

Université de Montréal

Utilisation de splines monotones afin de condenser  
des tables de mortalité dans un contexte bayésien

par

Valérie Patenaude

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

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

avril 2011

© Valérie Patenaude, 2010

# Université de Montréal

Faculté des études supérieures

Ce mémoire intitulé

## Utilisation de splines monotones afin de condenser des tables de mortalité dans un contexte bayésien

présenté par

**Valérie Patenaude**

a été évalué par un jury composé des personnes suivantes :

*Pierre Duchesne*

---

(président-rapporteur)

*Jean-François Angers*

---

(directeur de recherche)

*Louis Doray, Serge Dubuc*

---

(co-directeur)

*Manuel Morales*

---

(membre du jury)

Mémoire accepté le:

*Avril 2011*

---

## SOMMAIRE

---

Dans ce mémoire, nous cherchons à modéliser des tables à deux entrées monotones en lignes et/ou en colonnes, pour une éventuelle application sur les tables de mortalité. Nous adoptons une approche bayésienne non paramétrique et représentons la forme fonctionnelle des données par splines bidimensionnelles. L'objectif consiste à condenser une table de mortalité, c'est-à-dire de réduire l'espace d'entreposage de la table en minimisant la perte d'information. De même, nous désirons étudier le temps nécessaire pour reconstituer la table.

L'approximation doit conserver les mêmes propriétés que la table de référence, en particulier la monotonie des données. Nous travaillons avec une base de fonctions splines monotones afin d'imposer plus facilement la monotonie au modèle. En effet, la structure flexible des splines et leurs dérivées faciles à manipuler favorisent l'imposition de contraintes sur le modèle désiré. Après un rappel sur la modélisation unidimensionnelle de fonctions monotones, nous généralisons l'approche au cas bidimensionnel. Nous décrivons l'intégration des contraintes de monotonie dans le modèle *a priori* sous l'approche hiérarchique bayésienne. Ensuite, nous indiquons comment obtenir un estimateur *a posteriori* à l'aide des méthodes de Monte Carlo par chaînes de Markov. Finalement, nous étudions le comportement de notre estimateur en modélisant une table de la loi normale ainsi qu'une table  $t$  de distribution de Student. L'estimation de nos données d'intérêt, soit la table de mortalité, s'ensuit afin d'évaluer l'amélioration de leur accessibilité.

MOTS CLÉS : fonctions monotones bidimensionnelles, tables de mortalité, splines monotones, contraintes de monotonie, méthodes de Monte Carlo par chaînes de Markov

## SUMMARY

---

This master's thesis is about the estimation of bivariate tables which are monotone within the rows and/or the columns, with a special interest in the approximation of life tables. This problem is approached through a nonparametric Bayesian regression model, in particular linear combinations of regression splines. By condensing a life table, our goal is to reduce its storage space without losing the entries' accuracy. We will also study the reconstruction time of the table with our estimators.

The properties of the reference table, specifically its monotonicity, must be preserved in the estimation. We are working with a monotone spline basis since splines are flexible and their derivatives can easily be manipulated. Those properties enable the imposition of constraints of monotonicity on our model. A brief review on univariate approximations of monotone functions is then extended to bivariate estimations. We use hierarchical Bayesian modeling to include the constraints in the *prior* distributions. We then explain the Markov chain Monte Carlo algorithm to obtain a *posterior* estimator. Finally, we study the estimator's behaviour by applying our model on the Standard Normal table and the Student's t table. We estimate our data of interest, the life table, to establish the improvement in data accessibility.

KEYWORDS : monotone bivariate functions, life tables, constraints of monotonicity, monotone splines, Markov chain Monte Carlo techniques

# TABLE DES MATIÈRES

---

<b>Sommaire</b> .....	iii
<b>Summary</b> .....	iv
<b>Liste des figures</b> .....	vii
<b>Liste des tableaux</b> .....	ix
<b>Remerciements</b> .....	1
<b>Introduction</b> .....	2
<b>Chapitre 1. Modélisation avec splines</b> .....	5
1.1. Classes de fonctions splines.....	6
1.1.1. Bases polynomiales.....	7
1.1.2. Splines de lissage.....	10
1.1.3. Splines de régression.....	13
1.2. B-splines.....	18
1.3. Splines monotones.....	23
1.4. Choix de noeuds.....	26
1.5. Contraintes de monotonie.....	29
<b>Chapitre 2. Modélisation bidimensionnelle</b> .....	32
2.1. Contexte bidimensionnel.....	33
2.1.1. Monotonie.....	33
2.1.2. Splines bidimensionnelles.....	34

2.1.3.	Généralisation du choix de noeuds .....	37
2.2.	Contraintes .....	39
2.3.	Modèle hiérarchique de Bayes .....	40
2.4.	Modélisation .....	43
2.4.1.	Modèle <i>a priori</i> .....	43
2.4.2.	Méthodes numériques .....	47
2.4.3.	Modèle <i>a posteriori</i> .....	49
<b>Chapitre 3.</b>	<b>Cas pratiques .....</b>	<b>53</b>
3.1.	Table de la loi normale .....	53
3.1.1.	Modèle et résultats .....	55
3.1.2.	Étude de sensibilité des paramètres .....	60
3.2.	Table <i>t</i> de distribution de Student .....	63
3.3.	Table de mortalité .....	70
<b>Conclusion.</b>	.....	<b>81</b>
<b>Bibliographie</b>	.....	<b>83</b>
<b>Annexe A.</b>	<b>Tableaux des coefficients estimés .....</b>	<b>A-i</b>
A.1.	Table de la loi normale .....	A-i
A.2.	Table <i>t</i> de distribution de Student .....	A-ii
A.3.	Table de mortalité .....	A-ii

## LISTE DES FIGURES

---

1.1	Approximation par polynômes de Lagrange d'une courbe de croissance normalisée .....	9
1.2	Exemple de fonctions (a) B-splines et (b) T-splines.....	25
1.3	Approximation par T-splines d'une courbe de croissance normalisée ..	30
2.1	Exemple de fonctions bidimensionnelles (a) B-splines et (b) T-splines.	36
3.1	Sélection des noeuds intérieurs par le critère He et Shi pour la table normale .....	58
3.2	Cartographie des points chauds : les erreurs absolues associées à l'estimation de la table de la loi normale .....	59
3.3	(a) Moyenne et (b) écart type associés aux erreurs absolues moyennes obtenues en générant 600 estimations de la table de la loi normale, selon les valeurs de $\alpha$ .....	62
3.4	Sélection des noeuds intérieurs par le critère He et Shi pour la table $t$ de la loi de Student .....	65
3.5	Cartographie des points chauds : les erreurs absolues associées à l'estimation de la table $t$ de la loi de Student.....	69
3.6	Quantiles $t_{d.l.;0,95}$ en fonction des degrés de liberté : quantiles exacts, interpolation linéaire des données de la table et estimation par le modèle de splines .....	70
3.7	Sélection des noeuds intérieurs par le critère He et Shi pour la table de mortalité .....	75

3.8	Cartographie des points chauds : les erreurs absolues associées à la table de mortalité.....	80
-----	--	----

## LISTE DES TABLEAUX

---

1.1	Échantillon de données d'une courbe de croissance normalisée de la forme $(x_i, y_i)$ .....	8
1.2	Combinaison des points de cassure et des conditions de continuité pour obtenir les noeuds .....	20
3.1	Table de la loi normale $P(Z \leq z_p)$ .....	54
3.2	Table $t$ de la loi de Student .....	63
3.3	Reconstitution de la table $t$ de la loi de Student avec les coefficients estimés .....	66
3.4	Table de mortalité .....	72
3.5	Reconstitution de la table de mortalité avec les coefficients estimés ..	76

## REMERCIEMENTS

---

Voici tout un parcours qui s'achève! J'aimerais d'abord remercier mon directeur de recherche sans qui ce projet ne serait pas arrivé à terme et qui a fait preuve d'une grande patience. Ce fut un plaisir de travailler avec lui tout au long de ma maîtrise. J'aimerais également remercier mes co-directeurs de recherche, dont l'aide m'a été précieuse lors de la rédaction de ce mémoire. Je tiens à souligner l'aide du CRSNG qui m'a permis d'achever ma maîtrise sans souci financier en m'octroyant une bourse d'études supérieures du Canada Alexander-Graham-Bell.

Merci à mes amis qui ont su m'arracher à mes livres lorsque je n'avais que des chiffres en tête! Et à mes parents, qui ont cru en moi et qui m'ont poussée à poursuivre les études supérieures : merci pour le plus beau cadeau du monde.

# INTRODUCTION

---

Ce mémoire est en quelque sorte une extension d'un stage de recherche d'été sur l'estimation des fonctions non décroissantes. Notre intérêt sur la préservation de la monotonie dans l'estimateur est passé du cas unidimensionnel au cas bidimensionnel. En particulier, modéliser des tables monotones à deux entrées sans forme fonctionnelle peut faciliter l'accès aux données. Entre autres, une table de mortalité peut être condensée en l'estimant par un nombre limité de coefficients.

Plusieurs fonctions nécessitent un estimateur qui satisfait des conditions de forme en plus de ses propriétés de lissage des données. Une condition couramment rencontrée est la monotonie des fonctions, par exemple dans le cadre des courbes de croissance ou des fonctions de répartition. Ces fonctions monotones ont joué un rôle important en statistique depuis le développement de la régression isotonique (Bartholomew, 1959 ; Barlow *et al.*, 1972). Il s'agit d'une technique simple, mais l'estimateur résultant est une fonction en escaliers non différentiable et non continue. Il s'agit de propriétés recherchées dans les estimations, donc cet estimateur isotonique est non satisfaisant du point de vue du lissage. Wright et Wegman (1980) soulignent les considérations nécessaires à la combinaison du lissage et de la monotonie. Parmi les travaux sur l'estimation de fonctions monotones, notons Mammen (1991) qui a étudié les propriétés d'un estimateur isotonique combiné à une procédure de lissage par noyaux. Des méthodes de régression non paramétriques utilisant des bases de fonctions splines sous contraintes (Ramsay, 1988 ; Mammen et Thomas-Agnan, 1999) ont également été développées.

Dans ce mémoire, nous nous penchons sur le problème d'estimation monotone dans un contexte bidimensionnel. En particulier, nous sommes intéressés à modéliser des tables monotones à deux entrées, pour une application actuarielle ultime

sur des tables de mortalité. L'objectif est de condenser les tables tout en maintenant les propriétés de monotonie et un bon ajustement. En réduisant l'espace requis par la table, nous facilitons l'accès aux données tabulaires qui n'ont pas une définition fonctionnelle connue ou implémentée dans les logiciels. Le modèle présente également une approximation alternative à l'interpolation linéaire lorsque les utilisateurs de la table cherchent une valeur autre que les entrées observées.

Nous adoptons une approche bayésienne sur les techniques de régression non paramétrique populaires. Nous avons discuté avec M. Serge Dubuc (professeur émérite du département de mathématiques et statistique de l'Université de Montréal) sur le choix de la base de fonctions pour l'approximation des données. Les ondelettes, développées dans les années récentes, présentent des propriétés mathématiques attrayantes. Entre autres, elles combinent le paramètre de fréquence des séries de Fourier et le paramètre de localisation temporelle ou spatiale des splines. Une expansion d'ondelettes peut alors s'adapter aux changements abrupts dans les données. En général, une bonne estimation d'une fonction requiert une expansion d'ondelettes avec peu de coefficients non nuls. Les ondelettes possèdent une capacité de compression importante des données, due aux techniques de rétrécissement et de seuillage (*shrinkage* et *thresholding* respectivement) introduites par Donoho et Johnstone (1994, 1995, 1998). Cependant, les ondelettes sont de nature oscillatoire, alors elles ne sont pas monotones (voir le mémoire de Ubartas, 2007), ce qui complique l'imposition des contraintes. Après considérations, nous avons choisi la base de fonctions splines, dont le support compact permet un bon ajustement local. La monotonie des expansions par splines est possible, car les fonctions I-splines et T-splines sont monotones et peuvent facilement être dérivées.

L'interpolation des données monotones par splines bidimensionnelles a été développée dans la littérature (Carlson et Fritsch, 1985), mais l'estimation monotone est moins documentée. Dierckx (1995) suggère les splines de régression bidimensionnelles afin d'approximer des surfaces, outils repris par Beliakov (2000) en imposant la monotonie. Notre contribution consiste à estimer des tables monotones en sélectionnant une base de T-splines bidimensionnelles afin de conserver

la monotonie. Nous proposons un modèle *a priori* pour les observations. Nous illustrons notre modèle à travers trois exemples : une table de la loi normale, une table  $t$  de la loi de Student ainsi qu'une table de mortalité.

Ce mémoire est organisé de la manière suivante. Le premier chapitre est consacré à l'analyse fonctionnelle unidimensionnelle avec les fonctions splines. Les bases de splines  $y$  sont définies, en mettant l'emphase sur leurs propriétés et l'obtention des fonctions monotones. Nous construisons également l'algorithme pour la sélection du nombre et de l'emplacement des noeuds. Le second chapitre généralise l'expansion de splines au contexte bidimensionnel et introduit l'approche bayésienne. Nous construisons les splines bidimensionnelles et révisons leurs propriétés. L'algorithme du choix de noeuds est adapté au contexte. Les contraintes de monotonie en deux dimensions, plus compliquées qu'au chapitre précédent, constituent une section à elles seules. Nous discutons ensuite des modèles hiérarchiques de Bayes et des techniques de calcul numériques (notamment les méthodes de Monte Carlo par chaînes de Markov), sur lesquels se base notre modélisation. Les distributions *a priori* et les modèles *a posteriori* sont présentés, de même que l'algorithme d'échantillonnage de Gibbs afin d'obtenir l'estimateur final des coefficients. Finalement, au troisième chapitre, nous appliquons notre modèle bayésien à nos trois exemples. Nous vérifions la qualité de l'ajustement à l'aide d'une mesure d'erreur absolue, notamment l'erreur absolue maximale. La performance du modèle varie selon le degré de lissage de la table : nous obtenons une précision relative d'au moins 0,017, l'ordre pouvant aller jusqu'à  $10^{-5}$  pour les tables lisses. Nous présentons une étude de sensibilité des hyperparamètres des distributions *a priori*, dont les résultats indiquent une robustesse du modèle face au choix des densités *a priori*. De plus, nous regardons la quantité d'information condensée et le temps pour la récupérer. L'espace d'entreposage de notre estimateur est inférieur au tiers de l'espace de la table originale, et ce, sans diminuer la vitesse d'accès à l'information.

# Chapitre 1

---

## MODÉLISATION AVEC SPLINES

Dans ce mémoire, nous sommes intéressés à modéliser des tables à deux entrées dans un contexte bayésien non paramétrique. Nous travaillons donc dans un cadre bidimensionnel, c'est-à-dire avec des données de la forme

$$(x_i, y_j, f_{i,j}), \quad i = 1, \dots, n_x \text{ et } j = 1, \dots, n_y.$$

Les valeurs de  $x_i$  et de  $y_j$  réfèrent respectivement aux lignes et aux colonnes de la table tandis que les  $f_{i,j}$  correspondent aux valeurs de chaque cellule. Nous obtenons une table de  $n_x$  lignes et de  $n_y$  colonnes, pour un total de  $n = n_x \times n_y$  cellules observées.

Lorsque nous parlons de tables en statistique, nous pensons aux tables de probabilités et de quantiles usuelles. La forme fonctionnelle de ces distributions est programmée dans tout bon logiciel. Or, certaines lois sont plus compliquées, d'où la nécessité d'entreposer leurs valeurs sous forme tabulaire. Similairement, certaines données, telles les données de mortalité, ne sont pas associées à des fonctions explicites. Elles doivent donc être conservées dans des tables. Notre objectif vise à faciliter l'accès aux valeurs des tables lorsqu'elles sont appelées par des simulations ou autres méthodes statistiques. Dans un premier temps, nous cherchons à condenser l'information, c'est-à-dire de réduire l'espace d'entreposage de la table. Nous voulons attribuer aux données une fonction explicite estimée avec une certaine précision, leur attribuant ainsi une capacité d'interpolation autre que linéaire. La rapidité d'accès, soit l'obtention de la valeur estimée, est également considérée.

Lorsque nous procédons à l'estimation de tables, il est important de conserver les propriétés des données. Dans ce mémoire il est question de la monotonie des données (sur les lignes et/ou les colonnes de la table). Nous retenons en exemple la table de la loi normale et les quantiles de la loi de Student. Nous sommes particulièrement intéressés à modéliser une table de mortalité dont les données sont monotones avec l'âge.

Puisqu'un tel modèle bidimensionnel est une extension du modèle unidimensionnel, nous allons préalablement étudier l'estimation de courbes unidimensionnelles monotones. Ceci correspond à modéliser une colonne de la table à deux entrées qui est monotone, ou tout simplement des fonctions monotones telles les courbes de croissance et les fonctions de répartition. Nous disposons ainsi d'un jeu de données de la forme

$$(x_i, y_i), \quad i = 1, \dots, n,$$

où  $n$  est le nombre total d'observations.

Nous considérons le modèle non paramétrique suivant :

$$y_i = g(x_i) + \epsilon_i, \quad i = 1, \dots, n, \quad (1.1)$$

où  $g$  est une fonction inconnue, les  $x_i$  appartiennent à l'ensemble borné  $[a, b] \subset \mathbb{R}$  et les  $\epsilon_i$  symbolisent les termes d'erreurs du modèle. En se basant sur l'approche de Wahba (1990, chapitre 1), nous supposons les erreurs non corrélées et identiquement distribuées de moyenne 0 et de variance  $\sigma^2$ . En supposant également les erreurs normalement distribuées, nous notons  $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ .

Dans ce chapitre, nous portons une attention particulière sur l'utilisation des splines dans l'estimation de la fonction inconnue  $g$ . Nous introduisons les polynômes et les splines de façon détaillée, puis discutons de leurs avantages et inconvénients. Nous présentons également un modèle d'inférence classique pour nos données monotones.

## 1.1. CLASSES DE FONCTIONS SPLINES

L'approche générale préconisée afin d'approximer une fonction inconnue en se basant sur une quantité finie de données se résume en deux étapes :

- (1) le choix d'une classe de fonctions appropriée au problème d'approximation ;
- (2) l'élaboration d'un processus de sélection d'une fonction de la classe retenue.

Nous introduisons dans cette section différentes classes de fonctions utilisées pour résoudre des problèmes d'approximation : les fonctions polynomiales, les splines de lissage et de régression. Nous soulignons également les propriétés associées rendant chaque classe intéressante pour approximer une fonction  $g$  inconnue.

### 1.1.1. Bases polynomiales

Les bases polynomiales ont joué un rôle majeur dès les débuts de la théorie de l'approximation. Les polynômes sont en théorie idéaux pour approximer toute fonction  $g$  continue sur un intervalle  $[a, b]$ . En effet, le théorème de Weierstrass (comme énoncé par Schumaker, 1981, section 3.4) assure que toute fonction continue sur un intervalle fermé peut être estimée par un polynôme pour un certain niveau de précision.

**Théorème 1.1.1.** *Soit  $\varepsilon > 0$ . Pour toute fonction  $g \in \mathcal{C}[a, b]$ , où  $\mathcal{C}[a, b]$  est l'espace des fonctions continues sur l'intervalle  $[a, b]$ , il existe un polynôme  $p$  tel que  $\|g - p\|_\infty < \varepsilon$ .*

La norme *sup* d'une fonction  $g$  correspond à  $\|g\|_\infty = \sup_{x \in [a, b]} |g(x)|$ .

Un polynôme d'ordre  $m$  (ou de degré  $m-1$ ) est une fonction de la forme  $p(x) = \sum_{i=1}^m a_i x^{i-1}$ , où  $a_m \neq 0$  est le coefficient dominant du polynôme. Les polynômes sont faciles à manipuler numériquement et algébriquement, propriétés justifiant leur popularité. De plus, ils sont continûment différentiables, autrement dit lisses. Plusieurs méthodes existent dans la littérature pour construire les polynômes, telles que le développement en séries de Taylor et les polynômes de Lagrange que nous présentons ici. Les polynômes de Lagrange, couramment utilisés en analyse numérique, sont des polynômes d'interpolation satisfaisant le théorème suivant (voir De Boor, 1978, chapitre 1).

**Théorème 1.1.2.** *Si le jeu de données consiste en  $n$  points distincts  $(x_i, y_i)$ ,  $i = 1, \dots, n$ , alors il existe un et un seul polynôme  $p(x)$  d'ordre  $n$  passant par ces*

TABLEAU 1.1: Échantillon de données d'une courbe de croissance normalisée de la forme  $(x_i, y_i)$

$x$	4	6	7	8	9	10	11
$y$	0,273	0,398	0,459	0,518	0,574	0,626	0,675
$x$	12	13	14	15	16	17	18
$y$	0,724	0,785	0,871	0,0843	0,976	0,991	0,999

points. Ce polynôme est donné par

$$p(x) = \sum_{i=1}^n y_i l_i(x), \quad (1.2)$$

où les polynômes de Lagrange,  $l_i$ , sont définis par

$$l_i(x) = \prod_{\substack{f=1 \\ f \neq i}}^n \left( \frac{x - x_f}{x_i - x_f} \right).$$

Nous remarquons à l'équation (1.2) que plus  $n$  est élevé, plus l'ordre du polynôme devient élevé. En effet, le théorème de Weierstrass ne mentionne pas l'ordre du polynôme. En fait, si la fonction d'intérêt présente des changements prononcés dans sa courbe, une représentation fiable de la courbe requiert plusieurs points d'interpolation. Alors, un polynôme d'ordre très élevé est généralement nécessaire pour obtenir une bonne approximation. Or, les polynômes d'ordre élevé sont de nature oscillatoire. Autrement dit, il en résulte un problème d'oscillation aux bords de l'intervalle. Ce problème est particulièrement observé lorsque l'ordre  $m$  est supérieur à 3 ou 4. Il s'agit du phénomène de Runge, un phénomène qui se produit dans certains problèmes d'interpolation de fonctions lorsque le nombre de points d'interpolation est grand (pour plus de détails, voir Schumaker, 1981). Nous présentons un petit exemple numérique pour illustrer ce phénomène.

**Exemple 1.1.1.** Soit le jeu de données se trouvant dans le tableau 1.1. Nous voulons approximer la fonction sous-jacente aux données par des polynômes de Lagrange, c'est-à-dire

$$p(x) = \sum_{i=1}^{14} y_i l_i(x).$$

Nous fournissons un exemple de calcul d'un polynôme de Lagrange,  $l_5(x)$  :

$$l_5(x) = \left(\frac{x-6}{4-6}\right) \left(\frac{x-7}{4-7}\right) \cdots \left(\frac{x-18}{4-18}\right),$$

$$= 612 - 741,606x + 410,381x^2 - \cdots - \frac{1}{14!}x^{13}.$$

Les autres fonctions de base  $l_i$  se calculent de manière semblable. Le polynôme d'approximation  $p(x)$  qui interpole les données est représenté dans la figure 1.1. Ce graphique montre que l'interpolation des données n'est pas monotone même si les données le sont. En effet, les oscillations aux extrémités du domaine de définition ne sont pas représentatives d'une courbe de croissance. Nous pouvons attribuer cette instabilité à l'ordre 14 du polynôme obtenu.

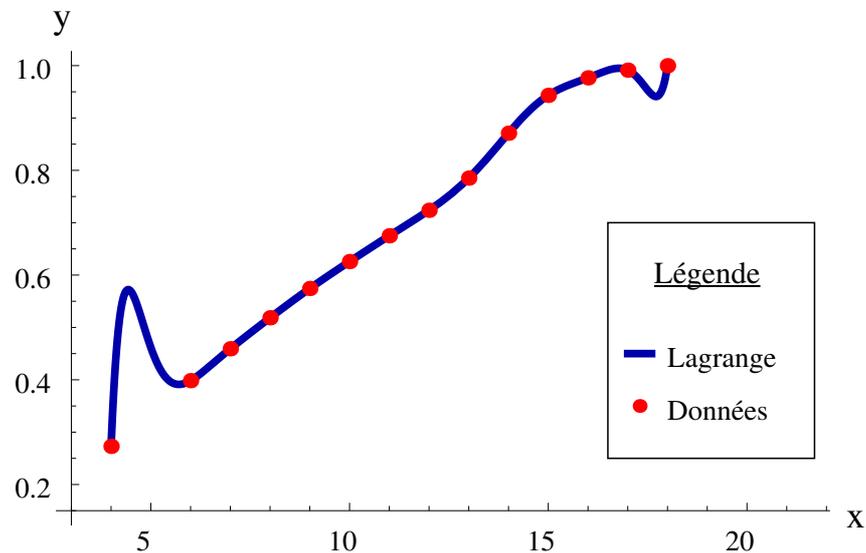


FIGURE 1.1: Approximation par polynômes de Lagrange d'une courbe de croissance normalisée

Ainsi, malgré leurs propriétés recherchées telles des dérivées et des intégrales facilement évaluables, les polynômes présentent un manque de flexibilité. En effet, un polynôme peut bien suivre une courbe sur une région, mais osciller fortement ailleurs : l'ajustement n'est bon que sur un petit intervalle. Afin d'obtenir une classe de fonctions plus flexibles, des polynômes de bas ordre sont nécessaires, et ce sur de petits intervalles. L'approche que nous considérons à la sous-section 1.1.3 consiste donc à diviser le domaine d'intérêt en plusieurs sous-intervalles pour que

sur chaque sous-intervalle, un polynôme de bas ordre estime bien  $g$ . Ce principe de polynôme par morceaux donne lieu aux splines, une classe de fonctions flexibles possédant les mêmes propriétés convoitées que les polynômes.

### 1.1.2. Splines de lissage

Le terme « fonctions splines » a été introduit par Schoenberg (1946a, b), bien que leur origine soit due aux travaux de Whittaker (1923) sur les méthodes de graduation de données. Ces recherches ont poussé Schoenberg (1964) et Reinsch (1967) à dériver les splines de lissage classiques comme classe d'estimateurs. Par ailleurs, Ahlberg, Nilson et Walsh (1964, 1967) mentionnent les splines dans un contexte d'interpolation de données sans bruit. Ce n'est donc qu'au début des années 1960 que la théorie des splines s'est développée. Les splines étaient alors considérées comme un outil d'analyse numérique pour modéliser des fonctions. L'utilisation des splines comme méthode de régression non paramétrique est attribuée à Wahba (1975), qui a démontré leurs propriétés statistiques. Notons également ses travaux en collaboration, soient Kimerldorf et Wahba (1971), Wahba et Wold (1975) et Craven et Wahba (1979), qui ont favorisé le développement des splines de lissage. Les splines ont depuis connu un essor fulgurant, notamment dans les domaines de l'économie et de la finance.

Supposons que la fonction  $g$  du modèle (1.1) soit lisse, autrement dit que  $g \in \mathcal{W}_2^m[a, b]$ . L'espace de Sobolev d'ordre  $m$ ,  $\mathcal{W}_2^m[a, b]$ , est défini comme

$$\mathcal{W}_2^m[a, b] = \left\{ g : g^{(j)} \text{ absolument continue, } j = 0, \dots, m-1; \right. \\ \left. g^{(m)} \in \mathcal{L}_2[a, b] \right\}.$$

L'espace de Lebesgue  $\mathcal{L}_2[a, b]$  correspond à l'ensemble des fonctions de carré intégrable sur l'intervalle  $[a, b]$ , soit

$$\mathcal{L}_2[a, b] = \left\{ g : \int_a^b |g(x)|^2 dx < \infty \right\}.$$

Afin d'approximer la fonction inconnue  $g$ , nous désirons un estimateur bien ajusté aux données, mais qui possède également un certain degré de régularité. En

combinant ces deux critères, l'estimateur optimal est la fonction  $g_\lambda$  qui minimise

$$\frac{1}{n} \sum_{j=1}^n (y_j - g(x_j))^2 + \lambda \int_a^b [g^{(m)}(x)]^2 dx, \quad \lambda > 0, \quad (1.3)$$

sur  $g \in \mathcal{W}_2^m[a, b]$ . L'estimateur  $g_\lambda$ , c'est-à-dire la spline de lissage, est donc un compromis entre le bon ajustement des données et l'obtention d'une fonction lisse.

Le paramètre  $\lambda$  contrôle le lissage de la fonction, soit le terme  $\int_a^b [g^{(m)}(x)]^2 dx$  de l'expression (1.3). Lorsque le paramètre de lissage,  $\lambda$ , est grand, les fonctions lisses sont privilégiées tandis que les fonctions avec une grande  $m^e$  dérivée sont fortement pénalisées. À l'inverse, pour des valeurs de  $\lambda$  proches de 0, l'emphase est mise sur la qualité de l'ajustement. En effet, le premier terme de l'expression (1.3) correspond à une pénalité pour l'ajustement des données. Par conséquent, l'estimateur  $g_\lambda$  se situe entre l'interpolation des données et la droite des moindres carrés selon la valeur de  $\lambda$ .

Nous voulons maintenant dériver la forme de l'estimateur spline. La proposition suivante garantit l'existence et l'unicité de l'estimateur (voir Eubank, 1988, section 5.3).

**Proposition 1.1.1.** *Si  $n \geq m$ , alors l'unique solution minimisant l'équation (1.3) prend la forme  $g_\lambda(x) = \sum_{j=1}^n \theta_{\lambda j} z_j(x)$ . Le vecteur de coefficients  $\boldsymbol{\theta}_\lambda = (\theta_{\lambda 1}, \dots, \theta_{\lambda n})^T$  est la solution de l'équation*

$$(\mathbf{Z}^T \mathbf{Z} + n\lambda \boldsymbol{\Omega}) \boldsymbol{\theta}_\lambda = \mathbf{Z}^T \mathbf{y}$$

avec  $\mathbf{Z} = \{z_j(x_i)\}_{i,j=1,\dots,n}$  et  $\boldsymbol{\Omega} = \left\{ \int_a^b z_i^{(m)}(t) z_j^{(m)}(t) dt \right\}_{i,j=1,\dots,n}$ . Les fonctions  $z_j$  forment la base de l'espace spline.

Les bases  $z_1, \dots, z_n$  peuvent être construites de différentes façons. Parmi les bases spécialement conçues pour les splines de lissage dans la littérature, nous retrouvons celle définie par Kimeldorf et Wahba (1970a, 1970b), par la suite simplifiée par Wahba (1978). Nous définissons les fonctions de base  $1, x, \dots, x^{m-1}, g_1(x), \dots, g_{n-m}(x)$  avec

$$(g_1(x), \dots, g_{n-m}(x)) = (Q(x, x_1), \dots, Q(x, x_n)) \mathbf{U},$$

où

$$Q(s, t) = [(m-1)!]^{-2} \int_a^b (s-u)_+^{m-1} (t-u)_+^{m-1} du, \quad a \leq s, t \leq b$$

et  $\mathbf{U}$  est la matrice de vecteurs propres de la matrice idempotente  $\mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$  pour  $\mathbf{X} = \{x_i^{j-1}\}_{\substack{i=1, \dots, n \\ j=1, \dots, m}}$ .

Le choix du paramètre de lissage,  $\lambda$ , dépend des données disponibles. Wahba et Wold (1975) proposent la validation croisée afin d'estimer  $\lambda$ . Posons  $g_\lambda^{[j]}$ , la fonction minimisant l'expression (1.3) lorsque la  $j^{\text{e}}$  observation est retirée de l'échantillon de données. Nous définissons ainsi le critère de validation croisée ordinaire

$$V_0(\lambda) = \frac{1}{n} \sum_{j=1}^n \left( y_j - g_\lambda^{[j]}(x_j) \right)^2.$$

En minimisant le critère  $V_0(\lambda)$ , nous obtenons un estimateur du risque de prévision qui ne dépend pas du paramètre  $\sigma^2$  du modèle (1.1). Un autre critère possédant cette propriété est la validation croisée généralisée, qui est en fait une version pondérée de  $V_0$  :

$$V(\lambda) = \frac{1}{n} \sum_{j=1}^n \left( y_j - g_\lambda^{[j]}(x_j) \right)^2 w_j(\lambda),$$

où  $w_j(\lambda) = [(1 - h_{jj}(\lambda))/[n^{-1} \text{tr}(I - \mathbf{H}(\lambda))]]^2$  tandis que  $\mathbf{H}(\lambda) = \mathbf{Z}(\mathbf{Z}^T \mathbf{Z} + n\lambda \mathbf{\Omega}) \mathbf{Z}^T$  et les matrices  $\mathbf{Z}$  et  $\mathbf{\Omega}$  sont définies à la proposition 1.1.1. Le terme  $h_{jj}(\lambda)$  correspond au  $j^{\text{e}}$  élément diagonal de  $\mathbf{H}(\lambda)$ . Les poids  $w_j(\lambda)$  compensent alors pour une distribution non uniforme des points  $x_j$  et la possibilité d'une fonction  $g$  non périodique. En effet, Craven et Wahba (1979) ont démontré que

$$y_j - g_\lambda^{[j]}(x_j) = (y_j - g_\lambda(x_j)) / (1 - h_{jj}(\lambda))$$

pour les splines de lissage. En réécrivant les critères comme

$$V_0(\lambda) = \frac{1}{n} \sum_{j=1}^n (y_j - g_\lambda(x_j))^2 / (1 - h_{jj}(\lambda))^2$$

et

$$V(\lambda) = \frac{1}{n} \sum_{j=1}^n (y_j - g_\lambda(x_j))^2 / \left( 1 - n^{-1} \sum_{j=1}^n h_{jj}(\lambda) \right)^2,$$

il ressort que les valeurs individuelles  $h_{11}(\lambda), \dots, h_{nn}(\lambda)$  de  $V_0(\lambda)$  sont substituées par leur moyenne dans  $V(\lambda)$ , ce qui atténue l'influence des points leviers. D'autres

résultats, dont l'optimalité du critère  $V(\lambda)$  dans un contexte de régression non paramétrique, se trouvent dans l'article de Craven et Wahba (1979).

### 1.1.3. Splines de régression

Comme mentionné précédemment, le concept d'une spline polynomiale de régression consiste à considérer des polynômes de bas ordre sur de petits sous-intervalles de  $[a, b]$ . Afin de définir l'espace des polynômes par morceaux, nous posons la partition de  $k + 1$  sous-intervalles

$$I_i = [\xi_i, \xi_{i+1}), \quad i = 0, \dots, k-1, \quad \text{et } I_k = [\xi_k, \xi_{k+1}],$$

où  $\boldsymbol{\xi} = (\xi_0, \dots, \xi_{k+1})^T$  est une séquence croissante telle que  $a = \xi_0 < \xi_1 < \dots < \xi_{k+1} = b$ . En définissant un polynôme  $P_i$  d'ordre  $m$  sur chaque sous-intervalle  $I_i$ , nous obtenons l'espace de polynômes par morceaux aux points de cassure  $\boldsymbol{\xi}$  :

$$\mathcal{P}_m(\boldsymbol{\xi}) = \{f : f(x) = P_i(x)\mathbb{1}_i; P_i \in \mathcal{P}_m, i = 0, \dots, k\}, \quad (1.4)$$

où  $\mathbb{1}_i$  est la fonction indicatrice pour l'intervalle  $I_i$  et  $\mathcal{P}_m$  est l'ensemble des polynômes d'ordre  $m$ . Bien que l'espace obtenu soit plus flexible que l'espace polynomial, une autre propriété est perdue : les fonctions ne sont plus nécessairement lisses. Des points de discontinuité peuvent notamment être observés aux points de cassure. Afin d'obtenir un certain niveau de régularité, nous imposons l'égalité des dérivées à chaque jonction des polynômes, c'est-à-dire

$$(D^{u-1}P_{i-1})(\xi_i) = (D^{u-1}P_i)(\xi_i), \quad u = 1, \dots, \nu_i, \quad i = 1, \dots, k, \quad (1.5)$$

où la notation  $(D^{u-1}P_i)(\xi)$  représente la dérivée  $d^{u-1}P_i/dx^{u-1}$  évaluée à  $\xi$  si  $u > 1$  et  $(D^0P_i)(\xi) = P_i(\xi)$  si  $u = 1$ . Les ordres de continuité  $\nu_i$  permettent de manipuler la nature de l'espace spline en contrôlant le lissage des splines aux points de cassure. D'un côté, si  $\nu_i = 0$ , alors la spline peut être discontinue au point  $\xi_i$ . D'un autre côté, si tous les  $\nu_i$  valent  $m$ , alors la fonction spline consiste en un seul polynôme d'ordre  $m$ .

Pour s'assurer d'obtenir l'espace de polynômes par morceaux d'ordre  $m$  le plus lisse possible sans effacer les points de cassure  $\boldsymbol{\xi}$ , les ordres de continuités

sont habituellement équivalents au degré  $m - 1$  du polynôme. Deux polynômes adjacents ont ainsi les mêmes dérivées jusqu'à l'ordre  $m - 2$ . Autrement dit, nous considérons une contrainte supplémentaire à l'espace (1.4), soit

$$\mathcal{S}_m(\boldsymbol{\xi}) = \mathcal{P}_m(\boldsymbol{\xi}) \cap \mathcal{C}^{m-2}[a, b],$$

où  $\mathcal{C}^{m-2}[a, b] = \{f : f^{(i)} \text{ existe et est continue sur } [a, b], i = 1, \dots, m - 2\}$ . L'espace  $\mathcal{S}_m(\boldsymbol{\xi})$  dénote l'espace des fonctions splines d'ordre  $m$  avec noeuds  $\boldsymbol{\xi}$  de multiplicité 1. Le terme « noeuds » réfère aux points de cassure où les contraintes de continuité sont imposées. Quand à la multiplicité des noeuds, elle découle directement du nombre de contraintes de continuité, c'est-à-dire de  $\nu_i$ , et est caractérisée par  $m - \nu_i$ . Par conséquent, lorsque l'ordre de continuité est différent de  $m - 1$ , nous observons des noeuds multiples (confondus). De manière générale, nous définissons l'espace de splines d'ordre  $m$  avec noeuds  $\boldsymbol{\xi}$ ,

$$\begin{aligned} \mathcal{S}_m(\boldsymbol{\xi}, \boldsymbol{\nu}) = \{s : s(x) = P_i(x) \text{ pour } x \in I_i, P_i \in \mathcal{P}_m, i = 0, \dots, k; \\ \forall P_i, \text{ les contraintes (1.5) sont satisfaites}\}, \end{aligned} \quad (1.6)$$

où la multiplicité dépend du vecteur  $\boldsymbol{\nu}$ .

Nous identifions d'abord la dimension de l'espace de fonctions splines afin de lui attribuer par la suite une base (Curry et Schoenberg, 1966).

**Théorème 1.1.3.** *Pour un vecteur de points de cassure  $\boldsymbol{\xi} = (\xi_0, \dots, \xi_{k+1})^T$  et un vecteur  $\boldsymbol{\nu} = (\nu_1, \dots, \nu_k)^T$  d'ordres de continuité,  $\mathcal{S}_m(\boldsymbol{\xi}, \boldsymbol{\nu})$  est un espace linéaire de dimension*

$$m + \sum_{i=1}^k (m - \nu_i) = m(k + 1) - \sum_{i=1}^k \nu_i.$$

Étant donné que nous travaillons avec des espaces linéaires dont la dimension diminue lorsque l'ordre de continuité augmente, il est évident que

$$\mathcal{P}_m \subseteq \mathcal{S}_m(\boldsymbol{\xi}) \subseteq \mathcal{S}_m(\boldsymbol{\xi}, \boldsymbol{\nu}) \subseteq \mathcal{P}_m(\boldsymbol{\xi}).$$

Il en ressort que toute base de  $\mathcal{S}_m$  doit au moins inclure une base de l'espace polynomial  $\mathcal{P}_m$ . Nous pouvons choisir les fonctions  $1, (x - a), \dots, (x - a)^{m-1}$ . Afin de trouver les  $\sum_{i=1}^k (m - \nu_i)$  fonctions de base additionnelles, il faut étudier l'idée derrière les splines de régression (également appelées splines des moindres carrés).

En utilisant le théorème de Taylor, le modèle (1.1) à estimer peut être réécrit comme

$$y_i = \sum_{j=1}^m \theta_j (x_i - a)^{j-1} + R(x_i) + \epsilon_i,$$

où

$$R(x) = [(m-1)!]^{-1} \int_a^b g^{(m)}(z) (x-z)_+^{m-1} dz \quad (1.7)$$

et  $x_i \in [a, b]$ ,  $\forall i$ .

L'estimateur spline utilise l'information contenue dans  $R(x)$  contrairement au modèle polynomial. Le lemme suivant (voir Eubank, 1988, section 7.2) nous permet de représenter l'intégrale (1.7) par une somme de puissances tronquées.

**Lemme 1.1.1.** *Soit  $g \in \mathcal{S}_m(\boldsymbol{\xi}, \boldsymbol{\nu})$ . Alors, l'intégrale (1.7) devient*

$$R(x) = \sum_{j=1}^{m-1} \gamma_j (x-a)^j + \sum_{i=1}^k \sum_{j=\nu_i}^{m-1} \beta_{i,j} (x-\xi_i)_+^j$$

pour certains ordres de continuité  $\nu_i$  et points de cassure  $\xi_i$  donnés.

DÉMONSTRATION. Une spline est définie comme un polynôme par morceaux, c'est-à-dire  $g(x) = P_i(x)$  pour  $x \in I_i = [\xi_i, \xi_{i+1})$ ,  $i = 0, \dots, k$ , avec les contraintes données à l'équation (1.5). Puisque les polynômes  $P_i$  varient sur chaque sous-intervalle  $I_i$ , nous séparons l'intégrale (1.7) en  $k+1$  parties :

$$R(x) = \sum_{i=0}^k \left( \int_{\xi_i}^{\xi_{i+1}} \frac{P_i^{(m)}(z)}{(m-1)!} (x-z)_+^{m-1} dz \right). \quad (1.8)$$

En intégrant par parties  $m-1$  fois, le premier terme à droite de l'équation (1.8) devient une somme de fonctions,

$$\begin{aligned} & \int_{\xi_0}^{\xi_1} \frac{P_0^{(m)}(z)}{(m-1)!} (x-z)_+^{m-1} dz \\ &= \frac{P_0^{(m-1)}(z)}{(m-1)!} (x-z)_+^{m-1} \Big|_{\xi_0}^{\xi_1} + \int_{\xi_0}^{\xi_1} \frac{P_0^{(m-1)}(z)}{(m-2)!} (x-z)_+^{m-2} dz \\ &= \dots \\ &= \sum_{j=0}^{m-1} \frac{P_0^{(j)}(\xi_1)}{j!} (x-\xi_1)_+^j - \sum_{j=0}^{m-1} \frac{P_0^{(j)}(\xi_0)}{j!} (x-\xi_0)_+^j. \end{aligned} \quad (1.9)$$

Il s'agit ensuite de considérer les contraintes (1.5) qui s'appliquent sur l'intervalle  $I_0$ , autrement dit  $P_0^{(j)}(\xi_1) = P_1^{(j)}(\xi_1)$  pour  $j = 0, \dots, \nu_1 - 1$ . Nous réécrivons alors l'expression (1.9) par

$$\begin{aligned} & \int_{\xi_0}^{\xi_1} \frac{P_0^{(m)}(z)}{(m-1)!} (x-z)_+^{m-1} dz \\ &= \sum_{j=0}^{\nu_1-1} \frac{P_1^{(j)}(\xi_1)}{j!} (x-\xi_1)_+^j + \sum_{j=\nu_1}^{m-1} \frac{P_0^{(j)}(\xi_1)}{j!} (x-\xi_1)_+^j - \sum_{j=0}^{m-1} \frac{P_0^{(j)}(\xi_0)}{j!} (x-\xi_0)_+^j. \end{aligned}$$

Similairement, nous obtenons pour le second terme de l'équation (1.8)

$$\begin{aligned} & \int_{\xi_1}^{\xi_2} \frac{P_1^{(m)}(z)}{(m-1)!} (x-z)_+^{m-1} dz \\ &= \sum_{j=0}^{m-1} \frac{P_1^{(j)}(\xi_2)}{j!} (x-\xi_2)_+^j - \sum_{j=0}^{m-1} \frac{P_1^{(j)}(\xi_1)}{j!} (x-\xi_1)_+^j \\ &= \sum_{j=0}^{\nu_2-1} \frac{P_2^{(j)}(\xi_2)}{j!} (x-\xi_2)_+^j + \sum_{j=\nu_2}^{m-1} \frac{P_1^{(j)}(\xi_2)}{j!} (x-\xi_2)_+^j \\ & \quad - \sum_{j=0}^{\nu_1} \frac{P_1^{(j)}(\xi_1)}{j!} (x-\xi_1)_+^j - \sum_{j=\nu_1}^{m-1} \frac{P_1^{(j)}(\xi_1)}{j!} (x-\xi_1)_+^j. \end{aligned}$$

Ainsi, en combinant les deux premiers termes de l'intégrale (1.8), quelques regroupements s'effectuent :

$$\begin{aligned} & \int_{\xi_0}^{\xi_2} \frac{g^{(m)}(z)}{(m-1)!} (x-z)_+^{m-1} dz \\ &= \sum_{j=\nu_1}^{m-1} \left( \frac{P_0^{(j)}(\xi_1) - P_1^{(j)}(\xi_1)}{j!} \right) (x-\xi_1)_+^j + \sum_{j=\nu_2}^{m-1} \frac{P_1^{(j)}(\xi_2)}{j!} (x-\xi_2)_+^j \\ & \quad + \sum_{j=0}^{\nu_2-1} \frac{P_2^{(j)}(\xi_2)}{j!} (x-\xi_2)_+^j - \sum_{j=0}^{m-1} \frac{P_0^{(j)}(\xi_0)}{j!} (x-\xi_0)_+^j. \end{aligned}$$

En répétant la procédure pour chaque terme, nous obtenons un reste (1.8) de la forme

$$\begin{aligned} R(x) &= \sum_{i=1}^k \sum_{j=\nu_i}^{m-1} \left( \frac{P_{i-1}^{(j)}(\xi_i) - P_i^{(j)}(\xi_i)}{j!} \right) (x-\xi_i)_+^j - \sum_{j=0}^{m-1} \frac{P_0^{(j)}(\xi_0)}{j!} (x-\xi_0)_+^j \\ &= \sum_{i=1}^k \sum_{j=\nu_i}^{m-1} \beta_{i,j} (x-\xi_i)_+^j + \sum_{j=0}^{m-1} \gamma_j (x-\xi_0)_+^j, \end{aligned}$$

où  $\xi_0 = a$ .

□

Ainsi, en approximant la fonction inconnue  $g$  par une spline, l'estimateur de  $g$  prend la forme

$$s(x) = \sum_{j=1}^m \alpha_j (x - a)^{j-1} + \sum_{i=1}^k \sum_{j=\nu_i}^{m-1} \beta_{i,j} (x - \xi_i)_+^j, \quad (1.10)$$

où la spline  $s \in \mathcal{S}_m(\boldsymbol{\xi}, \boldsymbol{\nu})$ . Les coefficients  $(\alpha_1, \dots, \alpha_m, \beta_{1,\nu_1}, \dots, \beta_{k,m-1})^T$  peuvent être déterminés par la méthode des moindres carrés, d'où le nom de la classe « splines de régression ». Lorsque les points de cassure  $\boldsymbol{\xi}$  sont de multiplicité 1, l'estimation (1.10) devient plutôt

$$s(x) = \sum_{j=1}^m \alpha_j (x - a)^{j-1} + \sum_{i=1}^k \beta_i (x - \xi_i)_+^{m-1}.$$

**Théorème 1.1.4.** *Une base pour l'espace  $\mathcal{S}_m(\boldsymbol{\xi}, \boldsymbol{\nu})$  est donnée par les fonctions*

$$\phi_{i,j}(x) = \begin{cases} (x - a)^j, & \text{si } i = 0; \\ (x - \xi_i)_+^j, & \text{si } i = 1, \dots, k; \end{cases}$$

pour  $i = 0, \dots, k$ ,  $j = \nu_i, \dots, m - 1$  et  $\nu_0 = 0$ .

En effet, l'ensemble des fonctions  $\phi_{i,j}(x)$  comporte une base polynomiale (lorsque  $i = 0$ ) ainsi que  $\sum_{i=1}^k (m - \nu_i)$  fonctions linéairement indépendantes que nous cherchions précédemment. Pour une preuve du théorème et de l'indépendance linéaire de chaque élément de la base, voir Schumaker (1981, section 4.1). Il en découle que chaque spline  $s \in \mathcal{S}_m(\boldsymbol{\xi}, \boldsymbol{\nu})$  a une représentation unique de la forme (1.10). Cette formulation est équivalente aux caractéristiques suivantes lorsque  $\nu_i = m - 1$  :

- (1)  $s$  est un polynôme par morceaux d'ordre  $m$  défini sur tout sous-intervalle  $[\xi_i, \xi_{i+1})$ ;
- (2)  $s$  a  $m - 2$  dérivées continues;
- (3) la  $(m - 1)^{\text{e}}$  dérivée de  $s$  est une fonction en escalier avec sauts aux points  $\xi_1, \dots, \xi_k$ .

Bien qu'adéquate pour des fins théoriques, cette base de puissances tronquées se traduit moins bien en applications numériques. Nous présenterons aux sections 1.2 et 1.3 d'autres bases générant l'espace spline qui sont plus pratiques à manipuler.

Pour plus d'informations sur les splines de régression et de lissage, nous recommandons l'ouvrage de Eubank (1988) ainsi que l'article de Wegman et Wright (1983) qui discute des applications statistiques des splines dans la littérature.

## 1.2. B-SPLINES

Les fonctions B-splines font partie des splines de régression. Le terme B-spline, introduit par Schoenberg (1946a, 1946b), signifie tout simplement une base spline. En effet, Curry et Schoenberg (1966) ont démontré que toute fonction spline  $s(x)$  d'ordre  $m$  peut être représentée par une combinaison linéaire,

$$s(x) = \sum_{i=1}^z a_i N_i^m(x), \quad \forall x \in [a, b], \quad (1.11)$$

où les  $N_i^m(x)$  sont une base de l'espace spline  $\mathcal{S}_m(\boldsymbol{\xi}, \boldsymbol{\nu})$  défini à l'équation (1.6), d'où le nom B-spline. Notons que les fonctions B-splines sont également dénotées par  $B_i^m$  dans la littérature.

Afin de construire ces bases splines, nous devons approfondir les notions suivantes. Nous commençons par définir le concept de différence divisée comme un quotient de déterminants.

**Définition 1.2.1.** *Soient les points distincts  $t_i, \dots, t_{i+k}$  et la fonction  $g$ . La différence divisée d'ordre  $k$  de la fonction  $g$  aux points  $t_i, \dots, t_{i+k}$  est définie par*

$$[t_i, \dots, t_{i+k}]g = \begin{vmatrix} 1 & t_i & \cdots & t_i^{k-1} & g(t_i) \\ 1 & t_{i+1} & \cdots & t_{i+1}^{k-1} & g(t_{i+1}) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & t_{i+k} & \cdots & t_{i+k}^{k-1} & g(t_{i+k}) \end{vmatrix} \begin{vmatrix} 1 & t_i & \cdots & t_i^k \\ 1 & t_{i+1} & \cdots & t_{i+1}^k \\ \vdots & \vdots & \vdots & \vdots \\ 1 & t_{i+k} & \cdots & t_{i+k}^k \end{vmatrix}^{-1}.$$

Nous présentons un exemple afin de mieux comprendre le concept de différences divisées.

**Exemple 1.2.1.** *En pratique, nous avons besoin d'évaluer des B-splines définies avec différents points de cassure. Pour ce faire, nous devons calculer les différences divisées à ces points. Par exemple, nous désirons calculer la 3<sup>e</sup> différence divisée d'une fonction  $g$  aux points  $(t_3, t_4, t_5, t_6) = (4, 12, 14, 16)$ . Nous considérons la fonction  $g_x(t) = (t - x)_+^2$  pour une valeur  $x \in [t_3, t_6]$  fixée : il s'agit donc d'un*

polynôme par morceaux d'ordre 3. Afin d'évaluer la différence divisée à  $x = 10$ , nous calculons le quotient de déterminants

$$[4, 12, 14, 16](t - 10)_+^2 = \frac{\begin{vmatrix} 1 & 4 & 4^2 & 0 \\ 1 & 12 & 12^2 & 2^2 \\ 1 & 14 & 14^2 & 4^2 \\ 1 & 16 & 16^2 & 6^2 \end{vmatrix}}{\begin{vmatrix} 1 & 4 & 4^2 & 4^3 \\ 1 & 12 & 12^2 & 12^3 \\ 1 & 14 & 14^2 & 14^3 \\ 1 & 16 & 16^2 & 16^3 \end{vmatrix}}^{-1},$$

où la dernière colonne du numérateur correspond à  $(t_j - 10)_+^2 = [\max(t_j - 10, 0)]^2$  pour  $j = 3, \dots, 6$ . Nous obtenons alors

$$[4, 12, 14, 16](t - 10)_+^2 = 576/15360 = 0,0375.$$

Nous pouvons normaliser ce résultat en le multipliant par  $(t_6 - t_3) = 12$ , soit

$$(t_6 - t_3)[t_3, t_4, t_5, t_6](t - 10)_+^2 = 12 \times 0,0375 = 0,45.$$

Le choix d'une fonction de puissance tronquée pour  $g$  sera justifié plus loin en définissant les  $B$ -splines.

Une autre méthode pour définir la  $k^e$  différence divisée consiste à prendre le coefficient dominant du polynôme d'ordre  $k + 1$  (c'est-à-dire le coefficient de  $x^k$ ) qui coïncide avec  $g$  aux points  $t_i, \dots, t_{i+k}$ . Une propriété importante qui découle de cette définition est l'écriture récursive possible des différences divisées. Étant donné  $t_i \neq t_{i+k}$ , nous obtenons la relation de récursion

$$[t_i, \dots, t_{i+k}]g = \frac{[t_{i+1}, \dots, t_{i+k}]g - [t_i, \dots, t_{i+k-1}]g}{t_{i+k} - t_i},$$

avec  $[t_i]g = g(t_i)$ . Dans le cas où  $t_i = \dots = t_{i+k}$  et  $g \in \mathcal{C}^k$ , la  $k^e$  différence divisée s'écrit

$$[t_i, \dots, t_{i+k}]g = \frac{g^{(k)}(t_i)}{k!}.$$

Une seconde propriété de la  $k^e$  différence divisée nous intéresse :

$$[t_i, \dots, t_{i+k}]g = 0 \quad \text{si } g \text{ est un polynôme d'ordre } \leq k. \quad (1.12)$$

En effet, le coefficient dominant du polynôme  $a_1 + \dots + a_k x^{k-1}$  est 0 s'il est considéré comme un polynôme d'ordre  $k + 1$ .

TABLEAU 1.2: Combinaison des points de cassure et des conditions de continuité pour obtenir les noeuds

Points de cassure	$\xi_0 = a$	$\xi_1$	$\cdots$	$\xi_k$	$\xi_{k+1} = b$
Nombre de contraintes de continuité	$\nu_0 = 0$	$\nu_1$	$\cdots$	$\nu_k$	$\nu_{k+1} = 0$
Multiplicité des noeuds	$m$	$m - \nu_1$	$\cdots$	$m - \nu_k$	$m$
Noeuds associés	$t_1$	$t_{m+1}$	$\cdots$	$\cdots$	$t_{z+1}$
	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
	$t_m$	$t_{2m-\nu_1}$	$\cdots$	$t_z$	$t_{z+m}$

Par ailleurs, nous définissons formellement la notion de noeuds mentionnés à la sous-section 1.1.3. La partition  $\boldsymbol{\xi}$  et les conditions de continuité sont incorporées dans la séquence de noeuds.

**Définition 1.2.2.** Soient les vecteurs des points de cassure  $\boldsymbol{\xi} = (\xi_0, \dots, \xi_{k+1})^T$  et des contraintes (ordres) de continuité  $\boldsymbol{\nu} = (\nu_1, \dots, \nu_k)^T$ . Alors, le vecteur de noeuds  $\boldsymbol{t} = (t_1, \dots, t_{z+m})^T$  satisfait les conditions suivantes :

- (1)  $t_1 \leq t_2 \leq \cdots \leq t_m \leq \xi_0$  et  $\xi_{k+1} \leq t_{z+1} \leq \cdots \leq t_{z+m}$  ;
- (2) pour  $i = 1, \dots, k$ , le point  $\xi_i$  se retrouve exactement  $m - \nu_i$  fois dans  $\boldsymbol{t}$ .

Notons que le choix des  $m$  premiers noeuds et des  $m$  derniers noeuds n'est pas spécifié. De Boor (1978) propose de choisir

$$t_1 = \dots = t_m = \xi_0 \text{ et } t_{z+1} = \dots = t_{z+m} = \xi_{k+1}.$$

Ce choix correspond à prendre  $\nu_0 = \nu_{k+1} = 0$ , c'est-à-dire que nous n'imposons aucune contrainte de continuité aux extrémités du domaine. Le tableau 1.2 récapitule la construction du vecteur de noeuds. Nous verrons à la section 1.4 comment choisir les  $z - m$  noeuds intérieurs, correspondant aux points  $\xi_1 < \dots < \xi_k$ .

Nous pouvons désormais définir les B-splines d'ordre  $m$  par la  $m^e$  différence divisée de fonctions de puissances tronquées. Ainsi, la  $i^e$  B-spline d'ordre  $m$  ayant pour noeuds le vecteur  $\boldsymbol{t}$  satisfait

$$N_i^m(x|\boldsymbol{t}) = (t_{i+m} - t_i)[t_i, \dots, t_{i+m}](\cdot - x)_+^{m-1}, \quad \forall x \in \mathbb{R}, \quad (1.13)$$

pour  $i = 1, \dots, z$ . La notation  $(\cdot - x)_+^{m-1}$  de l'équation (1.13) indique que la différence divisée est appliquée sur  $(a - x)_+^{m-1}$  comme une fonction de  $a$  pour un  $x$  fixé. Lorsque  $x \notin [t_i, t_{i+m}]$ , nous obtenons  $N_i^m(x|\mathbf{t}) = 0$ . En effet, si  $x > t_{i+m}$ , la puissance tronquée  $(\cdot - x)_+^{m-1}$  vaut 0 sur l'intervalle  $[t_i, t_{i+m}]$ . Si  $x < t_i$ , alors  $(\cdot - x)_+^{m-1}$  devient un polynôme d'ordre  $m$ ,  $(\cdot - x)^{m-1}$ , dont la différence divisée est 0, comme indiqué par la propriété (1.12). Notons aussi que le terme multipliant la différence divisée permet d'obtenir une base de fonctions normalisées. Pour le reste de ce mémoire, nous allégerons la notation des splines par  $N_i^m(\cdot)$  au lieu de  $N_i^m(\cdot|\mathbf{t})$ , bien que les splines dépendent des noeuds  $\mathbf{t}$  qui leur infèrent un certain degré de lissage.

Puisque les différences divisées peuvent s'écrire par récurrence, il en va de même pour les fonctions B-splines. De Boor (1978) a prouvé l'équivalence de l'expression (1.13) et de la définition récursive des B-splines.

**Définition 1.2.3.** Soit la séquence de noeuds  $t_1 = \dots = t_m < t_{m+1} < \dots < t_z < t_{z+1} = \dots = t_{z+m}$ .

Les B-splines (normalisées) d'ordre  $m$  sont définies de manière récursive

$$N_i^m(x) = \left( \frac{x - t_i}{t_{i+m-1} - t_i} \right) N_i^{m-1}(x) + \left( \frac{t_{i+m} - x}{t_{i+m} - t_{i+1}} \right) N_{i+1}^{m-1}(x),$$

avec

$$N_i^1(x) = \begin{cases} 1 & \text{si } t_i \leq x < t_{i+1}; \\ 0 & \text{sinon.} \end{cases}$$

Par convention, lorsque les noeuds sont confondus (multiplicité des noeuds), nous posons  $\frac{0}{0} = 0$ .

Cette construction de B-splines a l'avantage d'être plus facile à compiler numériquement que les splines par différences divisées. Nous évaluons, en exemple, une fonction B-spline par cette méthode récursive.

**Exemple 1.2.2.** Soit le vecteur de noeuds  $(t_1, \dots, t_9)^T = (4, 4, 4, 12, 14, 16, 18, 18, 18)^T$ . Nous voulons évaluer la 3<sup>e</sup> B-spline d'ordre 3 de la base,  $N_3^3(x)$ , au

point  $x = 10$ . Nous procédons par récurrence :

$$\begin{aligned} N_3^3(x) &= \left( \frac{x - t_3}{t_5 - t_3} \right) N_3^2(x) + \left( \frac{t_6 - x}{t_6 - t_4} \right) N_4^2(x), \\ &= \left( \frac{x - t_3}{t_5 - t_3} \right) \left[ \left( \frac{x - t_3}{t_4 - t_3} \right) N_3^1(x) + \left( \frac{t_5 - x}{t_5 - t_4} \right) N_4^1(x) \right] \\ &\quad + \left( \frac{t_6 - x}{t_6 - t_4} \right) \left[ \left( \frac{x - t_4}{t_5 - t_4} \right) N_4^1(x) + \left( \frac{t_6 - x}{t_6 - t_5} \right) N_5^1(x) \right]. \end{aligned}$$

En choisissant  $x = 10$ , seule la fonction  $N_3^1(x)$  est non nulle, car  $x \in [t_3, t_4) = [4, 12)$ . Nous avons ainsi

$$\begin{aligned} N_3^3(10) &= \left( \frac{10 - t_3}{t_5 - t_3} \right) \left( \frac{10 - t_3}{t_4 - t_3} \right) N_3^1(10), \\ &= \left( \frac{10 - 4}{14 - 4} \right) \left( \frac{10 - 4}{12 - 4} \right) = \frac{36}{80} = 0,45. \end{aligned}$$

Rappelons que nous avons obtenu le même résultat à l'exemple 1.2.1 lorsque nous avons normalisé la différence divisée. Ainsi, la forme récursive des B-splines est équivalente à sa notation par différences divisées.

Plusieurs propriétés des B-splines sont mises en évidence par la définition 1.2.3 :

- (i)  $N_i^m(x)$  est un polynôme par morceaux d'ordre  $m$  ;
- (ii)  $N_i^m(x) > 0$  pour  $t_i < x < t_{i+m}$  ;
- (iii)  $N_i^m(x) = 0$  pour  $x \notin [t_i, t_{i+m}]$  ;
- (iv)  $\sum_{i=1}^z N_i^m(x) = 1$ .

Bien qu'évidente, la propriété (i) indique que les B-splines sont en soi des splines. Par conséquent, les caractéristiques appréciées chez les polynômes se retrouvent également chez les fonctions  $N_i^m$ . Nous soulignons entre autres l'aisance des manipulations algébriques et numériques des fonctions et leur caractère relativement lisse. Les propriétés (ii) et (iii) révèlent que les  $N_i^m$  sont une famille de fonctions non négatives ayant un support compact. Ainsi, parmi toutes les B-splines d'ordre  $m$  avec noeuds  $\mathbf{t}$ , seulement  $m$  B-splines peuvent être non nulles sur l'intervalle  $[t_j, t_{j+1}]$ , soient  $N_{j-m+1}^m, N_{j-m+2}^m, \dots, N_j^m$ . De plus, comme nous l'avons mentionné plus tôt, les B-splines sont par définition normalisées. En combinant les propriétés (ii)-(iv), il en découle que la suite des  $N_i^m$  (en fonction de  $i$ ) crée une partition de

la fonction unité. L'intérêt du support compact est lié à la flexibilité de l'estimateur spline, tel qu'exprimé en (1.11). Un changement dans un coefficient  $a_i$  affecte la spline  $s(x)$  seulement sur l'intervalle  $[t_i, t_{i+m}]$ . Les coefficients présentent donc une sensibilité locale, c'est-à-dire une flexibilité locale recherchée pour une bonne approximation.

Tout comme les polynômes, nous pouvons facilement dériver et intégrer les B-splines. Nous présentons ici la dérivée de la  $i^{\text{e}}$  B-spline d'ordre  $m$  :

$$[N_i^m(x)]' = \left( \frac{m-1}{t_{i+m-1} - t_i} \right) N_i^{m-1}(x) - \left( \frac{m-1}{t_{i+m} - t_{i+1}} \right) N_{i+1}^{m-1}(x).$$

Remarquons que la dérivée des B-splines se construit également de manière récursive. De plus, cette dérivée demeure une combinaison linéaire de B-splines à support fini : elle fait donc encore partie de la famille des splines. Nous verrons à la section suivante d'autres bases splines obtenues à partir des B-splines et qui partagent les mêmes propriétés.

### 1.3. SPLINES MONOTONES

Nous avons montré à la section précédente que les B-splines sont des fonctions non négatives. Ainsi, il est facile d'imposer des contraintes sur une spline  $s(x) = \sum_{i=1}^z a_i N_i^m(x)$  afin d'obtenir une approximation non négative. Dans le cas où nous avons des données monotones, soit croissantes ou décroissantes, il est d'intérêt de travailler avec des fonctions elles-mêmes monotones afin de faciliter l'imposition de contraintes.

Avant de développer les bases monotones, mentionnons que les B-splines ont été originellement introduites sous une autre normalisation par Curry et Schoenberg (1966). Ces bases, dénotées par  $M_i^m$ , sont associées aux B-splines par la relation suivante :

$$M_i^m(x) = \frac{m}{t_{i+m} - t_i} N_i^m(x).$$

Ces M-splines satisfont donc la normalisation  $\int_{-\infty}^{\infty} M_i^m(x) dx = 1$  et sont également non négatives. Ramsay (1988) procède à l'intégration des M-splines afin d'obtenir des fonctions monotones. Nous notons ces splines intégrées, ou I-splines, par  $I_i^m(x) = \int_a^x M_i^m(t) dt$ . Si la M-spline est d'ordre  $m$ , alors la I-spline est un

polynôme par morceaux d'ordre  $m + 1$ . Une écriture plus pratique pour construire les I-splines est, pour  $j = i, \dots, i + m - 1$ ,

$$I_i^m(x) = \begin{cases} 0, & \text{si } x < t_i, \\ \sum_{k=i}^j \frac{t_{k+m+1}-t_k}{m+1} M_k^{m+1}(x), & \text{si } x \in [t_j, t_{j+1}] \\ 1, & \text{si } x > t_{i+m}. \end{cases}$$

Pour effectuer une estimation monotone, Beliakov (2000) utilise une base de fonctions étroitement liées aux I-splines, mais qui diffèrent selon leur mode de définition. Il définit les fonctions  $T_i^m(x)$  comme une combinaison linéaire de B-splines, soit

$$T_i^m(x) = \sum_{j=i}^z N_j^m(x), \quad i = 1, \dots, z.$$

Ainsi, pour l'ensemble de B-splines d'ordre  $m$  définies sur les noeuds  $t_1, \dots, t_{z+m}$ ,  $z$  étant le nombre de paramètres libres dans l'estimation, les fonctions  $T_i^m$  sont linéairement indépendantes. Ces fonctions, appelées T-splines dû à leur forme « trapézoïdale » qui est surtout apparente pour les splines linéaires, forment donc également une base de l'espace spline  $\mathcal{S}_m(\mathbf{t})$ . Toute spline peut donc être écrite comme une combinaison linéaire de T-splines. La figure 1.2 nous permet de comparer l'allure des bases de fonctions B-splines et de fonctions T-splines quadratiques, toutes deux définies sur l'intervalle  $[0, 1]$  avec les noeuds intérieurs 0,3 et 0,6.

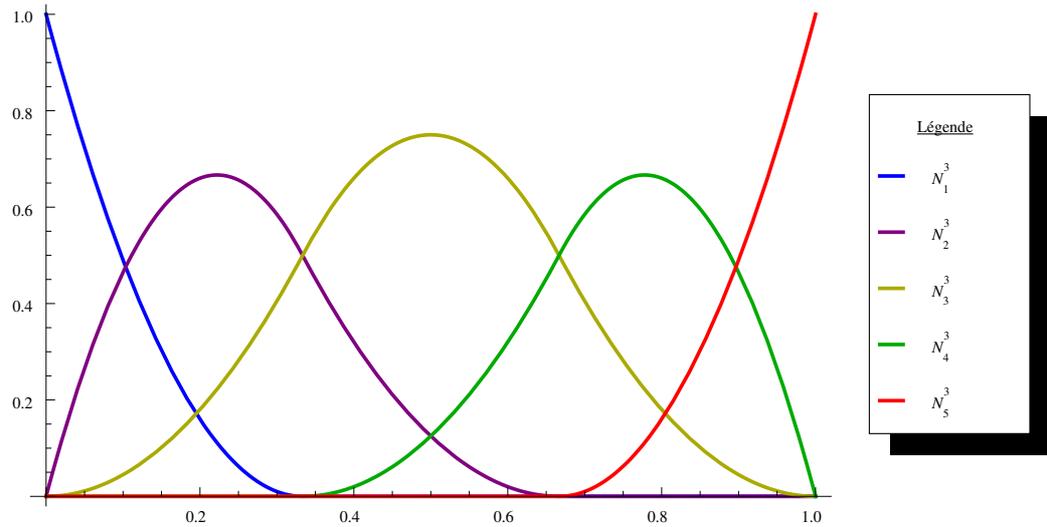
Cette nouvelle base de T-splines possède plusieurs propriétés intéressantes :

- (i) le support de  $T_i^m(x)$  est  $[t_i, t_{z+m}]$  ;
- (ii)  $T_i^m(x) = 1$  pour  $x \in [t_{i+m-1}, t_{z+1}]$  et  $i = 1, \dots, z - 1$  ;
- (iii) pour  $x \in [t_i, t_{i+1}]$ , la dérivée est

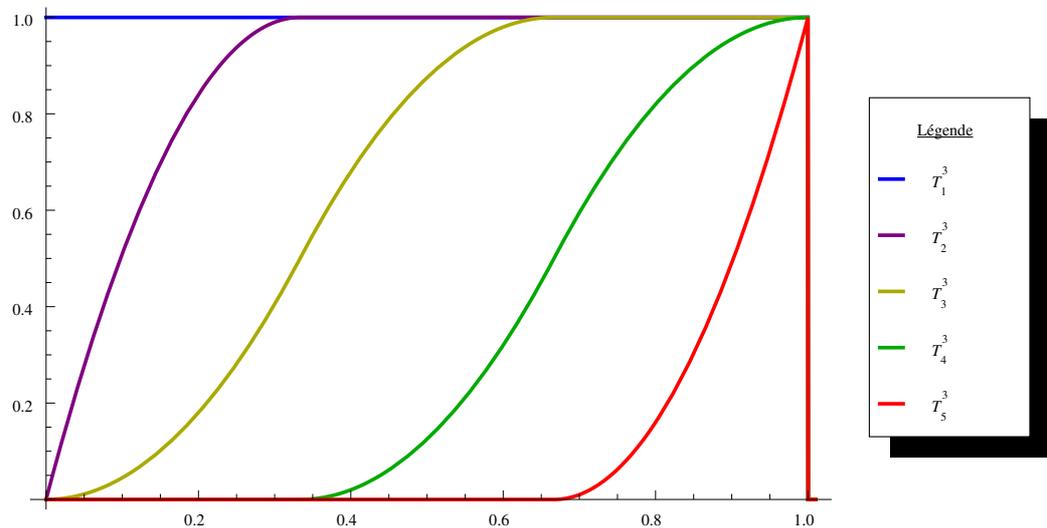
$$[T_i^m(x)]' = (m - 1) \left( \frac{N_i^{m-1}(x)}{t_{i+m-1} - t_i} - \frac{N_{z+1}^{m-1}(x)}{t_{z+m} - t_{z+1}} \right) \quad (1.14)$$

$$= \frac{m - 1}{t_{i+m-1} - t_i} N_i^{m-1}(x). \quad (1.15)$$

Les T-splines sont des fonctions monotones, mais contrairement aux I-splines, elles conservent un support fini. En effet, la propriété (i) laisse entendre que  $I_i^m$  et  $T_i^m$  coïncident sur l'intervalle  $[a, b]$ , mais diffèrent sur  $(b, \infty)$ . La propriété (ii)



(a)



(b)

FIGURE 1.2: Exemple de fonctions (a) B-splines et (b) T-splines

découle directement de la normalisation des B-splines. Nous obtenons l'équation (1.15) pour la dérivée grâce à la définition 1.2.3 des B-splines. En effet, puisque les noeuds  $t_{z+1} = \dots = t_{z+m} = b$  sont confondus et  $N_{z+1}^m = 0$  pour tout  $x \in [a, b]$ , alors le deuxième terme de l'expression (1.14) est nul par définition.

Les T-splines sont facilement calculables à partir des B-splines, donc elles sont pratiques pour effectuer des manipulations numériques. Il en est de même pour

leurs dérivées et leurs intégrales, ce qui facilite l'imposition de contraintes sur la forme de l'estimateur spline. Nous verrons à la section 1.5 comment appliquer ces contraintes de monotonie sur l'estimateur.

#### 1.4. CHOIX DE NOEUDS

Considérons le modèle  $y_i = g(x_i) + \epsilon_i$ ,  $i = 1, \dots, n$ , où la fonction inconnue  $g : [a, b] \rightarrow \mathbb{R}$  est approximée par la spline

$$s(x) = \sum_{i=1}^z \theta_i b_i^m(x),$$

où  $b_i^m$  est une base spline quelconque d'ordre  $m$  avec la suite de noeuds  $t_1 = \dots = t_m < \dots < t_{z+1} = \dots = t_{z+m}$ , dénotée par  $\mathbf{t}$ . Le choix des noeuds constitue une étape primordiale à la construction des splines. En effet, le nombre  $z$  de paramètres libres, qui spécifient la fonction spline, découle directement du nombre de noeuds. De plus, les noeuds intérieurs, soient  $t_{m+1}, \dots, t_z$ , déterminent la forme des bases splines et donc de l'approximation finale.

L'emplacement des noeuds est l'un des items qui différencie les splines d'interpolation et les splines de régression. Dans le cas de l'interpolation, les noeuds correspondent aux points d'observation, donc le nombre de paramètres est équivalent à la taille de notre échantillon. Pour les splines de régression, les noeuds sont positionnés à des valeurs fixes sur l'intervalle  $[a, b]$  et sont généralement moins nombreux. Par conséquent, les splines de régression nécessitent moins de paramètres libres à estimer à l'aide des données. Les applications statistiques nécessitant un nombre très restreint de paramètres privilégient les splines de régression. Bien sûr, l'obtention d'un modèle parcimonieux se fait aux dépens d'une certaine flexibilité, alors certains compromis doivent être respectés afin de conserver une bonne approximation.

Les deux principes de base qui régissent l'emplacement et le nombre de noeuds sont les suivants. Plus il y a de noeuds dans une région, plus la spline est flexible dans cette même région. En théorie, si une fonction présente de brusques changements sur une région donnée, l'ajout de noeuds est nécessaire afin d'y obtenir

un bon ajustement. Toutefois, en pratique, un nombre restreint de noeuds intérieurs donne assez de flexibilité pour modéliser une telle courbure sur un intervalle donné. Par ailleurs, plus il y a de points entre deux noeuds consécutifs, mieux sera définie la courbe sur cet intervalle. Entre autres, il doit toujours y avoir au moins une donnée entre chaque noeud. En effet, il est inutile de placer plusieurs noeuds dans une région où sa faible densité de points indique que la fonction  $y$  est moins bien définie.

Plusieurs procédures existent dans la littérature afin de déterminer l'emplacement et le nombre de noeuds. Une méthode simple consiste à choisir des noeuds équidistants pour un nombre  $z$  fixé d'avance. Cependant, cette technique n'offre pas un compromis optimal entre une grande flexibilité (donc un bon ajustement) et un faible nombre de noeuds. Plusieurs auteurs, dont Friedman et Silverman (1989), Shi et Li (1995) et He et Shi (1996, 1998), ont avancé une méthode pas à pas descendante (*backward selection*) pour sélectionner les noeuds. À partir d'un ensemble de noeuds équidistants, aussi appelés noeuds uniformes, l'approche consiste à enlever un noeud à la fois jusqu'à l'obtention d'un emplacement optimal. L'ensemble contenant un noeud entre chaque observation peut également servir de point de départ pour la méthode pas à pas descendante. Afin de déterminer si la suppression d'un noeud est optimale ou non, en termes de flexibilité et du nombre minimal de noeuds, la validation croisée généralisée (Friedman et Silverman, 1989; Shi et Li, 1995) et les critères d'information (He et Shi, 1996 et 1998; Kooperberg et Stone, 1992) sont fréquemment utilisés. Dans ce mémoire, nous considérons les critères d'information, les plus connus étant ceux de Akaike (AIC, 1974), de Schwarz (Bayesian Information Criterion - BIC, 1978) et de Hannan-Quinn (1979). De manière générale, un critère d'information prend la forme

$$C_n(z) = 2 \log(L_n(z))/n + z \phi(n)/n, \quad (1.16)$$

où  $L_n(z)$  est la vraisemblance maximale d'un modèle avec  $z$  paramètres basée sur un échantillon de  $n$  données. Le premier terme mesure la qualité de l'estimation, donc la flexibilité de la spline. Le second terme de l'équation (1.16) correspond à une pénalité pour le nombre de paramètres. Notons que  $\phi(n) = 2$

pour l'AIC,  $\phi(n) = \log(n)$  pour le BIC et  $\phi(n) = 2 \log(\log(n))$  pour le critère de Hannan-Quinn. En utilisant un de ces critères, le nombre de noeuds sélectionnés correspond à

$$\hat{z} = \arg \min_z C_n(z).$$

Notons que  $\hat{z}$  est nécessairement supérieur ou égal à  $m$ , l'ordre de la spline, puisque les  $m$  premiers noeuds sont confondus à l'extrémité du domaine pour respecter les conditions de lissage.

L'utilisation de noeuds uniformes est suffisante lorsque la fonction à approximer ne présente pas de changements brusques dans ses dérivées (autrement dit, la pente de la courbe ne doit pas changer brusquement). Nous nous intéressons plutôt à une séquence de noeuds dont l'emplacement n'est pas uniforme. Un tel espacement permet un meilleur ajustement des fonctions présentant des comportements locaux différents selon la région du domaine. Nous suivrons la stratégie pas à pas décrite précédemment, en appliquant quelques modifications proposées par He et Shi (1998). Bien que ces deux auteurs suggèrent de prendre les quantiles des valeurs distinctes de  $x_i$  comme noeuds, nous préférons diviser le domaine par rapport aux valeurs de  $y_i$  pour tenir compte de l'allure de la fonction. Nous débutons alors avec l'ensemble de noeuds intérieurs  $t_{j+m} = x_{[jn/k]}$ , autrement dit le point  $x_i$  associé au  $(j/k)^e$  quantile des valeurs de  $y_i$  pour  $j = 1, \dots, k$ . Puisque le nombre de paramètres libres est relié au nombre de noeuds intérieurs par  $z = m + k$  pour un ordre fixé, estimer  $z$  revient à estimer  $k$ . Pour sélectionner le paramètre  $k$ , He et Shi (1998) ont développé leur propre critère d'information à minimiser, soit

$$IC(k) = \log \left( \sum_{i=1}^n |y_i - \hat{g}(x_i)| \right) + 2(k+2)/n. \quad (1.17)$$

Si l'estimateur de  $k$  trouvé est supérieur à 2, nous procédons ensuite à la suppression itérative des noeuds, décrite précédemment. Nous conservons le sous-modèle présentant la valeur  $IC$  la plus faible. Remarquons qu'au lieu de procéder à une sélection pas à pas descendante, il est possible d'utiliser une méthode ascendante.

## 1.5. CONTRAINTES DE MONOTONIE

Considérons le modèle

$$y_i = g(x_i) + \epsilon_i, \quad i = 1, \dots, n.$$

Les données peuvent présenter certaines caractéristiques que nous désirons retrouver dans l'estimation. En effet, un défi souvent rencontré est de conserver l'allure ou la forme de la courbe de données lors de la modélisation. Nous nous intéressons particulièrement aux données monotones croissantes ou décroissantes, lorsqu'il est, par exemple, question de fonctions de répartition, de survie ou de courbes de croissance. Sans perte de généralité, nous ne considérons ici que les données non décroissantes : nos estimations doivent donc satisfaire  $\hat{g}(x^*) \geq \hat{g}(x)$  pour tout  $x^* \geq x$ .

Plusieurs travaux ont été publiés à cet effet. Entre autres, Fritsch et Carlson (1980) ont travaillé sur l'interpolation monotone avec des fonctions cubiques par morceaux. Similairement, Schumaker (1983) a considéré l'interpolation par splines quadratiques. Au lieu d'appliquer les contraintes de monotonie sur des interpolants, nous préférons travailler avec des splines de régression qui permettent la construction de modèles plus parcimonieux. Dans un tel contexte de régression, une base adéquate est nécessaire pour faciliter l'imposition de contraintes. Nous étudions ici l'aspect des contraintes de monotonie lorsque les T-splines sont utilisées.

Considérons l'estimation de la fonction  $g$  donnée par

$$\hat{g}(x) = \sum_{i=1}^z \hat{\theta}_i b_i^m(x),$$

où  $b_i^m$  est la fonction de base choisie et  $\hat{\theta}_i$  sont les coefficients estimés. Lorsque  $b_i^m = T_i^m$ , nous travaillons avec des fonctions non décroissantes. Ainsi, une condition suffisante, mais non nécessaire, pour obtenir une estimation monotone non décroissante est

$$\hat{\theta}_i \geq 0, \quad \forall i.$$

En effet, la dérivée de  $T_i^m$  est proportionnelle à  $N_i^{m-1}$ , c'est-à-dire qu'elle est non négative. Par conséquent, la dérivée de l'estimateur spline  $\hat{g}$  est une combinaison

linéaire de fonctions non négatives : il suffit alors de prendre une séquence de coefficients  $\hat{\theta}_i$  non négatifs. Cette démarche est basée sur le lien entre la monotonie d'une fonction et sa dérivée : si  $g'(x) \geq 0 \quad \forall x$ , alors la fonction  $g(x)$  est non décroissante.

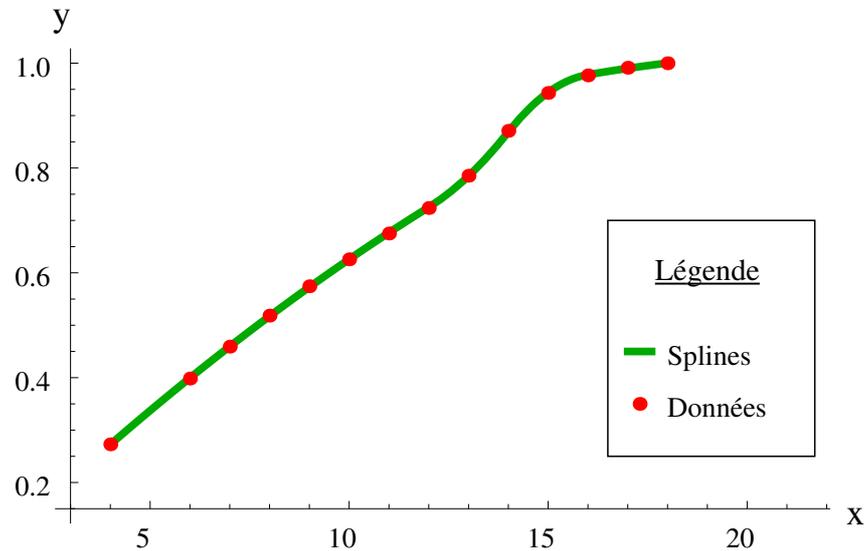


FIGURE 1.3: Approximation par T-splines d'une courbe de croissance normalisée

Les contraintes de monotonie se simplifient pour une spline d'ordre 2 ou 3, autrement dit pour une spline linéaire ou quadratique. Dans ce cas, la contrainte sur les coefficients est moins restrictive. Les coefficients  $\hat{\theta}_i$  doivent satisfaire

$$\hat{\theta}_i \geq 0 \text{ pour } i = 2, \dots, z.$$

Remarquons qu'il n'y a aucune contrainte imposée au premier coefficient  $\hat{\theta}_1$ , car ce dernier détermine la valeur de l'estimateur au point  $x = a$ . De plus, cette nouvelle contrainte est suffisante et nécessaire afin d'obtenir le meilleur estimateur monotone possible, contrairement aux splines d'ordre supérieur à 3.

**Exemple 1.5.1.** *Considérons le jeu de données monotones dans le tableau 1.1. Afin d'obtenir un estimateur monotone, nous modélisons les données avec des fonctions T-splines en imposant les contraintes de non-négativité sur les coefficients. En travaillant avec les T-splines quadratiques associées au vecteur de noeuds  $\mathbf{t} = (4, 4, 4, 12, 14, 16, 18, 18, 18)^T$ , nous obtenons une approximation de la*

forme

$$s(x) = \sum_{i=1}^6 \theta_i T_i^3(x),$$

où les  $\theta_i$  sont estimés par la méthode des moindres carrés sous contraintes d'inégalités. Nous obtenons ainsi le vecteur de coefficients estimés  $\hat{\theta} = (0,2726, 0,2633, 0,2366, 0,1925, 0,0261, 0,0091)^T$ . La figure 1.3 illustre l'approximation par splines qui, contrairement à l'interpolation de Lagrange dans l'exemple 1.1.1, préserve la monotonie sur l'intervalle  $[4, 18]$ . De plus, le modèle avec splines nécessite moins de coefficients que l'interpolation pour approximer la courbe de données. En fait, l'interpolation par polynômes de Lagrange considère les 14 observations comme des coefficients, tandis que la spline  $s(x)$  ne requiert que 6 coefficients, soit 43% des données initiales.

L'imposition de contraintes dans un modèle bidimensionnel est plus compliquée, comme nous le verrons à la section 2.2. Au prochain chapitre, nous étudierons des données monotones dans un contexte bidimensionnel. Entre autres, nous généraliserons les splines vues dans ce chapitre aux splines bivariées, de même que les contraintes. Nous passerons de la modélisation classique à une approche bayésienne afin d'ajuster nos données.

## Chapitre 2

---

### MODÉLISATION BIDIMENSIONNELLE

Dans ce chapitre, il est question de modéliser des tables à deux entrées telles que des tables de mortalité, des tables de quantiles et de la fonction de répartition de la loi normale. Nous sommes intéressés à approximer chaque entrée de la table en n'utilisant qu'un nombre restreint de paramètres. Donc, nous cherchons à réduire l'espace nécessaire afin de recréer une table de données. De plus, nous désirons une estimation qui conserve les propriétés de la table : nous considérons ici la monotonie des données. Notre contribution consiste à proposer un modèle *a priori* utilisant des splines et à obtenir un estimateur monotone de tables.

Nous avons ainsi des données de la forme

$$(x_i, y_j, f_{i,j}), \quad i = 1, \dots, n_x \text{ et } j = 1, \dots, n_y,$$

pour un total de  $n = n_x \times n_y$  observations. Nous considérons cette fois le modèle

$$f_{i,j} = g(x_i, y_j) + \epsilon_{i,j}, \quad i = 1, \dots, n_x, \quad j = 1, \dots, n_y, \quad (2.1)$$

où  $(x_i, y_j) \in [a, b] \times [c, d]$  et  $g$  est une fonction bidimensionnelle inconnue, et donc à estimer. Nous supposons la fonction  $g$  lisse, soit  $g \in \mathcal{W}_2^m([a, b] \times [c, d])$ . Les  $\epsilon_{i,j}$  sont des termes d'erreurs que nous supposons non corrélées et identiquement distribuées :  $\epsilon_{i,j} \sim \mathcal{N}(0, \sigma^2)$ .

Le prochaine section se veut une extension des notions vues précédemment afin d'introduire une modélisation pour les données bidimensionnelles. Nous introduisons l'approche bayésienne ainsi que les techniques de calculs numériques afin d'obtenir une estimation des données.

## 2.1. CONTEXTE BIDIMENSIONNEL

### 2.1.1. Monotonie

La monotonie pour une fonction bidimensionnelle est moins intuitive que pour le cas unidimensionnel. Carlson et Fritsch (1985) spécifient ce concept pour une fonction  $g$  en se basant sur les applications physiques qui génèrent des données monotones. Nous caractérisons alors la fonction  $g(x, y)$  de monotone si toutes les tranches unidimensionnelles de  $g$  prises parallèlement aux axes de coordonnées sont monotones.

**Définition 2.1.1.** *Une fonction  $g$  ayant le domaine  $\mathcal{D} = [a, b] \times [c, d]$  est monotone sur  $\mathcal{D}$  si*

(1) *Pour chaque point fixé  $x = x^*$  et  $y_1 \leq y_2$ ,*

$$s_1 \cdot g(x^*, y_1) \leq s_1 \cdot g(x^*, y_2),$$

où

$$s_1 = \begin{cases} +1, & \text{si } g \text{ est monotone croissante en } y, \\ -1, & \text{si } g \text{ est monotone décroissante en } y. \end{cases}$$

(2) *Pour chaque point fixé  $y = y^*$  et  $x_1 \leq x_2$ ,*

$$s_2 \cdot g(x_1, y^*) \leq s_2 \cdot g(x_2, y^*)$$

où

$$s_2 = \begin{cases} +1, & \text{si } g \text{ est monotone croissante en } x, \\ -1, & \text{si } g \text{ est monotone décroissante en } x. \end{cases}$$

Les valeurs  $s_1$  et  $s_2$  sont constantes, dans le sens que si  $s_1 = +1$  pour une valeur de  $x^*$ , alors  $s_1 = +1$  pour toutes les valeurs de  $x^* \in [a, b]$ . Notons qu'il est possible d'avoir  $s_1 \neq s_2$ , c'est-à-dire une fonction monotone croissante en  $x$  et décroissante en  $y$ , ou vice versa. Nous verrons à la section 2.2 comment les contraintes de monotonie sur l'estimateur s'adaptent dans une telle situation.

Pour certaines tables à deux entrées, seule la monotonie en  $y$  (respectivement en  $x$ ) est requise. En restant dans le contexte physique de Carlson et Fritsch

(1985), il s'agit d'une monotonie partielle, ce qui revient à dire que  $g(x, y)$  est une fonction unidimensionnelle monotone pour un  $x$  fixé (respectivement  $y$ ).

### 2.1.2. Splines bidimensionnelles

Il est possible de généraliser la théorie des splines unidimensionnelles à une application avec plusieurs variables. Parmi les méthodes développées, nous retrouvons le produit tensoriel (produit de Kronecker) de splines, comme introduit par de Boor (1978).

Les splines bidimensionnelles par produit de Kronecker sont définies comme suit.

**Définition 2.1.2.** *Soient les suites strictement croissantes*

$$a = \xi_0 < \xi_1 < \cdots < \xi_g < \xi_{g+1} = b$$

et

$$c = \nu_0 < \nu_1 < \cdots < \nu_h < \nu_{h+1} = d.$$

La fonction  $s(x, y)$  dénote une spline bidimensionnelle (par produit de Kronecker) sur  $\mathcal{D} = [a, b] \times [c, d]$  d'ordre  $k$  en  $x$  et  $l$  en  $y$ , avec les points de cassure  $\xi_0, \dots, \xi_{g+1}$  associés aux  $x$  et  $\nu_0, \dots, \nu_{h+1}$  aux  $y$ , si les conditions suivantes sont satisfaites :

(1) Sur chaque sous-rectangle  $\mathcal{D}_{i,j} = [\xi_i, \xi_{i+1}] \times [\nu_j, \nu_{j+1}]$ ,  $s(x, y)$  est donnée par un polynôme d'ordre  $k$  en  $x$  et  $l$  en  $y$  :

$$s|_{\mathcal{D}_{i,j}} \in \mathcal{P}_k \otimes \mathcal{P}_l, \quad i = 0, \dots, g; j = 0, \dots, h,$$

où  $\mathcal{P}_k$  est l'espace polynomial d'ordre  $k$ .

(2) La fonction  $s(x, y)$  et toutes ses dérivées partielles,  $\partial^{i+j} s(x, y) / \partial x^i \partial y^j$ , sont continues sur  $\mathcal{D}$ , c'est-à-dire

$$\frac{\partial^{i+j}}{\partial x^i \partial y^j} s(x, y) \in \mathcal{C}(\mathcal{D}), \quad i = 0, \dots, k - 2; j = 0, \dots, l - 2.$$

Cette définition, avec quelques modifications appropriées, est aussi valable pour des splines de dimensions supérieures. Toutefois, dans le cadre des tables de données, nous ne présentons que le cas bidimensionnel.

Il est préférable de travailler avec une base de fonctions splines, telles les B-splines, qu'avec une base de puissances tronquées pour simplifier les calculs

numériques. Nous introduisons alors les vecteurs de noeuds  $\mathbf{t}$  et  $\boldsymbol{\tau}$  en suivant la définition 1.2.2, où les noeuds sont confondus aux extrémités du domaine

$$\begin{aligned}\xi_0 = t_1 = \cdots = t_k < t_{k+1} < \cdots < t_{z_1} < t_{z_1+1} = \cdots = t_{z_1+k} = \xi_{g+1}, \\ \nu_0 = \tau_1 = \cdots = \tau_l < \tau_{l+1} < \cdots < \tau_{z_2} < \tau_{z_2+1} = \cdots = \tau_{z_2+l} = \nu_{h+1}.\end{aligned}$$

Chaque spline  $s(x, y)$  est alors une combinaison linéaire de tous les produits possibles entre la base de B-splines d'ordre  $k$  en  $x$  et la base de B-splines d'ordre  $l$  en  $y$ . Autrement dit, une spline bidimensionnelle a une représentation unique

$$s(x, y) = \sum_{i=1}^{z_1} \sum_{j=1}^{z_2} \theta_{i,j} N_i^k(x) N_j^l(y),$$

où  $N_i^k$  et  $N_j^l$  sont des B-splines normalisées définies respectivement sur les vecteurs de noeuds  $\mathbf{t}$  et  $\boldsymbol{\tau}$ .

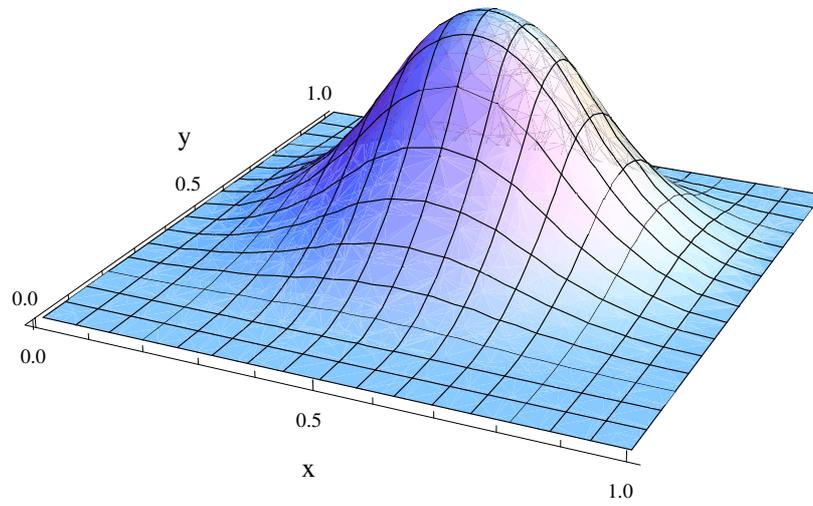
Les propriétés du produit de Kronecker de B-splines  $N_i^k(x) N_j^l(y)$ , également noté  $N_{i,j}^{k,l}(x, y)$ , découlent directement du cas unidimensionnel. En particulier,

- (i)  $N_i^k(x) N_j^l(y) > 0$  pour  $t_i < x < t_{i+k}$  et  $\tau_j < y < \tau_{j+l}$ ;
- (ii)  $N_i^k(x) N_j^l(y) = 0$  pour  $(x, y) \notin [t_i, t_{i+k}] \times [\tau_j, \tau_{j+l}]$ ;
- (iii)  $\sum_{i=1}^{z_1} \sum_{j=1}^{z_2} N_i^k(x) N_j^l(y) = 1, \quad \forall (x, y) \in \mathcal{D}$ .

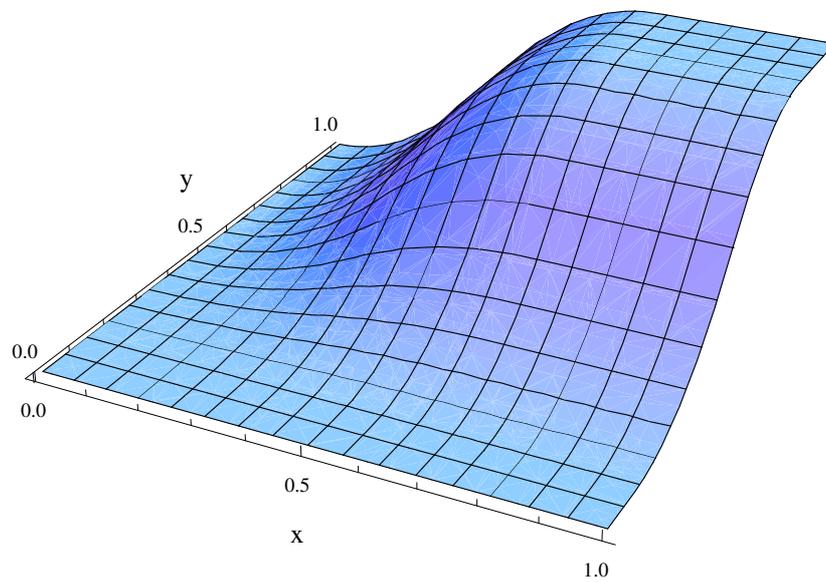
Les propriétés (i) et (ii) indiquent que les B-splines construites par produit de Kronecker sont encore une famille de fonctions non négatives à support compact. En effet, parmi les  $z_1 \times z_2$  fonctions de base  $N_{i,j}^{k,l}$ , seulement  $kl$  fonctions peuvent être non nulles sur la surface  $[t_i, t_{i+k}] \times [\tau_j, \tau_{j+l}]$ . L'estimateur spline obtenu conserve donc une certaine flexibilité locale. Par ailleurs, la normalisation des B-splines s'étend également au cas bidimensionnel, comme illustré en (iii).

Comme à la section 1.3, nous pouvons modifier la base de B-splines bidimensionnelles afin de travailler avec des surfaces monotones. Ainsi, en remplaçant les B-splines par des T-splines, nous facilitons l'obtention d'une estimation monotone. Les caractéristiques des T-splines sont généralisables au cas bidimensionnel. Le produit de Kronecker de T-splines,  $T_{i,j}^{k,l}(x, y) = T_i^k(x) T_j^l(y)$ , permet alors

- (i) le support de  $T_{i,j}^{k,l}(x, y)$  est  $[t_i, t_{z_1+k}] \times [\tau_j, \tau_{z_2+l}]$ ;
- (ii)  $T_{i,j}^{k,l}(x, y) = 1$  pour  $x, y \in [t_{i+k-1}, t_{z_1+k}] \times [\tau_{j+l-1}, \tau_{z_2+l}]$ .



(a)



(b)

FIGURE 2.1: Exemple de fonctions bidimensionnelles (a) B-splines et (b) T-splines

Bien que les T-splines bidimensionnelles aient un support fini, la figure 2.1 montre que les B-splines ont un domaine de définition plus restreint. Par ailleurs, il est à remarquer que les T-splines à deux variables forment des surfaces monotones croissantes qui plafonnent à 1.

L'avantage des splines construites par produit de Kronecker est la simplicité avec laquelle nous pouvons les évaluer, les dériver et les intégrer. Le principal désavantage de ces splines est que nous sommes restreints à des domaines d'approximation rectangulaires ou des domaines qui peuvent facilement être transformés en rectangle. De plus, la subdivision du domaine est aussi rectangulaire dû à la position des noeuds, ce qui réduit la flexibilité de la partition. Une alternative aux produits de Kronecker est l'usage de splines définies sur une triangulation arbitraire. Nous mentionnons entre autres les splines de Powell-Sabin et les splines du simplexe, qui sont surtout utilisées pour reconstruire des surfaces. Cependant, ces splines multidimensionnelles sont beaucoup plus complexes et moins attrayantes d'un point de vue computationnel. Un bref aperçu de ces fonctions est fourni par Dierckx (1995, chapitre 2).

Dans notre contexte de tables à deux entrées, nous travaillons avec un domaine rectangulaire. Par conséquent, l'utilisation de produits de Kronecker est justifiée. Afin d'alléger la notation, nous référerons aux splines construites par produit de Kronecker avec une base de T-splines par splines bidimensionnelles pour le reste de ce mémoire.

### 2.1.3. Généralisation du choix de noeuds

L'implémentation d'une approximation par T-splines requiert une sélection de noeuds. Puisque les splines bidimensionnelles que nous utilisons sont formées par le produit de deux bases de T-splines, il faut déterminer l'emplacement et le nombre de noeuds pour deux vecteurs,  $\mathbf{t}$  et  $\boldsymbol{\tau}$ . Nous retombons en quelque sorte sur l'approche unidimensionnelle détaillée à la sous-section 1.4. Dans cette section, nous survolons les modifications à y apporter pour tenir compte des deux vecteurs de noeuds.

He et Shi (1996) implémentent l'approximation par produits de Kronecker de B-splines dans un contexte de régression linéaire partielle. Ils considèrent la sélection de noeuds par validation croisée et par critère d'information. Nous retenons le critère de l'équation (1.17), dont l'extension à deux dimensions devient

$$IC(k_1, k_2) = \log \left( \sum_{i=1}^{n_x} \sum_{j=1}^{n_y} |f_{i,j} - \hat{g}(x_i, y_j)| \right) + 2(k_1 + k_2 + 2)/n, \quad (2.2)$$

où  $k_1$  et  $k_2$  sont les nombres de noeuds intérieurs des vecteurs  $\mathbf{t}$  et  $\boldsymbol{\tau}$  tandis que  $\hat{g}$  est l'estimation de la fonction  $g$ . L'emplacement des noeuds peut être uniforme ou non uniforme si la fonction  $g$  présente des comportements locaux très différents selon la région. L'avantage des noeuds uniformes dans le cas bidimensionnel est l'efficacité et la rapidité avec lesquelles la sélection est faite. En effet, puisque les noeuds intérieurs prennent la forme  $t_{k+i} = a + (b - a)\frac{i}{(k_1+1)}$ ,  $i = 1, \dots, k_1$ , et  $\tau_{l+j} = c + (d - c)\frac{j}{(k_2+1)}$ ,  $j = 1, \dots, k_2$ , alors seul le nombre de noeuds doit être déterminé. Remarquons aussi que le nombre de noeuds pour chaque vecteur peut être différent sans compliquer le processus de sélection.

Lorsque nous travaillons avec des noeuds non uniformes, le processus de sélection de noeuds se fait de manière progressive. Nous considérons un sous-ensemble d'emplacements à partir des valeurs distinctes du jeu de données. Les noeuds seront sélectionnés à partir de cet ensemble. Soit une suite initiale de noeuds,  $\{t_1, \dots, t_{m_1}, \tau_1, \dots, \tau_{m_2}\}$ . L'algorithme consiste à ajouter un noeud  $t_{m_1+1}$  ou  $\tau_{m_2+1}$  tel que le nouvel ensemble de noeuds soit préféré par le critère d'information à tout autre ensemble  $\{t_1, \dots, \tau_{m_2}, t_m$  ou  $\tau_m\}$ . Cette procédure de placement s'arrête lorsque l'addition d'un noeud n'optimise plus le critère (2.2). L'algorithme se poursuit en supprimant un noeud à la fois : le noeud supprimé est celui dont l'absence est préférée par le critère de sélection. L'algorithme prend donc fin lorsque la suppression de noeuds n'est plus optimale. Lors du choix de la suite de noeuds initiale, nous voulons tenir compte de l'allure de la fonction à estimer. Par conséquent, nous choisissons les noeuds  $\mathbf{t}$  (respectivement  $\boldsymbol{\tau}$ ) comme les valeurs  $x$  (respectivement  $y$ ) associées aux quantiles empiriques de  $f_{i,j}$ . Cette sélection favorise la flexibilité de la spline dans les régions où la surface présente de brusques changements, donc favorise un bon ajustement.

## 2.2. CONTRAINTES

La définition de monotonie 2.1.1 suggère l'imposition de contraintes sur chaque argument individuellement. La contrainte suivante est posée par Beliakov (2000) lorsque le produit de Kronecker des T-splines est utilisé.

**Théorème 2.2.1.** *Soit une spline bidimensionnelle de la forme*

$$s(x, y) = \sum_{i=1}^{z_1} \sum_{j=1}^{z_2} \theta_{i,j} T_{i,j}^{m,l}(x, y).$$

*De manière générale, les coefficients  $\theta_{i,j}$  doivent respecter l'inégalité suivante pour assurer une approximation monotone croissante par rapport à l'argument  $x$  :*

$$\sum_{i=q-m+2}^q \sum_{j=1}^{z_2} \theta_{i,j} \frac{N_i^{m-1}(t_q) T_j^l(\tau_u)}{t_{i+m-1} - t_i} \geq 0, \quad q = m, \dots, z_1, \quad u = l, \dots, z_2.$$

*En particulier, pour une spline d'ordre  $m \leq 3$ , la contrainte devient*

$$\sum_{j=1}^u \theta_{q,j} \geq 0, \quad u = 1, \dots, z_2, \quad q = 2, \dots, z_1.$$

*Similairement, la contrainte de monotonie par rapport à l'argument  $y$  est donnée par*

$$\sum_{i=1}^q \theta_{i,u} \geq 0, \quad q = 1, \dots, z_1, \quad u = 2, \dots, z_2.$$

DÉMONSTRATION. Soit la base de T-splines bidimensionnelles,  $T_{i,j}^{m,l}(x, y)$ , ayant pour noeuds  $t_1, \dots, t_{m+z_1}$  sur l'argument  $x$  et  $\tau_1, \dots, \tau_{l+z_2}$  sur les  $y$ .

La condition de monotonie implique que toutes les dérivées partielles de la spline soient non négatives, et ce en chaque point. Montrons que  $\frac{\partial}{\partial x} s(x, y) \geq 0$  pour la monotonie sur  $x$ .

Puisque les fonctions  $T_{i,j}^{m,l}(x, y)$  sont des produits de Kronecker,  $T_{i,j}^{m,l}(x, y) = T_i^m(x) T_j^l(y)$ , leurs dérivées partielles sont proportionnelles aux dérivées de T-splines à une variable. Ainsi,

$$\begin{aligned} \frac{\partial}{\partial x} s(x, y) &= \sum_{i=1}^{z_1} \sum_{j=1}^{z_2} \theta_{i,j} \frac{\partial}{\partial x} T_{i,j}^{m,l}(x, y) \\ &= \sum_{i=1}^{z_1} \sum_{j=1}^{z_2} \theta_{i,j} T_j^l(y) \frac{\partial}{\partial x} T_i^m(x), \end{aligned}$$

et en incorporant la dérivée (1.15), nous obtenons

$$\frac{\partial}{\partial x}s(x, y) = \sum_{i=1}^{z_1} \sum_{j=1}^{z_2} \theta_{i,j} \frac{(m-1)N_i^{m-1}(x)T_j^l(y)}{t_{i+m-1} - t_i}. \quad (2.3)$$

Puisque les B-splines ont un support compact, seules les fonctions  $N_{q-m+2}^{m-1}, \dots, N_q^{m-1}$  sont non nulles pour  $x \in [t_q, t_{q+1}]$ . Alors en évaluant la dérivée partielle à  $x \in [t_q, t_{q+1}]$ ,  $q = m, \dots, z_1$ , l'équation (2.3) se simplifie

$$\frac{\partial}{\partial x}s(x, y) \propto \sum_{i=q-m+2}^q \sum_{j=1}^{z_2} \theta_{i,j} \frac{N_i^{m-1}(x)T_j^l(y)}{t_{i+m-1} - t_i}.$$

En particulier, pour un ordre  $m = 3$ , seule la base  $N_q^{m-1}$  est non nulle lorsqu'évaluée à  $t_{q+1}$ ,

$$\begin{aligned} \frac{\partial}{\partial x}s(x, y) \Big|_{x=t_{q+1}} &\propto \sum_{j=1}^{z_2} \theta_{q,j} \frac{N_q^{m-1}(t_{q+1})T_j^l(y)}{t_{q+m-1} - t_q}, \quad q = 2, \dots, z_1 \\ &\propto \sum_{j=1}^{z_2} \theta_{q,j} T_j^l(y). \end{aligned}$$

Cette contrainte dépend d'un  $y$  fixé. Notons que pour  $y \in [\tau_i, \tau_{i+1}]$ ,  $T_{k+1}^l(y) = 0$ ,  $\forall k \geq i$ . Pour des valeurs de  $y$  se retrouvant dans n'importe quel sous-intervalle, nous généralisons la contrainte :

$$\sum_{j=1}^u \theta_{q,j} \geq 0, \quad u = 1, \dots, z_2, \quad q = 2, \dots, z_1.$$

Une argumentation semblable entraîne la contrainte de monotonie sur les  $y$ .  $\square$

### 2.3. MODÈLE HIÉRARCHIQUE DE BAYES

Afin de modéliser les données, nous voulons utiliser une approche bayésienne au lieu de l'approche classique. Les méthodes bayésiennes ont été introduites par Bayes (1763), mais développées et popularisées par Laplace durant le XVIII<sup>e</sup> siècle. Toutefois, la complexité des intégrations numériques associées aux modèles a interrompu l'essor de l'analyse bayésienne au profit de l'analyse fréquentiste (dite classique). Depuis une vingtaine d'années, il y a un regain d'intérêt pour les statistiques bayésiennes grâce à l'amélioration des techniques de calcul et à la performance des logiciels disponibles. Nous introduisons ainsi dans cette section

quelques notions de base afin de comprendre les modèles de Bayes (voir Robert, 2001, pour plus de détails).

Considérons l'approche paramétrique afin de modéliser un phénomène observé. Nous commençons par définir un modèle statistique paramétrique, qui est constitué des observations d'une variable aléatoire  $X$ , distribuée selon  $f(x|\theta)$ , où seul le paramètre  $\theta$  est inconnu et appartient à l'espace fini  $\Theta$ . L'objectif d'une telle modélisation est d'inférer sur le paramètre  $\theta$  en utilisant uniquement les observations  $x$ . Bayes et Laplace ont approfondi le modèle classique en considérant que l'incertitude du paramètre  $\theta$  pouvait être modélisée par une distribution de probabilité  $\pi$  sur  $\Theta$ , appelée distribution *a priori*. Cette addition nous mène à définir un modèle statistique bayésien.

**Définition 2.3.1.** *Un modèle statistique bayésien consiste en un modèle statistique paramétrique,  $f(x|\theta)$ , et d'une distribution a priori,  $\pi(\theta)$ , sur le paramètre  $\theta$ .*

Un modèle statistique de Bayes est donc caractérisé par l'information *a priori* du modèle, c'est-à-dire l'information connue avant de faire l'inférence, et par l'information contenue dans les données disponibles. L'intérêt d'une distribution *a priori* repose sur sa capacité de résumer l'information disponible, ou encore le manque d'information. L'inférence sur le paramètre  $\theta$  est basée sur le théorème de Bayes, également connu comme le théorème des probabilités inverses, qui stipule

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{\int_{\Theta} f(x|\theta)\pi(\theta) d\theta}. \quad (2.4)$$

Autrement dit, l'inférence est faite à partir de la distribution conditionnelle décrite à l'expression (2.4),  $\pi(\theta|x)$ , que nous appelons distribution *a posteriori*. Ainsi, le théorème de Bayes met à jour l'information sur  $\theta$  contenue dans  $\pi(\theta)$  en soutirant l'information contenue dans les observations  $x$ .

Le modèle bayésien peut prendre une forme hiérarchique lorsqu'il est question de décomposer nos connaissances *a priori*. En effet, une approche hiérarchique modélise le manque d'information sur les paramètres de la distribution *a priori* à travers une autre distribution *a priori* sur ces paramètres. Ce type de modèle *a priori* implique donc plusieurs niveaux de distributions conditionnelles.

**Définition 2.3.2.** *Un modèle hiérarchique de Bayes est un modèle statistique bayésien,  $(f(x|\theta), \pi(\theta))$ , où la distribution a priori  $\pi(\theta)$  est décomposée en distributions conditionnelles,*

$$\pi_1(\theta|\theta_1), \pi_2(\theta_1|\theta_2), \dots, \pi_n(\theta_{n-1}|\theta_n),$$

*et une distribution marginale  $\pi_{n+1}(\theta_n)$  telle que*

$$\pi(\theta) = \int_{\Theta_1 \times \dots \times \Theta_n} \pi_1(\theta|\theta_1) \pi_2(\theta_1|\theta_2) \cdots \pi_{n+1}(\theta_n) d\theta_1 \cdots d\theta_n. \quad (2.5)$$

*Les paramètres  $\theta_i$  sont appelés les hyperparamètres de niveau  $i$ ,  $1 \leq i \leq n$ .*

Autrement dit, pour un modèle hiérarchique

$$x \sim f(x|\theta), \theta \sim \pi_1(\theta|\theta_1), \dots, \theta_n \sim \pi_{n+1}(\theta_n), \quad (2.6)$$

nous pouvons récupérer le modèle usuel de Bayes  $x \sim f(x|\theta)$  et  $\theta \sim \pi(\theta)$  en utilisant pour distribution *a priori* la densité (2.5).

Les modèles hiérarchiques sont pratiques parce qu'ils permettent un conditionnement sur tous les niveaux, ce qui compense pour la complexité occasionnée par ces mêmes niveaux de conditionnement successifs. Par ailleurs, la décomposition facile de la densité *a priori* est transmise à la distribution *a posteriori*. Sans perte de généralité, considérons le modèle

$$x \sim f(x|\theta), \theta \sim \pi_1(\theta|\theta_1) \text{ et } \theta_1 \sim \pi_2(\theta_1)$$

afin d'illustrer ce résultat.

**Lemme 2.3.1.** *La distribution a posteriori de  $\theta$  est*

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

où

$$\pi(\theta|\theta_1, x) = \frac{f(x|\theta) \pi_1(\theta|\theta_1)}{m_1(x|\theta_1)},$$

$$m_1(x|\theta_1) = \int_{\Theta} f(x|\theta) \pi_1(\theta|\theta_1) d\theta,$$

$$\pi(\theta_1|x) = \frac{m_1(x|\theta_1) \pi_2(\theta_1)}{m(x)},$$

$$m(x) = \int_{\Theta_1} m_1(x|\theta_1) \pi_2(\theta_1) d\theta_1.$$

Ce résultat démontre que la distribution  $\pi(\theta|x)$  peut être simulée en générant  $\theta_1$  selon  $\pi(\theta_1|x)$ , puis  $\theta$  selon  $\pi(\theta|\theta_1, x)$ . Lorsque ces deux dernières distributions conditionnelles sont plus simples à manipuler que celle *a posteriori*, le calcul des estimateurs bayésiens est alors grandement simplifié. Ce résultat est entre autres utilisé par les méthodes de calcul de Monte Carlo. Nous présentons une autre caractéristique importante des modèles hiérarchiques qui influence l'obtention des estimateurs.

**Lemme 2.3.2.** *Pour le modèle hiérarchique de Bayes (2.6), la distribution conditionnelle complète de  $\theta_i$ , sachant  $x$  et les  $\theta_j$ , ( $j \neq i$ ), satisfait*

$$\pi(\theta_i|x, \theta, \theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_n) = \pi(\theta_i|\theta_{i-1}, \theta_{i+1})$$

avec  $\theta_0 = \theta$  et  $\theta_{n+1} = 0$  par convention.

L'importance de cette caractéristique est due au fait que les distributions conditionnelles d'un modèle hiérarchique dépendent uniquement d'hyperparamètres locaux. Les techniques de calcul comme l'échantillonnage de Gibbs sont alors adaptées afin de travailler avec des modèles hiérarchiques complexes. À la section 2.4.2, nous verrons de façon détaillée l'application de telles méthodes pour trouver l'estimateur du modèle.

## 2.4. MODÉLISATION

### 2.4.1. Modèle *a priori*

Nous désirons appliquer l'approche bayésienne vue à la section précédente sur notre modèle (2.1). Puisque nous ne connaissons pas la forme fonctionnelle de  $g$ , nous travaillons avec un modèle non paramétrique. En fait, plusieurs méthodes non paramétriques pour modéliser  $f_{i,j}$  comme une fonction non linéaire de variables explicatives supposent des combinaisons linéaires de fonctions de base des prédicteurs. Par conséquent, notre méthode d'estimation suppose que la fonction bidimensionnelle  $g$  peut se décomposer de la façon suivante :

$$g(x, y) = \sum_{i=1}^{z_1} \sum_{j=1}^{z_2} \theta_{i,j} b_{i,j}(x, y), \quad (2.7)$$

où  $\theta_{i,j}$  est le coefficient multipliant l'élément  $b_{i,j}$  de la base. Nous choisissons une base de T-splines,  $T_{i,j}^{k,l}$ , afin de profiter de leur prédisposition à la monotonie.

En définissant le coefficient vectoriel  $\boldsymbol{\theta}$  par

$$\boldsymbol{\theta} = (\theta_{1,1}, \theta_{1,2}, \dots, \theta_{z_1, z_2})^T,$$

puis la matrice de splines bidimensionnelles  $\mathbf{X}$  par

$$\mathbf{X} = \begin{bmatrix} T_{1,1}(x_1, y_1) & T_{1,2}(x_1, y_1) & \cdots & T_{z_1, z_2}(x_1, y_1) \\ T_{1,1}(x_1, y_2) & T_{1,2}(x_1, y_2) & \cdots & T_{z_1, z_2}(x_1, y_2) \\ \vdots & \vdots & \vdots & \vdots \\ T_{1,1}(x_{n_x}, y_{n_y}) & T_{1,2}(x_{n_x}, y_{n_y}) & \cdots & T_{z_1, z_2}(x_{n_x}, y_{n_y}) \end{bmatrix},$$

nous obtenons une écriture matricielle

$$\mathbf{f} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\epsilon},$$

où  $\mathbf{f} = (f_{1,1}, f_{1,2}, \dots, f_{n_x, n_y})^T$  est le vecteur des observations et  $\boldsymbol{\epsilon} = (\epsilon_{1,1}, \epsilon_{1,2}, \dots, \epsilon_{n_x, n_y})^T$  est le vecteur des erreurs. Nous retrouvons ainsi le modèle de régression linéaire dont les prédicteurs sont les colonnes de la matrice  $\mathbf{X}$ . Cette forme linéaire, lorsque nous considérons les erreurs distribuées normalement, implique la distribution suivante pour  $\mathbf{f}$  :

$$\mathbf{f} | \boldsymbol{\theta}, \sigma^2 \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\theta}, \sigma^2 \mathbf{I}_n), \quad (2.8)$$

où  $\mathbf{I}_n$  est la matrice identité de dimension  $n$ . Par hypothèse, nous pourrions donner une autre structure à la matrice de covariance, soit  $\boldsymbol{\Sigma}_\epsilon$ , de sorte que  $\boldsymbol{\epsilon} \sim \mathcal{N}_n(\mathbf{0}, \sigma^2 \boldsymbol{\Sigma}_\epsilon)$ .

Nous remarquons que deux paramètres restent inconnus dans la distribution de  $\mathbf{f}$  : l'ensemble des coefficients  $\theta_{i,j}$  et la quantité  $\sigma^2$  qui affecte la variance. Afin de travailler dans un contexte bayésien, nous considérons que ces paramètres sont des variables aléatoires distribuées selon une certaine densité *a priori*. Le paramètre d'intérêt, c'est-à-dire le paramètre sur lequel nous voulons faire de l'inférence, est le vecteur de coefficients  $\boldsymbol{\theta}$ . Quant au paramètre  $\sigma^2$ , il s'agit en fait d'un hyperparamètre de niveau 1 selon la définition 2.3.2. En d'autres mots, la distribution *a priori* de  $\boldsymbol{\theta}$  est conditionnelle à  $\sigma^2$  et nous devons utiliser un modèle hiérarchique de Bayes. Une distribution *a priori* doit tenir compte de

l'information disponible sur le paramètre. En particulier, les contraintes de monotonie sur les coefficients  $\theta_{i,j}$  doivent être incluses dans leur densité *a priori*. De plus, nous désirons une distribution qui assure des valeurs positives, car la quantité  $\sigma^2$  ne peut prendre des valeurs négatives.

Lorsque les observations sont normalement distribuées, il est courant d'utiliser une densité *a priori* conjuguée telle la loi normale. Ainsi, la distribution *a posteriori* appartient à la même famille que celle *a priori*. Toutefois, la loi multinormale sur les coefficients ne satisfait pas les contraintes de monotonie. Par conséquent, nous présumons que le vecteur  $\boldsymbol{\theta}$  est distribué selon une loi normale tronquée multidimensionnelle. Soit le sous-espace  $R$  donné par

$$R = \{\boldsymbol{\theta} \in \mathbb{R}^{z^1 \cdot z^2} : \mathbf{C}\boldsymbol{\theta} \geq \mathbf{0}\}, \quad (2.9)$$

où  $\mathbf{C}$  est la matrice des contraintes définies à la section 2.2. Notons qu'un vecteur est supérieur ou égal à  $\mathbf{0}$  si chacun de ses éléments est supérieur ou égal à 0. Nous illustrons les différentes structures possibles de la matrice  $\mathbf{C}$  à l'aide d'un exemple.

**Exemple 2.4.1.** Soit le vecteur de coefficients  $\boldsymbol{\theta} = (\theta_{1,1}, \dots, \theta_{3,2})^T$ . La matrice de contraintes  $\mathbf{C}$  sur ces coefficients peut entre autres prendre les formes suivantes :

$$\begin{aligned}
 a) \mathbf{C} &= \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}, & b) \mathbf{C} &= \begin{bmatrix} 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 & -1 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}, \\
 c) \mathbf{C} &= \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}.
 \end{aligned}$$

Le cas a) correspond à des données monotones croissantes en  $x$  et en  $y$  tandis qu'en b), les données sont décroissantes en  $x$  et croissantes en  $y$ . Finalement, le cas c) représente une monotonie partielle en  $x$  (données croissantes).

Un vecteur  $\mathbf{f}$  suit alors une distribution normale tronquée,  $\mathcal{NT}_R(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , si sa fonction de densité est proportionnelle à  $\phi(\mathbf{f}; \boldsymbol{\mu}, \boldsymbol{\Sigma})\mathbb{I}_R(\mathbf{f})$ , où  $\phi(\mathbf{f}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$  est la densité d'une loi multinormale et  $\mathbb{I}_R(\cdot)$  est la fonction indicatrice de  $R$ . Certaines propriétés de la loi multinormale sont aussi valides pour une loi multinormale tronquée. En effet, ses distributions conditionnelles suivent également une distribution normale tronquée (Horrace, 2005; Rodriguez-Yam *et al.*, 2004). Considérons les partitions de  $\mathbf{X}$ ,  $\boldsymbol{\mu}$  et  $\boldsymbol{\Sigma}$

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ X_2 \end{bmatrix}, \quad \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \mu_2 \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_1 \\ \boldsymbol{\Sigma}_1^T & \sigma_{22} \end{bmatrix}.$$

La distribution conditionnelle de  $X_2$  sachant  $\mathbf{X}_1 = \mathbf{x}_1$  est alors

$$X_2 | \mathbf{X}_1 = \mathbf{x}_1 \sim \mathcal{NT}_{R_2}(\mu_2^*, \sigma_{22}^*), \quad (2.10)$$

où

$$\begin{aligned} \mu_2^* &= \mu_2 + \boldsymbol{\Sigma}_1^T \boldsymbol{\Sigma}_{11}^{-1} (\mathbf{x}_1 - \boldsymbol{\mu}_1), \\ \sigma_{22}^* &= \sigma_{22} - \boldsymbol{\Sigma}_1^T \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_1, \\ R_k &= \{x_2 \in \mathbb{R} : (\mathbf{x}_1, x_2) \in R\}. \end{aligned}$$

Cette propriété des distributions conditionnelles est importante pour l'implantation de l'échantillonnage de Gibbs (voir section 2.4.2).

Alors, nous posons la distribution *a priori* des coefficients  $\theta_{i,j}$

$$\boldsymbol{\theta} | \sigma^2 \sim \mathcal{NT}_R(\boldsymbol{\theta}_0, \sigma^2 \boldsymbol{\Sigma}_\theta) \quad (2.11)$$

avec l'ensemble de contraintes exprimé en (2.9), tandis que  $\sigma^2$  suit plutôt une loi inverse-gamma,  $\mathcal{IG}$ ,

$$\sigma^2 \sim \mathcal{IG}\left(\frac{\alpha}{2}, \frac{\gamma}{2}\right). \quad (2.12)$$

Le choix des hyperparamètres  $\boldsymbol{\theta}_0$ ,  $\boldsymbol{\Sigma}_\theta$ ,  $\alpha$  et  $\gamma$  dépend de l'information *a priori* disponible. D'abord, nous voulons spécifier un ensemble de valeurs plausibles pour

$\theta_0$ . Nous choisissons  $\theta_0$  comme l'estimateur de  $\theta$  qui résout le problème suivant de moindres carrés avec des contraintes d'inégalités :

$$\min \|\mathbf{X}\theta - \mathbf{f}\|^2 \text{ sujet à } \mathbf{C}\theta \geq \mathbf{0}.$$

L'algorithme pour résoudre ce problème est expliqué par Lawson et Hanson (1974). Nous obtenons ainsi une série de coefficients satisfaisant les contraintes de monotonie. Quant à la matrice de covariance  $\Sigma_\theta$ , elle peut prendre la forme

$$\Sigma_\theta = \frac{1}{n_0}(\mathbf{X}^T \mathbf{X})^{-1},$$

où  $n_0$  représente le nombre de pseudo-observations. Cette structure de variance-covariance a été nommée densité *a priori* de type G par Zellner (1986). Nous justifierons son usage à la section suivante.

Nous limitons le choix des valeurs de  $n_0$  entre 0 et 1 par manque d'information *a priori*. De plus, le choix de  $\gamma$  et  $\alpha$  dépend des données observées. En effet, une table de données ayant une précision de  $10^{-3}$  nécessite une plus petite valeur de  $\sigma^2$  que des données ayant une précision à l'unité. Alors, nous considérons les valeurs de  $\gamma$  et  $\alpha$  qui maximisent la loi marginale des observations  $\mathbf{f}$ .

Avant de définir l'estimateur *a posteriori* de  $\theta$ , nous introduisons quelques notions de calculs numériques qui seront utiles par la suite.

#### 2.4.2. Méthodes numériques

Certains problèmes peuvent entraver l'évaluation d'estimateur bayésien en pratique. Entre autres, l'évaluation numérique explicite de la distribution *a posteriori*,  $\pi(\theta|x)$ , peut être très compliquée, voire impossible. Lorsque nous considérons une fonction de perte quadratique, l'estimateur obtenu est lié directement à la distribution *a posteriori* à travers son espérance

$$\begin{aligned} \delta^\pi(x) &= \int_{\Theta} \theta \pi(\theta|x) d\theta \\ &= \frac{\int_{\Theta} \theta \pi(\theta) f(x|\theta) d\theta}{\int_{\Theta} \pi(\theta) f(x|\theta) d\theta}. \end{aligned}$$

Similairement, la majorité des estimateurs hiérarchiques de Bayes ne peuvent pas être calculé analytiquement.

La méthode classique de Monte Carlo permet l'approximation numérique d'une intégrale de la forme

$$\int_{\Theta} g(\theta)\pi(\theta)f(x|\theta) d\theta. \quad (2.13)$$

Si  $h$  est une densité de probabilité dont le support inclut celui de  $f(x|\theta)\pi(\theta)$ , alors l'intégrale (2.13) peut s'exprimer comme une espérance par rapport à la densité  $h$ , soit

$$\int_{\Theta} \frac{g(\theta)\pi(\theta)f(x|\theta)}{h(\theta)} h(\theta) d\theta.$$

L'algorithme de Monte Carlo avec fonction d'importance  $h$  est le suivant :

- (1) générer  $\theta_1, \dots, \theta_m$  selon la densité  $h$ , choisie telle qu'il soit facile à générer les  $\theta_i$  ;
- (2) approximer (2.13) par

$$\frac{1}{m} \sum_{i=1}^m g(\theta_i)\omega(\theta_i),$$

avec les poids  $\omega(\theta_i) = f(x|\theta_i)\pi(\theta_i)/h(\theta_i)$ .

Par la loi forte des grands nombres, cette approximation converge presque assurément (voir Robert, 2001) vers l'équation (2.13). Outre sa capacité de générer rapidement et fiablement des valeurs, la fonction  $h$  doit être presque proportionnelle à  $g(x)\pi(\theta)f(x|\theta)$  afin de réduire la variabilité. En plus d'un support incluant celui de la distribution *a posteriori*, la fonction  $h$  ne peut pas décroître plus rapidement que  $\pi(\theta|x)$  au niveau des ailes.

Nous présentons également une méthode de Monte Carlo plus généralisée qui approxime la simulation de variables aléatoires d'une distribution *a posteriori*. L'avantage d'une telle méthode par rapport à la méthode classique est que la construction d'une fonction d'importance  $h$  adéquate est non nécessaire. Nous y référons par méthode de Monte Carlo par chaînes de Markov, connue également sous son abréviation MCMC. Nous nous intéressons en particulier à l'approche de l'échantillonnage de Gibbs, car elle s'applique facilement sur les structures hiérarchiques. Supposons qu'un modèle bayésien s'écrit

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

Lorsque les distributions  $\pi_1(\theta|x, \lambda)$  et  $\pi_2(\lambda|x)$  sont connues et peuvent être simulées, la simulation de  $\theta$  selon  $\pi(\theta|x)$  est équivalent à générer  $\lambda$  selon  $\pi_2(\lambda|x)$ , puis générer  $\theta$  selon  $\pi_1(\theta|x, \lambda)$ .

De manière générale, l'échantillonnage de Gibbs est multidimensionnel. Outre les modèles hiérarchiques, cette méthode permet de considérer un vecteur de paramètres d'intérêt et de générer chaque paramètre conditionnellement aux autres. Supposons que  $\mathbf{x}$  dénote le vecteur de données, que le vecteur de paramètres  $\boldsymbol{\theta}$  se décompose en  $(\theta_1, \dots, \theta_p)^T$  pour  $p > 1$  et que les densités conditionnelles correspondantes,  $\pi_1, \dots, \pi_p$ , peuvent facilement générer des observations. L'algorithme d'échantillonnage de Gibbs associé à cette décomposition génère la  $t^e$  itération :

- Sachant  $\boldsymbol{\theta}^{(t)}$ , simuler
- $\theta_1^{(t+1)} \sim \pi_1(\theta_1|\mathbf{x}, \theta_2^{(t)}, \dots, \theta_p^{(t)})$  ;
  - $\theta_2^{(t+1)} \sim \pi_2(\theta_2|\mathbf{x}, \theta_1^{(t+1)}, \theta_3^{(t)}, \dots, \theta_p^{(t)})$  ;
  - ...
  - $\theta_p^{(t+1)} \sim \pi_p(\theta_p|\mathbf{x}, \theta_1^{(t+1)}, \dots, \theta_{p-1}^{(t+1)})$ .

Les valeurs initiales  $\boldsymbol{\theta}^{(0)}$  sont choisies parmi le support de la densité *a posteriori*  $\pi(\boldsymbol{\theta}|\mathbf{x})$ . La distribution conjointe  $\pi$  est stationnaire à chaque étape de l'algorithme puisque les densités  $\pi_i$  sont complètement conditionnelles à  $\pi$ . En supposant que la chaîne pour chaque paramètre  $\theta_i$  est ergodique, l'indépendance des coefficients à l'intérieur d'une chaîne n'est pas importante. En effet, le théorème ergodique indique que la moyenne

$$\frac{1}{m} \sum_{k=1}^m g(\theta_i^{(k)})$$

converge vers  $E^\pi[g(\theta_i)|\mathbf{x}]$  lorsque  $m$  tend vers l'infini.

Nous utiliserons donc le méthode d'échantillonnage de Gibbs afin de déterminer les estimateurs *a posteriori* des coefficients du modèle.

### 2.4.3. Modèle *a posteriori*

Nous désirons obtenir une estimation de la fonction bidimensionnelle inconnue  $g$  qui prend la forme (2.7) en considérant une base de T-splines à deux variables à l'aide des observations  $\mathbf{f}$ . L'inférence porte alors sur les coefficients  $\theta_{i,j}$  et leurs estimateurs s'expriment par des quantités *a posteriori*. Comme mentionné à la

section précédente, l'estimateur recherché correspond à l'espérance *a posteriori* de  $\boldsymbol{\theta}|\mathbf{f}$  sous la fonction de perte quadratique.

Robert (2001) présente les calculs bayésiens standards pour obtenir les distributions *a posteriori* de modèles linéaires lorsque la famille de distributions *a priori* conjuguées sur  $(\boldsymbol{\theta}, \sigma^2)$  sont des densités multinormale et inverse-gamma. Une démarche semblable est possible lorsque nous combinons les distributions *a priori* de  $(\boldsymbol{\theta}, \sigma^2)$  données en (2.11)-(2.12) avec la fonction de vraisemblance du modèle exprimé en (2.8). Posons  $\tilde{\boldsymbol{\theta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{f}$  dans le but de simplifier la notation. Alors, les distributions *a posteriori* sont

$$\boldsymbol{\theta}|\mathbf{f}, \sigma^2 \sim \mathcal{NT}_R(\boldsymbol{\theta}^*, \sigma^2 \boldsymbol{\Sigma}^*), \quad (2.14)$$

$$\sigma^2|\mathbf{f} \sim \mathcal{IG}\left(\frac{\alpha^*}{2}, \frac{\gamma^*}{2}\right), \quad (2.15)$$

où

$$\boldsymbol{\theta}^* = \frac{n_0 \boldsymbol{\theta}_0 + \tilde{\boldsymbol{\theta}}}{n_0 + 1},$$

$$\boldsymbol{\Sigma}^* = \frac{1}{n_0} (\mathbf{X}^T \mathbf{X})^{-1},$$

$$\alpha^* = \alpha + n,$$

$$\gamma^* = \gamma + \|\mathbf{f} - \mathbf{X} \tilde{\boldsymbol{\theta}}\|^2 + \frac{n_0}{n_0 + 1} (\boldsymbol{\theta}_0 - \tilde{\boldsymbol{\theta}})^T \mathbf{X}^T \mathbf{X} (\boldsymbol{\theta}_0 - \tilde{\boldsymbol{\theta}}),$$

et l'ensemble de contraintes  $R$  est défini en (2.9). L'utilisation d'une distribution *a priori* de type G permet de simplifier la forme du modèle *a posteriori*. Les critiques associées au choix d'une telle distribution sont principalement fondées sur la dépendance de la distribution *a priori* sur  $\mathbf{X}$ . L'argument repose sur le fait que  $\mathbf{X}$  peut être considéré comme une variable aléatoire, donc qu'une modélisation *a priori* ne devrait pas dépendre de  $\mathbf{X}$ . Robert (2001, section 4.5) réfute cet argument, car le modèle de régression est lui-même conditionnel aux variables explicatives. La distribution *a priori* peut donc être vue comme une distribution *a posteriori* par rapport aux variables  $\mathbf{X}$ . Il s'agirait donc d'un conditionnement opéré en deux étapes.

Les distributions *a posteriori* (2.14)-(2.15) ont une structure hiérarchique, alors en les intégrant, nous retrouvons la distribution *a posteriori* de  $\boldsymbol{\theta}|\mathbf{f}$  :

$$\pi(\boldsymbol{\theta}|\mathbf{f}) = \int_0^\infty \pi_1(\boldsymbol{\theta}|\mathbf{f}, \sigma^2) \pi_2(\sigma^2|\mathbf{f}) d\sigma^2.$$

L'ensemble de contraintes  $R$  complique le calcul des moments de  $\pi(\boldsymbol{\theta}|\mathbf{f})$ . Par conséquent, nous sommes dans un contexte favorable à l'implantation de l'échantillonnage de Gibbs. En fait, il est même difficile de générer des valeurs de  $\boldsymbol{\theta}$  selon la distribution  $\pi_1(\boldsymbol{\theta}|\mathbf{f}, \sigma^2)$ . Nous considérons donc de conditionner complètement chaque paramètre  $\theta_{i,j}$  sur le reste des paramètres. Afin de faciliter la notation, modifions la notation du vecteur  $\boldsymbol{\theta}$ ,

$$\begin{aligned} \boldsymbol{\theta} &= (\theta_{1,1}, \theta_{1,2}, \dots, \theta_{z_1, z_2})^T, \\ &= (\theta_1, \theta_2, \dots, \theta_{z_1 \cdot z_2})^T, \end{aligned}$$

où  $\theta_i$  est le  $i^e$  élément du vecteur  $\boldsymbol{\theta}$ . La  $(t+1)^e$  itération de l'algorithme d'échantillonnage de Gibbs met à jour les valeurs  $(\theta_1^{(t)}, \dots, \theta_{z_1 \cdot z_2}^{(t)}, \sigma^{2(t)})$ . Nous simulons alors

$$\begin{aligned} \circ \theta_1^{(t+1)} &\sim p_1(\theta_1|\mathbf{f}, \theta_2^{(t)}, \dots, \theta_{z_1 \cdot z_2}^{(t)}, \sigma^{2(t)}), \\ \circ \theta_2^{(t+1)} &\sim p_2(\theta_2|\mathbf{f}, \theta_1^{(t+1)}, \theta_3^{(t)}, \dots, \theta_{z_1 \cdot z_2}^{(t)}, \sigma^{2(t)}), \\ &\dots \\ \circ \theta_{z_1 \cdot z_2}^{(t+1)} &\sim p_{z_1 \cdot z_2}(\theta_{z_1 \cdot z_2}|\mathbf{f}, \theta_1^{(t+1)}, \dots, \theta_{z_1 \cdot z_2 - 1}^{(t+1)}, \sigma^{2(t)}), \\ \circ \sigma^{2(t+1)} &\sim p(\sigma^2|\mathbf{f}, \theta_1^{(t+1)}, \dots, \theta_{z_1 \cdot z_2}^{(t+1)}), \end{aligned}$$

où pour  $i = 1, \dots, z_1 \cdot z_2$ , la distribution

$$p_i(\theta_i|\mathbf{f}, \theta_1^{(t+1)}, \dots, \theta_{i-1}^{(t+1)}, \theta_{i+1}^{(t)}, \dots, \theta_{z_1 \cdot z_2}^{(t)}, \sigma^{2(t)})$$

est une normale tronquée unidimensionnelle déterminée à partir des équations (2.14) et (2.10). Similairement, la distribution  $p(\sigma^2|\mathbf{f}, \theta_1^{(t+1)}, \dots, \theta_{z_1 \cdot z_2}^{(t+1)})$  peut être obtenue par (2.15).

Pour un  $t$  assez élevé, les valeurs  $(\theta_1^{(t)}, \dots, \theta_{z_1 \cdot z_2}^{(t)})^T$  sont approximativement des réalisations de la distribution *a posteriori*  $\pi(\boldsymbol{\theta}|\mathbf{f})$ . Cependant, la vitesse de convergence dépend souvent des valeurs initiales. Il est donc préférable, en termes de convergence, de prendre les résultats de l'itération  $t$  comme nouvelles valeurs initiales afin d'obtenir une suite de réalisations de la loi *a posteriori*. Ainsi, lors de

l'implantation de l'algorithme MCMC en pratique, nous itérons un grand nombre de fois, soit  $M$  fois, mais ne conservons habituellement que les  $\lfloor M/2 \rfloor$  dernières itérations. Il s'agit ensuite de prendre la moyenne des réalisations pour chaque coefficient afin d'obtenir nos estimateurs :

$$\hat{\theta}_i = \frac{1}{\lfloor M/2 \rfloor} \sum_{k=\lfloor M/2 \rfloor+1}^M \theta_i^{(k)}, \quad i = 1, \dots, z_1 \cdot z_2. \quad (2.16)$$

Nous sommes maintenant prêts à appliquer cette démarche de modélisation à des tables à deux entrées de données réelles.

# Chapitre 3

---

## CAS PRATIQUES

L'objectif du présent chapitre est d'étudier le comportement de l'estimateur construit avec des splines bidimensionnelles décrit précédemment. Nous appliquons ce modèle bayésien non paramétrique à différents jeux de données réelles, soient des tables à deux entrées présentant un certain degré de monotonie. Nous commençons d'abord nos exemples avec une table de probabilités d'une distribution normale, c'est-à-dire des données formant une surface sans changement abrupt. Nous poursuivons en modélisant une table de quantiles de la loi de Student, dont les données présentent des variations plus prononcées. Finalement, le modèle bayésien théorique est appliqué sur une table de mortalité à deux entrées. Afin d'étudier l'efficacité de notre modèle bayésien, nous comparons son ajustement à celui du modèle classique. Nous considérons également la capacité de condenser et de recréer les tables à partir de notre estimateur.

Nous avons effectué les simulations à l'aide du logiciel *R 2.11.1*. Les fonctions B-splines, les générateurs d'observations pseudo-aléatoires des lois normale tronquée et inverse gamma et l'algorithme des moindres carrés avec contraintes linéaires y sont prédéfinis (« packages » *pscl*, *mvtnorm* et *limSolve*). Nous avons programmé les algorithmes de sélection des noeuds et d'échantillonnage de Gibbs.

### 3.1. TABLE DE LA LOI NORMALE

Nous considérons dans notre premier exemple une table de la loi normale dont les entrées sont des probabilités de la forme  $P(Z \leq z_p)$ , où  $Z$  représente une variable aléatoire de densité normale standard. Nous considérons les quantiles  $z_p$

entre 0 et 3,59, pour un total de 360 probabilités cumulatives de la loi normale. Les valeurs de  $x$  avec lesquelles nous travaillons correspondent aux quantiles jusqu'à leur première décimale, allant de 0,0 à 3,5 par pas de 0,1. Nous identifions la deuxième décimale des quantiles  $z_p$  par les valeurs de  $y$  (de 0,00 à 0,09 par pas de 0,01). Le tableau 3.1 présente les entrées de la table de la loi normale. Il est apparent que la table est monotone au sens bidimensionnel. En effet, elle est monotone croissante en  $x$  ainsi qu'en  $y$ . Nous pouvons alors ajuster nos données à l'aide d'une base de fonctions T-splines. L'approche bayésienne décrite au chapitre 2 est utilisée afin d'estimer les coefficients du modèle (2.7) avec les contraintes de monotonie exprimées en (2.9).

TABLEAU 3.1: TABLE DE LA LOI NORMALE  $P(Z \leq z_p)$

$z_p$	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633

TABLEAU 3.1 (SUITE)

$z_p$	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998
3,5	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998

### 3.1.1. Modèle et résultats

Voici une récapitulation de la procédure à suivre :

- (1) La première étape consiste à fixer les hyperparamètres  $n_0$ ,  $\alpha$  et  $\gamma$  des distributions *a priori* introduites aux équations (2.11) et (2.12). Nous considérons les valeurs possibles de  $n_0$  entre 0 et 1. Les hyperparamètres  $\alpha$  et  $\gamma$  de la loi inverse-gamma sont associés à la variance des observations  $\mathbf{f}$  à travers la variable  $\sigma^2$ . L'information *a priori* que nous possédons sur cette distribution correspond à la précision de chacune des entrées tabulaires, laquelle est d'ordre  $10^{-4}$ . Le paramètre de forme  $\alpha$  ne reflète pas

directement nos connaissances préalables. Nous étudions alors les valeurs de  $\alpha$  supérieures à 0. Quant à  $\gamma$ , ce paramètre d'échelle régit la dispersion de la distribution. Nous désirons ainsi une valeur de  $\gamma$  relativement petite afin d'obtenir une distribution concentrée. Nous tenons compte de l'ordre de précision en sélectionnant une valeur de  $\gamma$  qui maximise la densité marginale des observations  $\mathbf{f}$ . Afin de simplifier les calculs, nous considérons la densité marginale sans les contraintes de monotonie. Nous obtenons ainsi une valeur de  $\gamma$  égale à  $n^{-1}\alpha(\mathbf{f} - \mathbf{X}\tilde{\boldsymbol{\theta}})^T(\mathbf{f} - \mathbf{X}\tilde{\boldsymbol{\theta}})$ , où  $\tilde{\boldsymbol{\theta}}$  est l'estimateur des moindres carrés usuel. Notons que l'hyperparamètre  $\gamma$  dépend de  $\alpha$ , donc l'effet de  $\alpha$  sur la distribution inverse-gamma est atténué. L'hyperparamètre repose également sur les éléments de la base spline, lesquels nous devons optimiser. Le choix des hyperparamètres  $n_0$  et  $\alpha$  est discuté en détails à la sous-section 3.1.2.

- (2) L'étape suivante consiste à sélectionner l'emplacement et le nombre de noeuds dans notre contexte bayésien, lesquels déterminent entièrement les éléments de la base spline. Nous choisissons ainsi l'ensemble de noeuds qui minimise le critère d'information He et Shi (1996, 1998) défini à l'équation (2.2). L'algorithme pour le choix des noeuds est résumé ci-dessous.
  - (a) Nous fixons les ensembles de noeuds intérieurs  $\mathbf{t}$  et  $\boldsymbol{\tau}$  comme étant les valeurs  $x$  et  $y$  associées aux quantiles empiriques de  $\mathbf{f}$ . Chaque élément de la base spline compte ainsi au moins une observation.
  - (b) Nous faisons varier le nombre de noeuds intérieurs en  $x$  ( $k_1$ ) et en  $y$  ( $k_2$ ). Pour chaque combinaison  $(k_1, k_2)$  possible, nous calculons l'hyperparamètre  $\gamma$  ainsi que les paramètres des distributions *a posteriori* (2.14)-(2.15). En utilisant la méthode d'échantillonnage de Gibbs pour 1000 itérations, nous obtenons l'estimateur  $\hat{\boldsymbol{\theta}}$  décrit à l'équation (2.16). Nous évaluons ensuite le critère He et Shi et conservons les nombres de noeuds  $k_1$  et  $k_2$  minimisant le critère.
  - (c) Lorsque les nombres  $k_1$  et  $k_2$  varient, l'emplacement des noeuds est prédéterminé, comme mentionné à l'étape (a). Ainsi, après avoir optimisé  $k_1$  et  $k_2$ , un dernier ajustement est effectué sur la position des

noeuds. Nous retirons le premier noeud intérieur et évaluons le critère He et Shi. Nous procédons de même pour tous les noeuds intérieurs et conservons l'ensemble de noeuds auquel est associé la plus petite valeur du critère.

Cette étape est répétée jusqu'à ce que le critère He et Shi ne puisse plus être diminué.

- (3) En déterminant le nombre et l'emplacement optimal des noeuds, nous avons fixé les éléments de la base spline. Nous calculons alors l'estimateur final selon la méthode de Monte Carlo par chaînes de Markov. Nous générons 5000 vecteurs de coefficients par l'algorithme d'échantillonnage de Gibbs (voir la sous-section 2.4.3). Nous ne conservons que les 2500 derniers vecteurs de coefficients et prenons leur moyenne afin d'obtenir l'estimateur désiré,  $\hat{\theta}$ . Notons que les contraintes de monotonie sont incluses dans l'algorithme d'échantillonnage de Gibbs puisque nous modélisons les données  $f$  selon une loi normale tronquée multivariée.

Nous pouvons maintenant discuter des résultats obtenus pour la table des probabilités cumulatives d'une loi normale. La figure 3.1 représente les valeurs du critère He et Shi selon les nombres de noeuds intérieurs  $t$  et  $\tau$ , soient  $k_1$  et  $k_2$ . En étudiant le graphique, nous constatons que seul le nombre de noeuds intérieurs par rapport à  $x$  affecte la valeur du critère d'information. En effet, le graphique semble plutôt constant en fonction de  $k_2$  tandis que les variations sont observées en  $k_1$ . Puisque le minimum est atteint à  $k_1 = 17$ , nous choisissons 17 noeuds intérieurs par rapport à  $x$  et aucun par rapport à  $y$ . Notons qu'à l'étape (2c), le retrait d'un noeud intérieur ne diminue pas le critère d'information pour cet exemple : l'ensemble des 17 noeuds intérieurs précédent est donc conservé. Étant donné que nous travaillons avec des fonctions T-splines quadratiques, les vecteurs de noeuds comprennent 3 noeuds confondus aux extrémités du domaine de définition des splines. Donc, en incluant les noeuds répétés, nous obtenons une base de fonctions splines comprenant 60 éléments. Autrement dit, nous cherchons à estimer 60 coefficients selon l'approche bayésienne.

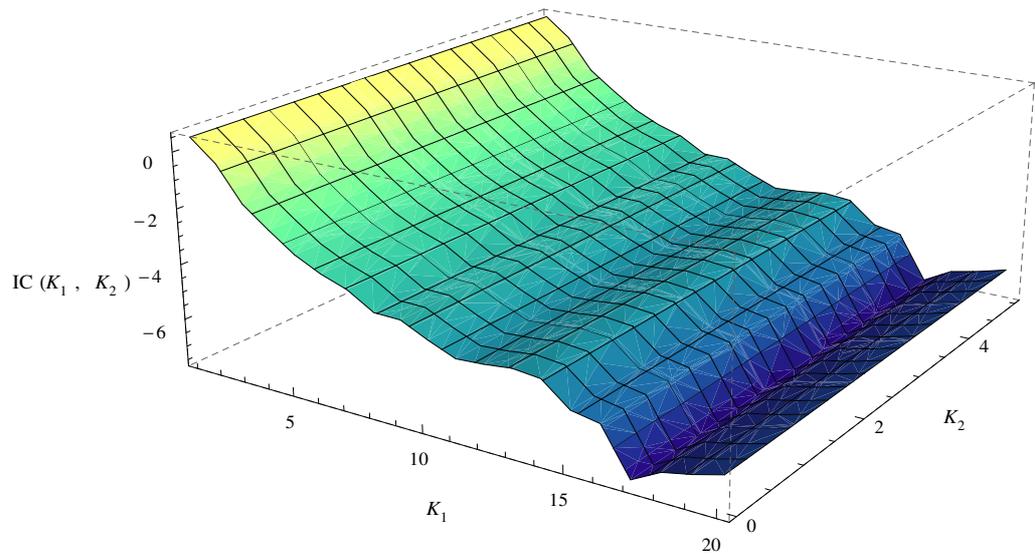


FIGURE 3.1: Sélection des noeuds intérieurs par le critère He et Shi pour la table normale

Avant d'estimer la table de la loi normale, notons que nous avons retenu les valeurs 0,45 et 500 pour les hyperparamètres  $n_0$  et  $\alpha$  respectivement. Ces mêmes valeurs ont été utilisées pour sélectionner la base spline selon la méthode décrite à l'étape (2). La modélisation finale des probabilités cumulatives de la loi normale nous donne, à la quatrième décimale près, exactement la même table que les données originales, soit le tableau 3.1 (pour l'estimation des coefficients, voir l'annexe A.1). En effet, la figure 3.2, qui illustre les erreurs absolues de l'estimation sous forme d'une carte de points chauds, indique en bleu une erreur maximale d'ordre  $5 \times 10^{-6}$ . Plus précisément, l'erreur absolue maximale est de  $4,4420 \times 10^{-6}$  pour le quantile  $z_p = 0,49$ , c'est-à-dire  $(x; y) = (0,4; 0,09)$ . Quant à l'erreur absolue minimale, elle est de  $2,2296 \times 10^{-9}$  pour le quantile  $z_p = 3,59$ . Nous remarquons également que les plus grandes erreurs d'estimation sont concentrées entre les quantiles 0,30 à 1,19. Ce phénomène coïncide avec la croissance plus marquée des probabilités pour ces mêmes quantiles. L'erreur absolue moyenne de notre modèle bayésien est de  $1,34991 \times 10^{-6}$  tandis qu'elle est de  $1,35003 \times 10^{-6}$

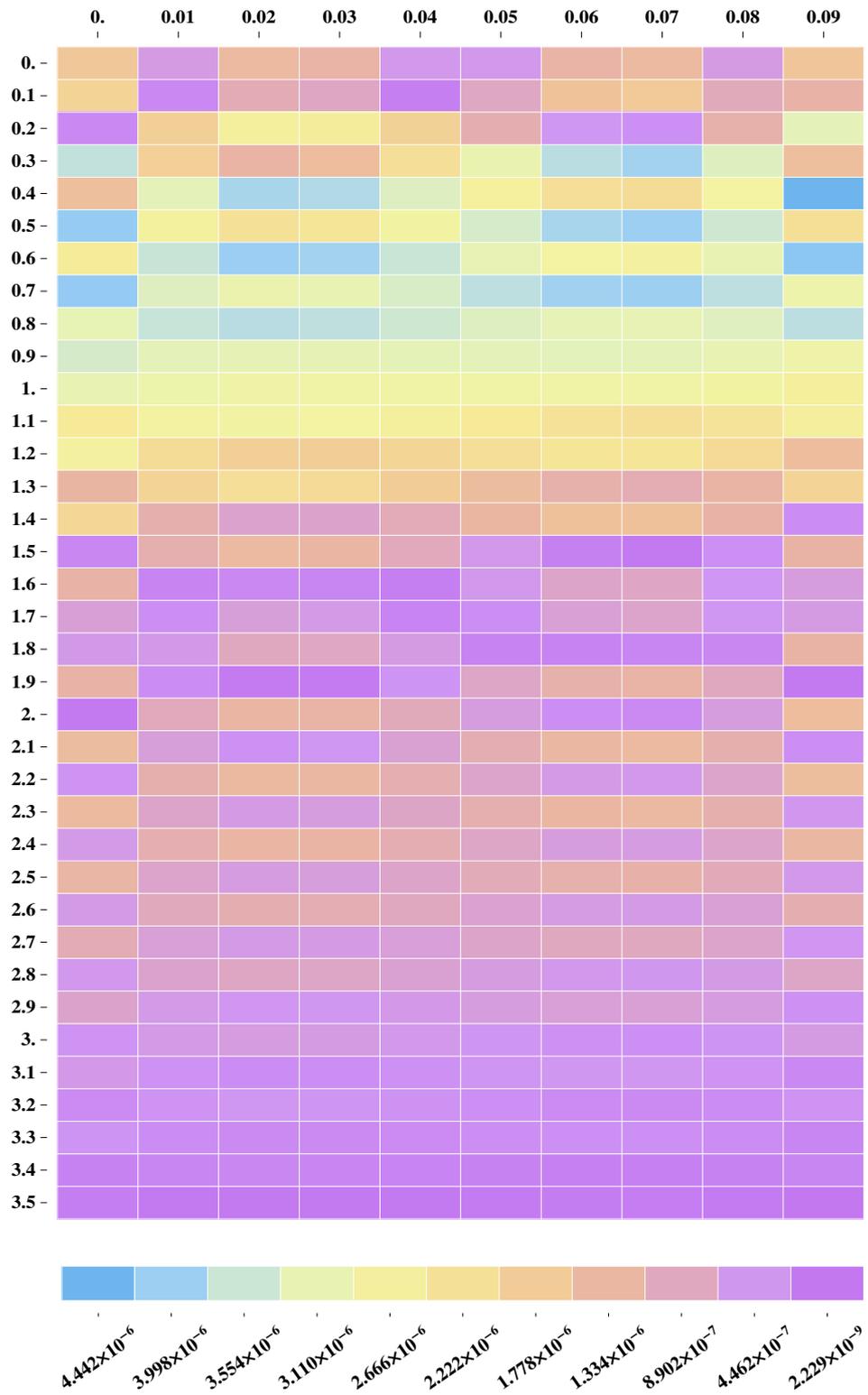


FIGURE 3.2: Cartographie des points chauds : les erreurs absolues associées à l'estimation de la table de la loi normale

pour le modèle classique. La différence d'ordre  $10^{-11}$  entre les deux modèles est donc négligeable. Pour la table normale, l'approche bayésienne donne d'aussi bons résultats que l'approche classique avec splines.

Nous sommes également intéressés par la réduction de l'espace de rangement qu'apporte la modélisation. Pour reconstituer les 360 entrées de la table de la loi normale, seuls 60 coefficients sont nécessaires. En d'autres mots, nous réduisons de 83,3% la quantité d'information à conserver. Selon les mesures informatiques d'espace, la taille des données en mémoire passe de 3636 octets à 841 octets, à savoir une diminution de 76,9%. De plus, nous désirons connaître le temps de reconstitution de la table de la loi normale en utilisant le logiciel *R 2.11.1*. Afin de construire la table à partir de la fonction des probabilités cumulatives déjà implantée en *R*, le logiciel prend 0,032 seconde. En utilisant les T-splines et les coefficients estimés, nous pouvons reconstruire la table en 0,031 seconde. Ainsi, pour un temps équivalent d'évaluation, la modélisation par T-splines permet de condenser la table à plus de 80%, ce qui justifie l'intérêt porté au modèle. L'entreposage de la table restreint et l'erreur absolue moyenne de l'estimation reflètent la performance du modèle bayésien face à des données lisses.

### 3.1.2. Étude de sensibilité des paramètres

Nous désirons étudier l'effet des hyperparamètres  $n_0$  et  $\alpha$  sur notre estimation. Nous voulons entre autres déterminer s'il est préférable d'utiliser une distribution *a priori* diffuse ou non pour les coefficients à travers les valeurs de  $n_0$ . Voici la procédure à suivre lorsque nous étudions  $n_0$  :

- (1) Nous fixons un ensemble de valeurs possibles pour l'hyperparamètre  $n_0$  entre 0 et 1. Nous considérons les valeurs suivantes :

$$n_0 = \{0,05; 0,10; 0,15; 0,20; \dots; 0,85; 0,90; 0,95\}.$$

- (2) Pour chacune des valeurs de  $n_0$ , nous procédons comme suit :
  - (a) Par souci de simplicité, nous fixons les éléments de la base spline ainsi que l'hyperparamètre  $\alpha$ . Nous calculons l'estimateur  $\hat{\theta}$  des coefficients en appliquant l'algorithme d'échantillonnage de Gibbs avec

1000 itérations. Nous retenons alors l'erreur absolue moyenne associée à l'estimation de la table de la loi normale.

- (b) Nous répétons l'étape précédente 600 fois afin de recueillir de l'information sur la qualité d'ajustement de notre estimateur à travers différentes simulations.
- (c) Nous calculons la moyenne et l'écart type de l'ensemble des 600 erreurs absolues moyennes. Nous tenons ainsi compte des variations des erreurs entre chaque simulation.

Pour chacune des 20 valeurs de  $n_0$  étudiées, les erreurs absolues des 600 estimations générées par l'algorithme MCMC se trouvent en moyenne entre  $1,35001 \times 10^{-6}$  et  $1,35004 \times 10^{-6}$ . Les variations de l'hyperparamètre entraînent donc des différences d'ajustement d'ordre  $10^{-11}$ . Or, les erreurs absolues varient du même ordre d'une simulation à l'autre pour un  $n_0$  fixe. En effet, les écarts types associés aux estimations se situent entre  $10^{-11}$  et  $10^{-10}$ , ce qui rend les variations moyennes négligeables.

La procédure est similaire lorsque nous faisons varier l'hyperparamètre  $\alpha$ . Nous considérons plutôt l'ensemble de valeurs possibles

$$\alpha = \{100; 200; 300; 400; 500; 600; 700; 800; 900\}.$$

Nous considérons ces grandes valeurs afin de réduire la variance de la distribution inverse-gamma *a priori*, laquelle est donnée par  $\frac{\gamma^2}{(\alpha-1)^2(\alpha-2)}$  pour  $\alpha > 2$ . La moyenne et l'écart type de l'erreur absolue des 600 estimations sont représentés graphiquement en fonction de l'hyperparamètre  $\alpha$  à la figure 3.3. Le comportement observé des erreurs pour  $\alpha$  est semblable à celui décrit pour  $n_0$ . En effet, les variations des erreurs moyennes entre les valeurs étudiées ainsi que des écarts types sont d'ordre  $10^{-12}$ .

L'ajustement du modèle bayésien varie de manière négligeable en fonction des deux hyperparamètres à l'étude. Autrement dit, le modèle est robuste aux variations des hyperparamètres  $n_0$  et  $\alpha$ . Une distribution *a priori* diffuse pour les coefficients  $\theta$ , à savoir une valeur de  $n_0$  proche de 0, est donc un choix aussi valide qu'une distribution informative accordant un poids supérieur aux connaissances

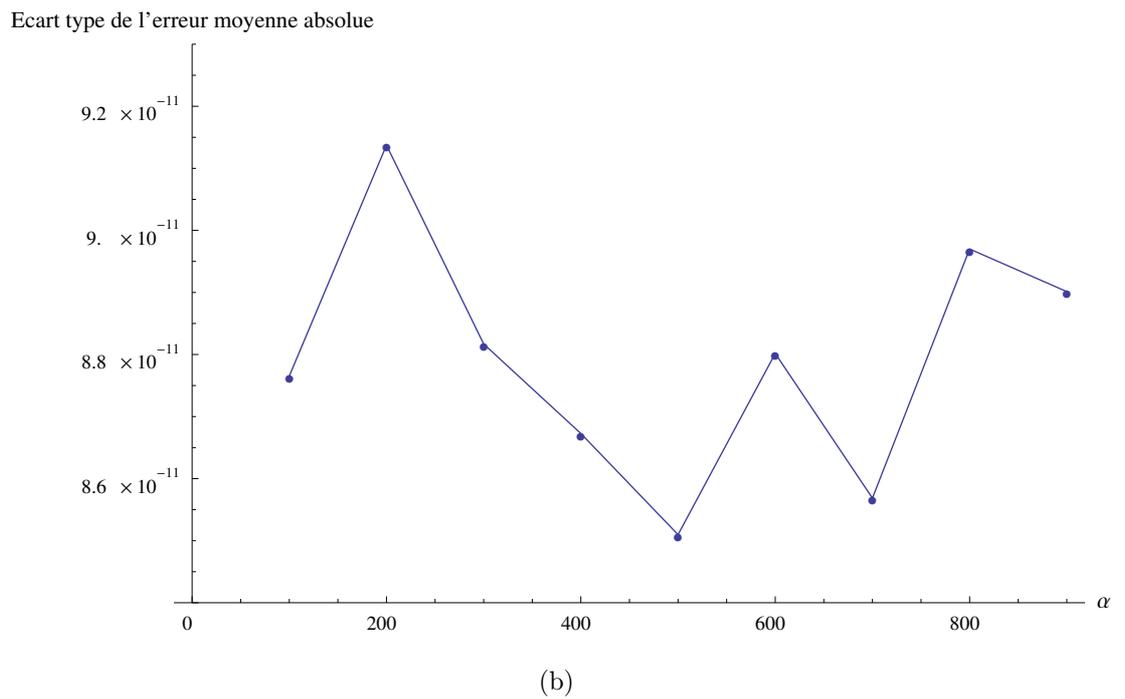
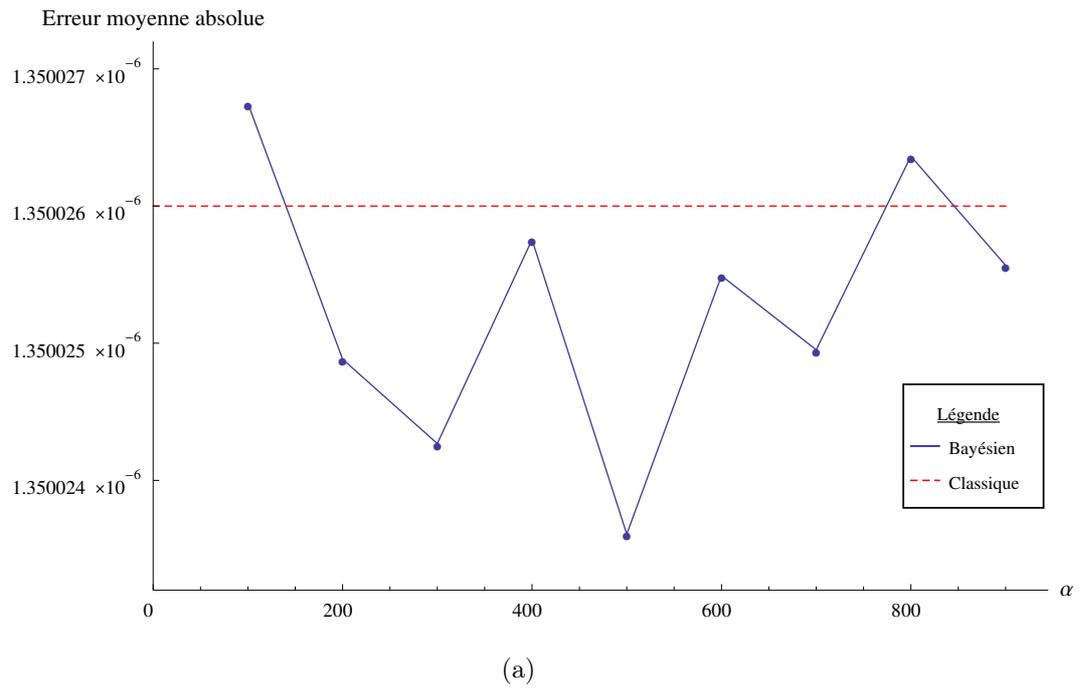


FIGURE 3.3: (a) Moyenne et (b) écart type associés aux erreurs absolues moyennes obtenues en générant 600 estimations de la table de la loi normale, selon les valeurs de  $\alpha$

*a priori*. Nous concluons également que le paramètre de forme  $\alpha$  n'ajoute aucune information à la modélisation.

### 3.2. TABLE $t$ DE DISTRIBUTION DE STUDENT

Nous appliquons notre modélisation sur un second exemple, soit la table  $t$  de la loi de Student. Les entrées de la table sont les quantiles de la distribution,  $t_{d.l.;1-\alpha}$ , associés à un certain degré de liberté (d.l.) et une probabilité cumulative  $(1-\alpha)$ . Les valeurs de  $x$  correspondent alors aux différents degrés de liberté tandis que les probabilités sont identifiées par  $y$ . Nous considérons 46 valeurs de  $x$  et 8 valeurs de  $y$  :

$$x = \{3; 4; 5; \dots; 30; 32; 34; \dots; 50; 55; 60; 65; 70; 80; 100; 150; 200\},$$

$$y = \{0,80; 0,90; 0,95; 0,975; 0,99; 0,995; 0,999; 0,9995\}.$$

TABLEAU 3.2: TABLE  $t$  DE LA LOI DE STUDENT

$1 - \alpha$ d.l.	0,80	0,90	0,95	0,975	0,99	0,995	0,999	0,9995
3	0,978	1,638	2,353	3,182	4,541	5,841	10,215	12,924
4	0,941	1,533	2,132	2,776	3,747	4,604	7,173	8,610
5	0,920	1,476	2,015	2,571	3,365	4,032	5,893	6,869
6	0,906	1,440	1,943	2,447	3,143	3,707	5,208	5,959
7	0,896	1,415	1,895	2,365	2,998	3,499	4,785	5,408
8	0,889	1,397	1,860	2,306	2,896	3,355	4,501	5,041
9	0,883	1,383	1,833	2,262	2,821	3,250	4,297	4,781
10	0,879	1,372	1,812	2,228	2,764	3,169	4,144	4,587
11	0,876	1,363	1,796	2,201	2,718	3,106	4,025	4,437
12	0,873	1,356	1,782	2,179	2,681	3,055	3,930	4,318
13	0,870	1,350	1,771	2,160	2,650	3,012	3,852	4,221
14	0,868	1,345	1,761	2,145	2,624	2,977	3,787	4,140
15	0,866	1,341	1,753	2,131	2,602	2,947	3,733	4,073
16	0,865	1,337	1,746	2,120	2,583	2,921	3,686	4,015
17	0,863	1,333	1,740	2,110	2,567	2,898	3,646	3,965
18	0,862	1,330	1,734	2,101	2,552	2,878	3,610	3,922

TABLEAU 3.2 (SUITE)

$d.l. \backslash 1 - \alpha$	0,80	0,90	0,95	0,975	0,99	0,995	0,999	0,9995
19	0,861	1,328	1,729	2,093	2,539	2,861	3,579	3,883
20	0,860	1,325	1,725	2,086	2,528	2,845	3,552	3,850
21	0,859	1,323	1,721	2,080	2,518	2,831	3,527	3,819
22	0,858	1,321	1,717	2,074	2,508	2,819	3,505	3,792
23	0,858	1,319	1,714	2,069	2,500	2,807	3,485	3,768
24	0,857	1,318	1,711	2,064	2,492	2,797	3,467	3,745
25	0,856	1,316	1,708	2,060	2,485	2,787	3,450	3,725
26	0,856	1,315	1,706	2,056	2,479	2,779	3,435	3,707
27	0,855	1,314	1,703	2,052	2,473	2,771	3,421	3,690
28	0,855	1,313	1,701	2,048	2,467	2,763	3,408	3,674
29	0,854	1,311	1,699	2,045	2,462	2,756	3,396	3,659
30	0,854	1,310	1,697	2,042	2,457	2,750	3,385	3,646
32	0,853	1,309	1,694	2,037	2,449	2,738	3,365	3,622
34	0,852	1,307	1,691	2,032	2,441	2,728	3,348	3,601
36	0,852	1,306	1,688	2,028	2,434	2,719	3,333	3,582
38	0,851	1,304	1,686	2,024	2,429	2,712	3,319	3,566
40	0,851	1,303	1,684	2,021	2,423	2,704	3,307	3,551
42	0,850	1,302	1,682	2,018	2,418	2,698	3,296	3,538
44	0,850	1,301	1,680	2,015	2,414	2,692	3,286	3,526
46	0,850	1,300	1,679	2,013	2,410	2,687	3,277	3,515
48	0,849	1,299	1,677	2,011	2,407	2,682	3,269	3,505
50	0,849	1,299	1,676	2,009	2,403	2,678	3,261	3,496
55	0,848	1,297	1,673	2,004	2,396	2,668	3,245	3,476
60	0,848	1,296	1,671	2,000	2,390	2,660	3,232	3,460
65	0,847	1,295	1,669	1,997	2,385	2,654	3,220	3,447
70	0,847	1,294	1,667	1,994	2,381	2,648	3,211	3,435
80	0,846	1,292	1,664	1,990	2,374	2,639	3,195	3,416
100	0,845	1,290	1,660	1,984	2,364	2,626	3,174	3,390
150	0,844	1,287	1,655	1,976	2,351	2,609	3,145	3,357
200	0,843	1,286	1,653	1,972	2,345	2,601	3,131	3,340

Nos 368 observations, représentées sous forme tabulaire en 3.2, sont monotones au sens bidimensionnel (décroissantes en  $x$  et croissantes en  $y$ ). Remarquons que les quantiles associés aux deux premiers degrés de liberté sont omis. Les variations extrêmes de ces quantiles en fonction des probabilités ne sont pas reflétées dans le reste de la table. En effet, la valeur  $t_{1;1-\alpha}$  varie de 1,376 à 636,6 pour les probabilités cumulatives à l'étude tandis que  $t_{2;1-\alpha}$  varie de 1,061 à 31,60. Afin d'éviter un mauvais ajustement dû à ces quantiles, nous les enlevons de notre modélisation.

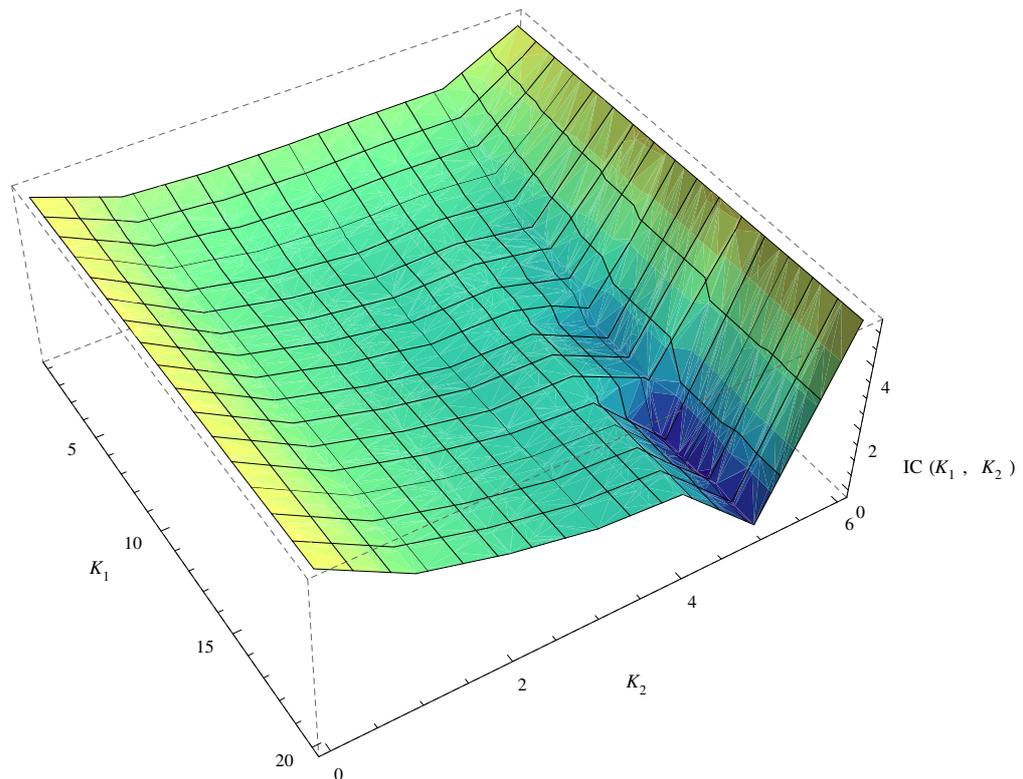


FIGURE 3.4: Sélection des noeuds intérieurs par le critère He et Shi pour la table  $t$  de la loi de Student

Nous implémentons l'algorithme décrit à la section précédente afin d'estimer nos données. Nous choisissons  $n_0 = 0,45$  et  $\alpha = 500$  comme valeurs initiales des hyperparamètres. Pour définir la base spline, nous optimisons les valeurs du critère He et Shi qui sont illustrées en fonction des nombres de noeuds intérieurs à la figure 3.4. Le minimum est obtenu en  $k_1 = 15$  et  $k_2 = 5$ , ce qui correspond à un ensemble de 15 noeuds intérieurs en  $x$  et 5 noeuds en  $y$  dont l'emplacement

est prédéterminé aux quantiles équidistants de  $f$ . L'optimalité de chaque position est vérifiée : le retrait de deux noeuds en  $x$  diminue le critère d'information. En considérant les noeuds confondus, ce nouvel ensemble de noeuds construit une base spline de 128 éléments.

TABLEAU 3.3: RECONSTITUTION DE LA TABLE  $t$  DE LA LOI DE STUDENT AVEC LES COEFFICIENTS ESTIMÉS

d.l. \ $1 - \alpha$	0,80	0,90	0,95	0,975	0,99	0,995	0,999	0,9995
3	0,978	1,632	2,369	3,160	4,556	5,828	10,199	12,900
4	0,941	1,533	2,138	2,775	3,763	4,622	7,235	8,707
5	0,919	1,475	2,012	2,563	3,350	4,005	5,813	6,746
6	0,906	1,441	1,945	2,451	3,150	3,719	5,234	5,995
7	0,896	1,415	1,894	2,365	2,998	3,502	4,796	5,427
8	0,889	1,396	1,859	2,305	2,895	3,354	4,501	5,044
9	0,884	1,383	1,834	2,263	2,823	3,251	4,298	4,782
10	0,879	1,372	1,813	2,228	2,763	3,168	4,138	4,578
11	0,875	1,363	1,796	2,201	2,718	3,105	4,021	4,431
12	0,873	1,356	1,782	2,179	2,681	3,055	3,931	4,320
13	0,870	1,350	1,771	2,160	2,650	3,013	3,854	4,224
14	0,868	1,345	1,761	2,145	2,624	2,977	3,788	4,142
15	0,866	1,341	1,753	2,132	2,603	2,947	3,733	4,073
16	0,865	1,337	1,746	2,120	2,583	2,921	3,685	4,013
17	0,863	1,333	1,740	2,110	2,567	2,898	3,645	3,964
18	0,862	1,330	1,734	2,101	2,552	2,879	3,611	3,922
19	0,861	1,328	1,729	2,093	2,539	2,861	3,580	3,884
20	0,860	1,325	1,725	2,086	2,528	2,845	3,552	3,850
21	0,859	1,323	1,721	2,080	2,518	2,831	3,527	3,819
22	0,858	1,321	1,717	2,074	2,508	2,819	3,505	3,792

TABLEAU 3.3 (SUITE)

$1 - \alpha$ d.l.	0,80	0,90	0,95	0,975	0,99	0,995	0,999	0,9995
23	0,858	1,319	1,714	2,069	2,500	2,807	3,485	3,768
24	0,857	1,318	1,711	2,064	2,492	2,797	3,467	3,746
25	0,856	1,316	1,708	2,060	2,485	2,787	3,450	3,725
26	0,856	1,315	1,706	2,056	2,479	2,779	3,435	3,707
27	0,855	1,314	1,703	2,052	2,473	2,771	3,421	3,689
28	0,855	1,312	1,701	2,048	2,467	2,763	3,408	3,674
29	0,854	1,311	1,699	2,045	2,462	2,756	3,396	3,659
30	0,854	1,310	1,697	2,042	2,457	2,750	3,385	3,646
32	0,853	1,309	1,694	2,037	2,449	2,738	3,365	3,622
34	0,852	1,307	1,691	2,032	2,441	2,728	3,348	3,601
36	0,852	1,306	1,688	2,028	2,434	2,719	3,332	3,582
38	0,851	1,304	1,686	2,024	2,429	2,712	3,319	3,566
40	0,851	1,303	1,684	2,021	2,423	2,705	3,307	3,551
42	0,850	1,302	1,682	2,018	2,419	2,698	3,296	3,538
44	0,850	1,301	1,680	2,015	2,414	2,692	3,286	3,526
46	0,849	1,300	1,679	2,013	2,410	2,687	3,277	3,515
48	0,849	1,299	1,677	2,011	2,406	2,682	3,269	3,505
50	0,849	1,299	1,676	2,009	2,403	2,678	3,261	3,496
55	0,848	1,297	1,673	2,004	2,396	2,668	3,245	3,476
60	0,848	1,296	1,671	2,000	2,390	2,660	3,232	3,460
65	0,847	1,295	1,669	1,997	2,385	2,654	3,221	3,447
70	0,847	1,294	1,667	1,995	2,381	2,648	3,211	3,435
80	0,846	1,292	1,664	1,990	2,374	2,639	3,195	3,416
100	0,845	1,290	1,660	1,984	2,364	2,626	3,174	3,390
150	0,844	1,287	1,655	1,976	2,352	2,609	3,145	3,357
200	0,843	1,286	1,653	1,972	2,345	2,601	3,131	3,340

Le modèle bayésien final des quantiles  $t$  de la distribution de Student nous donne l'estimation tabulée en 3.3, arrondie à la troisième décimale (les coefficients estimés se trouvent à l'annexe A.2). Contrairement aux données normales, nous observons des différences entre la table originale et celle estimée, notamment pour les premiers degrés de liberté. En effet, la figure 3.5, qui illustre les erreurs absolues par une carte des points chauds, indique un regroupement d'erreurs importantes d'ordre supérieur à  $10^{-2}$ . En particulier, l'erreur maximale de 0,122 est associée au quantile  $t_{5;0,9995} = 6,869$ . Similairement, l'erreur absolue minimale est de  $3,78 \times 10^{-9}$  pour le quantile  $t_{34;0,8} = 0,852$ . Quant à l'erreur absolue moyenne, elle est de  $2,10 \times 10^{-3}$  pour le modèle bayésien contre  $2,09 \times 10^{-3}$  pour le modèle classique. Notons toutefois que la majorité des entrées ont une erreur inférieure ou égale à  $10^{-4}$ , d'où nous obtenons une erreur médiane de  $6,80 \times 10^{-5}$ .

La modélisation bayésienne avec splines ne permet pas seulement d'estimer les entrées de la table, mais également d'interpoler entre les observations. En effet, le modèle avec splines peut estimer les valeurs associées à tous les  $x$  et  $y$  dans le domaine de définition. Il est plus pertinent d'étudier l'interpolation sur la table  $t$  de la loi de Student que sur la table normale, car les données sont moins lisses et donc moins favorables à l'interpolation linéaire. Par exemple, nous considérons les quantiles  $t_{d.l.;0,95}$  pour l'ensemble des degrés de liberté possibles. Lorsque les degrés sont inférieurs à 80, l'interpolation des données et l'estimation par splines présentent essentiellement le même comportement. Ce phénomène est expliqué par les observations disponibles à presque chaque degré (d'où l'interpolation efficace). Pour les degrés de liberté supérieurs à 80 pour lesquels l'information est largement omise de la table, nous comparons notre modèle et l'interpolation linéaire avec la courbe des quantiles exacts (figure 3.6). Nous constatons que notre estimateur ajuste mieux les quantiles que l'interpolation, bien que l'ordre des différences soit faible. Ainsi, dans cet exemple, le modèle bayésien ajoute une précision supplémentaire aux valeurs entre les entrées de la table.

La réduction d'espace de rangement apportée par notre modélisation est moins importante qu'au premier exemple. Nous reconstituons nos 368 observations à l'aide de 128 coefficients, donc nous diminuons la quantité d'information de 65,2%.

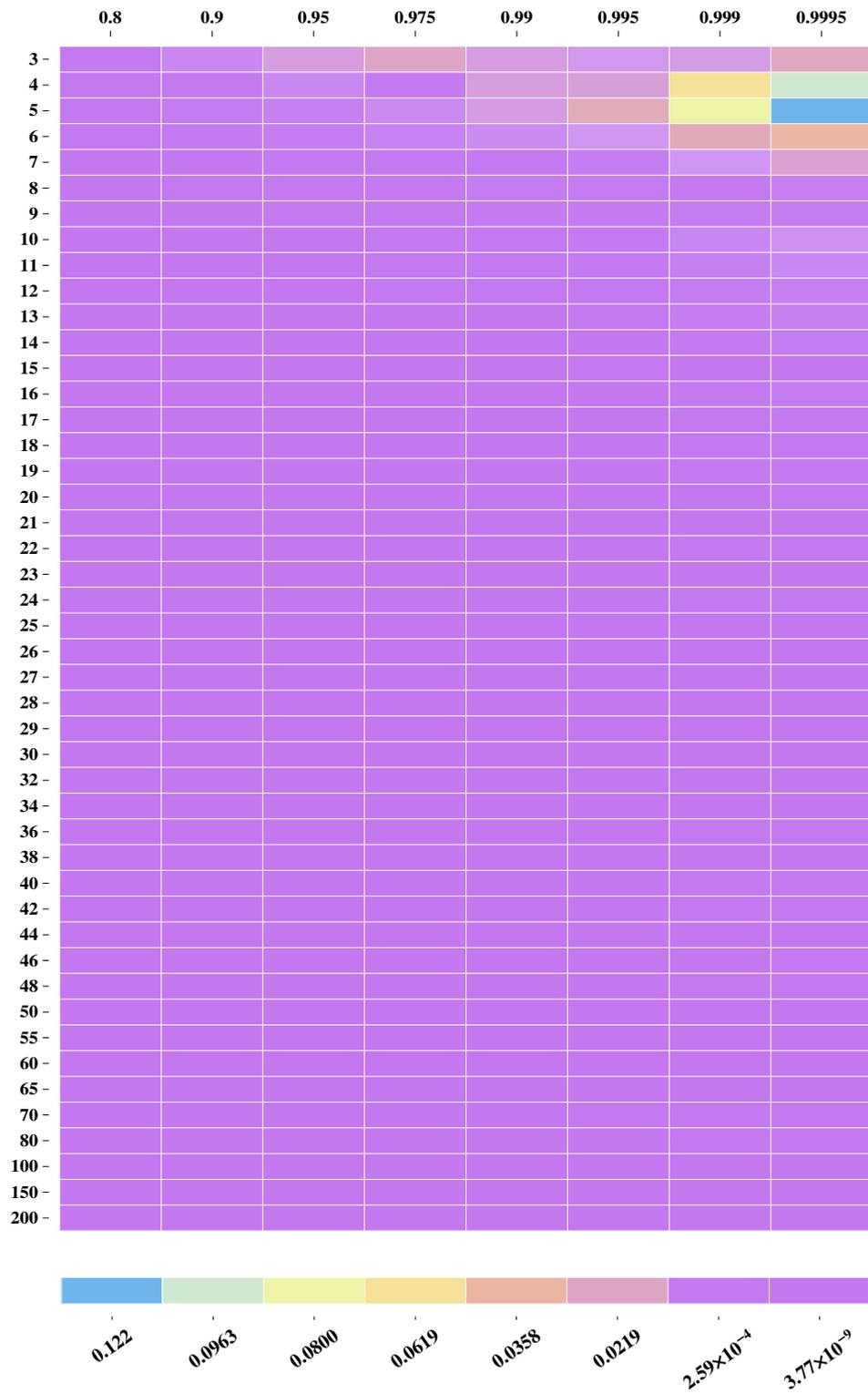


FIGURE 3.5: Cartographie des points chauds : les erreurs absolues associées à l'estimation de la table  $t$  de la loi de Student

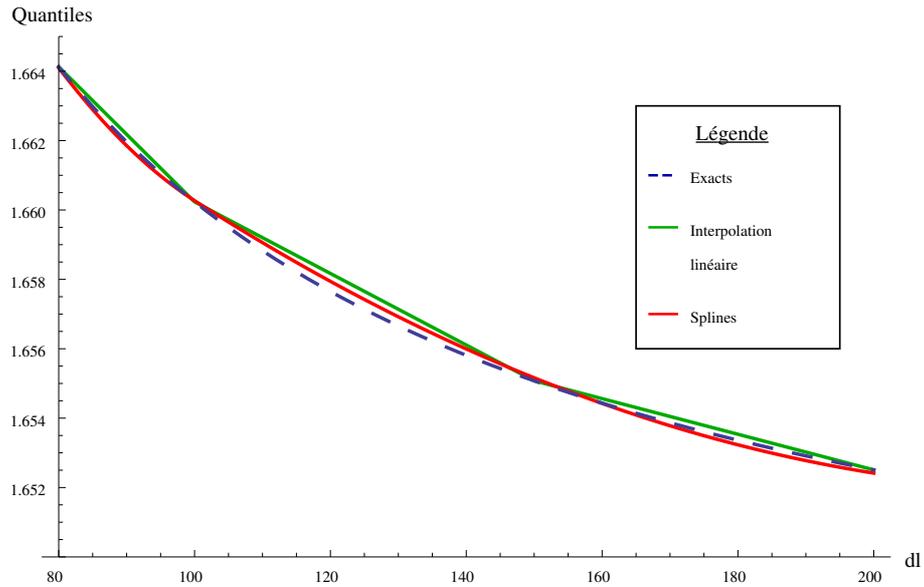


FIGURE 3.6: Quantiles  $t_{d.l.;0,95}$  en fonction des degrés de liberté : quantiles exacts, interpolation linéaire des données de la table et estimation par le modèle de splines

En termes informatiques, la taille des données passe de 4416 octets à 1896 octets, soit une diminution de 57,1%. De plus, le temps nécessaire pour reconstruire la table des quantiles  $t$  à partir des coefficients est semblable au temps requis par le logiciel  $R$  pour obtenir les données originales, c'est-à-dire environ 0,015 seconde. Les résultats moins performants de notre modèle bayésien par rapport au premier exemple peuvent être attribués aux grandes variations des données associées au troisième degré de liberté. En effet, plus de coefficients sont nécessaires afin de bien représenter le comportement des données.

### 3.3. TABLE DE MORTALITÉ

Les deux premiers exemples de ce chapitre constituent des tables de données dont les entrées peuvent être facilement obtenues à travers les fonctions prédéfinies des logiciels. Nous sommes plutôt intéressés par des données qui s'obtiennent seulement sous forme tabulaire. Il est important de faciliter l'accès aux entrées de ces tables lorsqu'elles sont requises pour un quelconque calcul. Notre modélisation bayésienne avec fonctions splines nous permet justement d'estimer chaque entrée individuelle de la table (ou bien la table au complet). Elle permet également

l'interpolation entre les données. L'estimation par un ensemble de coefficients réduit l'espace de rangement des grandes tables, une propriété désirable si plusieurs tables sont utilisées.

Notre source de données pour ce troisième exemple est la Base de données sur la longévité canadienne (BDLC). La BDLC est une réalisation du département de démographie de l'université de Montréal (voir site BDLC, 2010). Elle contient les informations nécessaires à l'établissement des tables de mortalité pour le Canada ainsi que chaque province et territoire. En général, les données comprises dans les tables de mortalité varient selon l'âge (dénnoté par  $x$ ) et la période/année. Les tables fournissent notamment le nombre de survivants à chaque âge,  $l_x$ , sur une cohorte initiale de 100 000 individus ( $l_0$ ). Parmi les autres informations se trouvant dans les tables de mortalité, il y a les probabilités de décès d'un individu d'âge  $x$  avant son prochain anniversaire,  $q_x$ . Notons que le nombre de survivants est strictement décroissant par rapport à l'âge  $x$ , mais que les probabilités de décès ne sont pas nécessairement monotones. En effet, le taux de mortalité infantile est plus élevé qu'aux âges subséquents. Cependant, les probabilités de décès  $q_x$  découlent du nombre de survivants  $l_x$ , ce qui rend ces informations interchangeables :

$$q_x = \frac{l_x - l_{x+1}}{l_x}.$$

Dans le cadre de ce mémoire, nous avons sélectionné une table de mortalité pour le Québec et pour les deux sexes réunis. Nous définissons alors l'âge des individus par  $x$  et l'année pour laquelle sont compilées les données par  $y$ . Afin de travailler avec des données monotones, nous considérons uniquement le nombre de survivants  $l_x$ . En divisant  $l_x$  par le nombre d'individus au départ, nous cherchons plutôt à modéliser la probabilité de survivre jusqu'à l'âge  $x$  (dénnoté  $S(x)$ ). Nous sélectionnons également un sous-ensemble de la table de mortalité disponible, soient les âges de 60 à 110 et les années 1996 à 2006 pour un total de 561 données, afin de faciliter la représentation visuelle des données au tableau 3.4. Notons toutefois que les données ne sont pas monotones en  $y$ , c'est-à-dire d'une période d'observation à l'autre.

TABLEAU 3.4: TABLE DE MORTALITÉ

année âge	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
60	0,8988	0,9011	0,9049	0,9039	0,9092	0,9112	0,9120	0,9134	0,9159	0,9168	0,9193
61	0,8912	0,8928	0,8973	0,8960	0,9020	0,9040	0,9048	0,9063	0,9087	0,9101	0,9129
62	0,8822	0,8840	0,8885	0,8877	0,8936	0,8958	0,8970	0,8987	0,9008	0,9027	0,9063
63	0,8725	0,8744	0,8790	0,8790	0,8847	0,8871	0,8889	0,8901	0,8929	0,8947	0,8990
64	0,8624	0,8638	0,8691	0,8690	0,8751	0,8784	0,8795	0,8812	0,8842	0,8859	0,8914
65	0,8512	0,8529	0,8589	0,8584	0,8650	0,8685	0,8695	0,8715	0,8751	0,8761	0,8826
66	0,8388	0,8410	0,8466	0,8465	0,8535	0,8572	0,8585	0,8614	0,8653	0,8666	0,8735
67	0,8250	0,8276	0,8323	0,8328	0,8420	0,8452	0,8466	0,8503	0,8546	0,8561	0,8633
68	0,8106	0,8125	0,8181	0,8195	0,8288	0,8321	0,8338	0,8382	0,8436	0,8446	0,8516
69	0,7949	0,7974	0,8028	0,8046	0,8142	0,8184	0,8198	0,8254	0,8302	0,8318	0,8386
70	0,7780	0,7804	0,7871	0,7887	0,7998	0,8030	0,8049	0,8120	0,8164	0,8180	0,8257
71	0,7594	0,7621	0,7689	0,7723	0,7841	0,7862	0,7875	0,7971	0,8014	0,8031	0,8113
72	0,7398	0,7433	0,7501	0,7535	0,7667	0,7679	0,7711	0,7794	0,7851	0,7876	0,7956
73	0,7195	0,7224	0,7295	0,7339	0,7474	0,7495	0,7530	0,7617	0,7676	0,7702	0,7790
74	0,6965	0,7001	0,7083	0,7119	0,7271	0,7302	0,7335	0,7428	0,7489	0,7512	0,7625
75	0,6724	0,6769	0,6847	0,6888	0,7048	0,7084	0,7123	0,7218	0,7296	0,7316	0,7433
76	0,6482	0,6508	0,6605	0,6648	0,6816	0,6862	0,6899	0,6986	0,7066	0,7103	0,7232

TABLEAU 3.4 (SUITE)

année âge	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
77	0,6217	0,6233	0,6349	0,6396	0,6566	0,6623	0,6660	0,6741	0,6833	0,6880	0,7007
78	0,5950	0,5958	0,6065	0,6122	0,6306	0,6358	0,6397	0,6490	0,6580	0,6637	0,6771
79	0,5650	0,5682	0,5759	0,5836	0,6018	0,6092	0,6122	0,6218	0,6317	0,6382	0,6517
80	0,5360	0,5362	0,5458	0,5539	0,5706	0,5791	0,5828	0,5939	0,6026	0,6110	0,6250
81	0,5047	0,5036	0,5130	0,5235	0,5387	0,5488	0,5517	0,5648	0,5732	0,5821	0,5964
82	0,4712	0,4686	0,4796	0,4894	0,5078	0,5155	0,5196	0,5328	0,5420	0,5516	0,5672
83	0,4374	0,4324	0,4445	0,4555	0,4738	0,4824	0,4848	0,4995	0,5087	0,5179	0,5345
84	0,4023	0,3972	0,4085	0,4190	0,4391	0,4470	0,4505	0,4646	0,4741	0,4837	0,5014
85	0,3653	0,3614	0,3720	0,3804	0,4039	0,4105	0,4131	0,4300	0,4386	0,4481	0,4659
86	0,3296	0,3255	0,3358	0,3441	0,3658	0,3740	0,3761	0,3923	0,4016	0,4100	0,4316
87	0,2954	0,2894	0,3016	0,3070	0,3296	0,3362	0,3396	0,3562	0,3612	0,3747	0,3957
88	0,2609	0,2537	0,2649	0,2723	0,2936	0,2977	0,2998	0,3191	0,3229	0,3361	0,3576
89	0,2282	0,2200	0,2291	0,2359	0,2581	0,2610	0,2624	0,2810	0,2862	0,2964	0,3190
90	0,1961	0,1878	0,1964	0,2020	0,2249	0,2271	0,2272	0,2435	0,2484	0,2584	0,2833
91	0,1653	0,1571	0,1635	0,1707	0,1906	0,1942	0,1928	0,2103	0,2138	0,2235	0,2472
92	0,1382	0,1297	0,1354	0,1420	0,1606	0,1622	0,1621	0,1770	0,1806	0,1880	0,2131
93	0,1125	0,1058	0,1094	0,1156	0,1320	0,1320	0,1335	0,1458	0,1492	0,1579	0,1809

TABLEAU 3.4 (SUITE)

année âge	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
94	0,0902	0,0828	0,0876	0,0918	0,1078	0,1046	0,1068	0,1192	0,1211	0,1303	0,1493
95	0,0705	0,0642	0,0678	0,0719	0,0834	0,0843	0,0844	0,0951	0,0954	0,1051	0,1208
96	0,0540	0,0485	0,0514	0,0548	0,0643	0,0645	0,0647	0,0738	0,0736	0,0820	0,0963
97	0,0404	0,0358	0,0379	0,0408	0,0484	0,0480	0,0484	0,0558	0,0553	0,0624	0,0751
98	0,0295	0,0257	0,0273	0,0296	0,0355	0,0348	0,0353	0,0411	0,0404	0,0462	0,0572
99	0,0210	0,0180	0,0190	0,0208	0,0253	0,0244	0,0250	0,0295	0,0287	0,0333	0,0424
100	0,0145	0,0122	0,0129	0,0143	0,0175	0,0167	0,0172	0,0205	0,0197	0,0233	0,0306
101	0,0097	0,0080	0,0085	0,0095	0,0118	0,0110	0,0115	0,0138	0,0131	0,0158	0,0215
102	0,0063	0,0051	0,0054	0,0061	0,0077	0,0070	0,0074	0,0090	0,0084	0,0104	0,0147
103	0,0040	0,0032	0,0033	0,0038	0,0048	0,0043	0,0046	0,0057	0,0052	0,0066	0,0097
104	0,0024	0,0019	0,0020	0,0023	0,0030	0,0026	0,0028	0,0035	0,0032	0,0041	0,0062
105	0,0014	0,0011	0,0011	0,0014	0,0018	0,0015	0,0016	0,0020	0,0018	0,0024	0,0038
106	0,0008	0,0006	0,0006	0,0008	0,0010	0,0008	0,0009	0,0012	0,0010	0,0014	0,0023
107	0,0005	0,0003	0,0003	0,0004	0,0006	0,0004	0,0005	0,0006	0,0006	0,0008	0,0013
108	0,0002	0,0002	0,0002	0,0002	0,0003	0,0002	0,0003	0,0003	0,0003	0,0004	0,0008
109	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0002	0,0001	0,0002	0,0004
110	0,0001	0,0000	0,0000	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0002

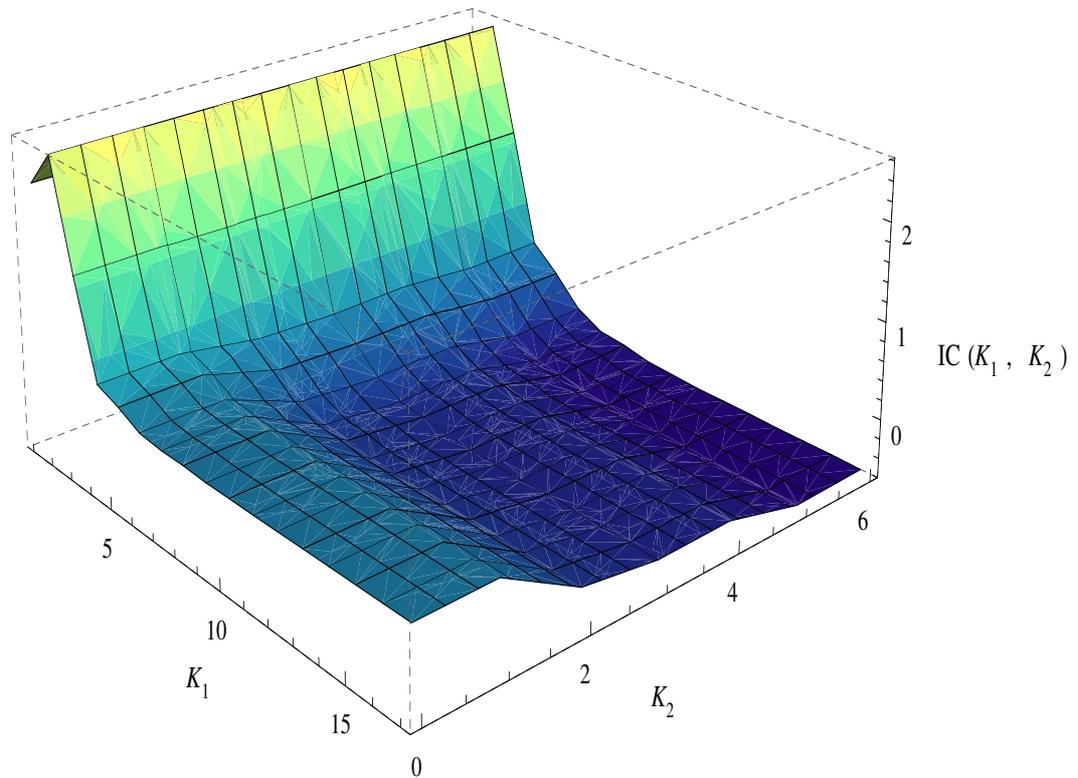


FIGURE 3.7: Sélection des noeuds intérieurs par le critère He et Shi pour la table de mortalité

Avant de démontrer l'efficacité de notre modèle bayésien, nous définissons la base de fonctions splines utilisée, en fixant les hyperparamètres  $n_0 = 0,45$  et  $\alpha = 500$ . Les valeurs du critère He et Shi obtenues en fonction des noeuds intérieurs sont illustrées dans la figure 3.7. Bien que le graphique semble atteindre un minimum vers  $k_1 = 7$  et  $k_2 = 2$ , il ne s'agit que d'un minimum local. Le minimum global se trouve plutôt à l'extrémité du graphique, pour  $k_1 = 11$  et  $k_2 = 5$ . Un nombre inférieur ou supérieur de noeuds intérieurs en  $x$  n'améliore pas le critère d'information. Par conséquent, nous obtenons une base composée de 112 éléments, donc 112 coefficients à estimer.

TABLEAU 3.5: RECONSTITUTION DE LA TABLE DE MORTALITÉ AVEC LES COEFFICIENTS ESTIMÉS

année âge	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
60	0,8989	0,9015	0,9034	0,9059	0,9085	0,9106	0,9123	0,9140	0,9155	0,9170	0,9193
61	0,8910	0,8932	0,8956	0,8983	0,9010	0,9033	0,9052	0,9067	0,9082	0,9101	0,9129
62	0,8823	0,8843	0,8871	0,8900	0,8928	0,8953	0,8974	0,8989	0,9005	0,9027	0,9062
63	0,8727	0,8747	0,8778	0,8809	0,8840	0,8867	0,8890	0,8907	0,8924	0,8947	0,8990
64	0,8623	0,8645	0,8676	0,8711	0,8745	0,8774	0,8799	0,8819	0,8838	0,8861	0,8914
65	0,8510	0,8534	0,8565	0,8603	0,8642	0,8673	0,8700	0,8724	0,8746	0,8768	0,8830
66	0,8386	0,8412	0,8443	0,8486	0,8530	0,8563	0,8591	0,8620	0,8647	0,8668	0,8736
67	0,8252	0,8278	0,8310	0,8359	0,8409	0,8442	0,8472	0,8508	0,8540	0,8560	0,8630
68	0,8107	0,8134	0,8167	0,8223	0,8279	0,8313	0,8343	0,8386	0,8426	0,8444	0,8514
69	0,7950	0,7978	0,8013	0,8075	0,8138	0,8173	0,8203	0,8255	0,8302	0,8319	0,8388
70	0,7779	0,7810	0,7846	0,7916	0,7986	0,8021	0,8053	0,8112	0,8166	0,8182	0,8254
71	0,7595	0,7628	0,7668	0,7746	0,7822	0,7859	0,7892	0,7958	0,8018	0,8033	0,8111
72	0,7398	0,7434	0,7478	0,7563	0,7646	0,7685	0,7720	0,7793	0,7858	0,7871	0,7960
73	0,7188	0,7227	0,7275	0,7367	0,7456	0,7498	0,7537	0,7614	0,7684	0,7698	0,7797
74	0,6965	0,7005	0,7058	0,7156	0,7252	0,7298	0,7339	0,7420	0,7494	0,7512	0,7621
75	0,6729	0,6769	0,6825	0,6930	0,7032	0,7084	0,7128	0,7211	0,7288	0,7315	0,7431
76	0,6482	0,6520	0,6579	0,6688	0,6797	0,6855	0,6904	0,6987	0,7066	0,7106	0,7227

TABLEAU 3.5 (SUITE)

année âge	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
77	0,6221	0,6255	0,6317	0,6432	0,6547	0,6611	0,6664	0,6747	0,6829	0,6882	0,7008
78	0,5946	0,5973	0,6039	0,6160	0,6282	0,6351	0,6408	0,6491	0,6578	0,6641	0,6772
79	0,5658	0,5676	0,5744	0,5872	0,6002	0,6075	0,6134	0,6221	0,6312	0,6383	0,6520
80	0,5356	0,5362	0,5434	0,5570	0,5706	0,5781	0,5842	0,5934	0,6031	0,6108	0,6251
81	0,5040	0,5031	0,5108	0,5251	0,5394	0,5472	0,5534	0,5633	0,5737	0,5816	0,5966
82	0,4711	0,4689	0,4770	0,4920	0,5069	0,5147	0,5210	0,5315	0,5426	0,5507	0,5664
83	0,4373	0,4340	0,4422	0,4577	0,4730	0,4808	0,4873	0,4983	0,5099	0,5181	0,5347
84	0,4023	0,3984	0,4066	0,4222	0,4377	0,4457	0,4522	0,4635	0,4753	0,4839	0,5015
85	0,3663	0,3621	0,3701	0,3856	0,4010	0,4091	0,4158	0,4271	0,4391	0,4479	0,4668
86	0,3302	0,3257	0,3334	0,3486	0,3639	0,3720	0,3786	0,3899	0,4019	0,4109	0,4310
87	0,2950	0,2901	0,2972	0,3121	0,3273	0,3350	0,3413	0,3524	0,3644	0,3732	0,3946
88	0,2608	0,2551	0,2614	0,2761	0,2912	0,2982	0,3039	0,3148	0,3266	0,3349	0,3578
89	0,2274	0,2208	0,2260	0,2406	0,2555	0,2616	0,2663	0,2770	0,2886	0,2959	0,3203
90	0,1956	0,1881	0,1923	0,2065	0,2211	0,2262	0,2298	0,2401	0,2514	0,2579	0,2833
91	0,1659	0,1580	0,1614	0,1749	0,1887	0,1929	0,1958	0,2055	0,2162	0,2221	0,2475
92	0,1384	0,1306	0,1333	0,1456	0,1583	0,1619	0,1643	0,1731	0,1830	0,1886	0,2130
93	0,1130	0,1057	0,1080	0,1188	0,1299	0,1331	0,1352	0,1428	0,1517	0,1575	0,1798
94	0,0902	0,0836	0,0857	0,0948	0,1042	0,1071	0,1089	0,1153	0,1231	0,1290	0,1488

TABLEAU 3.5 (SUITE)

année âge	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
95	0,0705	0,0646	0,0664	0,0740	0,0818	0,0843	0,0860	0,0912	0,0979	0,1037	0,1211
96	0,0538	0,0487	0,0502	0,0563	0,0627	0,0648	0,0662	0,0705	0,0760	0,0815	0,0966
97	0,0401	0,0358	0,0370	0,0418	0,0469	0,0486	0,0497	0,0530	0,0576	0,0624	0,0754
98	0,0295	0,0259	0,0268	0,0305	0,0344	0,0357	0,0365	0,0390	0,0425	0,0463	0,0575
99	0,0212	0,0183	0,0190	0,0217	0,0245	0,0254	0,0260	0,0278	0,0303	0,0332	0,0426
100	0,0146	0,0123	0,0128	0,0147	0,0167	0,0173	0,0177	0,0189	0,0207	0,0228	0,0304
101	0,0096	0,0079	0,0082	0,0096	0,0109	0,0113	0,0115	0,0123	0,0135	0,0151	0,0211
102	0,0063	0,0051	0,0053	0,0062	0,0071	0,0074	0,0075	0,0080	0,0089	0,0100	0,0145
103	0,0041	0,0033	0,0034	0,0041	0,0047	0,0049	0,0049	0,0053	0,0059	0,0067	0,0099
104	0,0024	0,0019	0,0020	0,0024	0,0028	0,0029	0,0030	0,0032	0,0036	0,0042	0,0063
105	0,0013	0,0010	0,0011	0,0013	0,0015	0,0016	0,0016	0,0018	0,0020	0,0024	0,0038
106	0,0008	0,0006	0,0006	0,0007	0,0009	0,0009	0,0009	0,0010	0,0011	0,0014	0,0023
107	0,0005	0,0004	0,0004	0,0005	0,0005	0,0005	0,0005	0,0006	0,0007	0,0008	0,0014
108	0,0003	0,0002	0,0002	0,0002	0,0003	0,0003	0,0003	0,0003	0,0003	0,0004	0,0007
109	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0002	0,0004
110	0,0000	0,0000	0,0000	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0002

Les données de mortalité évaluées en utilisant les coefficients (voir l'annexe A.3) et les T-splines sont présentées au tableau 3.5. Hormis quelques exceptions, la majorité des différences observées entre la table originale et celle reconstituée se trouvent à la quatrième décimale des entrées. En effet, l'erreur absolue moyenne de l'estimation est  $9,825 \times 10^{-4}$  tandis que la médiane est  $2,473 \times 10^{-5}$ . Quelques entrées présentent une erreur d'ordre  $10^{-3}$ , telle la survie jusqu'à 86 ans pour la période 1999 (erreur absolue maximale de  $5,194 \times 10^{-3}$  pour une valeur originale de 0,3441). L'erreur minimale, soit  $7,606 \times 10^{-8}$ , est associée à la probabilité de survie 0,7084 pour l'âge 75 et l'année 2001. La répartition détaillée des erreurs est illustrée à la figure 3.8.

Comme mentionné précédemment, 112 coefficients sont nécessaires afin de reconstruire les 561 entrées de la table : il s'agit d'une réduction de 80% de la quantité d'information à conserver. L'espace de rangement de la table de mortalité est de 4609 octets tandis que seulement 1417 octets sont requis pour les coefficients. Cette diminution de 69,3% est significative dans un contexte de travail avec de grandes tables de données. De plus, le logiciel *R* prend environ 0,032 seconde pour lire la table de mortalité tandis que le temps nécessaire à construire la table à partir des coefficients est de 0,016 seconde. Ce gain de vitesse de 50% favorise donc l'accès aux données de forme tabulaire.

Finalement, le modèle bayésien décrit dans ce mémoire procure un ajustement semblable au modèle classique. Notre modèle, robuste au choix des hyperparamètres des distributions *a priori*, donne assurément une estimation respectant les propriétés de monotonie des données initiales. En se restreignant aux tables qui sont suffisamment lisses (qui présentent un certain degré de régularité), nous pouvons ainsi condenser à plus de 65% les différentes tables bidimensionnelles. De plus, cette économie d'espace ne se fait pas au détriment de la vitesse d'évaluation. En particulier, pour des tables ne possédant pas de forme fonctionnelle préexistante dans les logiciels, le gain de temps de notre estimation peut atteindre jusqu'à 50%. L'accès à de telles données est donc grandement facilité par notre modélisation.

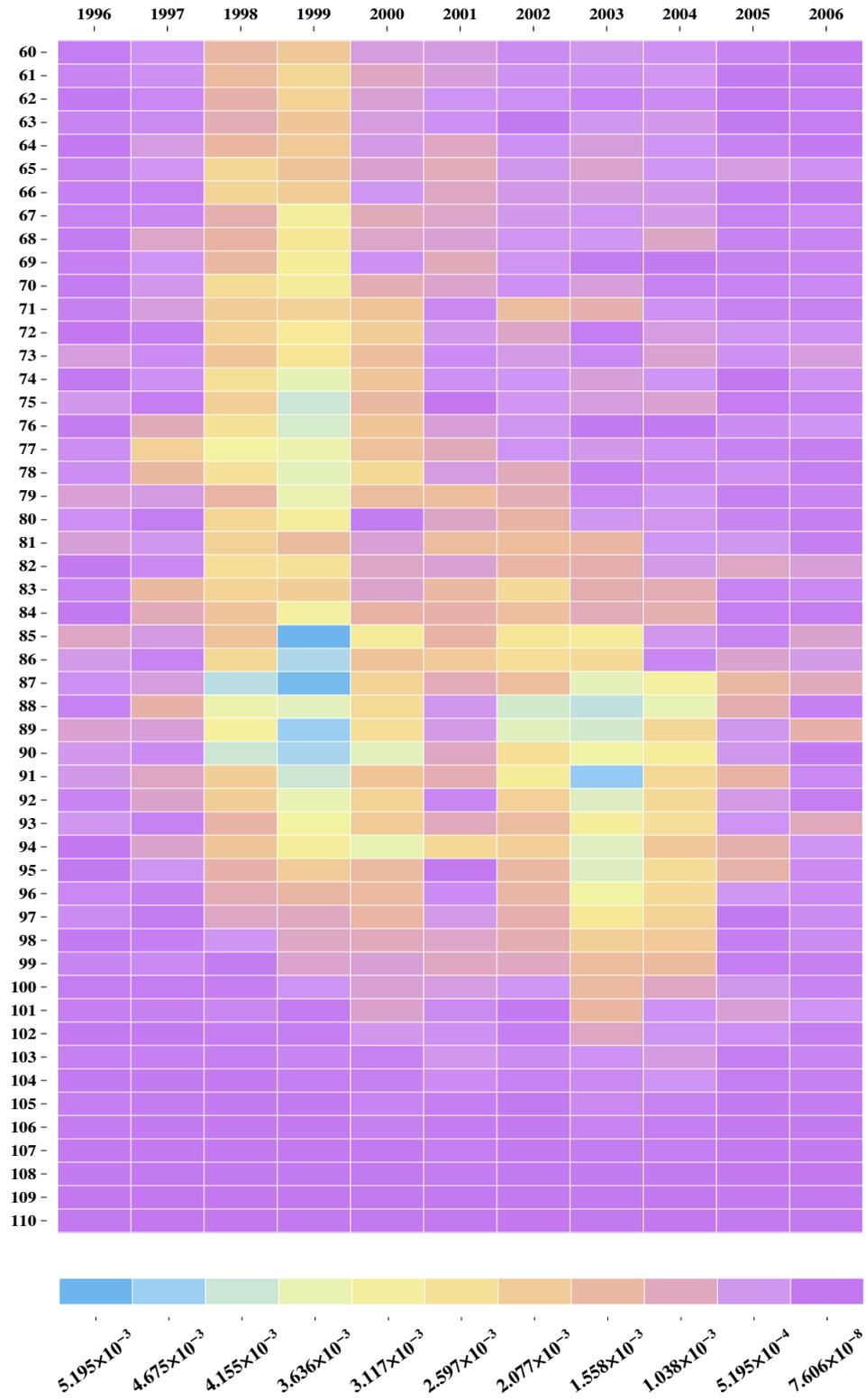


FIGURE 3.8: Cartographie des points chauds : les erreurs absolues associées à la table de mortalité

# CONCLUSION

---

Dans ce mémoire, nous avons étudié l'estimation de tables à deux entrées dont les colonnes et/ou les lignes sont monotones. Une approche bayésienne a été utilisée sur les méthodes de régression non paramétrique appliquée dans la littérature à la modélisation de surfaces. L'objectif final était de modéliser les tables de mortalité afin de réduire leur espace d'entreposage et d'y faciliter l'accès. Afin de conserver la propriété de monotonie des tables, nous nous sommes inspirés des contraintes linéaires proposées dans la publication de Beliakov (2000).

La première étape de notre procédure était de définir la base de fonctions monotones dont la combinaison représentait la forme fonctionnelle des données. Les fonctions splines ont été retenues à l'encontre des polynômes (manque de flexibilité) et des ondelettes (fonctions non monotones). Les T-splines, fonctions monotones dérivées des B-splines usuelles, avaient l'avantage d'un support compact et d'une flexibilité leur permettant un bon ajustement local. Le nombre de noeuds, en plus d'affecter la flexibilité des splines, déterminait leur capacité de compression des données (par le nombre de coefficients).

L'étape suivante consistait à l'élaboration d'un algorithme de sélection des noeuds dans un contexte bidimensionnel afin d'optimiser le nombre de coefficients en tenant compte de la qualité de l'ajustement. Une approche bayésienne par modèles hiérarchiques a été appliquée au modèle. Nous avons ainsi défini un modèle *a priori* pour les données, dans lequel nous avons incorporé les contraintes de monotonie. L'évaluation du modèle *a posteriori* a ensuite nécessité des méthodes de calcul numérique de Monte Carlo, à savoir l'algorithme d'échantillonnage de Gibbs. Ainsi, chaque vecteur de coefficients généré était conforme aux contraintes de monotonie.

Finalement, nous avons testé notre modèle bayésien sur de vraies données tabulaires afin d'étudier le comportement de l'estimateur final. Le premier test a été effectué sur une table de la loi normale, où les données sont lisses. Le modèle a ensuite été testé sur des données présentant plus de variations, soit une table  $t$  de distribution de Student. L'application finale s'est faite sur une table de mortalité du Québec provenant de la BDLC, dont les probabilités de survie étaient décroissantes avec l'âge.

À travers les deux premiers exemples, nous avons étudié la sensibilité des paramètres des distributions *a priori* ainsi que la capacité d'interpolation de l'estimateur. Puisque la variation des hyperparamètres n'affectait pas la qualité de l'ajustement de l'estimateur, nous avons conclu que le modèle était robuste au choix des distributions *a priori*. Quant à l'estimation de valeurs entre les entrées de la table  $t$  de la loi Student, notre estimateur semblait suivre de plus près les quantiles exacts que l'interpolation linéaire des observations. Nous pouvons extrapoler ces deux résultats aux autres tables de données, notamment à la table de mortalité considérée dans ce mémoire.

En général, notre modèle bayésien a bien performé : la qualité de l'ajustement est comparable au modèle classique, en s'assurant que la monotonie de la table soit respectée. Le pouvoir de compression du modèle est considérable pour les tables relativement lisses. Les tables un peu moins lisses peuvent tout de même être condensées à 65% (exemple de la table  $t$  de la loi de Student). L'objectif de réduire l'espace des tables à deux entrées a été satisfait sans nécessiter un surplus de temps pour calculer l'estimation. En effet, le temps de reconstruction de la table à partir des coefficients estimés était semblable au temps d'évaluation de la table originale lorsque les fonctions étaient incorporées au logiciel. Pour la lecture d'une table non incorporée au logiciel (telle que la table de mortalité), nous avons inscrit un important gain de vitesse pour l'évaluation de notre estimation.

# BIBLIOGRAPHIE

---

- [1] AHLBERG, J. H., NILSON, E. N. ET WALSH, J. L. (1964). Fundamental Properties of Generalized Splines. *Proceedings of the National Academy of Sciences of the United States of America*, **52** (6), 1412-1419.
- [2] AHLBERG, J. H., NILSON, E. N. ET WALSH, J. L. (1967). *The Theory of Splines and their Applications*. New York : Academic Press.
- [3] AKAIKE, H. (1974). A New Look at the Statistical Model Identification. *I.E.E.E. Transactions on Automatic Control*, **19** (6), 716-723.
- [4] BAYES, T. (1763). An Essay Toward Solving a Problem in the Doctrine of Chance. *Philosophical Transactions of the Royal Society of London*, **53**, 370-418.
- [5] BARLOW, R., BARTHOLEMEW, D., BREMNER, J. ET BRUNK, H. (1972). *Statistical Inference under Order Restrictions*. New York : Wiley.
- [6] BARTHOLEMEW, D. (1959). A test of homogeneity for ordered alternatives. *Biometrika*, **46**, 36-48.
- [7] BDCL (2010). *Base de données sur la longévité canadienne*. Département de démographie, Université de Montréal. Disponible au [www.bdlc.umontreal.ca](http://www.bdlc.umontreal.ca) (données téléchargées en juillet 2009).
- [8] BELIAKOV, G. (2000). Shape Preserving Approximation Using Least Squares Splines. *Approximation Theory and its Applications*, **16** (4), 80-98.
- [9] CARLSON, R. E. ET FRITSCH, F. N. (1985). Monotone Piecewise Bicubic Interpolation. *SIAM Journal of Numerical Analysis*, **22** (2), 386-400.
- [10] CRAVEN, P. ET WAHBA, G. (1979). Smoothing Noisy Data with Spline Functions : Estimating the Correct Degree of Smoothing by the Method of Generalized Cross-Validation. *Numerische Mathematik*, **31** (4), 377-403.

- [11] CURRY, H. B. ET SCHOENBERG, I. J. (1966). On Polya Frequency Functions, IV : The Fundamental Spline Functions and their Limits. *Journal d'analyse mathématique*, **17** (1), 71-107.
- [12] DE BOOR, C. (1978). *A Practical Guide to Splines*. New York : Springer-Verlag.
- [13] DIERCKX, P. (1995). *Curve and Surface Fitting with Splines*. Oxford : Clarendon Press.
- [14] DONOHO, D. L. ET JOHNSTONE, I. M. (1994). Ideal Spatial Adaptation by Wavelet Shrinkage. *Biometrika*, **81** (3), 425-455.
- [15] DONOHO, D. L. ET JOHNSTONE, I. M. (1995). Adapting to Unknown Smoothness Via Wavelet Shrinkage. *Journal of the American Statistical Association*, **90**, 1200-1224.
- [16] DONOHO, D. L. ET JOHNSTONE, I. M. (1998). Minimax Estimation Via Wavelet Shrinkage. *The Annals of Statistics*, **26** (3), 879-921.
- [17] EUBANK, R.L. (1988). *Spline Smoothing and Nonparametric Regression*. New York : Marcel Dekker, Inc.
- [18] FRIEDMAN, J. H. ET SILVERMAN, B. W. (1989). Flexible Parsimonious Smoothing and Additive Modeling. *Technometrics*, **31** (1), 3-39.
- [19] FRITSCH, F. N. ET CARLSON, R. E. (1980). Monotone Piecewise Cubic Interpolation. *SIAM Journal of Numerical Analysis*, **17** (2), 238-246.
- [20] HANNAN, E. J. ET QUINN, B. G. (1979). The Determination of the Order of an Autoregression. *Journal of the Royal Statistical Society, Series B*, **41** (2), 190-195.
- [21] HE, X. ET SHI, P. (1996). Bivariate Tensor-Product B-Splines in a Partly Linear Model. *Journal of Multivariate Analysis*, **58**, 162-181.
- [22] HE, X. ET SHI, P. (1998). Monotone B-Spline Smoothing. *Journal of the American Statistical Association*, **93** (442), 643-650.
- [23] HORRACE, W. C. (2005). Some results on the multivariate truncated normal distribution. *Journal of Multivariate Analysis*, **94** (1), 209-221.
- [24] KIMELDORF, G. S. ET WAHBA, G. (1970a). A Correspondence between Bayesian Estimation on Stochastic Processes and Smoothing by Splines. *The Annals of Mathematical Statistics*, **41** (2), 495-502.
- [25] KIMELDORF, G. S. ET WAHBA, G. (1970b). Spline Functions and Stochastic Processes. *Sankhyā : The Indian Journal of Statistics, Series A*, **32**, 173-180.

- [26] KIMELDORF, G. S. ET WAHBA, G. (1971). Some Results on Tchebycheffian Spline Functions. *Journal of Mathematical Analysis and Applications*, **33** (1), 82-95.
- [27] KOOPERBERG, C. ET STONE, C. J. (1992). Logspline Density Estimation for Censored Data. *Journal of Computational and Graphical Statistics*, **1** (4), 301-328.
- [28] LAWSON, C. L. ET HANSON, R. J. (1974). *Solving Least Squares Problems*. Philadelphia : SIAM.
- [29] MAMMEN, E. (1991). Estimating a Smooth Monotone Regression Function. *The Annals of Statistics*, **19** (2), 724-740.
- [30] MAMMEN, E. ET THOMAS-AGNAN, C. (1999). Smoothing Splines and Shape Restrictions. *Scandinavian Journal of Statistics*, **26** (2), 239-252.
- [31] RAMSAY, J. O. (1988). Monotone Regression Splines in Action. *Statistical Science*, **3** (4), 425-461.
- [32] REINSCH, C. H. (1967). Smoothing by Spline Functions. *Numerische Mathematik*, **10** (3), 177-183.
- [33] ROBERT, C. P. (2001). *The Bayesian Choice*, 2<sup>e</sup> édition. New York : Springer-Verlag.
- [34] RODRIGUEZ-YAM, G., DAVIS, R. A. ET SCHARF, L. L. (2004). Efficient Gibbs Sampling of Truncated Multivariate Normal with Application to Constrained Linear Regression. (*Submitted*).
- [35] SCHOENBERG, I. J. (1946a). Contributions to the Problem of Approximation of Equidistant Data by Analytic Functions. Part A : On the Problem of Smoothing or Graduation, a First Class of Analytic Approximation Formulae. *Quarterly of Applied Mathematics*, **4**, 45-99.
- [36] SCHOENBERG, I. J. (1946b). Contributions to the Problem of Approximation of Equidistant Data by Analytic Functions. Part B : On the Problem of Osculatory Interpolation, a Second Class of Analytic Approximation Formulae. *Quarterly of Applied Mathematics*, **4**, 112-141.
- [37] SCHOENBERG, I. J. (1964). Spline Functions and the Problem of Graduation. *Proceedings of the National Academy of Sciences*, **52** (4), 947-950.
- [38] SCHUMAKER, L. L. (1981). *Spline Functions : Basic Theory*. New York : Wiley.

- [39] SCHUMAKER, L. L. (1983). On Shape Preserving Quadratic Spline Interpolation. *SIAM Journal of Numerical Analysis*, **20** (4), 854-864.
- [40] SCHWARZ, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, **6** (2), 461-464.
- [41] SHI, P. ET LI, G. (1995). Global Convergence Rates of B-Spline M-Estimators in Nonparametric Regression. *Statistica Sinica*, **5**, 303-318.
- [42] UBARTAS, C. (2007). Estimation non paramétrique bayésienne de courbes de croissance. Mémoire de maîtrise, Université de Montréal.
- [43] WAHBA, G. (1975). Smoothing Noisy Data by Spline Functions. *Numerische Mathematik*, **24** (5), 383-393.
- [44] WAHBA, G. (1978). Improper Priors, Spline Smoothing and the Problem of Guarding Against Model Errors in Regression. *Journal of the Royal Statistical Society, Series B*, **40** (3), 364-372.
- [45] WAHBA, G. (1990). *Spline Models for Observational Data*. Philadelphia : Society for Industrial and Applied Mathematics.
- [46] WAHBA, G. ET WOLD, S. (1975). A Complete Automatic French Curve : Fitting Spline Functions by Cross Validation. *Communications in Statistics, Theory and Methods*, **4**, 1-17.
- [47] WEGMAN, E. J. ET WRIGHT, I. W. (1983). Splines in Statistics. *Journal of the American Statistical Association*, **78** (382), 351-365.
- [48] WHITTAKER, E. T. (1923). On a New Method of Graduation. *Proceedings of the Edinburgh Mathematical Society*, **41**, 63-75.
- [49] WRIGHT, I. W. ET WEGMAN, E. J. (1980). Isotonic, Convex and Related Splines. *The Annals of Statistics*, **8** (5), 1023-1035.
- [50] ZELLNER, A. (1986). On assessing prior distributions and Bayesian regression analysis with g-prior distributions. In *Bayesian Inference and Decision Techniques : Essays in Honor of Bruno de Finetti*, (eds : P.K. Goel and A. Zellner), pp.233-243. Amsterdam : North-Holland.

# Annexe A

---

## TABLEAUX DES COEFFICIENTS ESTIMÉS

### A.1. TABLE DE LA LOI NORMALE

Coefficients  $\theta_{i,j}$ ,  $i = 1, \dots, 20$ ,  $j = 1, \dots, 3$

$5,00 \times 10^{-1}$	$1,80 \times 10^{-2}$	$1,79 \times 10^{-2}$
$1,99 \times 10^{-2}$	$9,08 \times 10^{-7}$	$-7,93 \times 10^{-5}$
$5,97 \times 10^{-2}$	$-2,73 \times 10^{-4}$	$-5,09 \times 10^{-4}$
$7,65 \times 10^{-2}$	$-1,04 \times 10^{-3}$	$-1,31 \times 10^{-3}$
$7,06 \times 10^{-2}$	$-1,60 \times 10^{-3}$	$-1,80 \times 10^{-3}$
$6,26 \times 10^{-2}$	$-1,99 \times 10^{-3}$	$-2,10 \times 10^{-3}$
$5,33 \times 10^{-2}$	$-2,17 \times 10^{-3}$	$-2,20 \times 10^{-3}$
$4,35 \times 10^{-2}$	$-2,17 \times 10^{-3}$	$-2,12 \times 10^{-3}$
$3,42 \times 10^{-2}$	$-2,02 \times 10^{-3}$	$-1,91 \times 10^{-3}$
$2,58 \times 10^{-2}$	$-1,75 \times 10^{-3}$	$-1,62 \times 10^{-3}$
$1,87 \times 10^{-2}$	$-1,44 \times 10^{-3}$	$-1,29 \times 10^{-3}$
$1,30 \times 10^{-2}$	$-1,12 \times 10^{-3}$	$-9,83 \times 10^{-4}$
$8,70 \times 10^{-3}$	$-8,26 \times 10^{-4}$	$-7,09 \times 10^{-4}$
$5,58 \times 10^{-3}$	$-5,81 \times 10^{-4}$	$-4,88 \times 10^{-4}$
$3,44 \times 10^{-3}$	$-3,89 \times 10^{-4}$	$-3,20 \times 10^{-4}$
$2,04 \times 10^{-3}$	$-2,49 \times 10^{-4}$	$-2,00 \times 10^{-4}$
$1,16 \times 10^{-3}$	$-1,52 \times 10^{-4}$	$-1,20 \times 10^{-4}$
$6,35 \times 10^{-4}$	$-8,87 \times 10^{-5}$	$-6,86 \times 10^{-5}$
$3,33 \times 10^{-4}$	$-4,96 \times 10^{-5}$	$-3,75 \times 10^{-5}$
$8,38 \times 10^{-5}$	$-1,32 \times 10^{-5}$	$-9,81 \times 10^{-6}$

