. *Reports on Public Health and Medical Subjects* No.33 (1926).
- Hastings, WK. (1970) Monte Carlo sampling methods using markow chains and their application. *Biometrika* 57 : 97-109.
- Hinkley, D.V. (1970) Inference about the change-point in a sequence of random variables. *Biometrika* 57, 1-17. .
- Hjort NL. (1990) Nonparametric Bayes estimation based on beta processes in models for life history data. *The Annals of Statistics* 18 : 1259-94.
- Hubert, C. et Lecoutre, J.P. (1989). Estimation fonctionnelle dans les modèles de survie. Modélisation des données censurées. Paris : *Economica*.
- Kaplan, E.L., Meier, P. (1958). Non parametric estimation from incomplete observations. *Journal of the American Statistical Association* 53 : 457-481.
- Kalbfleisch, J.D.(1978). Nonparametric Bayesian analysis of survival time data. *Journal of the Royal Statistical Society, Serie B* 40 (2) :214-221.

- Kalbfleisch, J.D. et Prentice, R. (1980). *The Statistical Analysis of Failure Time Data*. New York : *John Wiley and Sons*.
- Klein, J.P., Moeschberger, M.L. (1997). *Survival Analysis*. *Springer, New York*.
- Kuo, L., Smith, Adrian F.M. (1992). Bayesian computations in survival models via the Gibbs sampler. *Survival Analysis : State of the Art*, ed. Klein, J.P. and Goel, P.K., *Dodrecht : Kluwer Academics*, pp. 11-24.
- Lawless JF. (1982) *Statistical Models and Method for Lifetime Data*. New York : *Wiley*.
- Lawless, J.F. (1992). *Statistical Methods for Survival Data Analysis*. New York : *John Wiley and Sons*.
- Lee, T.T. (1982). *Statistical Models and Methods for Lifetime Data*. New York : *John Wiley and Sons*.
- Lehmann, E.L. et Casella, Ce. (1998). *Theory of Point Estimation (second edition)*, *Springer-Verlag* New York.
- MacEachern, S. et Muller, P. (1999). Efficient estimation of mixture of Dirichlet process models, *Technical Report. Department of Statistics, Ohio State University*, Columbus OH.
- Marguire, B.A., Pearson, E.S. et Wynn, A.H.A. (1952). The time intervals between industrial accidents. *Biometrika*, 38, 168-180.
- Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. (1953) Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, 21 : 1087-92.
- Morris, C.N. (1983). Natural exponential families with quadratic variance functions : statistical theory. *Ann. Statist.* 11, 515-529.
- Raftery, A.E. et Akman, V.E. (1986). Bayesian analysis of a Poisson process with a change-point. *Biometrika*, 73, 85-89.
- Richardson S. 1996. Développements récents de la biostatistique. *Revue d'Epidémiologie et de Santé Publique* 44 : 482-493.
- Robert, C. (1992). *L'analyse statistique bayésienne*. Statistique mathématique et probabilité. *Economica*.
- Robert, C. (1996). *Méthodes Monte Carlo par chaînes de Markov*. Statistique mathématique et probabilité. *Economica*.
- Rolin, J.M. (1983). Nonparametric Bayesian Statistics : A Stochastic Process Approach, in Specifying Statistical Models, from Parametric to Nonparametric using Bayesian or Non-Bayesian Approaches. *Lecture Notes in Statistics* vol 16, Springer, Berlin, 108-133.
- Sedmak, D.D., Meineke, T.A., Knechtges, D.S. et Anderson, J. (1989). Prognostic Significance of Cytokeratin-Positive Breast Cancer Metastases. *Modern Pathology*, 2, 516-520.
- Siegmund, D. (1986). Boundary crossing probabilities and statistical applications. *Ann. Statist.*, 14, 361-404.

- Smith, A.F.M. (1975). A Bayesian approach to inference about a change-point in a sequence of random variables. *Biometrika*, 62, 407-416.
- Spiegelhalter, D.J., Thomas, A., Best, N.G. et Gilks, W.R. (1996a) BUGS :Bayesian Inference Using Gibbs Sampling, version 1.4.1. *Medical Research Council Biostatistics Unit*, Cambridge University. [http ://www.mrc-bsu.cam.ac.uk/bugs/\(2005\)](http://www.mrc-bsu.cam.ac.uk/bugs/(2005))
- Spiegelhalter, D.J., Thomas, A., Best, N.G. et Gilks, W.R. (1996b) BUGS :Examples, version 1.4.1. *Medical Research Council Biostatistics Unit*, Cambridge University. [http ://www.mrc-bsu.cam.ac.uk/bugs/\(2005\)](http://www.mrc-bsu.cam.ac.uk/bugs/(2005))
- Sursarla, V., Van Rayzin, J. (1976). Nonparametric Bayesian estimation of survival curves from incomplete observation. *Journal of the American Statistical Association*71 :897-902.
- Weibull, W., (1951) « A statistical distribution function of wide applicability », *Journal of Applied Mechanics*, vol. 18, 1951, pp. 293-297.
- Worsley, K.J. (1986). Confidence regions and tests for change-point in a sequence of exponential family random variables. *Biometrika*, 73, 91-104.