

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE MENTOURI – CONSTANTINE
FACULTE DES SCIENCES

DEPARTEMENT DE MATHEMATIQUES



BEL
3820

N° d'ordre :

N° de série :

THESE PRESENTEE POUR L'OBTENTION
DU
DIPLÔME DE DOCTORAT D'ETAT
EN
MATHEMATIQUES

Titre : MODELES DE STATISTIQUE SEMI-PARAMETRIQUE
EN FIABILITE

(Mesures prédictives dans le modèle de Cox-Dirichlet avec censures)

Par Fatiha MESSACI BELALOU

Spécialité : PROBABILITES - STATISTIQUES

Devant le jury :

M. N. BENKAFADAR	<i>Prof. Université de Constantine</i>	<i>Président</i>
J. P. RAOULT	<i>Prof. Université René Descartes Paris V</i>	<i>Rapporteur</i>
L. ABBAOUI	<i>Prof. Université de Sétif</i>	<i>Examineur</i>
M. ROUSSIGNOL	<i>Prof. Université de Marne la Vallée</i>	<i>Examineur</i>
D. AISSANI	<i>Prof. Université de Béjaia</i>	<i>Examineur</i>
Z. MOHDEB	<i>M. C. Université de Constantine</i>	<i>Examineur</i>

Soutenue le, 25 janvier 2003

NOM : Fatiha MESSACI BELALOU

TITRE : MODELES DE STATISTIQUE SEMI-PARAMETRIQUE EN FIABILITE

RESUME. L'objet de cette thèse est d'étendre, au cas d'observations censurées, l'étude des mesures prédictives dans le modèle bayésien de Cox-Dirichlet effectuée par N. Gouget et J. P. Raoult pour des observations non censurées, à l'aide de la notion de mesure régulière par portions dont ils ont calculé les densités par rapport aux mesures de Lebesgue sur des sous ensembles caractérisés par des égalités et des inégalités strictes entre les observations. Cette notion reste centrale dans notre étude. Nous considérons successivement des cas de censure aléatoire, où les sous ensembles remarquables sont inchangés mais le calcul de densités doit être modifié, puis ceux des censures fixes, où les sous ensembles remarquables sont astreints à être contenus dans les "faces" de l'ensemble des observations, c'est à dire les parties caractérisées par la fixation des variables censurées.

MOTS CLES : Densités prédictives, Modèle de Cox, A priori de Dirichlet, Censures aléatoires, Censures fixes.

DISCIPLINE : Mathématiques appliquées, option Probabilités -Statistiques
Département de mathématiques, Université Mentouri, Constantine, 25000, Algérie.

A la mémoire de mon père

A Lamia et Mehdi qui ont subi les absences de maman.

A mon mari et à ma sœur qui m'ont encouragée à m'engager dans ce travail et qui m'ont soutenue et aidée tout au long de sa réalisation.

A ma mère et à mes frères qui m'ont apportée leur précieuse aide.

A toute ma famille, en particulier à Khelifa, Lamine , Iness, Souheila , Samira et Kamel.

A toute ma belle famille, en particulier à mes beaux parents.

A tous mes amis en particulier à Fella et Kamla qui m'ont aidée et encouragée.

Remerciements

J'ai déjà pu apprécier les qualités professionnelles et humaines de Jean Pierre RAOULT alors qu'il était mon enseignant de statistique au D-E-A puis membre de mon jury de thèse de doctorat de 3^{ème} cycle que j'ai soutenu au laboratoire de probabilités – statistique de l'université de Rouen dont il était le directeur. Je le remercie très sincèrement de m'avoir fait l'honneur et le plaisir d'accepter de diriger ce travail, malgré les difficultés supplémentaires de l'encadrement à distance, et de l'avoir fait si efficacement.

Je remercie vivement M. N. BENKAFADAR de l'honneur qu'il me fait en présidant ce jury.

Ma profonde gratitude va à M. ROUSSIGNOL pour s'être déplacé afin de participer à ce jury ainsi que pour sa gentillesse durant mes séjours à l'université de Marne la Vallée.

J'adresse mes remerciements les plus sincères et les plus chaleureux à L. ABBAOUI, D. AISSANI et Z. MOHDEB pour avoir accepté de s'intéresser à ce travail et de participer à ce jury.

Bien sûr je ne peux oublier de remercier vivement l'équipe du laboratoire de mathématiques appliquées de l'université de Marne la Vallée, qui m'a accueillie pour des stages, et offert tous les moyens de travail à différentes reprises durant l'élaboration de ce travail. En particulier je remercie S. Korry, L. Smail et S. Mercier pour leur aide technique et M. Morvan pour son aide administrative et sa gentillesse.

Mes vifs remerciements vont également au chef de département de mathématiques, au personnel qui l'entoure ainsi qu'au vice recteur et à la directrice chargés de la post graduation et de la recherche pour la compréhension et l'aide que j'ai trouvées auprès d'eux.

Table des matières

Introduction

1	Rappels	9
1.1	Processus de Dirichlet	9
1.2	Cadre général de l'analyse bayésienne	12
1.3	Modèle bayésien de Cox-Dirichlet	14
1.3.1	Définitions et notations	16
1.3.2	Résultats	20
2	Calcul des densités prédictives dans le modèle bayésien de Cox-Dirichlet à censures aléatoires à droite	24
2.1	Le modèle	24
2.2	La fonction de survie prédictive	25
2.3	Calcul des densités	26
3	Mesures prédictives dans le modèle bayésien de Cox-Dirichlet à Censures aléatoires à gauche	34
3.1	Le modèle	34
3.2	Fonction de survie prédictive	35
4	Mesures prédictives dans le modèle bayésien de Cox-Dirichlet à censures aléatoires mixtes	39
4.1	Le modèle	39
4.1.1	Modèle symétrique	40

4.1.2	Modèle à censure initiale à droite par T_2	40
4.1.3	Modèle à censure initiale à gauche par T_3	40
4.2	Echantillon de taille n	41
4.2.1	Modèle symétrique	41
4.2.2	Modèle à censure initiale à droite par T_2	42
4.2.3	Modèle à censure initiale à gauche par T_3	43

5 Calcul des densités prédictives dans le modèle bayésien de Cox-Dirichlet

	à censures fixes	44
5.1	Introduction	44
5.2	Calcul de densités prédictives	45
5.2.1	Cas particulier en dimension 2 ou 3	46
5.2.2	Cas général	54
5.3	Retour au modèle de Cox-Dirichlet	69
5.4	Extension	75
.1	Appendice A	77
.2	Appendice B	82

Introduction

D'une manière générale les données de survie mesurent le temps d'occurrence d'un certain événement. Ce dernier peut être la guérison ou la mort suite à la prise d'un médicament dans le domaine biomédical ou la panne d'une machine en fiabilité. Mais en fait ces données ne se limitent pas au domaine biomédical et à la fiabilité; elles ont été étendues à la sociologie, la criminologie, le marketing etc. Des exemples concernant les différents domaines d'application précités sont respectivement la durée d'un premier mariage, le temps d'aveux d'un criminel et la durée d'abonnement à une revue (cf Lee 1994).

Cependant, l'observation de telles données ne peut pas toujours se faire d'une manière complète; l'exemple du temps de survie d'un individu atteint d'une certaine maladie est assez explicite à cet égard. D'une part, le temps d'étude doit être fixé et le patient peut décéder d'une autre cause que sa maladie (par exemple un accident), c'est la censure à droite. D'autre part, le temps du début de la maladie peut être inconnu, c'est la censure à gauche. Autrement dit les temps de survie, réellement observés, ne sont pas toujours les véritables. Nous sommes en présence de données censurées et il est clair que c'est le cas pour beaucoup d'exemples rencontrés dans la pratique. Pour cet exemple les données sont, en même temps, censurées à gauche et à droite, c'est la censure mixte. Les censures à gauche et à droite ne sont que des cas particuliers de la censure par intervalle puisque dans ce cas on sait simplement que les temps de survie, des individus censurés, appartiennent à un certain intervalle. Un autre type de mécanisme de censure est donné par les censures progressives (cf Bordes 2002 et Alvarez-Bordes 2002), dans ce cas on retire, au fur et à

mesure, des éléments de l'étude (pour une raison de coût ou autre). L'incomplétude peut aussi s'exprimer par la troncature, une variable n'est alors prise en compte que si elle est supérieure à une valeur donnée (troncature à gauche) ou inférieure à une valeur donnée (troncature à droite). Pour notre exemple cela peut vouloir dire que les personnes dont la mort est antérieure au moment du début de l'étude sont automatiquement exclus de l'étude (troncature à gauche) et celles pour lesquelles la naissance est postérieure au moment de la fin de l'étude sont aussi automatiquement exclus de cette étude (troncature à droite).

La théorie de l'analyse de survie a pris naissance avec la parution de l'article fondateur de Kaplan-Meier (1958) dans lequel un estimateur de la fonction de survie a été proposé pour des données censurées à droite. Cet estimateur a inspiré beaucoup de travaux mais il s'est avéré difficilement généralisable à d'autres types de données incomplètes (troncature ou censure mixte). Des détails sur ce point sont donnés dans la thèse de Beck (2000). Quant à nous signalons, tout de même, que l'estimateur de Kaplan-Meier a été généralisé au cas où l'indicateur de censure n'est pas toujours connu, c'est à dire qu'on ne sait pas, pour certaines observations, si elles sont censurées ou pas (cf Van Der Laan et McKeague 1998). Un ouvrage de base sur les données de survie reste celui de Kalbfleisch et Prentice (1980). L'usage des processus de comptage pour l'inférence concernant les données incomplètes a été initié par Aalen (1975). Pour les développements récents, on peut se référer à Andersen et autres (1993), Coccozza (1997) et Fleming-Harrington (1991). Ces méthodes permettent de déduire rigoureusement et assez simplement les propriétés asymptotiques des statistiques utilisées, propriétés qui étaient, pour certaines, inconnues jusqu'alors.

L'introduction de covariables peut s'avérer nécessaire afin de tenir compte d'une certaine hétérogénéité entre les individus, l'hypothèse de variables indépendantes et équidistribuées n'étant pas toujours raisonnable. Dans l'exemple précédent, cela peut être représenté par le sexe et l'âge. Les modèles les plus connus, dans ce cadre, sont le modèle à hasard proportionnel (ou modèle de Cox introduit par cet auteur en 1972) et le modèle

à hasard additif (ou modèle d'Aalen introduit par cet auteur en 1978). Dans le modèle de Cox (respectivement d'Aalen) ces covariables agissent multiplicativement (respectivement additivement) sur le taux de défaillance. Le modèle de Cox se prête bien au calcul sur les fonctions de survie, sur lesquelles les covariables interviennent comme exposants d'une fonction de survie de base.

Par ailleurs, la prise en compte de résultats antérieurs ou de dires d'experts peut se faire par l'intermédiaire d'un modèle bayésien, on peut se reporter à Robert (1992) et à Florens et autres (1990) pour l'étude de l'analyse statistique bayésienne. La théorie bayésienne non paramétrique ou semi paramétrique a connu un essor considérable depuis les travaux de Ferguson (1973 et 1974) qui a introduit l'a priori de Dirichlet, ce dernier ayant eu un succès considérable. Cela étant dû à la facilité de son emploi, car dans un cadre de modèle discret la famille des lois de Dirichlet est la conjuguée des lois multinomiales (cf Florens et autres 1998); dans un cadre général on s'y ramène en considérant des partitions finies de l'ensemble continu des observations.

Dans ce cadre N. Gouget et J. P. Raoult (1999) ont étudié la mesure prédictive dans le modèle bayésien de Cox-Dirichlet (la fonction de base du modèle de Cox est munie de l'a priori de Dirichlet). Ils ont montré que cette loi est un cas particulier de ce qu'ils ont appelé mesures régulières par portions. De telles mesures ne sont pas absolument continues par rapport à la mesure de Lebesgue, mais leurs restrictions aux sous ensembles caractérisés par des égalités et des inégalités strictes entre les observations le sont. Ils ont effectué le calcul des densités ainsi définies dans le cas général d'une mesure régulière par portions puis dans le cas particulier du modèle de Cox-Dirichlet.

Pour notre part, nous envisageons le calcul de densités prédictives en rajoutant différents types de censure à des observations toujours régies par un modèle de Cox-Dirichlet. Nous effectuons les calculs dans le modèle de Cox conditionnellement aux valeurs des covariables, la loi de base étant seule inconnue. Le travail effectué dans [15] visait à mettre en place les calculs de densités applicables aux modèles, justifiés par des problèmes économétriques, introduits dans les travaux de Ruggiero (1989 et 1994). Or dans la pratique,

ce type de données est en général inévitablement censuré.

Nous commençons, au chapitre 1, par rappeler le processus de Dirichlet, le cadre général de l'analyse bayésienne dans laquelle nous nous situons ainsi que quelques résultats sur le modèle de Cox-Dirichlet donnés par N. Gouget et J. P. Raoult (1999) puisque ce fut le point de départ de ce travail.

Au chapitres 2, 3 et 4 nous étendons au cas d'observations aléatoirement censurées l'étude des mesures prédictives dans le modèle non paramétrique bayésien de Cox-Dirichlet effectuée, pour des observations non censurées dans [15], la loi des points de censure étant elle même régie par un modèle de Cox-Dirichlet indépendant de celui des observations. Nous montrons que dans les différents cas de censure envisagés (à droite, à gauche ou mixte), les mesures prédictives restent régulières par portions. Il en résulte donc une partition de l'espace des observations en parties contenues dans des sous espaces vectoriels, sur chacune desquelles la mesure prédictive est absolument continue par rapport à la mesure de Lebesgue. Nous calculons au chapitre 2, dans le cas de la censure à droite, les densités ainsi définies. De plus un exemple de calcul de densité, en dimension 2, est effectué dans le cas de la censure à gauche.

Quant au chapitre 5, l'extension y est effectuée pour des données censurées par des constantes. Après avoir détaillé le calcul des densités en dimensions 2 et 3, le cas général a été traité en donnant les densités restreintes à des sous ensembles des faces, ensembles caractérisés par la fixation des variables censurées et comportant des égalités et des inégalités strictes entre les observations non censurées.

Signalons que ce dernier chapitre est indépendant des chapitres 2, 3 et 4 et peut, de ce fait, être lu avant ceux ci.

Chapitre 1

Rappels

1.1 Processus de Dirichlet

L'introduction de ce processus, comme son nom l'indique, fait appel à la loi de Dirichlet que nous rappelons ci après.

Soit $(Z_j)_{1 \leq j \leq k}$ des variables aléatoires indépendantes de lois respectives la loi gamma de paramètres $(\alpha_j, 1)$ où $\alpha_j \geq 0$ et $\sum_{j=1}^k \alpha_j > 0$. la loi de Dirichlet de paramètre $(\alpha_1, \dots, \alpha_k)$, notée $\mathcal{D}_{(\alpha_1, \dots, \alpha_k)}$, est la loi du vecteur aléatoire $(Y_1, \dots, Y_k)$ où $Y_j = \frac{Z_j}{\sum_{j=1}^k Z_j}$ pour tout j tel que $1 \leq j \leq k$. Si $\alpha_j > 0$ pour tout $1 \leq j \leq k$ alors la loi de $(Y_1, \dots, Y_{k-1})$ est de densité :

$$f(y_1, \dots, y_{k-1}) = \frac{\Gamma(\alpha_1 + \dots + \alpha_k)}{\Gamma(\alpha_1) \dots \Gamma(\alpha_k)} \left(\prod_{j=1}^{k-1} y_j^{\alpha_j-1} \right) \left(1 - \sum_{j=1}^{k-1} y_j \right)^{\alpha_k-1} 1_{S_k}(y_1, \dots, y_{k-1})$$

où S_k est le simplexe :

$$S_k = \{(y_1, \dots, y_{k-1}) : y_j \geq 0, \sum_{j=1}^{k-1} y_j \leq 1\}.$$

Soit $S = \{s_1, \dots, s_k\}$ un ensemble fini et $\mathcal{P}_S$ l'ensemble des probabilités sur S . $p \in$

$\mathcal{P}_S$ ssi $p = (p_1, \dots, p_k)$ avec $p_i \geq 0$ pour tout i et $\sum_{i=1}^k p_i = 1$ ($p_i = P(\{s_i\})$). Si dans un modèle bayésien, l'ensemble des paramètres est $\Theta = \mathcal{P}_S$, alors ce dernier est en bijection avec le simplexe S_k . On peut donc munir l'ensemble des probabilités $\mathcal{P}_S$ de la loi de probabilité de Dirichlet $\mathcal{D}_{(\alpha_1, \dots, \alpha_k)}$. Cette dernière a la propriété qu'en observant n variables aléatoires indépendantes et identiquement distribuées à valeurs dans S , si l'a priori est $\mathcal{D}_{(\alpha_1, \dots, \alpha_k)}$ alors l'a posteriori (sachant qu'on a observé $(x_1, \dots, x_n)$) est $\mathcal{D}_{(\alpha'_1, \dots, \alpha'_k)}$ où $\alpha'_i = \alpha_i + \text{Card}\{j : x_j = s_i\}$. Ceci veut dire que les a priori de Dirichlet constituent la famille conjuguée des lois multinomiales, ce qui en facilite l'emploi.

Maintenant, au lieu de s'intéresser à une loi de probabilité chargeant un ensemble fini, franchissons le pas vers une loi quelconque P sur $\mathbb{R}_+$. En ne s'intéressant aux valeurs de P que pour certains intervalles formant une partition finie de $\mathbb{R}_+$, on se ramène au cas fini. La transition de la loi de Dirichlet au processus de Dirichlet s'obtient alors en considérant n'importe quelle partition comme suit.

soit α^* une mesure finie non nulle sur l'espace mesurable $(\mathfrak{S}, \mathcal{A})$ et soit P un processus indéré par les ensembles mesurables de $\mathcal{A}$. P est dit processus de Dirichlet de paramètre α^* et on note $\mathcal{D}_{\alpha^*}$, si pour tout entier k et toute partition mesurable $(B_1, \dots, B_k)$ de $\mathfrak{S}$ (c.a.d. les (B_i) sont mesurables, disjoints et de réunion $\mathfrak{S}$), le vecteur aléatoire $(P_{B_1}, \dots, P_{B_k})$ suit la loi de Dirichlet $\mathcal{D}_{(\alpha^*(B_1), \dots, \alpha^*(B_k))}$. Nous allons décrire les principales étapes de la preuve, de l'existence de ce processus, donnée dans l'article [9] de Ferguson.

D'une manière générale, si pour tout ensemble fini J les lois des vecteurs $\{P_{B_j}, j \in J\}$ sont données pour toute partition mesurable $\{B_j, j \in J\}$ on peut en déduire, pour tout ensemble fini I , les lois des vecteurs $\{P_{A_i}, i \in I\}$ pour toute famille mesurable $\{A_i, i \in I\}$ en posant :

$$P_{A_i} = \sum_{j \in J_i} P_{B_j} \quad (1.1)$$

où

$$J_i = \{j / B_j \subset A_i\}.$$

La condition suivante permet de satisfaire la propriété de compatibilité imposée par le théorème de Kolmogorov.

Condition C

Pour toute partition finie mesurable $\{B_j, j \in J\}$ de $\mathfrak{S}$ et pour toutes les partitions $\{J_k, k \in K\}$ de J , les deux vecteurs

$$(P_{(\bigcup_{j \in J_k} B_j)}, k \in K), (\sum_{j \in J_k} P_{(B_j)}, k \in K)$$

ont même distribution.

Proposition (cf [9])

Si pour toute partition finie mesurable la condition C est satisfaite et si pour toute famille finie mesurable $\{A_i, i \in I\}$ la loi du vecteur $\{P_{A_i}, i \in I\}$ est donnée par la formule (1.1) de sorte que les lois des variables aléatoires soient indépendantes de leur ordre, alors il existe une loi de probabilité $\mathcal{P}$ sur $([0, 1])^{\mathcal{A}}, (\mathcal{B} \cap [0, 1])^{\otimes \mathcal{A}}$ engendrant toutes les lois en dimension finie où $\mathcal{B} \cap [0, 1]$ est la tribu borélienne sur l'ensemble $[0, 1]$ et $(\mathcal{B} \cap [0, 1])^{\otimes \mathcal{A}}$ est la tribu produit des tribus boréliennes sur $[0, 1]$.

Maintenant en revenant au cas particulier où $(P_{B_1}, \dots, P_{B_k})$ suit la loi de Dirichlet $\mathcal{D}_{(\alpha^*(B_1), \dots, \alpha^*(B_k))}$, la condition C est satisfaite du fait que si le vecteur $\{Y_i, i \in I\}$, pour I fini, suit la loi de Dirichlet de paramètres $\{\alpha_i, i \in I\}$ alors pour toute partition $\{I_j, 1 \leq j \leq l\}$ de I

$$(\sum_{i \in I_1} Y_i, \dots, \sum_{i \in I_l} Y_i) \sim \mathcal{D}_{(\sum_{i \in I_1} \alpha_i, \dots, \sum_{i \in I_l} \alpha_i)};$$

ceci venant du fait que si X_1 et X_2 sont indépendantes de lois respectives $\Gamma(\alpha_1, 1)$ et $\Gamma(\alpha_2, 1)$ alors $X_1 + X_2$ suit la loi $\Gamma(\alpha_1 + \alpha_2, 1)$.

L'existence du processus de Dirichlet est ainsi établie.

En particulier pour tout ensemble mesurable A , P_A suit la loi Béta de paramètre $(\alpha^*(A), \alpha^*(A^c))$, donc :

$$E(P_A) = \frac{\alpha^*(A)}{\alpha^*(\mathfrak{S})}, \quad \text{var}(P_A) = \frac{\alpha^*(A)\alpha^*(A^c)}{(\alpha^*(\mathfrak{S}))^2(\alpha^*(\mathfrak{S}) + 1)}.$$

Il apparaît alors que l'espérance de la probabilité aléatoire P est une loi de probabilité et que la masse totale de α^* est un paramètre de précision. Le processus de Dirichlet met donc du flou autour d'une loi centrale.

1.2 Cadre général de l'analyse bayésienne

En théorie de l'échantillonnage un modèle statistique est décrit par la donnée d'un espace mesurable $(E, \mathfrak{S})$ pour un nombre infini d'observations, appelé l'espace de l'échantillon et par la donnée, sur cet espace, d'une famille de probabilités indexée par un paramètre ϑ , noté $\{P_\vartheta, \vartheta \in \Theta\}$, et qui sont les distributions d'échantillonnage des observations.

En statistique bayésienne, l'espace des paramètres Θ est muni d'une tribu $\mathcal{A}$ telle que : $\forall S \in \mathfrak{S}$, l'application $\vartheta \longrightarrow P_\vartheta(S)$ soit mesurable sur $(\Theta, \mathcal{A})$. Autrement dit l'application $P.(.)$ est une probabilité de transition sur $\Theta \times \mathfrak{S}$. L'espace mesurable $(\Theta, \mathcal{A})$ est muni d'une loi de probabilité μ , appelée loi à priori et qui exprime l'information disponible à priori. Le paramètre θ est regardé, dans ce cadre, en tant qu'objet aléatoire. Un modèle statistique bayésien est alors la donnée de l'unique loi de probabilité π sur l'espace produit $(\Theta \times E, \mathcal{A} \otimes \mathfrak{S})$ définie par :

$$\forall A \in \mathcal{A}, \forall S \in \mathfrak{S}, \quad \pi(A \times S) = \int_A P_\vartheta(S) d\mu(\theta).$$

La loi marginale de π sur l'espace $(E, \mathfrak{S})$ est dite loi prédictive, elle est donnée par :

$$\forall S \in \mathfrak{S}, P(S) = \int_{\Theta} P_{\vartheta}(S) d\mu(\theta).$$

Nous observons n variables aléatoires indépendantes strictement positives $X_1, \dots, X_n$ mais non identiquement distribuées. Cependant les lois des (X_i) sont les modifications d'une même loi de base, à l'aide de coefficients multiplicateurs du taux de défaillance. Autrement dit, il existe une loi de base $\mathcal{P}$ de fonction de survie F et une suite de nombres strictement positifs $(c_i)_{1 \leq i \leq n}$ de nombres appelés contraintes ou stress tels que la fonction de survie de X_i soit F^{c_i} . Trois approches sont alors envisageables. Si une loi est spécifiée pour F il s'agit alors d'estimer ses paramètres, c'est le modèle paramétrique. Dans notre cas le modèle est non paramétrique dans le sens où F est complètement inconnue, de nature non paramétrique, et que la relation liant les lois des (X_i) à F est connue mais dépend de paramètres (c_i) . Dans ce travail, seule l'inférence relative à F est étudiée puisque les contraintes sont supposées connues. Signalons qu'un troisième modèle, intermédiaire entre les deux précédents, appelé modèle semi paramétrique consiste à ne spécifier aucune loi pour F mais à représenter paramétriquement l'effet des covariables sur la survie.

Dans notre cas l'espace des paramètres est l'ensemble de toutes les lois de probabilité sur $(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$ sur lequel il faut spécifier une probabilité. A cette fin, ce qui se fait classiquement, est de regarder une probabilité sur $(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$ comme une application de $\mathcal{B}(\mathbb{R}_+)$ sur $[0,1]$. L'espace des paramètres est alors $\Theta = [0, 1]^{\mathcal{B}(\mathbb{R}_+)}$, auquel il est d'usage d'associer la tribu $(\mathcal{B} \cap [0, 1])^{\otimes \mathcal{B}(\mathbb{R}_+)}$, $\mathcal{B} \cap [0, 1]$ étant la tribu borélienne de $[0,1]$. De plus l'espace des probabilités P , caractérisées par leurs fonctions de survie G , est muni de l'a priori de Dirichlet $\mathcal{D}_{\alpha^*}$ rappelé plus haut où α^* est une mesure définie sur $(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$ de fonction de survie α . Rappelons que ceci veut dire que pour toute partition mesurable $(A_i)_{1 \leq i \leq m}$ de $\mathbb{R}_+$, le vecteur aléatoire $(\mathcal{P}_{A_i})_{1 \leq i \leq m}$ suit la loi de Dirichlet de paramètre $(\alpha^*(A_1), \dots, \alpha^*(A_m))$. En particulier pour toute suite strictement croissante de nombres réels positifs $(u_1, \dots, u_m)$, la suite aléatoire $G(u_1), \dots, G(u_m)$ admet pour densité par

rapport à la mesure de Lebesgue sur $\mathbb{R}_+^m$ l'application :

$$(y_1, \dots, y_m) \longrightarrow \frac{\Gamma(\alpha_0)}{\Gamma(\alpha_1) \dots \Gamma(\alpha_{m+1})} (1 - y_1)^{\alpha_1 - 1} \prod_{i=1}^{m-1} (y_i - y_{i+1})^{\alpha_{i+1} - 1} \times y_m^{\alpha_{m+1} - 1} 1_A(y_1, \dots, y_m) \quad (1.2)$$

où $A = \{(y_i)_{1 \leq i \leq m} : 0 < y_m < \dots < y_1 < 1\}$ et où

Γ désigne la fonction gamma eulérienne,

$\alpha_0 = \alpha^*(\mathbb{R}^+)$, $\alpha_1 = \alpha^*([0, u_1])$, $\alpha_2 = \alpha^*([u_1, u_2])$, $\dots$, $\alpha_m = \alpha^*([u_{m-1}, u_m])$,

$\alpha_{m+1} = \alpha^*([u_m, +\infty[)$.

1.3 Modèle bayésien de Cox-Dirichlet

Nous reprenons, ici, quelques résultats de l'article [15] de N. Gouget et J. P. Raoult qui nous seront utiles dans les chapitres suivants.

Soient $T_1, \dots, T_n$, n variables aléatoires strictement positives indépendantes de fonctions de survie respectives $G_1, \dots, G_n$, qui sont les modifications d'une fonction de survie de référence G par la transformation suivante :

$$\forall i \ 1 \leq i \leq n, \forall t > 0, G_i(t) = [G(t)]^{c_i}$$

où $c = (c_1, \dots, c_n)$ est un vecteur de nombres réels strictement positifs. c est supposé fixé et on place sur G un a priori de Dirichlet D_{α^*} . Ce modèle est appelé le modèle de Cox-Dirichlet.

La fonction de survie de la loi prédictive du vecteur $(T_1, \dots, T_n)$ s'écrit :

$$S(t_1, \dots, t_n) = \int \prod_{i=1}^n G^{c_i}(t_i) D_{\alpha^*}(dG) \quad (1.3)$$

Afin de pouvoir appliquer la formule (1.2) pour le calcul de cette intégrale, rangeons la suite $(t_i)_{1 \leq i \leq n}$ par ordre croissant et regroupant les éventuels ex æquo. Nous obtenons

une suite $(u_1, \dots, u_m)$ strictement croissante de longueur $m \leq n$.

Notons d_j la somme de tous les c_i tels que $t_i = u_j$.

Soient $\alpha_1, \dots, \alpha_{m+1}$ définis en (1.2).

Alors

$$S(t_1, \dots, t_n) = \int \prod_{j=1}^m (G(u_j))^{d_j} D_{\alpha^\bullet}(dG)$$

$$= \frac{\Gamma(\alpha_0)}{\Gamma(\alpha_1) \dots \Gamma(\alpha_{m+1})} \int \prod_{j=1}^{m+1} (z_{j-1} - z_j)^{\alpha_j-1} \prod_{j=1}^m z_j^{d_j} 1_{\{0 < z_m < \dots < z_1 < 1\}} \prod_{j=1}^m dz_j$$

où nous posons $z_0 = 1$ et $z_{m+1} = 0$.

L'intégration se fait alors successivement selon $z_m, z_{m-1}, \dots, z_1$.

En utilisant le fait que

$$(\forall p > 0, q > 0) \quad \int_0^u t^{p-1} (u-t)^{q-1} dt = u^{p+q-1} \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}$$

et en remarquant que l'intégration par rapport à dz_j est donnée par :

$$\int_0^{z_{j-1}} (z_{j-1} - z_j)^{\alpha_j-1} z_j^{a(j)-1} dz_j$$

$$\text{où } a(j) = \sum_{l=j}^m (\alpha_{l+1} + d_l)$$

il vient

$$S(t_1, \dots, t_n) = \frac{\Gamma(\alpha_0)}{\Gamma(\alpha_1) \dots \Gamma(\alpha_{m+1})} \prod_{j=1}^m \frac{\Gamma(a(j))\Gamma(\alpha_j)}{\Gamma(a(j) + \alpha_j)} = \frac{\Gamma(\alpha_0)}{\Gamma(\alpha_{m+1})} \prod_{j=1}^m \frac{\Gamma(a(j))}{\Gamma(a(j) + \alpha_j)}$$

Ce qui s'écrit sous la forme

$$S(t_1, \dots, t_n) = \frac{\Gamma(\alpha(0))}{\Gamma(\alpha(0) + d)} \prod_{i=1}^n \frac{\Gamma(\alpha(t_{\sigma_i}) + d_i^\sigma)}{\Gamma(\alpha(t_{\sigma_i}) + d_{i+1}^\sigma)} \quad (1.4)$$

où :

- $\sigma = (\sigma_1, \dots, \sigma_n)$ est une permutation réordonnant en croissant la suite $(t_i)_{1 \leq i \leq n}$, c'est à dire telle que la suite $(t_{\sigma_j})_{1 \leq j \leq n}$ soit croissante,
- $d_i^\sigma = \sum_{h=i}^n c_{\sigma_h}$ avec $d_{n+1}^\sigma = 0$ et $d = d_1^\sigma = \sum_{i=1}^n c_i$.

Ce résultat a été donné par M. Ruggiero dans [23]. Mais c'est N. Gouget et J. P. Raoult qui ont mis en évidence les propriétés de la mesure qui est associée à cette fonction de survie. En effet cette loi prédictive n'est pas absolument continue par rapport à la mesure de Lebesgue sur $\mathbb{R}^n$. En revanche, ces auteurs ont fait apparaître une propriété "d'absolue continuité par portions" qui est la suivante : considérons les sous ensembles de $\mathbb{R}^n$ caractérisés par des égalités et des inégalités strictes entre les coordonnées ; si le nombre d'égalités est $n-\ell$, un tel sous ensemble est une partie ouverte d'un sous espace vectoriel de dimension ℓ dans $\mathbb{R}^n$. La restriction de la loi prédictive considérée à chacun de ces sous ensembles est absolument continue par rapport à la mesure de Lebesgue ℓ dimensionnelle. Nous renvoyons à [15] pour un calcul explicite des densités définies par ces relations d'absolue continuité par portions.

Nous allons préciser cette notion de régularité par portions, le résultat fondamental qui en découle et l'application de ce dernier au modèle de Cox-Dirichlet.

1.3.1 Définitions et notations

Dérivées partielles

Soit A une partie de $\mathbb{R}^k$ et f une application de A dans $\mathbb{R}$. Soit $s = (s_1, \dots, s_\ell)$ une suite d'entiers distincts dans $\{1, \dots, k\}$. La ℓ dérivée partielle $\frac{\partial^\ell f}{\partial x_{s_1} \dots \partial x_{s_\ell}}$ de f est notée

$\partial_s f$. Elle est classiquement définie par la succession de ℓ dérivations partielles aux points dans l'intérieur de A , puis par un prolongement continu aux points de la frontière chaque fois que cela est faisable.

Configuration d'ex æquo

Nous appelons ainsi toute suite de nombres entiers strictement positifs $\lambda = (\lambda_1, \dots, \lambda_l)$ telle que $\sum_{j=1}^l \lambda_j = n$. Nous notons $n_j = \sum_{h=1}^j \lambda_h$ (avec $n_0 = 0$).

Ensembles et applications remarquables

Pour chaque n -permutation $\sigma = (\sigma_1, \dots, \sigma_n)$ et chaque configuration d'ex æquo λ , nous introduisons :

- le sous ensemble de $\mathbb{R}^n$ défini par :

$$\Delta_\sigma^\lambda = \{(t_i)_{1 \leq i \leq n} : \forall j \in \{1, \dots, l\} t_{\sigma_{n_{j-1}+1}} = \dots = t_{\sigma_{n_j}}, \forall j \in \{1, \dots, l-1\} t_{\sigma_{n_j}} < t_{\sigma_{n_{j+1}}}\}$$

Sa fermeture est

$$\bar{\Delta}_\sigma^\lambda = \{(t_i)_{1 \leq i \leq n} : \forall j \in \{1, \dots, l\} t_{\sigma_{n_{j-1}+1}} = \dots = t_{\sigma_{n_j}}, \forall j \in \{1, \dots, l-1\} t_{\sigma_{n_j}} \leq t_{\sigma_{n_{j+1}}}\}$$

Δ_σ^λ est l'ensemble de toutes les suites dans $\mathbb{R}^n$ qui peuvent être réordonnées en croissant par l'usage de la permutation σ et dans lesquelles il y a exactement l blocs d'ex æquo qui, lorsqu'on les recense en croissant, ont les tailles successives $\lambda_1, \dots, \lambda_l$.

Pour $\lambda = (1, \dots, 1)$, la suite de longueur n dont tous les éléments sont égaux à 1,

nous obtenons le sous ensemble noté Δ_σ :

$$\Delta_\sigma = \left\{ (t_i)_{1 \leq i \leq n} : \forall j \in \{1, \dots, n-1\} \ t_{\sigma_j} < t_{\sigma_{j+1}} \right\};$$

la fermeture de Δ_σ est :

$$\bar{\Delta}_\sigma = \left\{ (t_i)_{1 \leq i \leq n} : \forall j \in \{1, \dots, n-1\} \ t_{\sigma_j} \leq t_{\sigma_{j+1}} \right\}.$$

Si σ est la permutation identité, nous obtenons l'ensemble noté

$$\Delta^\lambda = \left\{ (t_i)_{1 \leq i \leq n} : \forall j \in \{1, \dots, l\} \ t_{n_{j-1}+1} = \dots = t_{n_j}, \forall j \in \{1, \dots, l-1\} \ t_{n_j} < t_{n_{j+1}} \right\}.$$

Si en même temps $\lambda = (1, \dots, 1)$ et σ est la permutation identité, nous obtenons l'ensemble noté

$$\Delta_{in} = \left\{ (t_i)_{1 \leq i \leq n} : \forall j \in \{1, \dots, n-1\} \ t_j < t_{j+1} \right\}$$

et sa fermeture

$$\bar{\Delta}_{in} = \left\{ (t_i)_{1 \leq i \leq n} : \forall j \in \{1, \dots, n-1\} \ t_j \leq t_{j+1} \right\}$$

- la variété linéaire :

$$P_\sigma^\lambda = \left\{ (t_i)_{1 \leq i \leq n} \ \forall j \in \{1, \dots, l\} \ \forall i \in J_j \ \forall i' \in J_j, t_i = t_{i'} \right\}$$

où $J_j = \{i : n_{j-1} < \sigma^{-1}(i) \leq n_j\}$, avec pour σ la permutation identité,

$$P^\lambda = \left\{ (t_i)_{1 \leq i \leq n} \forall j \in \{1, \dots, l\}, t_{n_{j-1}+1} = \dots = t_{n_j} \right\}$$

• l'application ϕ_σ^λ de $\mathbb{R}^l$ dans $\mathbb{R}^n$:

qui à chaque $x = (x_j)_{1 \leq j \leq l}$ fait correspondre la suite $t = (t_i)_{1 \leq i \leq n}$ telle que pour tout i et tout j :

$$t_i = x_j \Leftrightarrow n_{j-1} < (\sigma^{-1})_i \leq n_j,$$

où σ^{-1} est la permutation inverse de σ .

ϕ_σ^λ induit une bijection de $\mathbb{R}^l$ sur la variété linéaire P_σ^λ et une bijection de Δ_{i_l} (resp $\bar{\Delta}_{i_l}$) sur Δ_σ^λ (resp $\bar{\Delta}_\sigma^\lambda$).

Dans ϕ_σ^λ on omet λ si $\lambda = (1, \dots, 1)$, de même on omet σ si σ est l'application identité

• la fonction de survie réordonnée :

Soit μ une mesure positive et finie sur $\mathbb{R}^n$ de fonction de survie S et soit $\widehat{\phi}_\sigma$ la restriction de ϕ_σ à $\bar{\Delta}_{i_n}$. L'application définie sur $\bar{\Delta}_{i_n}$ par :

$$S_\sigma = S \circ \widehat{\phi}_\sigma \text{ est la fonction de survie réordonnée de la mesure } \mu.$$

Mesures régulières par portions

Nous appelons ainsi toute mesure finie sur $\mathbb{R}^n$ telle que, pour toute n -permutation σ , la restriction de sa fonction de survie, S , à $\bar{\Delta}_\sigma$ notée $\hat{S}_\sigma$, admette la n -dérivée continue $\partial_{1,\dots,n} \hat{S}_\sigma$.

1.3.2 Résultats

- Si $t = \varphi^\lambda(x_1, \dots, x_l)$ où $(x_1, \dots, x_l) \in \Delta_{i_l}$ alors $t = (x_1^{[\lambda_1]}, \dots, x_l^{[\lambda_l]})$ où $x_i^{[\lambda_i]}$ désigne la répétition, λ_i fois, de x_i et la formule (1.4) s'écrit :

$$S(t) = \frac{\Gamma(\alpha(0))}{\Gamma(\alpha(0) + d)} \prod_{j=1}^l \frac{\Gamma(\alpha(x_j) + e_j)}{\Gamma(\alpha(x_j) + e_{j+1})} \quad (1.5)$$

où

$$e_j = d_{n_{j-1}+1}^\sigma = \sum_{h=n_{j-1}+1}^n c_{\sigma_h}$$

- Si $y = (y_1, \dots, y_n) \in \bar{\Delta}_{i_n}$ alors

$$S_\sigma(y) = S(\hat{\phi}_\sigma(y)) = S(y_{\sigma^{-1}(1)}, \dots, y_{\sigma^{-1}(n)}) = \frac{\Gamma(\alpha(0))}{\Gamma(\alpha(0) + d)} \prod_{i=1}^n \frac{\Gamma(\alpha(y_i) + d_i^\sigma)}{\Gamma(\alpha(y_i) + d_{i+1}^\sigma)}$$

Si $s = (s_1, \dots, s_k)$ et si α est dérivable on obtient alors :

$$\begin{aligned} \partial_s(S_\sigma)(y) = S_\sigma(y) \prod_{j=1}^k \alpha'(y_{s_j}) [& (Ln\Gamma)'(\alpha(y_{s_j}) + d_{s_j}^\sigma) \\ & - (Ln\Gamma)'(\alpha(y_{s_j}) + d_{s_j+1}^\sigma)] \end{aligned} \quad (1.6)$$

Lemme 1 (cf lemma1 page 322 de [15])

Soit ν une mesure positive et finie sur $]M, N]^2$ et soit (S', S'') le couple de ses fonctions de survie réordonnées, soient a et b (avec $M \leq a < b \leq N$) tels que S' et S'' soient continues et qu'il existe les dérivées $\partial_{1,2}S'$ et $\partial_{1,2}S''$ qui soient continues sur $\Delta' = \{(x', x''); a < x' \leq x'' < b\}$. Soit ν^* la mesure définie, pour chaque sous ensemble borélien A de $]M, N[$, par $\nu^*(A) = \nu(\{(t, t); t \in A\})$ et soit S^* la fonction de survie de ν^* . Alors S^* admet une dérivée continue ∂S^* sur $]a, b[$, avec :

$$\partial S^*(x) = (\partial_2 S' - \partial_1 S'')(x, x) = (\partial_1 S' - \partial_2 S'')(x, x) \quad (1.7)$$

Résultat fondamental sur la mesure régulière par portions

Le théorème 2 de [15], que nous redonnons ci dessous, explicite pour chaque Δ^λ , associé à une configuration d'ex æquo $\lambda = (\lambda_1, \dots, \lambda_l)$, la densité d'une mesure régulière par portions, relativement à la mesure de Lebesgue sur P^λ . Il met en particulier en évidence qu'interviennnent des dérivées partielles de fonctions déduites de la fonction de survie S par permutations, c'est à dire des termes du type $\partial_s S_\sigma$ et ce uniquement pour certains couples (s, σ) ; ceux ci sont indexés par l'ensemble des suites de 0 et 1 de longueur $n-l$, décomposés en sous suites de longueurs $\lambda_1-1, \dots, \lambda_l-1$, soit $a = (a_1, \dots, a_{n-l}) = ((a_m^j)_{1 \leq m \leq \lambda_j-1})_{1 \leq j \leq l}$. Ces couples sont donc notés $(s(a), \sigma(a))$.

Théorème 1 (Gouget-Raoult cf [15] page 331)

Soit μ une mesure régulière par portions et soit $\lambda = (\lambda_j)_{1 \leq j \leq l}$ une configuration d'ex æquo. Notons $A^\lambda = \{0, 1\}^{n-l}$. Alors la restriction de μ à Δ^λ admet pour densité par rapport à la mesure de Lebesgue sur P^λ la fonction g^λ définie par :

$$g^\lambda = \sum_{a \in A^\lambda} (-1)^{q(a)} \partial_{s(a)} S_{\sigma(a)} \quad (1.8)$$

où, si on note $a = ((a_m^j)_{1 \leq m \leq \lambda_j-1})_{1 \leq j \leq l}$, nous obtenons, avec $n_j = \sum_{h=1}^j \lambda_h$:

- $q(a) = n - \sum_{j=1}^l \sum_{m=1}^{\lambda_j-1} a_m^j$
- $s(a) = (s_j)_{1 \leq j \leq l}$, avec $s_j = 1 + n_{j-1} + \sum_{m=1}^{\lambda_j-1} a_m^j$

• $\sigma(a)$ est la permutation inverse de $\tau(a)$ qui se décompose en sous suites de longueur $\lambda_1, \dots, \lambda_l$ sous la forme $((\tau_\nu^j)_{1 \leq \nu \leq \lambda_j})_{1 \leq j \leq l}$, avec

$$\tau_\nu^j = n_{j-1} + 1 + \sum_{u=1}^{\nu-1} a_u^j + (1 - a_\nu^j)(\lambda_j - \nu).$$

On remarque que ce théorème porte sur des sous ensembles de $\mathbb{R}^n$ qui sont des suites croissantes d'observations. Le cas général (parties Δ_σ^λ) s'en déduit par permutation et nous l'explicitons dans le cas du modèle de Cox-Dirichlet à censures aléatoires à droite.

Application au modèle de Cox-Dirichlet

L'expression de S , donnée à la formule (1.4), montre que la mesure prédictive associée à cette fonction de survie est régulière par portions, le théorème 1 lui est donc applicable. Après avoir explicité dans ce cas les termes $\partial_\sigma S_\sigma$, donnés à la formule 1.8 du théorème 1, Gouget et Raoult dans [15] ont déduit le résultat suivant.

Théorème 2 (Gouget-Raoult cf [15] page 335)

Soit μ la loi prédictive de la distribution de n durées de survie dans le modèle de Cox-Dirichlet, avec la suite de contraintes $c = (c_i)_{1 \leq i \leq n}$ et l'a priori de Dirichlet défini par la mesure α^* de fonction de survie α qui est continûment dérivable sur $\mathbb{R}_+$. Soit $\lambda = (\lambda_j)_{1 \leq j \leq l}$ une configuration d'ex æquo, selon laquelle nous décomposons la suites de contraintes c en l sous suites successives $(c_\nu^j)_{1 \leq \nu \leq \lambda_j}$; nous introduisons les sommes partielles supérieures de contraintes $e_j = \sum_{h=j}^l \sum_{\nu=1}^{\lambda_h} c_\nu^h$.

Alors la restriction de μ à Δ^λ admet comme densité par rapport à la mesure de Lebesgue sur la variété linéaire P^λ , la fonction g^λ définie par :

$$g^\lambda(x_1^{[\lambda_1]}, \dots, x_l^{[\lambda_l]}) = (-1)^l \frac{\Gamma(\alpha_0)}{\Gamma(\alpha_0 + e_1)} \prod_{j=1}^l B_j(x_j) C_j(x_j)$$

avec

$$B_j(x_j) = \frac{\Gamma(\alpha(x_j) + e_j)}{\Gamma(\alpha(x_j) + e_{j+1})} \alpha'(x_j)$$

et, $\mathcal{P}_{\lambda_j-1}$ désignant l'ensemble de tous les sous ensembles de $\{1, \dots, \lambda_j - 1\}$,

$$C_j(x_j) =$$

$$\sum_{M \in \mathcal{P}_{\lambda_j-1}} (-1)^{\text{Card } M} [(Ln\Gamma)'(\alpha(x_j) + e_{j+1} + c_{\lambda_j}^j + \sum_{v \in M} c_v^j) - (Ln\Gamma)'(\alpha(x_j) + e_{j+1} + \sum_{v \in M} c_v^j)].$$

Chapitre 2

Calcul des densités prédictives dans le modèle bayésien de Cox-Dirichlet à censures aléatoires à droite

2.1 Le modèle

On observe n variables aléatoires strictement positives indépendantes, de la forme,

$$T_i = \inf(T_{i,1}, \dots, T_{i,m}) = T_{i,1} \wedge \dots \wedge T_{i,m},$$

les variables $T_{i,k}$ ($1 \leq k \leq m$) étant elles même indépendantes et, pour tout k , le vecteur aléatoire $(T_{i,k})_{1 \leq i \leq n}$ étant lui même régi par le modèle bayésien de Cox-Dirichlet étudié dans [15] et rappelé au chapitre précédent. De manière précise, il existe une famille de fonctions de survie de base, $(G_k)_{1 \leq k \leq m}$ et une famille doublement indexée de contraintes strictement positives $c = (c_{i,k})_{1 \leq i \leq n, 1 \leq k \leq m}$ de telle sorte que la fonction de survie de $T_{i,k}$ soit $G_k^{c_{i,k}}$. D'autre part chaque famille de lois de base est munie d'un a priori de Dirichlet $\mathcal{D}_{\alpha_k^*}$, caractérisé par une mesure finie α_k^* sur $\mathbb{R}_+$ dite mesure de Dirichlet, de fonction de survie α_k (i.e. $\alpha_k(t) = \alpha_k^*([t, +\infty[)$). Rappelons (voir formule (1.2)) que cela signifie

que pour tout k et pour toute suite finie strictement croissante de nombres réels positifs $(u_1, \dots, u_s)$, la suite aléatoire $(1 - G_k(u_1), G_k(u_1) - G_k(u_2), \dots, G_k(u_{s-1}) - G_k(u_s))$ admet pour densité par rapport à la mesure de Lebesgue sur $\mathbb{R}^s$ l'application :

$$(y_1, \dots, y_s) \longrightarrow \frac{\Gamma(\alpha_{0,k})}{\Gamma(\alpha_{1,k}) \dots \Gamma(\alpha_{s+1,k})} \prod_{i=1}^s y_i^{\alpha_{i,k}-1} \times (1 - \sum_{i=1}^s y_i)^{\alpha_{s+1,k}-1} 1_{A_s}(y_1, \dots, y_s)$$

où,

$$A_s = \{(y_i)_{1 \leq i \leq s} / \forall i : 0 < y_i < 1, \sum_{i=1}^s y_i < 1\},$$

$$\alpha_{0,k} = \alpha_k^*(\mathbb{R}^+), \alpha_{1,k} = \alpha_k^*([0, u_1]), \alpha_{2,k} = \alpha_k^*([u_1, u_2]), \dots, \alpha_{s,k} = \alpha_k^*([u_{s-1}, u_s]), \\ \alpha_{s+1,k} = \alpha_k^*([u_s, +\infty[).$$

Nous supposons enfin que les m fonctions de survie de base sont a priori indépendantes, autrement dit $(G_k)_{1 \leq k \leq m}$ est muni de l'a priori $\prod_{k=1}^m \mathcal{D}_{\alpha_k^*}$.

Ce modèle peut en particulier être utilisé si les variables $T_{i,1}$ sont des durées de vie indépendantes, subissant chacune $m - 1$ censures à droite $T_{i,2}, \dots, T_{i,m}$. C'est pourquoi nous l'appellerons modèle bayésien de Cox-Dirichlet à censures à droite multiples et indépendantes. Le modèle ainsi obtenu pour les variables T_i n'est plus un modèle de Cox. Nous allons établir qu'il possède cependant encore (si les fonctions α_k sont continûment dérivables) la propriété d'absolue continuité par portions et expliciter les densités correspondantes.

2.2 La fonction de survie prédictive

La fonction de survie prédictive s'écrit :

$$S_c(t) = \prod_{k=1}^m \int \prod_{i=1}^n (G_k(t_i))^{c_{i,k}} \mathcal{D}_{\alpha_k^*}(dG_k) \quad (t = (t_i)_{1 \leq i \leq n})$$

Chacune des intégrales de ce produit est de la même forme que la survie prédictive donnée dans les travaux de M. Ruggiero dans [23] (voir formule (1.3)). D'après [15] elles admettent donc toutes une n -dérivée continue sur chaque partie $\bar{\Delta}_\sigma$ si les fonctions α_k sont continûment dérivables, et il en est donc de même de S_c . La loi prédictive des n durées de vie dans le modèle de Cox-Dirichlet à censures à droite multiples et indépendantes est donc régulière par portions. La formule (1.4) permet d'expliciter S_c . On obtient :

$$S_c(t) = \prod_{k=1}^m \left[\frac{\Gamma(\alpha_{0,k})}{\Gamma(\alpha_{0,k} + d_k)} \prod_{i=1}^n \frac{\Gamma(\alpha_k(t_{\sigma_i}) + d_{i,k}^\sigma)}{\Gamma(\alpha_k(t_{\sigma_i}) + d_{i+1,k}^\sigma)} \right] \quad (2.1)$$

où

- $\alpha_{0,k} = \alpha_k(0)$ (masse totale de α_k^*)
- $\sigma = (\sigma_1, \dots, \sigma_n)$ est une permutation telle que la suite $(t_{\sigma_i})_{1 \leq i \leq n}$ soit non décroissante.
- $d_{i,k}^\sigma = \sum_{h=i}^n c_{\sigma(h),k}$; $d_{n+1,k}^\sigma = 0$
- $d_k = d_{1,k}^\sigma = \sum_{i=1}^n c_{i,k}$

2.3 Calcul des densités

Pour appliquer le théorème 1 à la fonction de survie prédictive donnée en (2.1), il nous faut expliciter $\partial_{s(a)}(S_c)_{\sigma(a)}$. Rappelons que pour tout $s = (s_1, \dots, s_l)$ et tout σ , l'existence de $\partial_s(S_c)_\sigma$ et sa continuité sont assurées par l'hypothèse de continuité des dérivées α'_k . Des simplifications interviennent dans $\partial_{s(a)}(S_c)_{\sigma(a)}$, que nous allons présenter en omettant temporairement l'indexation par a . Tenant compte de la configuration d'ex æquo

$(\lambda_1, \dots, \lambda_l)$, nous sommes amenés à considérer $t = (x_1^{[\lambda_1]}, \dots, x_l^{[\lambda_l]})$, où $x_i^{[\lambda_i]}$ désigne la suite composée de λ_i termes tous égaux à x_i . Nous utilisons le lien entre $s(a)$ et λ indiqué au théorème 1. A cet effet, nous reprenons la formule (2.1) en utilisant la simplification :

$$(S_c)_\sigma(x_1^{[\lambda_1]}, \dots, x_l^{[\lambda_l]}) = \prod_{k=1}^m \frac{\Gamma(\alpha_{0,k})}{\Gamma(\alpha_{0,k} + d_k)} \prod_{j=1}^l \Phi_j(x_j, c, \sigma),$$

avec

$$\Phi_j(x_j, c, \sigma) = \prod_{k=1}^m \frac{\Gamma(\alpha_k(x_j) + e_{j,k})}{\Gamma(\alpha_k(x_j) + e_{j+1,k})},$$

où

$$e_{j,k} = d_{n_{j-1}+1,k}^\sigma = \sum_{h=n_{j-1}+1}^n c_{\sigma_h,k}$$

Par ailleurs, pour $y_1 \leq y_2 \leq \dots \leq y_n$,

$$(S_c)_\sigma(y_1, \dots, y_n) = \prod_{k=1}^m \left[\frac{\Gamma(\alpha_{0,k})}{\Gamma(\alpha_{0,k} + d_k)} \prod_{i=1}^n \frac{\Gamma(\alpha_k(y_i) + d_{i,k}^\sigma)}{\Gamma(\alpha_k(y_i) + d_{i+1,k}^\sigma)} \right]$$

et

$$\partial_s(S_c)_\sigma(y_1, \dots, y_n) = (S_c)_\sigma(y_1, \dots, y_n) \prod_{j=1}^l \left\{ \sum_{k=1}^m \alpha'_k(y_{s_j}) \left[(Ln\Gamma)'(\alpha_k(y_{s_j}) + d_{s_j,k}^\sigma) - (Ln\Gamma)'(\alpha_k(y_{s_j}) + d_{s_j+1,k}^\sigma) \right] \right\}.$$

En remarquant que, pour tout j tel que $1 \leq j \leq l$, on a $n_{j-1} < s_j \leq n_j$ nous obtenons :

$$\partial_s(S_c)_\sigma(x_1^{[\lambda_1]}, \dots, x_l^{[\lambda_l]}) = \prod_{k=1}^m \frac{\Gamma(\alpha_{0,k})}{\Gamma(\alpha_{0,k} + d_k)} \prod_{j=1}^l \Phi_j(x_j, c, \sigma) \Psi_j(x_j, c, \sigma, s),$$

où

$$\Psi_j(x_j, c, \sigma, s) = \sum_{k=1}^m \alpha'_k(x_j) \left[(Ln\Gamma)'(\alpha_k(x_j) + d_{s_j, k}^\sigma) - (Ln\Gamma)'(\alpha_k(x_j) + d_{s_j+1, k}^\sigma) \right].$$

Nous remarquons que vu la forme de $\sigma(a)$, qui ne modifie que des indices du même bloc d'ex æquo, $e_{j,k}$ s'écrit :

$$e_{j,k} = \sum_{i=n_{j-1}+1}^n c_{i,k}.$$

Avec la convention $e_{l+1,k} = 0$ pour tout k , il vient :

$$d_{s_j, k}^\sigma = \sum_{h=s_j}^{n_j} c_{\sigma_h, k} + e_{j+1, k}$$

En nous inspirant de la preuve du théorème 2 de Gouget-Raoult, nous allons écrire Ψ_j en fonction de $a = (a^j)_{1 \leq j \leq l}$ où $a^j = (a_m^j)_{1 \leq m \leq \lambda_j-1}$. Pour cela, il est pratique de décomposer les suites de contraintes $c_k = (c_{i,k})_{1 \leq i \leq n}$, pour chaque $1 \leq k \leq m$, en l blocs selon la configuration d'ex æquo λ comme cela a été fait pour la permutation $\tau(a)$ dans l'énoncé du théorème 1. Nous notons :

$$c_k = (c_k^j)_{1 \leq j \leq l} \text{ avec } c_k^j = (c_{w,k}^j)_{1 \leq w \leq \lambda_j} \text{ où } c_{w,k}^j = c_{n_{j-1}+w, k}.$$

Il vient alors :

$$e_{j,k} = \sum_{h=j}^l \sum_{w=1}^{\lambda_h} c_{w,k}^h$$

De même nous décomposons la permutation $\sigma(a)$ en l permutations $\sigma(a^j)$ où $\sigma(a^j)$ est la permutation sur l'ensemble $\{1, \dots, \lambda_j\}$, impact de $\sigma(a)$ sur $\{n_{j-1} + 1, \dots, n_j\}$. Ceci

veut dire que :

$$\forall w \in \{1, \dots, \lambda_j\}$$

$$(\sigma(a^j))_w = (\sigma(a))_{n_{j-1}+w} - n_{j-1},$$

Pour $s = s(a)$ et $\sigma = \sigma(a)$, $\partial_s(S_c)_\sigma(x_1^{[\lambda_1]}, \dots, x_l^{[\lambda_l]})$ ne dépend de a qu'à travers les termes $b_{j,k}(a) = \sum_{h=s_j}^{n_j} c_{\sigma_h,k}$ et $b'_{j,k}(a) = \sum_{h=s_j+1}^{n_j} c_{\sigma_h,k}$ qui s'écrivent avec nos notations :

$$b'_{j,k}(a) = \sum_{w=s_j+1-n_{j-1}}^{\lambda_j} c_{\sigma(a^j)_w,k}^j$$

et

$$b_{j,k}(a) = \sum_{w=s_j-n_{j-1}}^{\lambda_j} c_{\sigma(a^j)_w,k}^j = b'_{j,k}(a) + c_{\sigma(a^j)_{s_j-n_{j-1}},k}^j$$

$\tau(a)$ étant la permutation inverse de $\sigma(a)$, nous avons $(\sigma(a^j))^{-1} = \tau(a^j)$ qui est l'impact de $\tau(a)$ sur $\{n_{j-1}+1, \dots, n_j\}$. $b'_{j,k}(a)$ apparaît alors comme la somme de tous $c_{\nu,k}^j$ pour ν vérifiant $(\tau(a^j))_\nu > s_j - n_{j-1}$ et $b_{j,k}(a)$ est la somme de $b'_{j,k}(a)$ et de $c_{\nu,k}^j$ avec ν_j vérifiant $(\tau(a^j))_{\nu_j} = s_j - n_{j-1}$.

Le théorème 1 montre que pour tout $1 \leq v \leq \lambda_j$, nous avons :

$$\tau(a^j)_\nu = 1 + \sum_{u=1}^{\nu-1} a_u^j + (1 - a_\nu^j)(\lambda_j - \nu)$$

(Puisque $\tau(a^j)_\nu = \tau_\nu^j - n_{j-1}$)

et

$$s_j - n_{j-1} = 1 + \sum_{m=1}^{\lambda_j-1} a_m^j.$$

Pour la comparaison de $\tau(a^j)_\nu$ et de $s_j - n_{j-1}$, nous distinguons trois cas :

- $\nu = \lambda_j$ alors $\tau(a^j)_\nu = 1 + \sum_{u=1}^{\nu-1} a_u^j = s_j - n_{j-1}$, l'égalité voulant dire que $\nu_j = \lambda_j$.

- $\nu < \lambda_j$ et $a_\nu^j = 0$ alors

$$\begin{aligned} \tau(a^j)_\nu - (s_j - n_{j-1}) &= 1 + \sum_{u=1}^{\nu-1} a_u^j + \lambda_j - \nu - (1 + \sum_{m=1}^{\lambda_j-1} a_m^j) \\ &= \lambda_j - \nu - \sum_{m=\nu+1}^{\lambda_j-1} a_m^j > \lambda_j - \nu - (\lambda_j - 1 - \nu) > 0. \end{aligned}$$

- $\nu < \lambda_j$ et $a_\nu^j = 1$ alors

$$\tau(a^j)_\nu - (s_j - n_{j-1}) = 1 + \sum_{u=1}^{\nu-1} a_u^j - (1 + \sum_{m=1}^{\lambda_j-1} a_m^j) = -1 - \sum_{m=\nu+1}^{\lambda_j-1} a_m^j < 0.$$

Notons, comme dans [15],

$$M(a^j) = \{\nu : 1 \leq \nu < \lambda_j, a_\nu^j = 0\},$$

il vient :

$$b'_{j,k}(a) = \sum_{\nu \in M(a^j)} c_{\nu,k}^j$$

et

$$b_{j,k}(a) = b'_{j,k}(a) + c_{\lambda_j,k}^j.$$

Nous remarquons que, l'ensemble des (a^j) étant l'ensemble $\{0, 1\}^{\lambda_j-1}$ tout entier, l'ensemble des $M(a^j)$ est l'ensemble de tous les sous ensembles de $\{1, \dots, \lambda_j - 1\}$ noté $\mathcal{P}_{\lambda_j-1}$.

D'autre part l'exposant $q(a)$ utilisé dans l'expression de la densité au théorème 1

s'écrit ici :

$$q(a) = l + \sum_{j=1}^l \sum_{m=1}^{\lambda_j-1} (1 - a_m^j) = l + \sum_{j=1}^l \text{Card } M(a^j).$$

Nous substituons maintenant aux $\partial_{s(a)}(S_c)_{\sigma(a)}$, dans la formule donnant g^λ au théorème 1, les expressions des $\partial_s(S_c)_\sigma$ que nous venons d'obtenir. Nous appliquons la formule de distributivité générale, qui se présente ici sous la forme :

$$\sum_{(a^1, \dots, a^l) \in A^\lambda} \beta \prod_{j=1}^l \delta(a_j) = \beta \prod_{j=1}^l \sum_{a^j \in A^{\lambda, j}} \delta(a_j)$$

avec $A^\lambda = \prod_{j=1}^l \{0, 1\}^{\lambda_j-1}$ et $A^{\lambda, j} = \{0, 1\}^{\lambda_j-1}$

Nous arrivons alors au résultat suivant qui généralise le théorème 2 de Gouget-Raoult, ce dernier correspondant au cas $m = 1$.

Théorème 3 Soit μ la loi prédictive de n durées de survie dans le modèle bayésien de Cox-Dirichlet à censures à droite multiples et indépendantes. Les fonctions de survie des mesures de Dirichlet sont supposées continûment dérivables. Soit $\lambda = (\lambda_j)_{1 \leq j \leq l}$ une configuration d'ex æquo, selon laquelle nous décomposons les suites de contraintes $c_k = (c_{i,k})_{1 \leq i \leq n}$ en l sous suites successives $(c_k^j)_{1 \leq j \leq l}$ avec $c_k^j = (c_{\nu,k}^j)_{1 \leq \nu \leq \lambda_j}$, nous introduisons les sommes partielles supérieures de contraintes $e_{j,k} = \sum_{h=j}^l \sum_{\nu=1}^{\lambda_h} c_{\nu,k}^h$.

Alors la restriction de μ à Δ^λ admet comme densité par rapport à la mesure de Lebesgue sur la variété linéaire P^λ , la fonction g^λ définie par :

$$g^\lambda(x_1^{[\lambda_1]}, \dots, x_l^{[\lambda_l]}) = (-1)^l \prod_{k=1}^m \frac{\Gamma(\alpha_{0,k})}{\Gamma(\alpha_{0,k} + e_{1,k})} \prod_{j=1}^l B_j(x_j) C_j(x_j)$$

avec

$$B_j(x_j) = \prod_{k=1}^m \frac{\Gamma(\alpha_k(x_j) + e_{j,k})}{\Gamma(\alpha_k(x_j) + e_{j+1,k})}$$

et, $\mathcal{P}_{\lambda_j-1}$ désignant l'ensemble de tous les sous ensembles de $\{1, \dots, \lambda_j - 1\}$,

$$C_j(x_j) = \sum_{M \in \mathcal{P}_{\lambda_j-1}} (-1)^{\text{Card } M} \sum_{k=1}^m \alpha'_k(x_j) \times \\ [(Ln\Gamma)'(\alpha_k(x_j) + e_{j+1,k} + c_{\lambda_j,k}^j + \sum_{v \in M} c_{v,k}^j) - (Ln\Gamma)'(\alpha_k(x_j) + e_{j+1,k} + \sum_{v \in M} c_{v,k}^j)].$$

Le théorème ci dessus fournit l'expression des densités uniquement pour des suites croissantes d'observations, en mettant en évidence le rôle des ex æquo. Le corollaire suivant étend ce calcul à toutes les parties Δ_σ^λ . Il s'établit en remarquant que la densité g_σ^λ sur Δ_σ^λ s'obtient en appliquant le théorème 3 à la probabilité déduite de la loi prédictive par réarrangement croissant des coordonnées, autrement dit à la fonction de survie :

$$(t_1, \dots, t_n) \longrightarrow S_c(t_{(\sigma^{-1})_1}, \dots, t_{(\sigma^{-1})_n}).$$

Corollaire 1 la restriction de μ à Δ_σ^λ admet comme densité par rapport à la mesure de Lebesgue sur la variété linéaire P_σ^λ , la fonction g_σ^λ définie par :

$$g_\sigma^\lambda(t_1, \dots, t_n) = (-1)^l \prod_{k=1}^m \frac{\Gamma(\alpha_{0,k})}{\Gamma(\alpha_{0,k} + e_{1,k})} \prod_{j=1}^l B_j(x_j) C_j(x_j),$$

où x_j est la valeur commune des observations t_i appartenant au j^{ieme} bloc d'ex æquo, une fois rangées en ordre croissant, pour tout j et tout k les contraintes $c_{v,k}^j$ sont celles relatives aux observations appartenant au j^{ieme} bloc d'ex æquo, une fois rangées en ordre croissant et $B_j(x_j)$ et $C_j(x_j)$ sont définis comme au théorème 3.

Exemples. Nous reprenons dans notre cas, les exemples donnés dans [15], c'est à dire que nous écrivons les expressions de $C_j(x_j)$ pour $\lambda_j = 1$, $\lambda_j = 2$ et $\lambda_j = 3$.

1) $\lambda_j = 1$

$$C_j(x_j) = \sum_{k=1}^m \alpha'_k(x_j) \left[(Ln\Gamma)'(\alpha_k(x_j) + e_{j+1,k} + c_{1,k}^j) - (Ln\Gamma)'(\alpha_k(x_j) + e_{j+1,k}) \right]$$

2) $\lambda_j = 2$

$$C_j(x_j) = \sum_{k=1}^m \alpha'_k(x_j) \left[(Ln\Gamma)'(\alpha_k(x_j) + e_{j+1,k} + c_{2,k}^j) - (Ln\Gamma)'(\alpha_k(x_j) + e_{j+1,k}) \right. \\ \left. - (Ln\Gamma)'(\alpha_k(x_j) + e_{j+1,k} + c_{2,k}^j + c_{1,k}^j) \right. \\ \left. + (Ln\Gamma)'(\alpha_k(x_j) + e_{j+1,k} + c_{1,k}^j) \right]$$

3) $\lambda_j = 3$

$$C_j(x_j) = \sum_{k=1}^m \alpha'_k(x_j) \left[(Ln\Gamma)'(\alpha_k(x_j) + e_{j+1,k} + c_{3,k}^j) - (Ln\Gamma)'(\alpha_k(x_j) + e_{j+1,k}) \right. \\ \left. - (Ln\Gamma)'(\alpha_k(x_j) + e_{j+1,k} + c_{3,k}^j + c_{2,k}^j) \right. \\ \left. + (Ln\Gamma)'(\alpha_k(x_j) + e_{j+1,k} + c_{2,k}^j) \right. \\ \left. - (Ln\Gamma)'(\alpha_k(x_j) + e_{j+1,k} + c_{3,k}^j + c_{1,k}^j) \right. \\ \left. + (Ln\Gamma)'(\alpha_k(x_j) + e_{j+1,k} + c_{1,k}^j) \right. \\ \left. + (Ln\Gamma)'(\alpha_k(x_j) + e_{j+1,k} + c_{3,k}^j + c_{2,k}^j + c_{1,k}^j) \right. \\ \left. - (Ln\Gamma)'(\alpha_k(x_j) + e_{j+1,k} + c_{2,k}^j + c_{1,k}^j) \right].$$

Chapitre 3

Mesures prédictives dans le modèle bayésien de Cox-Dirichlet à Censures aléatoires à gauche

3.1 Le modèle

On observe n variables aléatoires de la forme :

$$T_i = \sup(T_{i,1}, T_{i,2}) = T_{i,1} \vee T_{i,2},$$

où les $2n$ variables aléatoires $(T_{i,k})$ ($1 \leq k \leq 2$) sont indépendantes et où, chacun des deux vecteurs aléatoires $(T_{1,k}, \dots, T_{n,k})$ est régi par un modèle de Cox-Dirichlet. Ceci veut dire qu'il existe deux fonctions de survie de base G_1 et G_2 et une famille doublement indexée de contraintes $(c_{i,k})_{1 \leq i \leq n, 1 \leq k \leq 2}$ de telle sorte que la fonction de survie de $T_{i,k}$ soit $G_k^{c_{i,k}}$. D'autre part, chacune des deux lois de base G_k est munie d'un a priori de Dirichlet $\mathcal{D}_{\alpha_k^*}$, la fonction de survie de α_k^* est notée α_k . On suppose enfin que les deux lois de base sont a priori indépendantes.

3.2 Fonction de survie prédictive

La fonction de survie prédictive s'écrit :

$$S_c(t_1, \dots, t_n) = \int \prod_{i=1}^n \left(G_1^{c_{i,1}}(t_i) + G_2^{c_{i,2}}(t_i) - G_1^{c_{i,1}}(t_i) G_2^{c_{i,2}}(t_i) \right) \mathcal{D}_{\alpha_1^*}(dG_1) \mathcal{D}_{\alpha_2^*}(dG_2).$$

Le produit à l'intérieur de cette intégrale vaut :

$$\sum_{\substack{I, J, K \subseteq \{1, \dots, n\} \\ I \cap J = \emptyset, I \cap K = \emptyset, J \cap K = \emptyset \\ I \cup J \cup K = \{1, \dots, n\}}} \prod_{i \in I} \prod_{j \in J} \prod_{k \in K} (-1)^{\text{Card} K} G_1^{c_{i,1}}(t_i) G_2^{c_{j,2}}(t_j) G_1^{c_{k,1}}(t_k) G_2^{c_{k,2}}(t_k)$$

L'application du théorème de Fubini donne alors :

$$\begin{aligned} S_c(t_1, \dots, t_n) &= \sum_{\substack{I, J, K \subseteq \{1, \dots, n\} \\ I \cap J = \emptyset, I \cap K = \emptyset, J \cap K = \emptyset \\ I \cup J \cup K = \{1, \dots, n\}}} \int \prod_{i \in I} \prod_{k \in K} (-1)^{\text{Card} K} G_1^{c_{i,1}}(t_i) G_1^{c_{k,1}}(t_k) \mathcal{D}_{\alpha_1^*}(dG_1) \\ &\quad \int \prod_{j \in J} \prod_{k \in K} G_2^{c_{j,2}}(t_j) G_2^{c_{k,2}}(t_k) \mathcal{D}_{\alpha_2^*}(dG_2) \\ &= \sum_{\substack{I, J, K \subseteq \{1, \dots, n\} \\ I \cap J = \emptyset, I \cap K = \emptyset, J \cap K = \emptyset \\ I \cup J \cup K = \{1, \dots, n\}}} (-1)^{\text{Card} K} \int \prod_{i \in I \cup K} G_1^{c_i}(t_i) \mathcal{D}_{\alpha_1^*}(dG_1) \int \prod_{j \in J \cup K} G_2^{c_j}(t_j) \mathcal{D}_{\alpha_2^*}(dG_2) \end{aligned}$$

On obtient ainsi S_c à partir de sommes et de produits de fonctions du type,

$$\mathbb{R}_+^{I'} \longrightarrow \mathbb{R}_+$$

$$(t_i)_{i \in I'} \longmapsto \int \prod_{i \in I'} G^{c_i}(t_i) \mathcal{D}_\alpha^*(dG),$$

où $I' \subset \{1, \dots, n\}$.

On retrouve la forme des intégrales apparaissant dans la formule (1.3) de sorte que chacune d'entre elles possède une I' -dérivée partielle continue sur $\tilde{\Delta}_\sigma$. Il en résulte que

S_c admet encore une $\{1, \dots, n\}$ -dérivée partielle continue sur $\bar{\Delta}_\sigma$, dont le calcul repose sur la remarque élémentaire suivante :

Soit I, J, K une partition de $\{1, \dots, n\}$ incluant la décomposition de toute suite x de $\mathbb{R}^n$ en 3 sous suites x_I, x_J et x_K ; soit par ailleurs une fonction φ définie sur $\mathbb{R}^{I \cup K}$ et une fonction Ψ définie sur $\mathbb{R}^{J \cup K}$; soit $x \in \mathbb{R}^n$ tel que φ admette en $x_{I \cup K}$ une $I \cup K$ -dérivée partielle $\partial_{I \cup K} \varphi$ et Ψ admette en $x_{J \cup K}$ une $J \cup K$ -dérivée partielle $\partial_{J \cup K} \Psi$; alors

$$\partial_{1, \dots, n} [\varphi(x_I, x_K) \Psi(x_J, x_K)] = \partial_{I \cup K} \varphi(x_I, x_K) \partial_J \Psi(x_J, x_K) + \partial_I \varphi(x_I, x_K) \partial_{J \cup K} \Psi(x_J, x_K). \quad (3.1)$$

Nous allons maintenant expliciter S_c en appliquant la formule (1.4)

$$S_c(t_1, \dots, t_n) = \sum_{\substack{I, J, K \subseteq \{1, \dots, n\} \\ I \cap J = \emptyset, I \cap K = \emptyset, J \cap K = \emptyset \\ I \cup J \cup K = \{1, \dots, n\}}} (-1)^{\text{Card} K} a_{I, K} \prod_{i \in I \cup K} \frac{\Gamma(\alpha_1(t_{(\sigma_1)_i}) + C_{i, I, K}^{\sigma_1})}{\Gamma(\alpha_1(t_{(\sigma_1)_i}) + C_{i+1, I, K}^{\sigma_1})} \\ b_{J, K} \prod_{j \in J \cup K} \frac{\Gamma(\alpha_2(t_{(\sigma_2)_j}) + D_{j, J, K}^{\sigma_2})}{\Gamma(\alpha_2(t_{(\sigma_2)_j}) + D_{j+1, J, K}^{\sigma_2})}$$

où,

$$\bullet a_{I, K} = \frac{\Gamma(\alpha_1(0))}{\Gamma(\alpha_1(0) + \sum_{i \in I \cup K} c_{i,1})}, \quad b_{J, K} = \frac{\Gamma(\alpha_2(0))}{\Gamma(\alpha_2(0) + \sum_{j \in J \cup K} c_{j,2})},$$

• σ_1 est une permutation sur l'ensemble $I \cup K$ telle que $(t_{(\sigma_1)_i})_{i \in I \cup K}$ soit non décroissante,

• σ_2 est une permutation sur l'ensemble $J \cup K$ telle que $(t_{(\sigma_2)_j})_{j \in J \cup K}$ soit non décroissante,

$$\bullet C_{i, I, K}^{\sigma_1} = \sum_{h \geq i} c_{(\sigma_1)_h}, \quad D_{j, J, K}^{\sigma_2} = \sum_{h > j} d_{(\sigma_2)_h}.$$

Pour $t \in \bar{\Delta}_\sigma$, $S_c(t_1, \dots, t_n)$ s'écrit :

$$S_c(t_1, \dots, t_n) = \sum_{\substack{I, J, K \subseteq \{1, \dots, n\} \\ I \cap J = \emptyset, I \cap K = \emptyset, J \cap K = \emptyset \\ I \cup J \cup K = \{1, \dots, n\}}} (-1)^{\text{Card} K} a_{I, K} \prod_{i \in I \cup K} \frac{\Gamma(\alpha_1(t_i) + C_{i, I, K}^{\sigma(\sigma^{-1})})}{\Gamma(\alpha_1(t_i) + C_{i+1, I, K}^{\sigma(\sigma^{-1})})} \\ b_{J, K} \prod_{j \in J \cup K} \frac{\Gamma(\alpha_2(t_j) + D_{j, J, K}^{\sigma(\sigma^{-1})})}{\Gamma(\alpha_2(t_j) + D_{j+1, J, K}^{\sigma(\sigma^{-1})})}$$

avec

$$C_{i,I,K}^\sigma = \sum_{\substack{h \geq i \\ h \in I \cup K}} c_{\sigma_h}, \quad D_{j,J,K}^\sigma = \sum_{\substack{h \geq j \\ h \in J \cup K}} d_{\sigma_h}$$

car σ rangeant en ordre croissant la suite $(t_1, \dots, t_n)$, elle va aussi ranger en ordre croissant toute sous suite de $(t_1, \dots, t_n)$, de sorte que :

$$C_{i,I,K}^\sigma = C_{i,I,K}^{\sigma_1}, \quad \text{et} \quad D_{j,J,K}^\sigma = D_{j,J,K}^{\sigma_2}.$$

Pour le calcul de $\partial_s \hat{S}_\sigma(t_1, \dots, t_n)$, tenant compte (voir (3.1)) du fait que pour $i \in K$, t_i apparaît deux fois dans les produits, nous obtenons alors :

$$\begin{aligned} \partial_s \hat{S}_\sigma(t_1, \dots, t_n) = & \sum_{\substack{I, J, K \subseteq \{1, \dots, n\} \\ I \cap J = \emptyset, I \cap K = \emptyset, J \cap K = \emptyset \\ I \cup J \cup K = \{1, \dots, n\}}} (-1)^{\text{Card} K} a_{I,K} b_{J,K} \\ & \prod_{i \in I \cup K} \frac{\Gamma(\alpha_1(t_i) + C_{(\sigma^{-1})i, I, K}^\sigma)}{\Gamma(\alpha_1(t_i) + C_{(\sigma^{-1})i+1, I, K}^\sigma)} \prod_{j \in J \cup K} \frac{\Gamma(\alpha_2(t_j) + D_{(\sigma^{-1})j, J, K}^\sigma)}{\Gamma(\alpha_2(t_j) + D_{(\sigma^{-1})j+1, J, K}^\sigma)} \\ & \prod_{s_k \in I} \alpha'_1(t_{s_k}) [(Ln\Gamma)'(\alpha_1(t_{s_k}) + C_{(\sigma^{-1})s_k, I, K}^\sigma) - (Ln\Gamma)'(\alpha_1(t_{s_k}) + C_{(\sigma^{-1})s_k+1, I, K}^\sigma)] \\ & \prod_{s_k \in J} \alpha'_2(t_{s_k}) [(Ln\Gamma)'(\alpha_2(t_{s_k}) + D_{(\sigma^{-1})s_k, J, K}^\sigma) - (Ln\Gamma)'(\alpha_2(t_{s_k}) + D_{(\sigma^{-1})s_k+1, J, K}^\sigma)] \\ & \prod_{s_k \in K} \{ \alpha'_1(t_{s_k}) [(Ln\Gamma)'(\alpha_1(t_{s_k}) + C_{(\sigma^{-1})s_k, I, K}^\sigma) - (Ln\Gamma)'(\alpha_1(t_{s_k}) + C_{(\sigma^{-1})s_k+1, I, K}^\sigma)] \\ & + \alpha'_2(t_{s_k}) [(Ln\Gamma)'(\alpha_2(t_{s_k}) + D_{(\sigma^{-1})s_k, J, K}^\sigma) - (Ln\Gamma)'(\alpha_2(t_{s_k}) + D_{(\sigma^{-1})s_k+1, J, K}^\sigma)] \} \end{aligned}$$

qui existe et qui est continue si les dérivées α'_1 et α'_2 existent et sont continues.

L'application du théorème 1 devrait alors permettre le calcul des densités. Cependant, les calculs pratiques s'avèrent très compliqués. Cette difficulté est mise en évidence par le calcul de la densité sur la diagonale dans le cas $n = 2$, qui est détaillé dans l'appendice et dont l'expression est donnée par :

$$\begin{aligned}
x &\longrightarrow a \frac{A_c(x)}{A_0(x)} [\ln A_{c_2}(x) - \ln A_c(x) - \ln A_0(x) + \ln A_{c_1}(x)] \\
&+ b \frac{B_d(x)}{B_0(x)} [\ln B_{d_2}(x) - \ln B_d(x) - \ln B_0(x) + \ln B_{d_1}(x)] \\
&+ a \frac{A_c(x)}{A_0(x)} b \frac{B_d(x)}{B_0(x)} [\ln A_{c_2}(x) - \ln A_0(x) + \ln B_{d_2}(x) - \ln B_0(x) \\
&\quad - \ln A_c(x) + \ln A_{c_1}(x) - \ln B_d(x) + \ln B_{d_1}(x)]' \\
&+ a \frac{A_c(x)}{A_0(x)} \frac{B_0(0)}{B_0(x)} \left(\frac{B_{d_2}(x)}{B_{d_2}(0)} + \frac{B_{d_1}(x)}{B_{d_1}(0)} \right) [\ln A_c(x) - \ln A_{c_1}(x) - \ln A_{c_2}(x) + \ln A_0(x)]' \\
&+ b \frac{B_d(x)}{B_0(x)} \frac{A_0(0)}{A_0(x)} \left(\frac{A_{c_2}(x)}{A_{c_2}(0)} + \frac{A_{c_1}(x)}{A_{c_1}(0)} \right) [\ln B_d(x) - \ln B_{d_1}(x) - \ln B_{d_2}(x) + \ln B_0(x)]'
\end{aligned}$$

où $c = c_{1,1} + c_{2,1}$, $d = c_{1,2} + c_{2,2}$, $A_h(t) = \Gamma(\alpha_1(t) + h)$, $B_h(t) = \Gamma(\alpha_2(t) + h)$,

$$a = \frac{\Gamma(\alpha_1(0))}{\Gamma(\alpha_1(0) + c)}, \quad b = \frac{\Gamma(\alpha_2(0))}{\Gamma(\alpha_2(0) + d)}$$

Chapitre 4

Mesures prédictives dans le modèle bayésien de Cox-Dirichlet à censures aléatoires mixtes

4.1 Le modèle

Si T_1 est la variable d'intérêt, T_2 est la variable de censure à droite et T_3 est la variable de censure à gauche, alors la définition de l'observation doublement censurée peut être prise soit comme

$$X_g = \inf(\sup(T_1, T_3), T_2) = (T_1 \vee T_3) \wedge T_2,$$

soit comme

$$X_d = \sup(\inf(T_1, T_2), T_3) = (T_1 \wedge T_2) \vee T_3,$$

selon que la censure initiale est celle à gauche ou celle à droite. Puisque $X_d = (T_1 \vee T_3) \wedge (T_2 \vee T_3)$, on a toujours $X_d \geq X_g$ et il apparaît que, si $T_3 \leq T_2$, alors $X_g = X_d$ et il

n'y a plus d'ambiguïté sur l'observation. Mais supposer ce lien entre les deux variables de censure compromettrait l'hypothèse de leur indépendance. Ceci nous conduit à la considération d'un troisième modèle, dit symétrique, où les censures à droite et à gauche sont assurées respectivement par $T_2 \vee T_3$ et $T_2 \wedge T_3$ et où donc l'observation est

$$X_s = [T_1 \vee (T_2 \wedge T_3)] \wedge (T_2 \vee T_3) = [T_1 \wedge (T_2 \vee T_3)] \vee (T_2 \wedge T_3).$$

On remarque qu'on peut réécrire X_g (resp X_d) sous la forme équivalente :

$$X_g = (T_1 \vee (T_2 \wedge T_3)) \wedge T_2 \text{ (resp } X_d = (T_1 \wedge (T_3 \vee T_2)) \vee T_3).$$

Les dernières formules donnant X_s , X_g et X_d ont mis à chaque fois en évidence un couple (U_2, U_3) tel que $U_2 \leq U_3$, et l'observation est censurée à gauche par U_2 et à droite par U_3 , sans qu'il soit besoin de préciser quelle est la censure initiale.

Si pour tout k , tel que $1 \leq k \leq 3$, G_k est la fonction de survie de la variable aléatoire T_k et si les variables $(T_k)_{1 \leq k \leq 3}$ sont indépendantes, on obtient alors les fonctions de survie des trois modèles ci-dessus.

4.1.1 Modèle symétrique

$$P(X_s > t) = (G_1 G_2 + G_1 G_3 + G_2 G_3 - 2G_1 G_2 G_3)(t).$$

4.1.2 Modèle à censure initiale à droite par T_2

$$P(X_d > t) = (G_3 + G_1 G_2 - G_1 G_2 G_3)(t).$$

4.1.3 Modèle à censure initiale à gauche par T_3

$$P(X_g > t) = (G_1 G_2 + G_2 G_3 - G_1 G_2 G_3)(t).$$

4.2 Echantillon de taille n

Soit $(T_{i,k})_{1 \leq i \leq n, 1 \leq k \leq 3}$ une famille de variables aléatoires indépendantes, où pour tout k tel que $1 \leq k \leq 3$, le vecteur aléatoire $(T_{1,k}, \dots, T_{n,k})$ est régi par un modèle de Cox-Dirichlet. Ceci veut dire qu'il existe trois fonctions de survie de base G_1, G_2 et G_3 et une famille doublement indexée de contraintes $c = (c_{i,k})_{1 \leq i \leq n, 1 \leq k \leq 3}$ de telle sorte que la fonction de survie de $T_{i,k}$ soit $G_k^{c_{i,k}}$. Par ailleurs chaque loi de base est munie d'un a priori de Dirichlet $\mathcal{D}_{\alpha_k^*}$, caractérisé par la mesure positive finie α_k^* de fonction de survie notée α_k . On suppose enfin que les trois lois de base sont a priori indépendantes. Les différents modèles ci dessus donnent lieu aux calculs que nous présentons maintenant.

Dans chaque cas on obtient, comme dans le cas de la censure à gauche, la fonction de survie à partir de sommes et de produits de fonctions du type

$$\int \prod_{i \in I'} G^{c_i}(t_i) \mathcal{D}_{\alpha^*}(dG),$$

d'où il résulte que S_c admet une $\{1, \dots, n\}$ -dérivée partielle continue sur chaque $\bar{\Delta}_\sigma$ et donc est régulière par portions.

4.2.1 Modèle symétrique

Ici l'observation d'indice i est :

$$T_i = [T_{i,1} \wedge (T_{i,2} \vee T_{i,3})] \vee (T_{i,2} \wedge T_{i,3}).$$

La fonction de survie prédictive s'écrit alors :

$$\prod_{k=1}^3 \mathcal{D}_{\alpha_k^*}(dG_k)$$

$$\begin{aligned}
&= \sum_{\substack{I, J, K, L \subseteq \{1, \dots, n\} \\ I \cap J = \emptyset, I \cap K = \emptyset, I \cap L = \emptyset \\ J \cap K = \emptyset, J \cap L = \emptyset, K \cap L = \emptyset \\ I \cup J \cup K \cup L = \{1, \dots, n\}}} (-2)^{\text{Card}L} \int \prod_{i \in I} \prod_{j \in J} \prod_{k \in K} \prod_{l \in L} (G_1^{c_{i,1}} G_2^{c_{i,2}})(t_i) (G_1^{c_{j,1}} G_3^{c_{j,3}})(t_j) \\
&\quad (G_2^{c_{k,2}} G_3^{c_{k,3}})(t_k) (G_1^{c_{l,1}} G_2^{c_{l,2}} G_3^{c_{l,3}})(t_l) \prod_{k=1}^3 \mathcal{D}_{\alpha_k^*}(dG_k) \\
&= \sum_{\substack{I, J, K, L \subseteq \{1, \dots, n\} \\ I \cap J = \emptyset, I \cap K = \emptyset, I \cap L = \emptyset \\ J \cap K = \emptyset, J \cap L = \emptyset, K \cap L = \emptyset \\ I \cup J \cup K \cup L = \{1, \dots, n\}}} (-2)^{\text{Card}L} \int \prod_{i \in I \cup J \cup L} G_1^{c_{i,1}}(t_i) \mathcal{D}_{\alpha_1^*}(dG_1) \\
&\quad \int \prod_{j \in I \cup K \cup L} G_2^{c_{j,2}}(t_j) \mathcal{D}_{\alpha_2^*}(dG_2) \\
&\quad \int \prod_{k \in J \cup K \cup L} G_3^{c_{k,3}}(t_k) \mathcal{D}_{\alpha_3^*}(dG_3).
\end{aligned}$$

4.2.2 Modèle à censure initiale à droite par T_2

Ici l'observation d'indice i est :

$$T_i = (T_{i,1} \vee T_{i,3}) \wedge (T_{i,2} \vee T_{i,3}).$$

La fonction de survie prédictive s'écrit alors :

$$\begin{aligned}
S_c(t_1, \dots, t_n) &= \int \prod_{i=1}^n (G_1^{c_{i,1}} G_2^{c_{i,2}} + G_3^{c_{i,3}} - G_1^{c_{i,1}} G_2^{c_{i,2}} G_3^{c_{i,3}})(t_i) \prod_{k=1}^3 \mathcal{D}_{\alpha_k^*}(dG_k) \\
&= \sum_{\substack{I, J, K \subseteq \{1, \dots, n\} \\ I \cap J = \emptyset, I \cap K = \emptyset, J \cap K = \emptyset \\ I \cup J \cup K = \{1, \dots, n\}}} (-2)^{\text{Card}K} \int \prod_{i \in I} \prod_{j \in J} \prod_{k \in K} (G_1^{c_{i,1}} G_2^{c_{i,2}})(t_i) (G_3^{c_{j,3}})(t_j) (G_1^{c_{k,1}} G_2^{c_{k,2}} G_3^{c_{k,3}})(t_k) \\
&\quad \prod_{k=1}^3 \mathcal{D}_{\alpha_k^*}(dG_k) \\
&= \sum_{\substack{I, J, K \subseteq \{1, \dots, n\} \\ I \cap J = \emptyset, I \cap K = \emptyset, J \cap K = \emptyset \\ I \cup J \cup K = \{1, \dots, n\}}} (-2)^{\text{Card}K} \int \prod_{i \in I \cup K} G_1^{c_{i,1}}(t_i) \mathcal{D}_{\alpha_1^*}(dG_1) \\
&\quad \int \prod_{j \in I \cup K} G_2^{c_{j,2}}(t_j) \mathcal{D}_{\alpha_2^*}(dG_2) \\
&\quad \int \prod_{k \in J \cup K} G_3^{c_{k,3}}(t_k) \mathcal{D}_{\alpha_3^*}(dG_3).
\end{aligned}$$

4.2.3 Modèle à censure initiale à gauche par T_3

Ici l'observation d'indice i est :

$$T_i = [T_{i,1} \vee (T_{i,2} \wedge T_{i,3})] \wedge T_{i,2}.$$

La fonction de survie prédictive s'écrit alors :

$$\begin{aligned} S_c(t_1, \dots, t_n) &= \int \prod_{i=1}^n (G_1^{c_{i,1}} G_2^{c_{i,2}} + G_2^{c_{i,2}} G_3^{c_{i,3}} - G_1^{c_{i,1}} G_2^{c_{i,2}} G_3^{c_{i,3}})(t_i) \prod_{k=1}^3 \mathcal{D}_{\alpha_k^*}(dG_k) \\ &= \sum_{\substack{I, J, K \subseteq \{1, \dots, n\} \\ I \cap J = \emptyset, I \cap K = \emptyset, J \cap K = \emptyset \\ I \cup J \cup K = \{1, \dots, n\}}} (-2)^{\text{Card}K} \int \prod_{i \in I} \prod_{j \in J} \prod_{k \in K} (G_1^{c_{i,1}} G_2^{c_{i,2}})(t_i) (G_2^{c_{j,2}} G_3^{c_{j,3}})(t_j) (G_1^{c_{k,1}} G_2^{c_{k,2}} G_3^{c_{k,3}})(t_k) \\ &\quad \prod_{k=1}^3 \mathcal{D}_{\alpha_k^*}(dG_k) \\ &= \sum_{\substack{I, J, K \subseteq \{1, \dots, n\} \\ I \cap J = \emptyset, I \cap K = \emptyset, J \cap K = \emptyset \\ I \cup J \cup K = \{1, \dots, n\}}} (-2)^{\text{Card}K} \int \prod_{i \in I \cup K} G_1^{c_{i,1}}(t_i) \mathcal{D}_{\alpha_1^*}(dG_1) \\ &\quad \int \prod_{j \in I \cup J \cup K} G_2^{c_{j,2}}(t_j) \mathcal{D}_{\alpha_2^*}(dG_2) \\ &\quad \int \prod_{k \in J \cup K} G_3^{c_{k,3}}(t_k) \mathcal{D}_{\alpha_3^*}(dG_3). \end{aligned}$$

Chapitre 5

Calcul des densités prédictives dans le modèle bayésien de Cox-Dirichlet à censures fixes

5.1 Introduction

Dans ce chapitre, nous étudions le cas des censures fixes appliquées à des temps de survie toujours régis par le modèle bayésien de Cox-Dirichlet. De façon précise, nous observons n variables aléatoires positives censurées $X_i = T_i \wedge \tau_i$. τ_i est fixé pour tout i et il existe une fonction de survie de base G et une famille de contraintes strictement positives $c = (c_1, \dots, c_n)$ de sorte que la fonction de survie de T_i soit la puissance G^{c_i} . De plus G est munie de l'a priori de Dirichlet $\mathcal{D}_{\alpha^*}$, caractérisé par la mesure finie α^* de fonction de survie α . La mesure prédictive est alors concentrée sur le pavé $\prod_{i=1}^n [0, \tau_i]$. Sa fonction de survie est donnée par :

$$t = (t_1, \dots, t_n) \longrightarrow \int P(X_1 > t_1, \dots, X_n > t_n) \mathcal{D}_{\alpha^*}(dG).$$

Elle est égale à :

$$S_c(t) = \begin{cases} S(t) & \text{si } t_i < \tau_i \quad \forall 1 \leq i \leq n \\ 0 & \text{sinon} \end{cases}$$

où $S(t)$ est donnée par la formule (1.4).

Cette loi prédictive perd la propriété de régularité par portions. Cependant à l'intérieur du pavé $\prod_{i=1}^n [0, \tau_i]$ l'étude est la même que dans le cas non censuré effectuée par N. Gouget et J. P. Raoult dans [15]. C'est pourquoi l'étude ultérieure portera sur le calcul des densités sur des sous ensembles des faces, ensembles caractérisés par la fixation des variables censurées et par des égalités et des inégalités strictes entre les variables non censurées.

Remarque 1

Pour une mesure régulière par portions, toute variété linéaire affine parallèle à un des axes de coordonnées a une mesure nulle. En effet, il suffit de vérifier cette propriété pour les variétés de dimension $n - 1$; toute partie de $\mathbb{R}^n$ d'équation $\varphi(x) = a$, où $a \in \mathbb{R}$ et φ est une forme linéaire, admet avec toute partie Δ_σ^λ une intersection de dimension $l - 1$, si l est la dimension de Δ_σ^λ (c'est à dire le nombre de blocs d'ex æquo dans la configuration λ) ; or sur chaque Δ_σ^λ la mesure est absolument continue par rapport à la mesure de Lebesgue l dimensionnelle (voir théorème 1, le passage des parties Δ^λ aux parties Δ_σ^λ se faisant par permutation).

5.2 Calcul de densités prédictives

Soit μ une mesure régulière par portions sur $\mathbb{R}^n$, de fonction de survie S , à laquelle on applique n censures (à droite) fixes $(\tau_i)_{1 \leq i \leq n}$. Autrement dit nous nous intéressons à la loi multi-dimensionnelle des variables $X_i = T_i \wedge \tau_i$ où μ la loi de $(T_1, \dots, T_n)$ est régulière par portions. La mesure obtenue est concentrée sur $\prod_{i=1}^n [0, \tau_i]$. Sur $\prod_{i=1}^n [0, \tau_i[$ la

mesure n'est pas modifiée par rapport au cas non censuré, dont l'étude a été faite par N. Gouget et J. P. Raoult dans [15]. Il reste donc à effectuer l'étude sur les faces $(n - 1)$ dimensionnelles d'équation $t_i = \tau_i$, c'est à dire les parties :

$$\prod_{j=1}^{i-1} [0, \tau_j] \times \{\tau_i\} \times \prod_{j=i+1}^n [0, \tau_j],$$

qui sont de mesure en général non nulle ; remarquons que même après censure, toutes les variétés affines de dimension $n - 1$ du type :

$$\prod_{j=1}^{i-1} \mathbb{R}_+ \times \{\rho_i\} \times \prod_{j=i+1}^n \mathbb{R}_+,$$

où $\rho_i \neq \tau_i$ restent de mesure nulle.

Afin de mieux comprendre l'étude à effectuer sur chacune des faces, nous la présentons tout d'abord pour $n = 2$ et $n = 3$.

5.2.1 Cas particulier en dimension 2 ou 3

$n = 2$

On observe $X_1 = T_1 \wedge \tau_1$ et $X_2 = T_2 \wedge \tau_2$.

Nous supposons, sans perte de généralité, que $\tau_1 \leq \tau_2$.

La restriction de la probabilité prédictive au "rectangle ouvert à droite" $[0, \tau_1[\times [0, \tau_2[$ n'est pas modifiée par rapport au cas où il n'y a pas de censure. Elle est donc absolument continue, par rapport à la mesure de Lebesgue, sur chacune des parties $\Delta_{1,2} \cap ([0, \tau_1[\times [0, \tau_2[)$, $\Delta_{2,1} \cap ([0, \tau_1[\times [0, \tau_2[)$ et $\Delta^2 \cap ([0, \tau_1[\times [0, \tau_2[)$ et sa densité est connue. On doit donc étudier cette probabilité prédictive sur les intervalles ouverts à droite $\{\tau_1\} \times [0, \tau_2[$ et $[0, \tau_1[\times \{\tau_2\}$ et au point $\{\tau_1, \tau_2\}$ (voir fig 1).

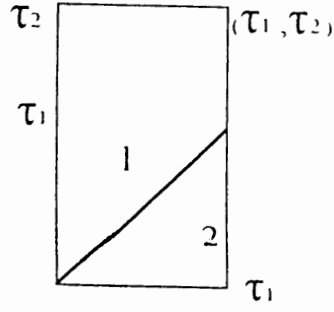


fig 1

$$1 : \Delta_{1,2} \cap ([0, \tau_1[\times [0, \tau_2[)$$

$$2 : \Delta_{2,1} \cap ([0, \tau_1[\times [0, \tau_2[)$$

Au point $\{\tau_1, \tau_2\}$ se trouve la masse $S(\tau_1, \tau_2)$.

Sur l'intervalle $[0, \tau_1[\times \{\tau_2\}$, qui est tout entier inclus dans $\Delta_{1,2}$, la restriction de la probabilité prédictive admet pour fonction de survie unidimensionnelle $S_{1,2}(\cdot, \tau_2)$ qui est continûment dérivable par hypothèse, et sa densité s'obtient par dérivation.

En revanche l'intervalle $\{\tau_1\} \times [0, \tau_2[$ a, sauf si $\tau_1 = \tau_2$, des intersections non vides à la fois avec $\Delta_{1,2}$ et $\Delta_{2,1}$. L'étude de la restriction à $\{\tau_1\} \times [0, \tau_2[$ impose donc de considérer séparément ses restrictions à :

$$\Delta_{1,2}^{\tau_1} = \Delta_{1,2} \cap (\{\tau_1\} \times [0, \tau_2[) \text{ (qui est vide si } \tau_1 = \tau_2)$$

et à :

$$\Delta_{2,1}^{\tau_1} = \Delta_{2,1} \cap (\{\tau_1\} \times [0, \tau_2[).$$

La première est $S_{1,2}(\tau_1, \cdot)$ et la seconde est $S_{2,1}(\tau_1, \cdot)$, d'où l'on déduit, par dérivation, la densité sur $\Delta_{1,2}^{\tau_1}$ et sur $\Delta_{2,1}^{\tau_1}$. Cette densité est ainsi obtenue, dans le cas $\tau_1 < \tau_2$, sauf au point (τ_1, τ_1) ; or en ce point la masse est $\mu(\{\tau_1, \infty[\times \{\tau_1\})$ qui est nulle (cf remarque 1). La densité est donc bien obtenue presque sûrement.

$n = 3$

On observe $X_1 = T_1 \wedge \tau_1$, $X_2 = T_2 \wedge \tau_2$, $X_3 = T_3 \wedge \tau_3$ et on suppose que $\tau_1 \leq \tau_2 \leq \tau_3$.

La probabilité prédictive censurée est alors concentrée sur le pavé fermé $[0, \tau_1] \times [0, \tau_2] \times [0, \tau_3]$ (voir fig 2) et sa restriction au "pavé ouvert à droite" $[0, \tau_1[\times [0, \tau_2[\times [0, \tau_3[$

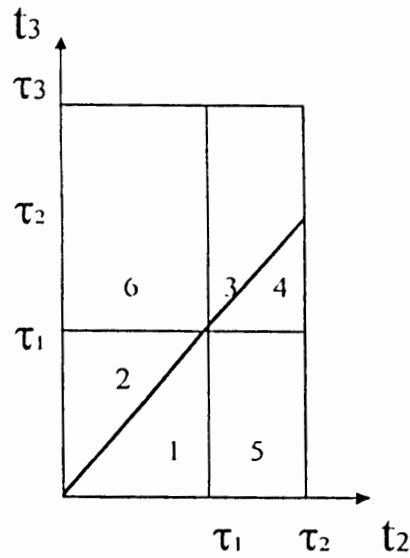


fig 3

$$1 : F_1 \cap \Delta_{3,2,1}$$

$$2 : F_1 \cap \Delta_{2,3,1}$$

$$3 : F_1 \cap \Delta_{1,2,3} \text{ si } \tau_1 < \tau_2$$

$$4 : F_1 \cap \Delta_{1,3,2} \text{ si } \tau_1 < \tau_2$$

$$5 : F_1 \cap \Delta_{3,1,2} \text{ si } \tau_1 < \tau_2$$

$$6 : F_1 \cap \Delta_{2,1,3} \text{ si on n'a pas } \tau_1 = \tau_2 = \tau_3$$

Evidemment sur chacune des parties $F_1 \cap \Delta_\sigma$ ci dessus la densité prédictive, par rapport à la mesure de Lebesgue bidimensionnelle, s'obtient par la $\{2,3\}$ dérivation de $S_\sigma(\tau_1, \dots)$.

Le complémentaire, dans la face F_1 , de $\cup_\sigma (\Delta_\sigma \cap F_1)$ se compose de deux types de parties.

1. celles parallèles aux axes

$$\{\tau_1\} \times \{\tau_1\} \times [0, \tau_3] \text{ si } \tau_1 < \tau_2$$

$$\{\tau_1\} \times [0, \tau_2] \times \{\tau_1\} \text{ si } \tau_1 < \tau_3$$

2. les "intervalles diagonaux"

$$\Delta_{2,3,1}^{2,1;\tau_1} = \{(t_1, t_2, t_3) \in F_1 ; t_2 = t_3 < \tau_1\} \text{ et si } \tau_1 < \tau_2$$

$$\Delta_{1,2,3}^{1,2;\tau_1} = \{(t_1, t_2, t_3) \in F_1 ; \tau_1 < t_2 = t_3\}.$$

Les intervalles de type 1 sont de mesure nulle (voir remarque 1). Il reste à calculer la densité, par rapport à la mesure de Lebesgue unidimensionnelle, sur les "intervalles diagonaux" ; or la mesure de fonction de survie $S(\tau_1, \cdot, \cdot)$ est régulière par portions sur $([0, \tau_1])^2$ et sur $(]\tau_1, \tau_2])^2$ (si $\tau_1 < \tau_2$). Donc la densité prédictive se calcule dans chaque cas selon la technique standard, pour la mesure en dimension 2, appliquée au couple $(S_{1,2}(\tau_1, \cdot, \cdot), S_{2,1}(\tau_1, \cdot, \cdot))$ (formule (1.7) du lemme 1) où

$$\begin{aligned} \text{sur } \Delta_{2,3,1}^{2,1;\tau_1} \quad S_{1,2}(\tau_1, t_2, t_3) &= S_{2,3,1}(\tau_1, t_2, t_3) \\ \text{et } S_{2,1}(\tau_1, t_2, t_3) &= S_{3,2,1}(\tau_1, t_2, t_3). \end{aligned}$$

tandis que

$$\begin{aligned} \text{sur } \Delta_{1,2,3}^{1,2;\tau_1} \quad S_{1,2}(\tau_1, t_2, t_3) &= S_{1,2,3}(\tau_1, t_2, t_3) \\ \text{et } S_{2,1}(\tau_1, t_2, t_3) &= S_{1,3,2}(\tau_1, t_2, t_3). \end{aligned}$$

$$\text{b) } F_2 = [0, \tau_1[\times \{\tau_2\} \times [0, \tau_3[$$

En tenant compte du fait que $t_1 < \tau_1 \leq \tau_2 \leq \tau_3$, F_2 a au plus trois intersections, non vides, avec les Δ_σ ; elles sont représentées dans la figure 4.

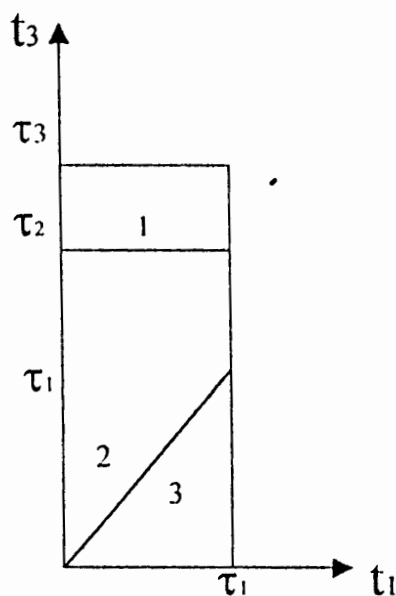


fig 4

$$1 : F_2 \cap \Delta_{1,2,3} \text{ si } \tau_2 < \tau_3$$

$$2 : F_2 \cap \Delta_{1,3,2}$$

$$3 : F_2 \cap \Delta_{3,1,2}.$$

Sur chacune des parties $F_2 \cap \Delta_\sigma$ ci dessus la densité prédictive, par rapport à la mesure de Lebesgue bidimensionnelle, s'obtient par la $\{1, 3\}$ dérivation de $S_\sigma(., \tau_2, .)$.

Le complémentaire, dans la face F_2 , de $\bigcup_\sigma (\Delta_\sigma \cap F_2)$ se compose, ici aussi, de deux types de parties.

1 : celle parallèle à un axe

$$[0, \tau_1[\times \{\tau_2\} \times \{\tau_2\}$$

2 : "l'intervalle diagonal"

$$\Delta_{1,3,2}^{2,1;\tau_2} = \{(t_1, t_2, t_3) \in F_2 ; t_1 = t_3 < \tau_2\}.$$

La partie de type 1 est de mesure nulle (voir remarque 1).

Il reste à calculer la densité, par rapport à la mesure de Lebesgue unidimensionnelle, sur "l'intervalle diagonal" $\Delta_{1,3,2}^{2,1;\tau_2}$; or ici aussi la mesure de fonction de survie $S(., \tau_2, .)$ est régulière par portions sur $([0, \tau_1])^2$. Il suffit donc de lui appliquer, comme précédemment, la technique standard donnée au lemme 1.

$$c) F_3 = [0, \tau_1[\times [0, \tau_2[\times \{\tau_3\}$$

Ici les possibilités d'égalité des τ_i n'interviennent pas et F_3 a toujours deux intersections, non vides, avec les Δ_σ . Elles sont représentées dans la figure 5.

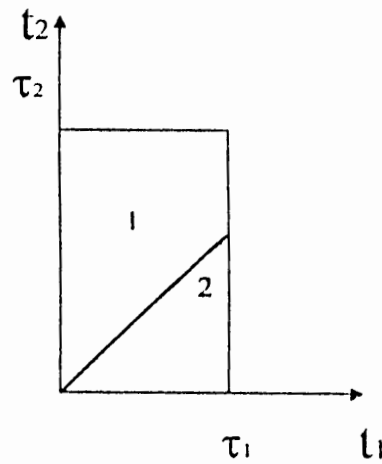


fig 5

$$1 : F_3 \cap \Delta_{1,2,3}$$

$$2 : F_3 \cap \Delta_{2,1,3}$$

Sur chacune des deux parties précédentes la densité, par rapport à la mesure de Lebesgue bidimensionnelle, s'obtient par la $\{1, 2\}$ dérivation de $S_\sigma(\cdot, \cdot, \tau_3)$.

Le complémentaire, dans F_3 , de $(F_3 \cap \Delta_{1,2,3}) \cup (F_3 \cap \Delta_{2,1,3})$ ne se compose que de la partie de type 2 "intervalle diagonal" suivante :

$$\Delta_{1,2,3}^{2,1;\tau_3} = \{(t_1, t_2, t_3) \in F_3 ; t_1 = t_2 < \tau_3\}.$$

La densité prédictive, par rapport à la mesure de Lebesgue unidimensionnelle, sur $\Delta_{1,2,3}^{2,1;\tau_3}$ s'obtient alors par la technique standard du lemme 1.

Faces de dimension 1

$$a) F_4 = \{\tau_1\} \times \{\tau_2\} \times [0, \tau_3[$$

F_4 a au plus trois intersections, non vides, avec les Δ_σ et elles sont représentées dans la figure 6

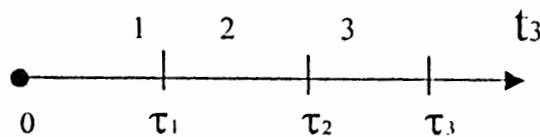


fig 6

$$1 : F_4 \cap \Delta_{3,1,2}$$

$$2 : F_4 \cap \Delta_{1,3,2} \text{ si } \tau_1 < \tau_2$$

$$3 : F_4 \cap \Delta_{1,2,3} \text{ si } \tau_2 < \tau_3.$$

Sur chacune des parties $F_4 \cap \Delta_\sigma$ ci dessus la densité prédictive, par rapport à la mesure de Lebesgue unidimensionnelle, s'obtient par dérivation de $S_\sigma(r_1, r_2, \cdot)$.

Par ailleurs remarquons que le complémentaire, dans F_4 , de $\bigcup_{\sigma} (F_4 \cap \Delta_\sigma)$ sont les points $\{\tau_1\} \times \{\tau_2\} \times \{\tau_1\}$ et $\{\tau_1\} \times \{\tau_2\} \times \{\tau_2\}$ qui sont de mesure nulle.

$$b) F_5 = \{\tau_1\} \times [0, \tau_2[\times \{\tau_3\}$$

F_5 a au plus deux intersections, non vides, avec les Δ_σ et elles sont représentées à la figure 7.

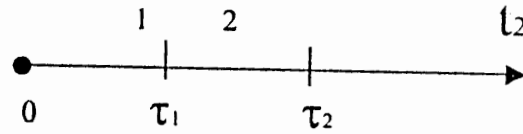


fig 7

$$1 : F_5 \cap \Delta_{2,1,3}$$

$$2 : F_5 \cap \Delta_{1,2,3} \text{ si } \tau_1 < \tau_2$$

Sur chacune des deux parties précédentes la densité prédictive, par rapport à la mesure de Lebesgue unidimensionnelle, s'obtient respectivement par dérivation de $S_{2,1,3}(\tau_1, \cdot, \tau_3)$ et de $S_{1,2,3}(\tau_1, \cdot, \tau_3)$.

Le complémentaire, dans F_5 , de $(F_5 \cap \Delta_{2,1,3}) \cup (F_5 \cap \Delta_{1,2,3})$, si $\tau_1 < \tau_2$, est $\{\tau_1\} \times \{\tau_1\} \times \{\tau_3\}$ qui est de mesure nulle.

$$c) F_6 = [0, \tau_1[\times \{\tau_2\} \times \{\tau_3\}$$

F_6 est entièrement incluse dans $\bar{\Delta}_{1,2,3}$, la densité restreinte à F_6 s'obtient donc par la dérivation de $S_{123}(\cdot, \tau_2, \tau_3)$.

5.2.2 Cas général

Rappelons que notre but est de calculer la densité de la restriction de la mesure prédictive à chaque face. Chacune de ces faces est caractérisée par l'ensemble des indices i pour lesquels x_i est fixée à la valeur de la censure τ_i . Soit donc

$$\mathcal{C} = \{(c_1, \dots, c_s) \subseteq \{1, \dots, n\}, c_1 < c_2 < \dots < c_s\}.$$

et soit

$$\mathcal{F}_{\mathcal{C}} = \{(x_1, \dots, x_n) / \forall i \in \mathcal{C} \ x_i = \tau_i, \forall i \notin \mathcal{C} \ x_i < \tau_i\}$$

la face de dimension $n - s$.

Remarquons les inégalités strictes ($x_i < \tau_i$, si $i \notin \mathcal{C}$) dans la définition de $\mathcal{F}_{\mathcal{C}}$. Autrement dit nous considérons des faces ouvertes à droite; en effet si certains des x_i (où $i \notin \mathcal{C}$) valent τ_i le point considéré se trouve dans une face de dimension plus faible.

En tenant compte du fait que si $i \notin \mathcal{C}$ alors $x_i < \tau_i$, nous déduisons que $\mathcal{F}_{\mathcal{C}}$ a une intersection vide avec certains ensembles Δ_{σ} (comme on a pu déjà le constater sur les cas $n = 2$ et $n = 3$). Cependant pour toute permutation σ , telle que $\Delta_{\sigma} \cap \mathcal{F}_{\mathcal{C}}$ n'est pas vide, la restriction de la mesure prédictive à $\Delta_{\sigma} \cap \mathcal{F}_{\mathcal{C}}$ admet, pour densité par rapport aux variables non censurées, le produit par $(-1)^{n-s}$ de la $\{r_1, \dots, r_{n-s}\}$ dérivée de S_{σ} où $r_j \notin \mathcal{C}$ pour tout j tel que $1 \leq j \leq n - s$.

Le complémentaire dans $\mathcal{F}_{\mathcal{C}}$ de $\bigcup_{\sigma} (\Delta_{\sigma} \cap \mathcal{F}_{\mathcal{C}})$ se compose de deux types de parties :

1. des parties pour lesquelles une variable non censurée prend la valeur d'une variable censurée; ces parties sont de mesure nulle (voir remarque 1), et il n'y a pas de calcul de densité à effectuer dans ce cas;
2. des parties comprenant des ex æquo d'une part entre des valeurs de variables non censurées et (ou) d'autre part entre des variables censurées, et pour lesquelles le calcul de

densités va être effectué ci dessous.

On se restreint dans cette partie à considérer des suites croissantes d'observations (le cas général s'en déduisant par permutation).

Pour tout p tel que $1 \leq p \leq s + 1$, on va considérer, la suite des observations strictement supérieures à la $(p - 1)^{ieme}$ censure (si $p > 1$) et inférieures à la p^{ieme} censure si $(p \leq s)$.

a. Pour $1 \leq p \leq s$, on considère ainsi, en posant $c_0 = 0$,

$$S_p = (x_{c_{p-1}+1}, \dots, x_{c_p})$$

- Si cette suite n'est pas réduite à x_{c_p} ($= \tau_{c_p}$), les observations qui y figurent et qui sont d'indice strictement inférieur à c_p sont non censurées et strictement inférieures à x_{c_p} (car il n'y a pas ici d'égalité entre valeurs censurées et valeurs non censurées). Notons λ^p la configuration d'ex æquo définie par S_p ; elle se termine nécessairement par 1.

Notons par ailleurs δ^p la suite, de même longueur que λ^p , dont le dernier élément est 1 et tous les autres, s'ils existent, sont 0; c'est la suite d'indicateurs de censure de S_p : tout élément de cette suite vaut 0 si le regroupement d'ex æquo est constitué d'observations non censurées, et 1 s'il s'agit d'une observation censurée.

- Si la suite S_p est réduite à x_{c_p} , sa configuration d'ex æquo, λ^p , est (1) et son indicateur de censure δ^p , est (1), ceci peut se produire dans trois cas.

- il y a deux censures en des valeurs distinctes, et aucune observation non censurée entre elles (alors $x_{c_{p-1}} = \tau_{c_{p-1}} < x_{c_p} = \tau_{c_p}$)

- il y a deux censures égales (alors $x_{c_{p-1}} = \tau_{c_{p-1}} = x_{c_p} = \tau_{c_p}$)

- la première observation x_1 est censurée.

b. Pour $p = s + 1$, on considère alors la suite

$$S_{s+1} = (x_{c_s+1}, \dots, x_n),$$

éventuellement vide si $x_{c_s} = x_n$. λ^{s+1} est la configuration d'ex æquo qui lui est associée, éventuellement vide; la suite d'indicateurs de censure δ^{s+1} est soit vide, soit composée de 0.

On définit alors

λ : concaténée de $\lambda^1, \dots, \lambda^{s+1}$

δ : concaténée de $\delta^1, \dots, \delta^{s+1}$,

Posons

$$\lambda = (\lambda_1, \dots, \lambda_l) \quad (5.1)$$

et $\delta = (\delta_1, \dots, \delta_l)$.

Interprétons les suites λ (qui a la structure d'une configuration d'ex æquo), que nous appellerons "suite de regroupements" et δ (suite de 0 et de 1).

Soit λ' la configuration d'ex æquo de la suite croissante d'observations $(x_1, \dots, x_n)$, λ se déduit de λ' de la manière suivante :

Si $\lambda' = (\lambda'_1, \dots, \lambda'_l)$, il peut y avoir deux types de paquets d'ex æquo,

- ceux concernant des valeurs censurées, sur lesquels nous n'effectueront pas de regroupements,
- ceux concernant des valeurs non censurées, que nous regroupant.

Pour passer de λ' à λ , on conserve comme regroupements les blocs d'ex æquo de valeurs non censurées mais, si $\lambda'_h \neq 1$ et caractérise des valeurs censurées ex æquo, on le remplace par $(1, \dots, 1)$ (suite de longueur λ'_h).

δ indique la nature des regroupements de valeurs par λ : 1 s'il s'agit d'une observation censurée, 0 s'il s'agit d'observations non censurées. Remarquons que chaque fois que $\delta_h = 1$ ($1 \leq h \leq l$) alors $\lambda_h = 1$; insistons sur le fait que des regroupements consécutifs distincts, d'effectif 1, peuvent concerner des valeurs égales, ceci ne pouvant se produire

que s'il s'agit de variables censurées.

Exemple 1

Soit $\mathcal{C} = \{4, 5, 8\}$ et soit une suite $(t_1, t_2, \dots, t_9)$ vérifiant les propriétés suivantes :

$$t_1 = t_2 < t_3 < t_4 = t_5 < t_6 = t_7 < t_8 < t_9;$$

$t_1, t_2, t_3, t_6, t_7, t_9$ sont non censurées et t_4, t_5, t_8 sont censurées.

Alors $\lambda' = (2, 1, 2, 2, 1, 1)$ où $\lambda'_3 = 2$ et $\lambda'_5 = 1$ sont des effectifs de paquets d'ex æquo de valeurs censurées, donc

$$\lambda = (2, 1, 1, 1, 2, 1, 1), \quad \delta = (0, 0, 1, 1, 0, 1, 0).$$

Notons que si au lieu de l'égalité $t_4 = t_5$, on avait eu l'inégalité stricte $t_4 < t_5$, λ' en aurait été affecté mais non λ et δ .

La longueur commune aux suites λ et δ est ici $l = 7$.

L'ensemble des suites d'observations $(t_1, t_2, \dots, t_9)$ conduisant au couple (λ, δ) considéré ici est :

$$\Delta^{\lambda, \delta} = \{(t_1, \dots, t_9); t_1 = t_2 < t_3 < t_4 \leq t_5 < t_6 = t_7 < t_8 < t_9\}$$

vérifiant en plus

$$\forall i \notin \{4, 5, 8\} t_i < \tau_i; \quad t_4 = \tau_4 = t_5 = \tau_5, t_8 = \tau_8.$$

L'ensemble des suites croissantes de valeurs prises par ces observations, regroupant des valeurs égales quand il s'agit d'observations non censurées, mais en ne les regroupant pas

dans le cas contraire est

$$\Delta^{7,\delta} = \{(x_1, \dots, x_7); x_1 < x_2 < x_3 \leq x_4 < x_5 < x_6 < x_7\}$$

vérifiant en plus

$$x_1 < \min(\tau_1, \tau_2), x_2 < \tau_3, x_5 < \min(\tau_6, \tau_7), x_7 < \tau_9; x_3 = \tau_4, x_4 = \tau_4, x_6 = \tau_8$$

les valeurs de censure se retrouvent ici aux emplacements (3, 4, 6) qui sont les indices des éléments égaux à 1 dans la suite des indicateurs de censure.

L'application Φ^λ consiste en la duplication de chacune des valeurs x_i associée à un $\lambda_i > 1$, de manière à obtenir la suite de λ_i valeurs toutes égales à x_i ; elle applique $\Delta^{7,\delta}$ sur $\Delta^{\lambda,\delta}$.

Revenons au cas général.

Soit une suite d'observations $(t_i)_{1 \leq i \leq n}$, pour laquelle

- les valeurs censurées sont aux emplacements de $\mathcal{C} = \{c_1, \dots, c_s\}$
- la suite des regroupements est $\lambda = (\lambda_1, \dots, \lambda_l)$
- la suite des indicateurs de censure des regroupements est $\delta = (\delta_1, \dots, \delta_l)$

Sur cette suite on peut effectuer deux opérations :

- (1) identifier les valeurs égales quand il s'agit d'observations non censurées; l'inverse de cette application est notée Φ^λ ; soit $(x_j)_{1 \leq j \leq l}$ l'image de $(t_i)_{1 \leq i \leq n}$ par cette application.
- (2) supprimer les valeurs censurées; cette application est notée $\Psi^{\mathcal{C}}$; soit $(u_1, \dots, u_{n-s})$ l'image de $(t_i)_{1 \leq i \leq n}$ par cette application; soit

$$\lambda_* = (\lambda_{*1}, \dots, \lambda_{*l-s}) \tag{5.2}$$

la suite des blocs d'ex æquo de $(u_1, \dots, u_{n-s})$; elle se déduit de $\lambda = (\lambda_1, \dots, \lambda_l)$ en en supprimant tous les 1 qui étaient associés à des δ_i égaux à 1 (valeurs censurées).

A $(x_j)_{1 \leq j \leq l}$, où les valeurs censurées restent présentes, on peut à son tour appliquer une opération de type (2) de suppression des valeurs censurées; soit θ^o cette application et

soit $(v_1, \dots, v_{l-s})$ l'image de $(x_j)_{1 \leq j \leq l}$.

D'autre part, à $(u_1, \dots, u_{n-s})$ on peut appliquer une opération de type (1) d'identification des valeurs égales ; l'inverse de cette application est Φ^{λ_*} .

Précisons ces opérations sur l'exemple 1 ci dessus.

L'opération d'identification des égalités entre les variables non censurées (passage de $\Delta^{\lambda, \delta}$ à $\Delta^{l, \delta}$) a été déjà effectuée ci-dessus par l'opération inverse de Φ^{λ} ,

à $(t_1, t_2, t_3, t_4, t_5, t_6, t_7, t_8, t_9)$ elle associe $(x_1, x_2, x_3, x_4, x_5, x_6, x_7)$, par ailleurs à $(t_1, t_2, t_3, t_4, t_5, t_6, t_7, t_8, t_9)$, on associe par l'opération Ψ^C (suppression des censures) le point de l'ensemble $\Delta_*^{\lambda, \delta}$ donné par :

$$(u_1, u_2, u_3, u_4, u_5, u_6) = (t_1, t_2, t_3, t_6, t_7, t_9).$$

Enfin au point $(x_1, x_2, x_3, x_4, x_5, x_6, x_7)$ de $\Delta^{l, \delta}$, on associe par θ^{δ} (qui est aussi une opération de suppression des censures) le point de $\Delta_*^{l, \delta}$ donné par :

$$(v_1, v_2, v_3, v_4) = (x_1, x_2, x_5, x_7).$$

Ici

$$\lambda = (2, 1, 1, 1, 2, 1, 1), \lambda_* = (2, 1, 2, 1)$$

On a donc le schéma suivant où toutes les applications sont bijectives.

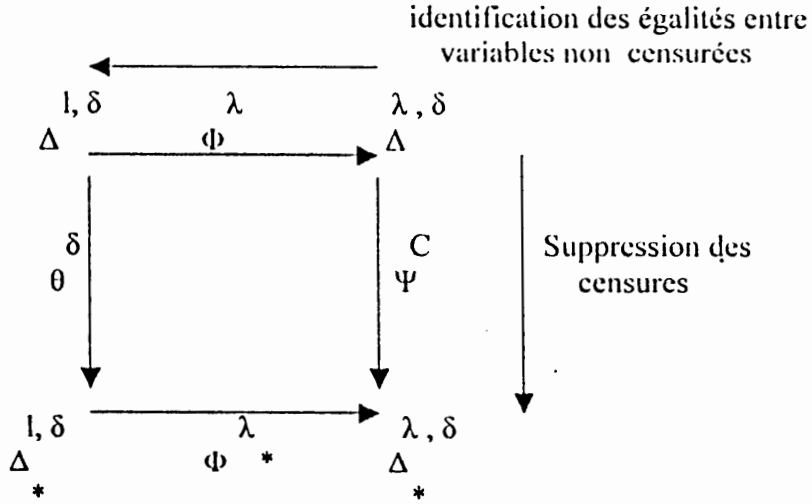


fig 8

En rappelant que $n_j = \sum_{i=1}^j \lambda_i$ nous introduisons, dans le cas général, les ensembles suivants :

- $\Delta^{\lambda, \delta}$ qui est l'ensemble des suites croissantes $(t_1, \dots, t_n)$ telles que :

- (i) pour tout j ($1 \leq j \leq l$) $t_{n_{j-1}+1} = \dots = t_{n_j}$,
- (ii) pour tout j ($1 \leq j \leq l-1$) tel que $(\delta_j, \delta_{j+1}) \neq (1, 1)$ $t_{n_j} < t_{n_{j+1}}$,
- (iii) pour tout j ($1 \leq j \leq l$) et tout i ($n_{j-1} < i \leq n_j$)
 - si $\delta_j = 0$ alors $t_i < \tau_i$
 - si $\delta_j = 1$ alors $t_i = \tau_i$.

- $\Delta^{l, \delta}$ qui est l'ensemble des suites croissantes $(x_1, \dots, x_l)$ telles que :

- (i') pour tout j ($1 \leq j \leq l-1$), si $(\delta_j, \delta_{j+1}) \neq (1, 1)$ $x_j < x_{j+1}$.
- (i'') pour tout j ($1 \leq j \leq l$) et tout i ($n_{j-1} < i \leq n_j$)
 - si $\delta_j = 0$ alors $x_i < \tau_i$
 - si $\delta_j = 1$ alors $x_i = \tau_i$.

Par définition, ϕ^λ est une bijection de $\Delta^{l, \delta}$ sur $\Delta^{\lambda, \delta}$ (rappelons que l'inverse de ϕ^λ associe à toute suite de $\Delta^{\lambda, \delta}$ la suite de $\Delta^{l, \delta}$ qui s'en déduit en regroupant les valeurs égales

quand il s'agit d'observations non censurées, mais en ne les regroupant pas dans le cas contraire). La propriété des suites appartenant à $\Delta^{\lambda,\delta}$, selon laquelle les t_i sont égaux à des valeurs de censure si et seulement si $i \in \mathcal{C}$ se traduit pour les suites appartenant à $\Delta^{l,\delta}$, par le fait que x_j est égale à une valeur de censure si et seulement si $\delta_j = 1$. On note par

$$\mathcal{C}_* = \{p_j, 1 \leq j \leq l - s\}$$

le complémentaire de l'ensemble de ces indices dans $\{1, \dots, l\}$, c'est à dire que p_j désigne le $j^{\text{ième}}$ bloc des variables non censurées et vérifie $\delta_{p_j} = 0$.

Remarquons que

- la relation (iii) signifie que l'ensemble $\Delta^{\lambda,\delta}$ est contenu dans la face $\mathcal{F}_{\mathcal{C}}$ (c'est à dire qu'on a $t_i = \tau_i$ si et seulement si $i \in \mathcal{C}$).
- par rapport à Δ^{λ} (défini dans les rappels), $\Delta^{\lambda,\delta}$ peut comporter plus d'égalités dues à d'éventuelles égalités entre des variables censurées se trouvant dans des blocs distincts.
- en revenant à la configuration d'ex æquo λ_* définie en (5.2), on a

$$\forall j \in \{1, \dots, l - s\} \quad \lambda_{*,j} = \lambda_{p_j}, \quad (5.3)$$

ceci provenant du fait que le $j^{\text{ième}}$ bloc après suppression des censures n'est que le $p_j^{\text{ième}}$ bloc dans la suite de départ.

Notons dans le cas général, comme pour l'exemple précédent, $\Delta_*^{\lambda,\delta}$ l'image de l'ensemble $\Delta^{\lambda,\delta}$ par l'application $\Psi^{\mathcal{C}}$ (suppression des censures) et $\Delta_*^{l,\delta}$ l'image de l'ensemble $\Delta^{l,\delta}$ par l'application θ^{δ} (qui est aussi une opération de suppression des censures). En fait $\Delta_*^{\lambda,\delta}$ n'est autre que l'image de $\Delta_*^{l,\delta}$ par l'application Φ^{λ_*} .

Considérons maintenant les probabilités déduites par les applications bijectives Φ^{λ} , $\Psi^{\mathcal{C}}, \Phi^{\lambda_*}$ et θ^{δ} de la probabilité prédictive étudiée sur $\Delta^{\lambda,\delta}$. Les fonctions de survie des

probabilités sur $\Delta_{\star}^{\lambda,\delta}$ et $\Delta_{\star}^{l,\delta}$ se déduisent respectivement de celles de $\Delta^{\lambda,\delta}$ et $\Delta^{l,\delta}$ en y fixant aux valeurs de censure les variables correspondant à ces censures (c'est à dire les $i \in \mathcal{C}$ pour $\Delta^{\lambda,\delta}$ et les j tels que $\delta_j = 1$ pour $\Delta^{l,\delta}$). On se trouve alors dans l'ouvert caractérisé par des inégalités strictes de majoration pour les variables non censurées, exactement dans la situation étudiée dans [15] en l'absence de censure.

Si on note $S_{\star}$ la fonction de survie du vecteur $(T_i)_{i \notin \mathcal{C}}$ et $g_{\star}^{\lambda}$ la densité sur $\Delta_{\star}^{\lambda,\delta}$ (qui est inclu dans $\Delta^{\lambda,\delta}$), on obtient par application de la formule (1.8) du théorème 1 :

$$g_{\star}^{\lambda} = \sum_{a \in A^{\lambda_{\star}}} (-1)^{q(a)} \partial_{s_{\star}(a)} S_{\star \sigma_{\star}(a)} \quad (5.4)$$

où en posant $k = l - s$ et $a = ((a_m^j)_{1 \leq m \leq \lambda_{\star,j}-1})_{1 \leq j \leq k}$

- $A^{\lambda_{\star}} = \{0, 1\}^{n-l}$
- $q(a) = n - s - \sum_{j=1}^k \sum_{m=1}^{\lambda_{\star,j}-1} a_m^j$,
- $s_{\star}(a) = (s_{\star,j})_{1 \leq j \leq k}$, avec $s_{\star,j} = 1 + n_{\star,j-1} + \sum_{m=1}^{\lambda_{\star,j}-1} a_m^j$ où $n_{\star,j} = \sum_{i=1}^j \lambda_{\star,i}$
- $(\sigma_{\star}(a))^{-1} = ((\pi_{\star,\nu}^j)_{1 \leq \nu \leq \lambda_{\star,j}})_{1 \leq j \leq k}$
avec $\pi_{\star,\nu}^j = n_{\star,j-1} + 1 + \sum_{u=1}^{\nu-1} a_u^j + (1 - a_{\nu}^j)(\lambda_{\star,j} - \nu)$.

Si $g^{\lambda,\delta}$ est la densité de la mesure prédictive sur $\Delta^{\lambda,\delta}$ alors

$$g^{\lambda,\delta} = g_{\star}^{\lambda} \circ \Psi^{\mathcal{C}}. \quad (5.5)$$

Nous allons expliciter cette densité en utilisant, d'une part le lien entre les deux configurations d'ex æquo $\lambda_{\star}$ et λ et d'autre part le lien entre la fonction de survie sur $\Delta_{\star}^{\lambda,\delta}$ et la fonction de survie sur $\Delta^{\lambda,\delta}$. On obtient, en rappelant que $\lambda_{\star,j} = \lambda_{p_j}$ (voir (5.3)),

$$q(a) = n - s - \sum_{j=1}^l \sum_{m=1}^{\lambda_j-1} a_m^j \quad (5.6)$$

puisque $\lambda_j = 1$ si $j \notin C_*$,

$$s_{*j} = 1 + \sum_{i=1}^{j-1} \lambda_{p_i} + \sum_{m=1}^{\lambda_{p_j}-1} a_m^j, \quad (5.7)$$

$$\pi_{* \nu}^j = \sum_{i=1}^{j-1} \lambda_{p_i} + 1 + \sum_{u=1}^{\nu-1} a_u^j + (1 - a_\nu^j)(\lambda_{p_j} - \nu). \quad (5.8)$$

Par ailleurs la fonction de survie sur $\Delta^{\lambda, \delta}$ se déduit de celle sur $\Delta_*^{\lambda, \delta}$, en rajoutant, à leurs emplacements, les variables censurées, autrement dit :

$$\partial_{s_*(a)} S_{*_{\sigma_*(a)}}(t_i)_{i \in C} = \partial_{s(a)} S_{\sigma(a)}(t_i)_{1 \leq i \leq n} \quad (5.9)$$

où $s(a)$ et $\sigma(a)$ sont définis de la manière suivante.

- $s(a)$ s'obtient de $s_*(a)$ en tenant compte du décalage subi par les indices de dérivation après suppression des censures.
- $\sigma(a)$ est une permutation sur l'ensemble $\{1, \dots, n\}$ qui s'obtient à partir de $\sigma_*(a)$ en tenant compte d'abord du décalage des indices de dérivation dû à la suppression des censures, et ensuite en y intercalant les indices de censure à leurs positions de départ.

En fait on a le schéma suivant :

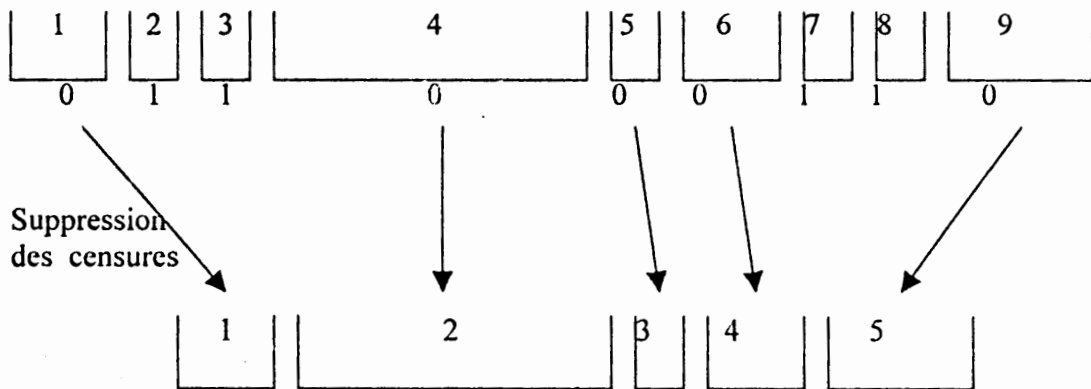


fig 9

où les chiffres à l'intérieur des blocs correspondent à leurs numéros et les autres sont respectivement leurs indicateurs de censure.

Ici on a

$$p_1 = 1, p_2 = 4, p_3 = 5, p_4 = 6, p_5 = 9$$

Il est clair que le bloc numéro j , après suppression des censures, provient en fait du bloc numéro p_j dans la suite de départ.

En posant $s(a) = (s_j)_{1 \leq j \leq k}$ et $n_j = \sum_{i=1}^j \lambda_i$, il devient clair que s_j est le "correspondant" de s_{*j} qui se trouve, dans la suite de départ, dans le bloc p_j , autrement dit :

$$s_j = 1 + n_{p_j-1} + \sum_{m=1}^{\lambda_{p_j}-1} a_m^{p_j}. \quad (5.10)$$

D'une part, nous pouvons faire remarquer que

$$n_{p_j-1} < s_j \leq n_{p_j}, \quad (5.11)$$

et d'autre part que l'utilisation de la formule (5.7) nous conduit au même résultat, en effet :

il est clair que s_j est la somme de s_{*j} et du nombre de censures dont les indices lui sont inférieurs; or ce nombre s'écrit $\sum_{i \in \mathcal{C}, i < p_j} \lambda_i$. Mais

$$\sum_{i=1}^{j-1} \lambda_{p_i} + \sum_{i \in \mathcal{C}, i < p_j} \lambda_i = n_{p_j-1},$$

ce qui redonne la formule (5.10).

Par ailleurs, en posant $(\sigma(a))^{-1} = ((\pi_\nu^j)_{1 \leq \nu \leq \lambda_j})_{1 \leq j \leq l}$, nous avons les propriétés suivantes - du fait que les variables de censure sont remises à leurs positions initiales et forment

des blocs d'une seule valeur, on a

$$\forall j \delta_j = 1, \pi_1^j = n_{j-1} + 1 \quad (5.12)$$

- dans un bloc p_j de variables censurées, la permutation opère comme sa correspondante dans le bloc j (après suppression des censures) en lui rajoutant le nombre de censures se trouvant avant ce bloc p_j , donc

$$\forall 1 \leq j \leq k \pi_\nu^{p_j} = \pi_{\star_\nu}^j + \sum_{i \in \mathcal{C}, i < p_j} \lambda_i,$$

ce qui s'écrit

$$\pi_\nu^{p_j} = n_{p_j-1} + 1 + \sum_{u=1}^{\nu-1} a_u^j + (1 - a_\nu^j)(\lambda_{p_j} - \nu). \quad (5.13)$$

On remarque enfin que, à cause des cas où $\lambda_j = 1$ et donc $\nu = 1$, les deux formules (5.12) et (5.13) s'écrivent sous la même forme suivante

$$\forall 1 \leq j \leq l, \forall 1 \leq \nu \leq \lambda_j \pi_\nu^j = n_{j-1} + 1 + \sum_{u=1}^{\nu-1} a_u^j + (1 - a_\nu^j)(\lambda_j - \nu). \quad (5.14)$$

Par utilisation des formules (5.4), (5.5), (5.6), (5.9), (5.10) et (5.14), nous arrivons au résultat suivant en faisant remarquer que $\Delta^{\lambda, \delta}$ est en bijection avec un sous ensemble de $\mathbb{R}^{l-s}$ par l'application : $(t_i)_{1 \leq i \leq n} \mapsto (t_{n_j})_{j/\delta_j=0}$, ce qui permet de définir la mesure de Lebesgue sur $\Delta^{\lambda, \delta}$ par transport de la mesure de Lebesgue sur $\mathbb{R}^{l-s}$.

Théorème 4 Soit μ une mesure régulière par portions sur $\mathbb{R}^n$, de fonction de survie S , à laquelle on applique n censures (à droite) fixes $\tau_1, \dots, \tau_n$. Soient s le nombre de valeurs censurées, $\lambda = (\lambda_i)_{1 \leq i \leq l}$ la suite de regroupements définie en (5.1) et $\delta = (\delta_i)_{1 \leq i \leq l}$ la suite d'indicateurs de censure qui lui est associée. Notons $A^\lambda = \{0, 1\}^{n-l}$.

Alors la restriction de μ à $\Delta^{\lambda, \delta}$ admet pour densité, par rapport à la mesure de Lebesgue, la fonction :

$$g^{\lambda, \delta} = \sum_{a \in A^\lambda} (-1)^{q(a)} \partial_{s(a)} S_{\sigma(a)}, \quad (5.15)$$

où, si on note $a = ((a_m^j)_{1 \leq m \leq \lambda_j - 1})_{1 \leq j \leq l}$, on obtient avec $n_j = \sum_{h=1}^j \lambda_h$:

- $q(a) = n - \sum_{j=1}^l \sum_{m=1}^{\lambda_j - 1} a_m^j - s$,
- $s(a) = (s_j)_{1 \leq j \leq k}$, $k = l - s$
avec $s_j = 1 + n_{p_j - 1} + \sum_{m=1}^{\lambda_{p_j} - 1} a_m^{p_j}$ où p_j est le $j^{\text{ième}}$ indice tel que $\delta_{p_j} = 0$
- $(\sigma(a))^{-1} = ((\pi_\nu^j)_{1 \leq \nu \leq \lambda_j})_{1 \leq j \leq l}$
avec $\pi_\nu^j = n_{j-1} + 1 + \sum_{u=1}^{\nu-1} a_u^j + (1 - a_\nu^j)(\lambda_j - \nu)$.

Remarquons que ce résultat constitue une généralisation du théorème 1 de Gouget-Raoult que nous retrouvons dans le cas où $\forall 1 \leq j \leq l \delta_j = 0$, ce qui veut dire que $s = 0$ (il n'y a aucune variable censurée).

Exemple 2

Soit la suite $(t_1, \dots, t_9)$ telle que

$$t_1 < t_2 < t_3 = t_4 = t_5 < t_6 = t_7 < t_8 = t_9,$$

où l'ensemble des indices de censure est $\mathcal{C} = \{1, 2, 6, 7\}$.

Cette suite est représentée dans le schéma suivant où les observations se trouvent sur l'axe et les censures correspondant aux variables non censurées sont données au dessus de cet axe.

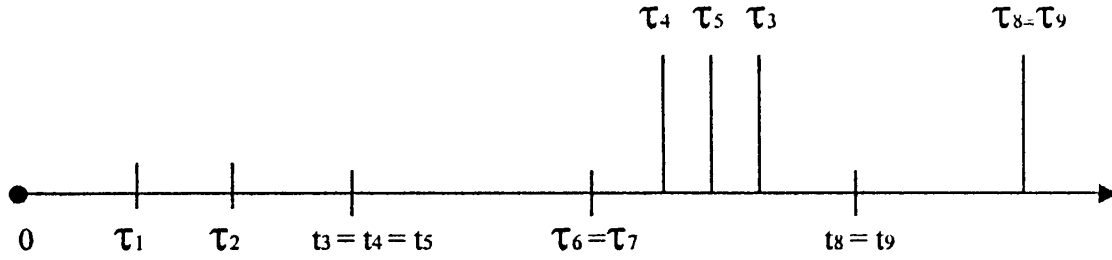


fig 10

- Pour $p = 1$, c'est à dire jusqu'à la première censure, qui intervient en τ_1 , nous avons la suite réduite à un élément $S_1 = (\tau_1)$ à laquelle est associée la configuration d'ex æquo $\lambda^1 = (1)$ et la suite d'indicateurs de censure $\delta^1 = (1)$.

- Pour $p = 2$, c'est à dire au delà de la première censure et jusqu'à la deuxième censure, qui interviennent respectivement en τ_1 et τ_2 ne se trouve aucune observation non censurée, nous obtenons la suite réduite à un élément $S_2 = (\tau_2)$ à laquelle est associée la configuration d'ex æquo $\lambda^2 = (1)$ et la suite d'indicateurs de censure $\delta^2 = (1)$.

- Pour $p = 3$, c'est à dire au delà de la deuxième censure et jusqu'à la troisième, qui interviennent respectivement en τ_2 et τ_6 , se trouvent trois observations non censurées et égales, nous obtenons la suite $S_3 = (t_3, t_4, t_5, \tau_6)$ à laquelle est associée la configuration d'ex æquo $\lambda^3 = (3, 1)$ et la suite d'indicateurs de censure $\delta^3 = (0, 1)$.

- Pour $p = 4$, c'est à dire au delà de la troisième censure et jusqu'à la quatrième, qui interviennent respectivement en des valeurs égales τ_6 et τ_7 , ne peut se trouver aucune observation non censurée, nous obtenons la suite $S_4 = (\tau_7)$ à laquelle est associée la configuration d'ex æquo $\lambda^4 = (1)$ et la suite d'indicateurs de censure $\delta^4 = (1)$.

- Pour $p = 5$, c'est à dire au delà de la dernière censure, qui intervient en τ_7 se trouvent deux observations non censurées et égales, nous obtenons la suite $S_5 = (t_8, t_9)$ à laquelle est associée la configuration d'ex æquo $\lambda^5 = (2)$ et la suite d'indicateurs de censure $\delta^5 = (0)$.

Il vient alors :

$\lambda' = (1, 1, 3, 2, 2)$, $\lambda = (1, 1, 3, 1, 1, 2)$, $\delta = (1, 1, 0, 1, 1, 0)$ et $\lambda_* = (3, 2)$.

$$\Delta^{\lambda,\delta} = \{(t_1, \dots, t_9), t_1 < t_2 < t_3 = t_4 = t_5 < t_6 = t_7 < t_8 = t_9\},$$

vérifiant

$$\forall i \notin \{1, 2, 6, 7\} \ t_i < \tau_i, \forall i \in \{1, 2, 6, 7\} \ t_i = \tau_i$$

par l'application inverse de Φ^λ , $\Delta^{\lambda,\delta}$ est envoyé dans

$$\Delta^{6,\delta} = \{(x_1, \dots, x_6), x_1 < x_2 < x_3 < x_4 \leq x_5 < x_6\}$$

vérifiant

$$x_1 = \tau_1, x_2 = \tau_2, x_4 = \tau_6 = x_5 = \tau_7, x_3 < \min(\tau_3, \tau_4, \tau_5), x_6 < \min(\tau_8, \tau_9)$$

Par l'application Ψ^c , $\Delta^{\lambda,\delta}$ est envoyé dans

$$\Delta_*^{\lambda,\delta} = \{(u_1, \dots, u_5), u_1 = t_3, u_2 = t_4, u_3 = t_5, u_4 = t_8, u_5 = t_9\}.$$

Et par l'application inverse de Φ_*^λ , $\Delta_*^{\lambda,\delta}$ est envoyé dans

$$\Delta_*^{6,\delta} = \{(v_1, v_2), v_1 = t_3, v_2 = t_8\}.$$

L'application du théorème précédent conduit à l'expression suivante de la densité restreinte à $\Delta^{\lambda,\delta}$ (voir l'appendice B pour les détails des calculs).

$$\begin{aligned} g^{\lambda,\delta}(\tau_1, \tau_2, x_3, x_3, x_3, \tau_6, \tau_7, x_6, x_6) = & (-\partial_{38} S_{125436798} + \partial_{39} S_{125436789} + \partial_{48} S_{124536789} + \\ & \partial_{48} S_{123546798} - \partial_{49} S_{124536789} - \partial_{58} S_{123456798} - \partial_{49} S_{123546789} \\ & + \partial_{59} S_{123456789})(\tau_1, \tau_2, x_3, x_3, x_3, \tau_6, \tau_7, x_6, x_6). \end{aligned}$$

On remarque que les indices des variables censurées, qui appartiennent ici à $\{1, 2, 6, 7\}$,

gardent toujours leurs positions pour la permutation σ .

5.3 Retour au modèle de Cox-Dirichlet

Revenons au cas où la mesure μ est la mesure prédictive dans le modèle de Cox-Dirichlet. La formule (1.4) montre que pour les permutations σ telles que $\Delta_\sigma \cap \mathcal{F}_C$ n'est pas vide, la survie de la mesure prédictive restreinte à $\Delta_\sigma \cap \mathcal{F}_C$, est pour $t = (t_1, \dots, t_n)$:

$$S_\sigma(t) = \frac{\Gamma(\alpha(0))}{\Gamma(\alpha(0) + d)} \prod_{i=1}^n \frac{\Gamma(\alpha(t_i) + d_{(\sigma^{-1})_i}^\sigma)}{\Gamma(\alpha(t_i) + d_{(\sigma^{-1})_i+1}^\sigma)}$$

et la densité par rapport aux variables non censurées s'écrit :

$$(-1)^{n-s} \partial_{r_1, \dots, r_{n-s}} S_\sigma(t) = (-1)^{n-s} S_\sigma(t) \prod_{j=1}^{n-s} \alpha'(t_{r_j}) [(Ln\Gamma)'(\alpha(t_{r_j}) + d_{(\sigma^{-1})_{r_j}}^\sigma) - (Ln\Gamma)'(\alpha(t_{r_j}) + d_{(\sigma^{-1})_{r_j}+1}^\sigma)]$$

où $r_j \in \{1, \dots, n\} \setminus C \ \forall j \in \{1, \dots, n-s\}$.

Nous allons maintenant étudier les parties de $\mathcal{F}_C$ présentant des ex æquo, d'une part entre les valeurs des variables non censurées et (ou) d'autre part entre des variables censurées. Afin d'appliquer la formule (1.8) du théorème 1, nous devons expliciter les termes $\partial_s(S_c)_\sigma$. Par application des formules (1.5) et (1.6) et en tenant compte du fait que $n_{p_j-1} < s_j \leq n_{p_j}$ (voir (5.11)), il vient :

$$\begin{aligned} \partial_s(S_c)_\sigma(x_i^{[\lambda_i]})_{1 \leq i \leq l} &= \frac{\Gamma(\alpha(0))}{\Gamma(\alpha(0) + d)} \prod_{j=1}^l \frac{\Gamma(\alpha(x_j) + e_j)}{\Gamma(\alpha(x_j) + e_{j+1})} \\ &\quad \prod_{j=1}^k \alpha'(x_{p_j}) [(Ln\Gamma)'(\alpha(x_{p_j}) + d_{s_j}^\sigma) - (Ln\Gamma)'(\alpha(x_{p_j}) + d_{s_j+1}^\sigma)]. \end{aligned}$$

$\sigma(a)$ étant une permutation ne modifiant que des indices d'un même bloc entre eux, e_j peut s'écrire

$$\forall j \quad e_j = \sum_{i=n_{j-1}+1}^n c_i,$$

ce qui montre que e_j est indépendant de σ . Rappelons que, par convention, $e_{l+1} = 0$.

Par ailleurs nous avons :

$$d_{s_j}^\sigma = b_j(a) + e_{p_j+1}$$

où

$$b_j(a) = \sum_{h=s_j}^{n_{p_j}} c_{\sigma_h}$$

et

$$d_{s_j+1}^\sigma = b'_j(a) + e_{p_j+1}$$

où

$$b'_j(a) = \sum_{h=s_j+1}^{n_{p_j}} c_{\sigma_h}$$

$b_j(a)$ et $b'_j(a)$ ne dépendant que de quelques contraintes d'indices appartenant au bloc numéro p_j , nous allons décomposer la suite de contraintes $c = (c_i)_{1 \leq i \leq n}$ en l sous suites successives (c^j) selon la configuration d'ex æquo λ . De même $\sigma(a)$ est décomposée en l permutations $\sigma(a^j)$ où $\sigma(a^j)$ est la permutation sur l'ensemble $\{1, \dots, \lambda_j\}$, impact de $\sigma(a)$ sur $\{n_{j-1} + 1, \dots, n_j\}$.

Nous notons :

$$c = (c^j)_{1 \leq j \leq l}, \quad c^j = (c_w^j)_{1 \leq w \leq \lambda_j} \text{ où } c_w^j = c_{n_{j-1}+w},$$

$$\forall w \in \{1, \dots, \lambda_j\}$$

$$(\sigma(a^j))_w = (\sigma(a))_{n_{j-1}+w} - n_{j-1},$$

(on se ramène à une permutation sur $\{1, \dots, \lambda_j\}$).

Il s'ensuit que :

$$e_j = \sum_{h=j}^l \sum_{w=1}^{\lambda_h} c_w^h$$

Nous obtenons alors :

$$b'_j(a) = \sum_{w=s_j+1-n_{p_j}-1}^{\lambda_{p_j}} c_{\sigma(a^{p_j})_w}^{p_j}$$

et

$$b_j(a) = \sum_{s_j-n_{p_j}-1}^{\lambda_{p_j}} c_{\sigma(a^{p_j})_w}^{p_j} = b'_j(a) + c_{\sigma(a^{p_j})_{s_j-n_{p_j}-1}}^{p_j}$$

$\pi(a)$ étant la permutation inverse de $\sigma(a)$, nous avons $(\sigma(a^{p_j}))^{-1} = \pi(a^{p_j})$ qui est l'impact de $\pi(a)$ sur $\{n_{p_j-1} + 1, \dots, n_{p_j}\}$. Il s'ensuit que $b'_j(a)$ est la somme de tous $c_{\nu_j}^{p_j}$ pour ν_j vérifiant $(\pi(a^{p_j}))_{\nu_j} > s_j - n_{p_j-1}$ et $b_j(a)$ est la somme de $b'_j(a)$ et de $c_{\nu_j}^{p_j}$ avec ν_j vérifiant $(\pi(a^{p_j}))_{\nu_j} = s_j - n_{p_j-1}$. Puisque $\pi(a^{p_j})_{\nu_j} = \pi_{\nu_j}^{p_j} - n_{p_j-1}$, il vient du théorème 4 :

$$\pi(a^{p_j})_\nu = 1 + \sum_{u=1}^{\nu-1} a_u^{p_j} + (1 - a_\nu^{p_j})(\lambda_{p_j} - \nu)$$

et

$$s_j - n_{p_j-1} = 1 + \sum_{m=1}^{\lambda_{p_j}-1} a_m^{p_j}.$$

Pour la comparaison de $\pi(a^{p_j})_\nu$ et de $s_j - n_{p_j-1}$, nous reprenons la même démarche donnée dans [15] en remplaçant, quand il le faut, j par p_j . Nous distinguons trois cas :

- $\nu = \lambda_{p_j}$ alors

$$\pi(a^{p_j})_\nu = 1 + \sum_{u=1}^{\nu-1} a_u^{p_j} = s_j - n_{p_j-1},$$

l'égalité voulant dire que $\nu_j = \lambda_{p_j}$.

$$\begin{aligned} & - \nu < \lambda_{p_j} \text{ et } a_\nu^{p_j} = 0 \text{ alors } \pi(a^{p_j})_\nu - (s_j - n_{p_j-1}) = 1 + \sum_{u=1}^{\nu-1} a_u^{p_j} + \lambda_{p_j} - \nu - (1 + \sum_{m=1}^{\lambda_{p_j}-1} a_m^{p_j}) \\ & = \lambda_{p_j} - \nu - \sum_{m=\nu+1}^{\lambda_{p_j}-1} a_m^{p_j} > \lambda_{p_j} - \nu - (\lambda_{p_j} - 1 - \nu) > 0. \end{aligned}$$

- $\nu < \lambda_{p_j}$ et $a_\nu^{p_j} = 1$ alors

$$\pi(a^{p_j})_\nu - (s_j - n_{p_j-1}) = 1 + \sum_{u=1}^{\nu-1} a_u^{p_j} - (1 + \sum_{m=1}^{\lambda_{p_j}-1} a_m^{p_j}) = -1 - \sum_{m=\nu+1}^{\lambda_{p_j}-1} a_m^{p_j} < 0.$$

Notons, comme dans [15], $M(a^j) = \{\nu : 1 \leq \nu < \lambda_j, a_\nu^j = 0\}$ et $\mathcal{P}_j$ l'ensemble de tous les sous ensembles de $\{1, \dots, \lambda_j - 1\}$, il vient

$$b'_j(a) = \sum_{\nu \in M(a^{p_j})} c_\nu^{p_j}$$

et

$$b_j(a) = b'_j(a) + c_{\lambda_{p_j}}^{p_j}.$$

Par ailleurs les indices j pour lesquels $\delta_j = 1$ ayant un $\lambda_j = 1$, nous pouvons nous en passer dans $q(a)$, donné au théorème 4, et obtenir :

$$q(a) = k + \sum_{j=1}^l \sum_{m=1}^{\lambda_j-1} (1 - a_m^j) = k + \sum_{j=1}^k \text{Card} M(a^{p_j}).$$

Nous arrivons ainsi au résultat suivant par application du théorème 4 et de la formule de distributivité générale qui s'écrit dans notre cas :

$$\sum_{(a^1, \dots, a^l) \in A^\lambda} \beta \prod_{j=1}^l \delta(a^j) = \beta \prod_{j=1}^l \sum_{a^j \in A^{\lambda, j}} \delta(a^j),$$

où $A^\lambda = \prod_{j=1}^l \{0, 1\}^{\lambda_j-1}$ et $A^{\lambda, j} = \{0, 1\}^{\lambda_j-1}$.

Rappelons que les indices qui ne sont pas des p_j ont un λ_j égal à 1 et que dans ce cas $\{0, 1\}^{\lambda_j-1}$ est vide.

Théorème 5 Soit μ la loi prédictive de n durées de survie dans le modèle de Cox-Dirichlet censuré, à droite, par les constantes $\tau_1, \dots, \tau_n$, avec la suite de contraintes $c = (c_i)_{1 \leq i \leq n}$ et l'a priori de Dirichlet défini par la mesure α^* , de fonction de survie α , qui est continûment dérivable sur $\mathbb{R}_+$. Soit s le nombre de variables censurées, $\lambda = (\lambda_i)_{1 \leq i \leq l}$ la suite de regroupements définie en (5.1) et $\delta = (\delta_i)_{1 \leq i \leq l}$ la suite d'indicateurs de censure qui lui est associée. Nous décomposons la suite de contraintes c , selon la configuration d'ex æquo λ , en l sous suites successives (c_j^i) ; nous introduisons les sommes partielles supérieures de contraintes $e_j = \sum_{h=j}^l \sum_{\nu=1}^{\lambda_h} c_\nu^h$.

Alors la restriction de μ aux parties non vides $\Delta^{\lambda,\delta}$ admet pour densité, la fonction :

$$g^{\lambda,\delta}(x_1^{[\lambda_1]}, \dots, x_l^{[\lambda_l]}) = (-1)^k \frac{\Gamma(\alpha(0))}{\Gamma(\alpha(o) + d)} \prod_{j=1}^l B_j(x_j) \prod_{j=1}^k C_j(x_{p_j})$$

où $k = l - s$,

$$B_j(x_j) = \frac{\Gamma(\alpha(x_j) + e_j)}{\Gamma(\alpha(x_j) + e_{j+1})}$$

et, $\mathcal{P}_{\lambda_j-1}$ désignant l'ensemble de tous les sous ensembles de $\{1, \dots, \lambda_j - 1\}$,

$$C_j(x_{p_j}) = \alpha'(x_{p_j}) \sum_{M \in \mathcal{P}_{\lambda_{p_j}-1}} (-1)^{\text{Card} M} [(Ln\Gamma)'(\alpha(x_{p_j}) + e_{p_j+1} + \sum_{\nu \in M} c_{\lambda_{p_j}}^{p_j} + \sum_{\nu \in M} c_{\nu}^{p_j}) - (Ln\Gamma)'(\alpha(x_{p_j}) + e_{p_j+1} + \sum_{\nu \in M} c_{\nu}^{p_j})],$$

où p_j est le $j^{\text{ième}}$ indice tel que $\delta_{p_j} = 0$.

Remarquons que ce résultat constitue (tout comme le théorème 3) une généralisation du théorème 2 de Gouget-Raoult que nous retrouvons dans le cas où $\forall 1 \leq j \leq l \delta_j = 0$, ce qui veut dire que $s = 0$ (il n'y a aucune variable censurée).

Le théorème ci dessus fournit l'expression des densités pour les suites croissantes des observations, en mettant en évidence le rôle des ex æquo. Le corollaire suivant étend ce résultat aux suites non croissantes en les rangeant par ordre croissant par usage d'une permutation σ et l'introduction de la suite des regroupements de (t_{σ_j}) et la suite d'indicateurs qui lui est associée δ .

Soit donc $\Delta_{\sigma}^{\lambda,\delta}$ l'ensemble des suites $(t_1, \dots, t_n)$ telle que

- (i) (t_{σ_j}) est une suite croissante,
- (ii) pour tout j ($1 \leq j \leq l$) $t_{\sigma_{n_{j-1}+1}} = \dots = t_{\sigma_{n_j}}$,
- iii) pour tout j ($1 \leq j \leq l-1$) tel que $(\delta_j, \delta_{j+1}) \neq (1, 1)$ $t_{\sigma_{n_j}} < t_{\sigma_{n_{j+1}}}$,
- (vi) pour tout j ($1 \leq j \leq l$) et tout i ($n_{j-1} < i \leq n_j$)

si $\delta_j = 0$ alors $t_{\sigma_i} < \tau_{\sigma_i}$

si $\delta_j = 1$ alors $t_{\sigma_i} = \tau_{\sigma_i}$

Nous rappelons que les censures ont pour effet que certaines permutations entre les variables ne sont plus possibles. La densité $g_{\sigma}^{\lambda, \delta}$ sur les parties non vides $\Delta_{\sigma}^{\lambda, \delta}$ s'obtient en appliquant le théorème 5 à la probabilité déduite de la loi prédictive par réarrangement croissant des coordonnées, autrement dit à la fonction de survie :

$$(t_1, \dots, t_n) \rightarrow S_c(t_{(\sigma^{-1})_1}, \dots, t_{(\sigma^{-1})_n})$$

Corollaire 2 *La restriction de μ à toute partie non vide $\Delta_{\sigma}^{\lambda, \delta}$ admet pour densité la fonction :*

$$g_{\sigma}^{\lambda}(t_1, \dots, t_n) = (-1)^k \frac{\Gamma(\alpha(0))}{\Gamma(\alpha(o) + d)} \prod_{j=1}^l B_j(x_j) \prod_{j=1}^k C_j(x_{p_j})$$

où x_j est la valeur commune des observations t_i appartenant au $j^{\text{ième}}$ bloc d'ex æquo une fois rangées en ordre croissant, et $B_j(x_j)$ et $C_j(x_{p_j})$ sont définis comme au théorème 5 avec pour tout j les contraintes c_j^i sont celles relatives aux observations appartenant au $j^{\text{ième}}$ bloc d'ex æquo une fois rangées en ordre croissant et λ est la suite de regroupements de la suite croissante (t_{σ_j}) .

5.4 Extension

Le travail effectué dans ce chapitre porte sur des censures fixes à droite. L'extension à laquelle nous allons nous intéresser concerne toujours des censures fixes mais à gauche et en mixte.

Nous nous limiterons à donner des indications pour le traitement de ces deux cas puisque l'idée principale, qui est la suppression des censures, a été explicitée pour les

censures à droite.

1) censure à gauche

Après l'introduction des bonnes définitions et notations " calquées " de celles données pour les censures à droite mais adaptées à ce cas, nous supprimons les censures pour pouvoir appliquer le théorème 1. Le problème ici est que le rapport entre les fonctions de survie sans censure et celles avec censure n'est pas aussi simple. Nous utilisons alors la formule de Poincaré pour exprimer les fonctions de survie données par le théorème 1 en fonctions des différentes fonctions de répartitions des vecteurs de dimensions inférieures ou égale. Puis nous remplaçons ces dernières par leurs homologues avec censures puisqu'elles s'en déduisent en y fixant aux valeurs de censures les variables correspondant à ces censures.

2) Censure mixte

Ici aussi, il faut commencer par introduire les bonnes définitions et notations relatives à ce cas. Ensuite par suppression des censures à gauche, nous nous ramenons au cas où n'interviennent que des censures à droite pour lequel la densité est donnée par le théorème 4. Puis le cheminement est le même que pour le cas précédent. C'est à dire que nous utilisons la formule de Poincaré pour se ramener des fonctions de survie aux fonctions de répartitions qui s'exprimeront alors en fonction des fonctions de répartition qui font intervenir les variables censurées.

.1 Appendice A

Calcul de la densité prédictive sur la diagonale, dans le cas de la censure à gauche, pour $n=2$.

Pour simplifier les notations, nous notons (G, H) au lieu de (G_1, G_2) , (c_i, d_i) au lieu de $(c_{i,1}, c_{i,2})$ et (α, β) au lieu de (α_1, α_2) . D'après le lemme 1, cette densité s'écrit :

$\partial S^*(x) = (\partial_2 S' - \partial_1 S'')(x, x)$ où pour $x \leq y$ $S'(x, y) = S_c(x, y)$ et $S''(x, y) = S_c(y, x)$, donc

$$\begin{aligned} S'(x, y) &= \int G^{c_1}(x) G^{c_2}(y) \mathcal{D}_{\alpha^*}(dG) + \int G^{c_1}(x) \mathcal{D}_{\alpha^*}(dG) \int H^{d_2}(y) \mathcal{D}_{\beta^*}(dH) \\ &- \int G^{c_1}(x) G^{c_2}(y) \mathcal{D}_{\alpha^*}(dG) \int H^{d_2}(y) \mathcal{D}_{\beta^*}(dH) + \int H^{d_1}(y) \mathcal{D}_{\beta^*}(dH) \int G^{c_2}(y) \mathcal{D}_{\alpha^*}(dG) \\ &+ \int H^{d_1}(x) H^{d_2}(y) \mathcal{D}_{\beta^*}(dH) - \int H^{d_1}(x) H^{d_2}(y) \mathcal{D}_{\beta^*}(dH) \int G^{c_2}(y) \mathcal{D}_{\alpha^*}(dG) \\ &- \int G^{c_1}(x) G^{c_2}(y) \mathcal{D}_{\alpha^*}(dG) \int H^{d_1}(x) \mathcal{D}_{\beta^*}(dH) - \int G^{c_1}(x) \mathcal{D}_{\alpha^*}(dG) \int H^{d_1}(x) H^{d_2}(y) \mathcal{D}_{\beta^*}(dH) \\ &+ \int G^{c_1}(x) G^{c_2}(y) \mathcal{D}_{\alpha^*}(dG) \int H^{d_1}(x) H^{d_2}(y) \mathcal{D}_{\beta^*}(dH) \\ &= \sum_{j=1}^9 S'_j(x, y) \\ \text{et,} \end{aligned}$$

$$\begin{aligned} S''(x, y) &= \int G^{c_1}(y) G^{c_2}(x) \mathcal{D}_{\alpha^*}(dG) + \int G^{c_1}(y) \mathcal{D}_{\alpha^*}(dG) \int H^{d_2}(x) \mathcal{D}_{\beta^*}(dH) \\ &- \int G^{c_1}(y) G^{c_2}(x) \mathcal{D}_{\alpha^*}(dG) \int H^{d_2}(x) \mathcal{D}_{\beta^*}(dH) + \int H^{d_1}(y) \mathcal{D}_{\beta^*}(dH) \int G^{c_2}(x) \mathcal{D}_{\alpha^*}(dG) \\ &+ \int H^{d_1}(y) H^{d_2}(x) \mathcal{D}_{\beta^*}(dH) - \int H^{d_1}(y) H^{d_2}(x) \mathcal{D}_{\beta^*}(dH) \int G^{c_2}(x) \mathcal{D}_{\alpha^*}(dG) \\ &- \int G^{c_1}(y) G^{c_2}(x) \mathcal{D}_{\alpha^*}(dG) \int H^{d_1}(y) \mathcal{D}_{\beta^*}(dH) - \int G^{c_1}(y) \mathcal{D}_{\alpha^*}(dG) \int H^{d_1}(y) H^{d_2}(x) \mathcal{D}_{\beta^*}(dH) \\ &+ \int G^{c_1}(y) G^{c_2}(x) \mathcal{D}_{\alpha^*}(dG) \int H^{d_1}(y) H^{d_2}(x) \mathcal{D}_{\beta^*}(dH) \\ &= \sum_{j=1}^9 S''_j(x, y) \end{aligned}$$

Nous remarquons que $S'_2(x, y) = f(x)g(y)$ et que $S''_2(x, y) = f(y)g(x)$ donc

$\partial_2 S'_2(x, x) = \partial_1 S''_2(x, x)$ et pour la même raison $\partial_2 S'_4(x, x) = \partial_1 S''_4(x, x)$.

Par ailleurs la formule de Ruggiero donne :

$$\int G^{c_1}(x) G^{c_2}(y) \mathcal{D}_{\alpha^*}(dG) = \frac{\Gamma(\alpha(0))}{\Gamma(\alpha(0) + c_1 + c_2)} \frac{\Gamma(\alpha(t_1) + c_1 + c_2)}{\Gamma(\alpha(t_1) + c_2)} \frac{\Gamma(\alpha(t_2) + c_2)}{\Gamma(\alpha(t_2))}$$

et

$$\int G^c(x) \mathcal{D}_{\alpha^*}(dG) = \frac{\Gamma(\alpha(0)) \Gamma(\alpha(x) + c)}{\Gamma(\alpha(x)) \Gamma(\alpha(0) + c)}$$

(bien sûr les mêmes formules sont valables pour H et β^* en remplaçant α par β)

Introduisons les notations suivantes :

$$c = c_1 + c_2 \text{ et } d = d_1 + d_2$$

$$a = \frac{\Gamma(\alpha(0))}{\Gamma(\alpha(0)+c)}, \quad b = \frac{\Gamma(\beta(0))}{\Gamma(\beta(0)+d)}, \quad A_h(t) = \Gamma(\alpha(t) + h) \text{ et } B_h(t) = \Gamma(\beta(t) + h)$$

Nous obtenons alors :

$$S'_1(x, y) = a \frac{A_c(x)}{A_{c_2}(x)} \frac{A_{c_2}(y)}{A_0(y)};$$

en remplaçant dans le second membre de cette égalité c_2 par c_1 , il vient :

$$S''_1(x, y) = a \frac{A_c(x)}{A_{c_1}(x)} \frac{A_{c_1}(y)}{A_0(y)}.$$

Par dérivation des deux termes précédents, nous obtenons :

$$(\partial_2 S'_1 - \partial_1 S''_1)(x, x) = a \frac{A_c(x)}{A_0(x)} [Ln A_{c_2}(x) - Ln A_c(x) - Ln A_0(x) + Ln A_{c_1}(x)]'$$

En remarquant que $S'_5(x, y)$ (resp $S''_5(x, y)$) s'obtient de $S'_1(x, y)$ (resp $S''_1(x, y)$) en remplaçant α par β (et donc c_i par d_i , a par b et A par B), nous obtenons :

$$(\partial_2 S'_5 - \partial_1 S''_5)(x, x) = b \frac{B_d(x)}{B_0(x)} [LnB_{d_2}(x) - LnB_d(x) - LnB_0(x) + LnB_{d_1}(x)]$$

Par ailleurs,

$$S'_9(x, y) = S'_1(x, y)S'_5(x, y),$$

et donc

$$\partial_2 S'_9(x, x) = S'_1(x, x)\partial_2 S'_5(x, x) + S'_5(x, x)\partial_2 S'_1(x, x).$$

Or,

$$S'_1(x, x)\partial_2 S'_5(x, x) = a \frac{A_c(x)}{A_0(x)} b \frac{B_d(x)}{B_0(x)} [LnB_{d_2}(x) - LnB_0(x)]'.$$

Mais $S'_5(x, x)\partial_2 S'_1(x, x)$ s'obtient de $S'_1(x, x)\partial_2 S'_5(x, x)$ en intervertissant α et β , d'où :

$$\partial_2 S'_9(x, x) = a \frac{A_c(x)}{A_0(x)} b \frac{B_d(x)}{B_0(x)} [LnA_{c_2}(x) - LnA_0(x) + LnB_{d_2}(x) - LnB_0(x)]'.$$

De même,

$$\partial_1 S''_9(x, x) = S''_1(x, x)\partial_1 S''_5(x, x) + S''_5(x, x)\partial_1 S''_1(x, x),$$

le second terme de la somme se déduisant du premier et $S''_5(x, x)$ de $S''_1(x, x)$, en intervertissant α et β , d'où :

$$\begin{aligned} (\partial_2 S'_9 - \partial_1 S''_9)(x, x) = & a \frac{A_c(x)}{A_0(x)} b \frac{B_d(x)}{B_0(x)} [LnA_{c_2}(x) - LnA_0(x) + LnB_{d_2}(x) \\ & - LnB_0(x) - LnA_c(x) + LnA_{c_1}(x) - LnB_d(x) + LnB_{d_1}(x)]' \end{aligned}$$

Il reste à calculer les termes $\partial_2(S'_3 + S'_7)$ et $\partial_1(S''_3 + S''_7)$ et d'en déduire les termes qui manquent.

$$(S'_3 + S'_7)(x, y) = -S'_1(x, y) \left[\int H^{d_2}(y) \mathcal{D}_{\beta^*}(dH) + \int H^{d_1}(x) \mathcal{D}_{\beta^*}(dH) \right],$$

et donc

$$\begin{aligned} \partial_2(S'_3 + S'_7)(x, x) = & -a \frac{A_c(x)}{A_0(x)} \left[\frac{B_0(0)}{B_0(x)} \frac{B_{d_2}(x)}{B_{d_2}(0)} (Ln B_{d_2}(x) - Ln B_0(x))' \right. \\ & \left. + \left(\frac{B_0(0)}{B_0(x)} \frac{B_{d_2}(x)}{B_{d_2}(0)} + \frac{B_0(0)}{B_0(x)} \frac{B_{d_1}(x)}{B_{d_1}(0)} \right) (Ln A_{c_2}(x) - Ln A_0(x))' \right]. \end{aligned}$$

$$(S''_3 + S''_7)(x, y) = -S''_1(x, y) \left[\int H^{d_2}(x) \mathcal{D}_{\beta^*}(dH) + \int H^{d_1}(y) \mathcal{D}_{\beta^*}(dH) \right],$$

et donc

$$\begin{aligned} \partial_1(S''_3 + S''_7)(x, x) = & -a \frac{A_c(x)}{A_0(x)} \left[\frac{B_0(0)}{B_0(x)} \frac{B_{d_2}(x)}{B_{d_2}(0)} (Ln B_{d_2}(x) - Ln B_0(x))' \right. \\ & \left. + \left(\frac{B_0(0)}{B_0(x)} \frac{B_{d_2}(x)}{B_{d_2}(0)} + \frac{B_0(0)}{B_0(x)} \frac{B_{d_1}(x)}{B_{d_1}(0)} \right) (Ln A_c(x) - Ln A_{c_1}(x))' \right]. \end{aligned}$$

Nous en déduisons :

$$\begin{aligned} \partial_2(S'_3 + S'_7) - \partial_1(S''_3 + S''_7)(x, x) = & a \frac{A_c(x)}{A_0(x)} \frac{B_0(0)}{B_0(x)} \left(\frac{B_{d_2}(x)}{B_{d_2}(0)} + \frac{B_{d_1}(x)}{B_{d_1}(0)} \right) \\ & [Ln A_c(x) - Ln A_{c_1}(x) - Ln A_{c_2}(x) + Ln A_0(x)]'. \end{aligned}$$

Vu que S'_6 (resp S'_8) se déduit de S'_3 (resp S'_7) et que de même S''_6 (resp S''_8) se déduit de S''_3 (resp S''_7) en intervertissant α et β , il vient :

$$\begin{aligned} (\partial_2(S'_6 + S'_8) - \partial_1(S''_6 + S''_8))(x, x) = & b \frac{B_d(x)}{B_0(x)} \frac{A_0(0)}{A_0(x)} \left(\frac{A_{c_2}(x)}{A_{c_2}(0)} + \frac{A_{c_1}(x)}{A_{c_1}(0)} \right) \\ & [Ln B_d(x) - Ln B_{d_1}(x) - Ln B_{d_2}(x) + Ln B_0(x)]'. \end{aligned}$$

La récapitulation de tous ces résultats conduit à l'expression de la densité sur la diagonale donnée par :

$$\begin{aligned}
\partial S^*(x) = & a \frac{A_c(x)}{A_0(x)} [LnA_{c_2}(x) - LnA_c(x) - LnA_0(x) + LnA_{c_1}(x)] \\
& + b \frac{B_d(x)}{B_0(x)} [LnB_{d_2}(x) - LnB_d(x) - LnB_0(x) + LnB_{d_1}(x)] \\
& + a \frac{A_c(x)}{A_0(x)} b \frac{B_d(x)}{B_0(x)} [LnA_{c_2}(x) - LnA_0(x) + LnB_{d_2}(x) - LnB_0(x) \\
& \quad - LnA_c(x) + LnA_{c_1}(x) - LnB_d(x) + LnB_{d_1}(x)]' \\
& + a \frac{A_c(x)}{A_0(x)} \frac{B_0(0)}{B_0(x)} \left(\frac{B_{d_2}(x)}{B_{d_2}(0)} + \frac{B_{d_1}(x)}{B_{d_1}(0)} \right) [LnA_c(x) - LnA_{c_1}(x) - LnA_{c_2}(x) + LnA_0(x)]' \\
& + b \frac{B_d(x)}{B_0(x)} \frac{A_0(0)}{A_0(x)} \left(\frac{A_{c_2}(x)}{A_{c_2}(0)} + \frac{A_{c_1}(x)}{A_{c_1}(0)} \right) [LnB_d(x) - LnB_{d_1}(x) - LnB_{d_2}(x) + LnB_0(x)]'
\end{aligned}$$

.2 Appendice B

Détails du calcul de la densité pour l'exemple 2 du chapitre 5.

Ici $n = 9$, $s = 4$, $l = 6$, $k = 2$ et

$$A^\lambda = \{0, 1\}^3 = \{(0, 0, 0), (0, 0, 1), (1, 0, 0), (0, 1, 0), (0, 1, 1), ((1, 0, 1), (1, 1, 0), (1, 1, 1))\}$$

- $a = (0, 0, 0)$ alors

$$q(a) = 5.$$

$$s_1 = 3 \text{ et } s_2 = 8.$$

$$\pi_1^1 = 1, \pi_1^2 = 2, \pi_1^3 = 5, \pi_2^3 = 4, \pi_3^3 = 3, \pi_1^4 = 6, \pi_1^5 = 7, \pi_1^6 = 9, \pi_2^6 = 8.$$

Ce qui donne le terme $-\partial_{38}S_{125436798}$.

- $a = (0, 0, 1)$ alors

$$q(a) = 4.$$

$$s_1 = 3 \text{ et } s_2 = 9.$$

$$\pi_1^1 = 1, \pi_1^2 = 2, \pi_1^3 = 5, \pi_2^3 = 4, \pi_3^3 = 3, \pi_1^4 = 6, \pi_1^5 = 7, \pi_1^6 = 8, \pi_2^6 = 9.$$

Ce qui donne le terme $+\partial_{39}S_{125436789}$.

- $a = (0, 1, 0)$ alors

$$q(a) = 4.$$

$$s_1 = 4 \text{ et } s_2 = 8.$$

$$\pi_1^1 = 1, \pi_1^2 = 2, \pi_1^3 = 5, \pi_2^3 = 3, \pi_3^3 = 4, \pi_1^4 = 6, \pi_1^5 = 7, \pi_1^6 = 9, \pi_2^6 = 8.$$

Ce qui donne le terme $+\partial_{48}S_{124536789}$.

- $a = (1, 0, 0)$ alors

$$q(a) = 4.$$

$$s_1 = 4 \text{ et } s_2 = 8.$$

$$\pi_1^1 = 1, \pi_1^2 = 2, \pi_1^3 = 3, \pi_2^3 = 5, \pi_3^3 = 4, \pi_1^4 = 6, \pi_1^5 = 7, \pi_1^6 = 9, \pi_2^6 = 8.$$

Ce qui donne le terme $\partial_{48}S_{123546798}$.

- $a = (0, 1, 1)$ alors

$$q(a) = 3.$$

$$s_1 = 4 \text{ et } s_2 = 9.$$

$$\pi_1^1 = 1, \pi_1^2 = 2, \pi_1^3 = 5, \pi_2^3 = 3, \pi_3^3 = 4, \pi_1^4 = 6, \pi_1^5 = 7, \pi_1^6 = 8, \pi_2^6 = 9.$$

Ce qui donne le terme $-\partial_{49}S_{124536789}$.

- $a = (1, 1, 0)$ alors

$$q(a) = 3.$$

$$s_1 = 5 \text{ et } s_2 = 8.$$

$$\pi_1^1 = 1, \pi_1^2 = 2, \pi_1^3 = 3, \pi_2^3 = 4, \pi_3^3 = 5, \pi_1^4 = 6, \pi_1^5 = 7, \pi_1^6 = 9, \pi_2^6 = 8.$$

Ce qui donne le terme $-\partial_{58}S_{123456798}$.

- $a = (1, 0, 1)$ alors

$$q(a) = 3.$$

$$s_1 = 4 \text{ et } s_2 = 9.$$

$$\pi_1^1 = 1, \pi_1^2 = 2, \pi_1^3 = 3, \pi_2^3 = 5, \pi_3^3 = 4, \pi_1^4 = 6, \pi_1^5 = 7, \pi_1^6 = 8, \pi_2^6 = 9.$$

Ce qui donne le terme $-\partial_{49}S_{123546789}$.

- $a = (1, 1, 1)$ alors

$$q(a) = 2.$$

$$s_1 = 5 \text{ et } s_2 = 9.$$

$$\pi_1^1 = 1, \pi_1^2 = 2, \pi_1^3 = 3, \pi_2^3 = 4, \pi_3^3 = 5, \pi_1^4 = 6, \pi_1^5 = 7, \pi_1^6 = 8, \pi_2^6 = 9.$$

Ce qui donne le terme $+\partial_{59}S_{123456789}$.

Bibliographie

- [1] O. Aalen. *Statistical inference for a family of counting processes*. Ph. D. dissertation, University of California, Berkely, 1975.
- [2] O. Aalen. Nonparametric inference for a family of counting processes. *Ann. Statist.*, 6(4) : 701-726, 1978.
- [3] P. Andersen, O.Borgan, R.Gill and N. Keiding. *Statistical Models Based on Counting Processes*. Spring Series in Statistics. Springer, 1993.
- [4] B. H. L. Beck. *Nonparametric Bayesian Analysis for Special Patterns of Incompleteness*. Thèse, Université de Louvain-La-Neuve, Belgique, 2000.
- [5] L. Bordes. Nonparametric estimation under progressive censoring. à paraître dans *Journal of Statistical Planning and inference*, 2002.
- [6] S. Avarez and L. Bordes. *Empirical quantile under type I progressive censoring*. Rapport interne LMAC n° 2002.03.
- [7] C. Coccozza-Thivent. *Processus stochastiques et fiabilité des systèmes*. Springer-Verlag Berlin Heidelberg, 1997.
- [8] D. R. Cox. Regression models and life tables (with discussion). *J. R. Stat. Soc. B*, 34, 187-220, 1972.
- [9] T. S. Ferguson. A Bayesian analysis of some nonparametric problems, *Ann.Statist.*1, 209-230, 1973.
- [10] T. S. Ferguson. Prior distributions on spaces of probability measures, *Ann.Statist.*, 2, 615-629, 1974.

- [11] T. Fleming and D. Harrington. *Counting Processes & Survival Analysis*. Wiley Series in Probability and Mathematical Statistics : Applied Probability and Statistics Section. Wiley & Sons, 1991.
- [12] J. P. Florens, M. Mouchart and J. M. Rolin. *Elements of bayesian Statistics*. Number 134 in Pure and Applied Mathematics; A series of Monographs and Textbooks. Marcel Dekker, Inc, 1990.
- [13] J. P. Florens, M. Mouchart and J. M. Rolin. *Semi- and Non parametric Bayesian Analysis of Duration Models with Dirichlet Priors : A survey*. Technical report, University of Louvain la Neuve, Belgium, december 1998.
- [14] N. Gouget. *Statistique semi- paramétrique bayésienne de durées de vie. Résultats théoriques et mise en oeuvre en fiabilité industrielle*. Thèse, Université de Marne la Vallée, Champs-sur-Marne, France, 1999.
- [15] N. Gouget and J. P. Raoult. Computation of predictive densities in the semi - parametric Bayesian Cox-Dirichlet model. *Non parametric Statistics*. **10**, 307,341, 1999.
- [16] J. D. Kalbfleisch and R. L. Prentice. *The Statistical Analysis of Failure Time Data*. Wiley & Sons, 1980.
- [17] E. L. Kaplan and P. Meier. Nonparametric estimation from incomplete observations. *Jasa*, 457-481, june 1958.
- [18] J. P. Klein and M. L. Moeschberger. *Survival Analysis. Techniques for Censored and Truncated Data*. Springer, 1997.
- [19] E. T. Lee. *Statistical Methods for Survival Data Analysis*. John Wiley & Sons. Inc, 1992.
- [20] F. Messaci. Mesures prédictives dans le modèle de Cox-Dirichlet à censures aléatoires. *C. R. Acad. Sci. Paris*, Ser. I335,1-4, 2002.
- [21] C. Robert. *L'analyse statistique bayésienne*. Economica, Paris, 1992.

- [22] J. M. Rolin . *Survival data with explanatory process. A full non parametric bayesian analysis*. Discussion paper, Institut de statistique, Université Catholique de Louvain, Louvain-la-Neuve, Belgique, 1999.
- [23] M. Ruggiero. Bayesian semiparametric estimation of proportionnel hazard model, *Journal of econometrics*, **62**, 272-300, 1994.
- [24] M. Ruggiero. Analyse semi paramétrique des modèles de durées. L'apport des méthodes bayésiennes. PhD thesis, Université Aix-Marseille.II, Marseille, France, 1989.
- [25] M. J. Van Der Laan and I.W. McKeague. Efficient estimation from Right-censoring data when failure indicators are missing at random. *The Annals of Statistics*. Vol 26, No 1, 164-182, 1998.

الإسم : فتيحة مساسي بلعوي

العنوان : مناويل للإحصاء النصف وسيطي في الوثوقية

ملخص: الغرض من هذا العمل هو تمديد، إلى حالة ملاحظات مراقبة، دراسة القياسات التكهنية في المنوال البايزي لكوكس دركل، التي تناولها غوج و رولت من أجل ملاحظات غير مراقبة و ذلك بواسطة مفهوم القياس المرن بتقطع، حيث حسبوا كثافته بالنسبة لقياسات لوبيغ على مجموعات مميزة بواسطة مساوات و متباينات تامة فيما بين الملاحظات. إن هذا المفهوم يبقى مركزي في دراستنا. نعتبر الواحدة تلوى الأخرى حالات من المراقبة العشوائية أين تكون المجموعات الشهيرة غير مبدلة لكن يجب أن نغير حساب الكثافة، ثم من المراقبة الثابتة أين تكون المجموعات الشهيرة مقيدة بالإحتواء في الأوجه، أي الأجزاء المميزة بتثبيت المتغيرات المراقبة.

الكلمات المفتاحية : الكثافات التكهنية، منوال كوكس، قبلية دركل، المراقبة العشوائية، المراقبة الثابتة.

NOM : Fatiha MESSACI BELALOU

TITRE : MODELES DE STATISTIQUE SEMI-PARAMETRIQUE EN FIABILITE

RESUME. L'objet de cette thèse est d'étendre, au cas d'observations censurées, l'étude des mesures prédictives dans le modèle bayésien de Cox-Dirichlet effectuée par N. Gouget et J. P. Raoult pour des observations non censurées, à l'aide de la notion de mesure régulière par portions dont ils ont calculé les densités par rapport aux mesures de Lebesgue sur des sous ensembles caractérisés par des égalités et des inégalités strictes entre les observations. Cette notion reste centrale dans notre étude. Nous considérons successivement des cas de censure aléatoire, où les sous ensembles remarquables sont inchangés mais le calcul de densités doit être modifié, puis ceux des censures fixes, où les sous ensembles remarquables sont astreints à être contenus dans les "faces" de l'ensemble des observations, c'est à dire les parties caractérisées par la fixation des variables censurées.

MOTS CLES : Densités prédictives, Modèle de Cox, A priori de Dirichlet, Censures aléatoires, Censures fixes.

DISCIPLINE : Mathématiques appliquées, option Probabilités -Statistiques

Département de mathématiques, Université Mentouri, Constantine, 25000, Algérie.

TITLE : MODELS OF SEMI-PARAMETRIC STATISTIC IN RELIABILITY

ABSTRACT. This thesis is devoted to the generalization, to censored observations, of the study of predictive measures in the bayesian Cox-Dirichlet model, which have been performed by N. Gouget and J. P. Raoult, for non censored observations, using the notion of piecewise regular measure, whose density they have computed w.r.t. Lebesgue measures, on remarkable subsets characterized by equalities and strict inequalities between observations. This notion remains central in our study. We consider successively cases of random censoring, for which the remarkable subsets are kept unchanged but the computation of densities has to be modified, then cases of fixed censoring where the remarkable subsets are constrained to be included in the "sides" of the set of observations, i.e. the subsets characterized by fixing the censored variables.

KEY-WORDS : Predictive densities, Cox-Model, Dirichlet Prior, Random censoring, Fixed censoring.

NOM : Fatiha MESSACI BELALOU

TITRE : MODELES DE STATISTIQUE SEMI-PARAMETRIQUE EN FIABILITE

RESUME. L'objet de cette thèse est d'étendre, au cas d'observations censurées, l'étude des mesures prédictives dans le modèle bayésien de Cox-Dirichlet effectuée par N. Gouget et J. P. Raoult pour des observations non censurées, à l'aide de la notion de mesure régulière par portions dont ils ont calculé les densités par rapport aux mesures de Lebesgue sur des sous ensembles caractérisés par des égalités et des inégalités strictes entre les observations. Cette notion reste centrale dans notre étude. Nous considérons successivement des cas de censure aléatoire, où les sous ensembles remarquables sont inchangés mais le calcul de densités doit être modifié, puis ceux des censures fixes, où les sous ensembles remarquables sont astreints à être contenus dans les "faces" de l'ensemble des observations, c'est à dire les parties caractérisées par la fixation des variables censurées.

MOTS CLES : Densités prédictives, Modèle de Cox, A priori de Dirichlet, Censures aléatoires, Censures fixes.

DISCIPLINE : Mathématiques appliquées, option Probabilités -Statistiques
Département de mathématiques, Université Mentouri, Constantine, 25000, Algérie.