



HAL
open science

Régression polynomiale par morceaux pour la propagation de fissures

Florine Greciet

► **To cite this version:**

Florine Greciet. Régression polynomiale par morceaux pour la propagation de fissures. Mathématiques [math]. Université de Lorraine, 2020. Français. NNT : 2020LORR0004 . tel-02510850

HAL Id: tel-02510850

<https://hal.univ-lorraine.fr/tel-02510850>

Submitted on 18 Mar 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : ddoc-theses-contact@univ-lorraine.fr

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4

Code de la Propriété Intellectuelle. articles L 335.2- L 335.10

http://www.cfcopies.com/V2/leg/leg_droi.php

<http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm>



UNIVERSITÉ
DE LORRAINE



Régression polynomiale par morceaux pour la propagation de fissures

THÈSE

présentée et soutenue publiquement le 20 décembre 2019

pour l'obtention du

Doctorat de l'Université de Lorraine

(mention mathématiques)

par

Florine Greciet

Composition du jury

<i>Rapporteurs :</i>	Faïcel Chamroukhi	Université de Caen
	Mitra Fouladirad	Université de Technologie de Troyes
<i>Examineurs :</i>	Madalina Deaconu	Inria Nancy - Grand Est
	Nikolaos Limnios	Université de Technologie de Compiègne
	Stéphane Pierret	Safran Aircraft Engine
	Marie Touzet-Cortina	Université de Bordeaux
<i>Encadrants :</i>	Romain Azaïs	Inria Grenoble-Rhône Alpes
	Maëva Biret	Safran Aircraft Engines
	Anne Gégout-Petit	Université de Lorraine

« *À ma maman* »

Remerciements

« Ressentir de la gratitude et ne pas l'exprimer est comme envelopper un présent et ne pas l'offrir. »

William Arthur Ward

Je tiens tout d'abord à remercier mes directeurs de thèse Anne Gégout-Petit et Romain Azaïs pour leur présence durant toute cette thèse. Grâce à vos conseils j'ai beaucoup appris et grandi à vos côtés. Vous avez su me guider, tout en me faisant confiance et en me permettant de laisser libre cours à mon imagination. Je vous remercie sincèrement pour cela.

Un grand merci à ma tutrice Maëva Biret sans qui cette thèse n'aurait pas été la même. Merci pour tes précieux conseils scientifiques comme personnels, pour ton soutien et ta confiance infaillible mais également pour l'amitié qui nous lie aujourd'hui.

Je tiens à remercier chaleureusement Faicel Chamroukhi et Mitra Fouladirad d'avoir accepté de rapporter ma thèse. Merci également aux autres membres du Jury Nikolaos Limnios, Marie Touzet, Madalina Deaconu et Stéphane Pierret, d'avoir accepté de participer à mon jury de thèse.

Je remercie également tous mes collègues du laboratoire de l'Université de Lorraine et de Safran Aircraft Engines avec une pensée particulière pour Clémence Karmann, Clémence Papi-Goutenègre, Lauréline Riandet, Sonia Valencian, Krystel Roux, Thibault Levy et Georgia Deaconu qui ont été très présents pendant ces années de thèse. Merci pour tous ces bons moments, pour les pauses cafés, les repas, les fous rires et tout simplement pour votre amitié.

Enfin, mille merci à mes amis et à ma famille d'avoir cru en moi pendant toutes ces années. J'ai pu compter sur vous dans les bons moments comme dans les plus difficiles. C'est grâce à votre soutien sans faille que j'écris ces derniers mots aujourd'hui. Papa, merci car c'est grâce à toi si j'en suis là aujourd'hui, je sais que tu es fier de moi et j'en suis sincèrement touchée. Justine, je ne te remercierais jamais assez pour tes messages quotidiens. Vincent, merci pour tes précieuses relectures et tes conseils sur la fin de la thèse. Gaëtan merci d'être à mes côtés au quotidien, de m'avoir soutenue et supportée pendant toute cette thèse et enfin merci de m'avoir donné la plus belle merveille du monde, notre fils Louis, dont les sourires émerveillent chacune de mes journées.

Cette thèse est l'un des plus grands projets de ma vie alors Un Grand Merci à Tous!

Table des matières

Remerciements	v
Table des matières	vii
Liste des figures	xi
Liste des tableaux	xv
Résumé	1
Abstract	3
1 Contexte industriel et état de l'art	5
1.1 Contexte industriel	6
1.1.1 Phénomène de propagation de fissures de fatigue	6
1.1.2 Problématique industrielle	8
1.2 Modèles de propagation de fissure et source d'aléa	9
1.2.1 Paramètres intrinsèques et extrinsèques des matériaux	9
1.2.2 Modèles déterministes	11
1.2.3 Présence d'aléa	13
1.2.4 Modèles stochastiques	15
1.3 Processus de Markov déterministes par morceaux	18
1.3.1 Modèle multiplicatif	19
1.3.2 Modèles à un saut	20
1.4 Régression polynomiale par morceaux	20
1.4.1 Modèle de régression polynomiale par morceaux	21
1.4.2 Méthode d'inférence et algorithmes	22
2 Régression polynomiale par morceaux sous contrainte de régularité	23
2.1 Introduction	24
2.2 Modèle de régression polynomiale par morceaux	24
2.2.1 Fonction de lien	24
2.2.2 Contrainte de régularité	25
2.2.3 Log-vraisemblance	25
2.2.4 Approximation par un modèle à variable latente	26
2.3 Maximum de vraisemblance à deux régimes	28
2.3.1 Modèle sans variable latente	29
2.3.2 Modèle à variable latente : algorithme EM	31
2.4 Généralisation à un nombre quelconque de régimes	33
2.4.1 Modèle sans variable latente	33
2.4.2 Modèle à variable latente et algorithme EM	35

2.5	Approche récursive : algorithme de type Bellman	36
2.5.1	Modèle sans variable latente	36
2.5.2	Lissage autour des instants de transition	39
3	Méthodes d'inférence sans a priori sur le nombre de régimes	41
3.1	Introduction	42
3.2	Approche de type segmentation binaire	42
3.2.1	Modèle sans variable latente	42
3.2.2	Modèle à variable latente et algorithme EM	45
3.3	Approche de segmentation binaire pénalisée	45
3.3.1	Vraisemblance pénalisée	45
3.3.2	Vraisemblance pénalisée et algorithme EM	48
3.4	Tentative d'étude théorique de la pénalité	48
3.4.1	Expressions de \hat{a}_{n+m} , \hat{b}_{n+m} et $\hat{\sigma}_{n+m}^2$	49
3.4.2	Espérances et variances de \hat{a}_{n+m} , \hat{b}_{n+m} et $\hat{\sigma}_{n+m}^2$	50
3.4.3	Comportement asymptotique de $\mathbb{E}[\hat{\sigma}_{n+m}^2]$	51
4	Package R et simulations	55
4.1	Package R : HSPOR	56
4.1.1	Fonction H2SPOR	56
4.1.2	Fonction HKSPOR	57
4.1.3	Fonction HKSPOR_DynProg	57
4.1.4	Fonction H2SPOR_DynProg	59
4.2	Simulations	59
4.2.1	Analyse de l'algorithme H2SPOR pour des modèles à deux régimes	59
4.2.2	Analyse des algorithmes HKSPOR, HKSPOR_DynProg, H2SPOR_DynProg et Bin_Pen	66
4.2.3	Exemple de cas difficiles pour H2SPOR_DynProg et Bin_Pen	76
4.2.4	Estimation a posteriori du nombre de régimes pour HKSPOR_DynProg	78
4.3	Discussion	78
5	Application sur données réelles	81
5.1	Données de fluage et de traction	82
5.2	Données de propagation de fissures	86
5.2.1	Données Safran Aircraft Engines	86
5.2.2	Données de Virkler	87
5.2.3	Données de recherche du seuil d'amorçage	89
6	Conclusion et perspectives	91
6.1	Conclusion	92
6.2	Perspectives	93
	Bibliographie	98
A	Complément d'information	I
A.1	Essai de propagation de fissure de fatigue	II
A.2	Essai de traction sur un matériau	III
A.3	Deux méthodes d'optimisation du chapitre 2	IV
A.3.1	Recherche trichotomique	IV
A.3.2	Programmation dynamique	V
A.4	Estimation d'un régime sous contraintes	VI
A.4.1	Modèle de régression polynomiale par morceaux sur un régime contraint	VI
A.4.2	Maximum de vraisemblance à un régime contraint	VII

A.5	Résolution des problèmes d'optimisation sous contrainte de régularité sur le modèle sans variable latente.	IX
A.5.1	Optimisation à deux régimes	IX
A.5.2	Optimisation à un nombre de régimes K quelconque	XI
A.5.3	Optimisation à un régime	XII
A.6	Résolution des problèmes d'optimisation sous contrainte de régularité pour le modèle à variable latente.	XIV
A.7	Complément d'étude du comportement asymptotique de $\mathbb{E}[\hat{\sigma}_{n+m}^2]$	XV
A.8	Documentation package HSPOR	XXIII

Liste des figures

1.1	Exemples de sollicitations subies par l'avion.	6
1.2	Tolérance au dommage.	7
1.3	Illustration schématique des différents régimes de propagation des fissures de fatigue. Les lignes verticales en pointillés indiquent les transitions entre les régimes.	7
1.4	Représentation schématique de plusieurs essais de vitesse de propagation de fissure.	8
1.5	Représentation bruitée d'un essai de vitesse de propagation de fissure.	9
1.6	Évolution d'une éprouvette soumise à un essai de traction.	10
1.7	68 courbes expérimentales de longueur de fissure en fonction du nombre de cycles. Ces courbes sont issues des essais de propagation réalisés dans [59].	13
1.8	68 courbes de vitesse de propagation de fissure en fonction de ΔK_I . Ces courbes sont calculées à partir des essais de propagation réalisés dans [59].	14
2.1	Exemple de loi de Z sachant X autour de τ_1	27
2.2	Schéma de la résolution d'un problème à K régimes.	38
3.1	Schéma du principe de segmentation de la méthode.	43
3.2	Schéma de l'étape 1 de la méthode quand réponse «non» à la règle d'arrêt. L'intervalle en gris n'est pas retenu par la méthode.	43
3.3	Schéma de l'étape 2 de la méthode quand réponses «oui» (à gauche) et «non» (à droite) aux règles d'arrêts. Les intervalles en gris ne sont pas retenus par la méthode.	44
3.4	Données simulées selon un modèle à cinq régimes où les droites verticales représentent les instants de transition (à gauche). Représentation simultanée de $\mathcal{L}(T, \mathbf{a}_2^*(T), \mathbf{S}_2^{2*}(T))$ et de sa version pénalisée $\mathcal{L}_p(T, \mathbf{a}_2^*(T), \mathbf{S}_2^{2*}(T))$ en fonction de T avec les droites verticales noir et rouge correspondent aux instants pour lesquels la log-vraisemblance et la log-vraisemblance pénalisée sont respectivement maximales (à droite).	46
3.5	Schéma du principe de segmentation de la méthode.	46
3.6	Schéma de l'étape 1 de la méthode quand réponse «non» à la règle d'arrêt. L'intervalle en gris n'est pas retenu par la méthode.	47
3.7	Schéma de l'étape 2 de la méthode quand réponse «non» à la règle d'arrêt. L'intervalle en gris n'est pas retenu par la méthode.	47
3.8	Schéma de l'étape 3 de la méthode quand réponse «oui» à la règle d'arrêt. Les intervalles en gris ne sont pas retenus par la méthode.	47
4.1	Sortie obtenue en 2.7 secondes par la fonction H2SPOR dans la console de Rstudio (à gauche) et dans l'onglet plot de Rstudio (à droite), lorsqu'elle estime un modèle de degré 2 à deux régimes.	56
4.2	Sortie obtenue en 47.6 minutes par la fonction HKSPOR dans la console de Rstudio (à gauche) et dans l'onglet plot de Rstudio (à droite), lorsqu'elle estime un modèle de degré 2 à trois régimes.	57
4.3	Sortie obtenue en 3.6 minutes par la fonction HKSPOR_DynProg dans la console de Rstudio (à gauche) et dans l'onglet plot de Rstudio (à droite), lorsqu'elle estime un modèle de degré 2 à trois régimes.	58

4.4	Sortie obtenue en 3.6 minutes par la fonction <code>HKSPOR_DynProg</code> dans la console de Rstudio (en haut à gauche) et dans l'onglet plot de Rstudio (les trois autres graphiques), lorsqu'elle estime un modèle de degré 2 à trois régimes, quand l'option <code>verbose=True</code>	58
4.5	Sortie obtenue en 2.1 minutes par la fonction <code>H2SPOR_DynProg</code> dans la console de Rstudio (à gauche) et dans l'onglet plot de Rstudio (à droite), lorsqu'elle estime un modèle de degré 2.	59
4.6	Données simulées selon le modèle A pour $n = 100$, $n = 200$ et $n = 500$ observations. Le vrai modèle est en trait plein et l'instant de transition τ_1 est représenté par la droite verticale.	60
4.7	Boxplots des estimateurs des coefficients des polynômes \mathbf{a}_2 , des variances \mathbf{S}_2^2 et de l'instant de transition T_1 obtenus par <code>H2SPOR(EM=False)</code> , en fonction du nombre d'observations n pour $B = 100$ répétitions indépendantes du modèle A . La droite horizontale représente la vraie valeur du paramètre.	61
4.8	Modèle de régression polynomiale à deux régimes contenant 2 points levier (+). Le vrai modèle est en trait plein et l'instant de transition τ_1 est représenté par la droite verticale.	62
4.9	Boxplots des valeurs estimées a_1 , T_1 et S_1^2 par <code>H2SPOR(EM=False)</code> et <code>H2SPOR(EM=True)</code> pour $B = 100$ répétitions indépendantes du modèle B . La droite horizontale représente la vraie valeur du paramètre.	63
4.10	Données simulées selon le modèle C (σ_2^2) lorsque $\sigma_2^2 = 0.09$ (en haut à gauche), $\sigma_2^2 = 0.25$ (en haut à droite), $\sigma_2^2 = 0.64$ (en bas à gauche) et $\sigma_2^2 = 1$ (en bas à droite). Le vrai modèle est en trait plein (droite horizontale) et l'instant de transition τ_1 est représenté par la droite verticale.	64
4.11	Boxplots des estimations des coefficients constants des polynômes a_1^0 et a_2^0 , de l'instant de transition T_1 et de la variance du bruit du premier régime S_1^2 , ainsi que les erreurs d'estimations relatives absolues faites sur la variance du second régime par <code>H2SPOR(EM=True)</code> , pour $B = 100$ répétitions indépendantes du modèle C (σ_2^2) lorsque $\sigma_2^2 = 0.09$, $\sigma_2^2 = 0.25$, $\sigma_2^2 = 0.64$ et $\sigma_2^2 = 1$. La droite horizontale représente soit la vraie valeur du paramètre, soit une erreur nulle d'estimation.	65
4.12	Réalisations du modèle D (points) et ajustement des données par <code>H2SPOR(EM=True)</code> (trait en pointillés) sous contrainte de continuité (à gauche) ou sous contrainte de dérivabilité (à droite). Le vrai modèle est en trait plein et l'instant de transition τ_1 est représenté par la droite verticale en trait plein alors que son estimation T_1 est en pointillé.	65
4.13	Données simulées selon le modèle E . Le vrai modèle est en trait plein et les instants de transition τ_1 et τ_2 sont représentés par les droites verticales.	67
4.14	Boxplots des erreurs relatives absolues des coefficients a_1 , a_2 et a_3 des polynômes de chaque régime estimés par les méthodes <code>HKSPOR</code> et <code>HKSPOR_DynProg</code> , ainsi que les estimations des variances \mathbf{S}_3^2 et des instants de transition \mathbf{T}_2 obtenues par ces deux mêmes méthodes pour $B = 100$ répétitions indépendantes du modèle E . La droite horizontale représente soit une erreur nulle, soit la vraie valeur du paramètre.	68
4.15	Boxplots des temps de calculs nécessaires à <code>HKSPOR</code> (en heures) et à <code>HKSPOR_DynProg</code> (en minutes) pour estimer les paramètres α_3 , τ_2 et σ_3^2 , pour $B = 100$ répétitions indépendantes du modèle E	69
4.16	Données simulées selon le modèle F . Le vrai modèle est en trait plein et l'instant de transition τ_1 est représenté par la droite verticale.	69
4.17	Boxplots des estimations de l'instant de transition τ_1 obtenues par <code>HKSPOR_DynProg</code> , <code>H2SPOR_DynProg</code> et <code>Bin_Pen</code> dans le cas où les méthodes trouvent le bon nombre de régimes du modèle F (soit $B = 100$) et distribution des RMSE pour $B = 100$ répétitions indépendantes du modèle F . La droite horizontale représente la vraie valeur de l'instant de transition τ_1	69

4.18	Boxplots des temps de calculs nécessaires à HKSPOR_DynProg, H2SPOR_DynProg et Bin_Pen pour estimer les paramètres du modèle, pour B = 100 répétitions indépendantes du modèle F	70
4.19	Distributions des RMSE calculés pour les trois méthodes pour B = 100 répétitions indépendantes du modèle E et boxplots des estimations des instants de transition τ_1 et τ_2 obtenues par HKSPOR_DynProg pour B = 100 répétitions indépendantes du modèle E et par H2SPOR_DynProg et Bin_Pen dans le cas où elles trouvent le bon nombre de régimes, c'est-à-dire pour B = 76 et B = 100 simulations du modèle E respectivement. La droite horizontale représente la vraie valeur du paramètre.	71
4.20	Boxplots des temps de calculs nécessaires à HKSPOR_DynProg, H2SPOR_DynProg et Bin_Pen pour estimer les paramètres du modèle, pour B = 100 répétitions indépendantes du modèle E	71
4.21	Données simulées selon le modèle G . Le vrai modèle est en trait plein et les instants de transition τ_1 , τ_2 et τ_3 sont représentés par les droites verticales.	72
4.22	Distributions des RMSE calculés pour les trois méthodes pour B = 100 répétitions indépendantes du modèle G et boxplots des estimations des instants de transition T_1 , T_2 et T_3 obtenues par HKSPOR_DynProg pour B = 100 répétitions indépendantes du modèle G et par H2SPOR_DynProg et Bin_Pen lorsqu'elles estiment le bon nombre de régimes, c'est-à-dire pour B = 76 et B = 1 réalisations du modèle G respectivement. La droite horizontale représente la vraie valeur du paramètre.	73
4.23	Boxplots des temps de calculs nécessaires à HKSPOR_DynProg, H2SPOR_DynProg et Bin_Pen pour estimer les paramètres du modèle, pour B = 100 répétitions indépendantes du modèle G	73
4.24	Données simulées selon le modèle H . Le vrai modèle est en trait plein et les instants de transition τ_1 à τ_5 sont représentés par les droites verticales.	74
4.25	Boxplots des RMSE calculés pour les trois méthodes pour B = 100 répétitions indépendantes du modèle H et distributions des estimations des instants de transition T_1 à T_5 obtenues par la méthode HKSPOR_DynProg, pour B = 100 répétitions indépendantes du modèle H et par les méthodes H2SPOR_DynProg et Bin_Pen lorsqu'elles estiment le bon nombre de régimes du modèle H , c'est-à-dire pour B = 89 et B = 0 simulations respectivement. La droite horizontale représente la vraie valeur du paramètre.	75
4.26	Boxplot des estimations de l'instant de transition obtenues par Bin_Pen lorsqu'elle estime un modèle à deux régimes au lieu de six (B = 100). La droite horizontale représente la valeur du vrai saut le plus proche soit τ_5	75
4.27	Boxplots des temps de calculs nécessaires à HKSPOR_DynProg, H2SPOR_DynProg (en minutes, à gauche) et Bin_Pen (en secondes, à droite) pour estimer les paramètres du modèle, pour B = 100 répétitions indépendantes du modèle H	76
4.28	Exemples de données simulées (points) pour lesquels H2SPOR_DynProg n'estime pas le bon nombre de régimes. Le modèle estimé par H2SPOR_DynProg est en trait plein. Les droites verticales noires correspondent aux vrais instants de transition, les rouges aux instants estimés par H2SPOR_DynProg.	77
4.29	Exemples de données simulées (points) pour lesquels Bin_Pen n'estime pas le bon nombre de régimes. Le modèle estimé par Bin_Pen est en trait plein. Les droites verticales noires correspondent aux vrais instants de transition, les rouges aux instants estimés par Bin_Pen.	77
5.1	Illustration schématique des différents stades du fluage. La courbe A correspond à un essai sous chargement constant; la courbe B correspond à un essai sous contrainte constante.	82

5.2	Ajustements et segmentations de quatre essais de fluage (un essai par ligne) par les méthodes HKSPOR_DynProg (à gauche), H2SPOR_DynProg (au milieu) et Bin_Pen (à droite) lorsqu'elles estiment un modèle de degré 2. Les courbes d'ajustements et les instants de transition estimés par les méthodes sont représentés en rouge.	83
5.3	Distributions des RMSE calculés à partir des valeurs prédites par les modèles estimés par HKSPOR_DynProg, H2SPOR_DynProg et Bin_Pen, pour B = 4 essais de fluage. . .	84
5.4	Ajustements et segmentations de trois essais de traction (un essai par ligne) par HKSPOR_DynProg (à gauche), H2SPOR_DynProg (au milieu) et Bin_Pen (à droite) lorsqu'elles estiment un modèle de degré 2. Les courbes d'ajustements et les instants de transition estimés par les méthodes sont représentés en rouge.	85
5.5	Distributions des RMSE calculés à partir des valeurs prédites par les modèles estimés par HKSPOR_DynProg, H2SPOR_DynProg et Bin_Pen, pour B = 3 essais de traction. .	85
5.6	Ajustements et segmentations de trois essais de propagation (un essai par ligne) en vitesse en fonction du facteur d'intensité des contraintes par les méthodes HKSPOR_DynProg (à gauche), H2SPOR_DynProg (au milieu) et Bin_Pen (à droite), lorsqu'elles estiment un modèle de degré 2. Ces essais sont représentés en échelle bi-logarithmique. Les courbes d'ajustements et les instants de transition estimés par les méthodes sont représentés en rouge.	86
5.7	Distributions des RMSE calculés à partir des valeurs prédites par les modèles estimés par les méthodes HKSPOR_DynProg, H2SPOR_DynProg et Bin_Pen, pour B = 19 essais de propagation de fissures de fatigue.	87
5.8	Ajustements et segmentations de trois essais de propagation (un essai par ligne) en vitesse en fonction du facteur d'intensité des contraintes par les méthodes HKSPOR_DynProg (à gauche), H2SPOR_DynProg (au milieu) et Bin_Pen (à droite), lorsqu'elles estiment un modèle de degré 1. Ces essais sont représentés en échelle bi-logarithmique. Les courbes d'ajustements et les instants de transition estimés par les méthodes sont représentés en rouge.	88
5.9	Boxplots des RMSE calculés à partir des valeurs prédites par les modèles estimés par les méthodes HKSPOR_DynProg, H2SPOR_DynProg et Bin_Pen, pour B = 68 essais de propagation de fissures de fatigue.	88
5.10	Nuage de points des instants de transition estimés par HKSPOR_DynProg (en abscisse) ou par Azaïs et al. dans [3] (en ordonnée).	89
5.11	Ajustements et segmentations de deux essais de seuil de non propagation (un essai par ligne) en vitesse en fonction du facteur d'intensité des contraintes par les méthodes HKSPOR_DynProg (à gauche), H2SPOR_DynProg (au milieu) et Bin_Pen (à droite), lorsqu'elles estiment un modèle de degré 2. Ces essais sont représentés en échelle bi-logarithmique. Les courbes d'ajustements et les instants de transition estimés par les méthodes sont représentés en rouge.	90
5.12	Distributions des RMSE calculés à partir des valeurs prédites par les modèles estimés par les méthodes HKSPOR_DynProg, H2SPOR_DynProg et Bin_Pen pour B = 2 essais de seuil de non propagation.	90
A.1	Plan de référence d'une éprouvette de propagation de fissures.	II
A.2	Illustration des exigences à satisfaire.	III
A.3	Éprouvettes normalisées.	III
A.4	Diagramme conventionnel d'un essai de traction.	IV
A.5	Évolution schématique de l'éprouvette durant l'essai de traction.	IV
A.6	Pyramide de nombres avec son chemin optimal en rouge.	VI
A.7	Séquence des calculs effectués par la relation de récursivité.	VI

Liste des tableaux

1.1	Notations.	11
4.1	Tableau des caractéristiques des différentes méthodes.	59
4.2	Distribution du nombre de régimes estimés par optimisation du critère BIC pour $B = 100$ réalisations indépendantes du modèle F à deux régimes.	78
5.1	Statistiques concernant les longueurs de fissure au instants de transition, les instants de transition en termes de nombre de cycles et le facteur d'intensité de contrainte correspondant.	87

Résumé

Un moteur d'avion est constitué de plusieurs familles de matériaux qui subissent de multiples mécanismes de dégradation dès leur fabrication mais également pendant leur cycle de vol (décollage, atterrissage, pressurisation, manoeuvre du pilote,...), où encore lors de son repos au sol. Un de ces mécanismes est lié au phénomène de fatigue, qui représente la dégradation subie par une pièce du fait de sollicitations cycliques répétées. Cette dégradation se traduit par l'amorçage d'une fissure et sa propagation jusqu'à rupture de la pièce. La prédiction de la durée de vie en propagation des pièces est donc un point très sensible puisqu'elle impacte à la fois le dimensionnement (étape préalable à la conception des pièces) et les procédures de maintenance (réparation ou changement de la pièce). Les calculs de durées de vie en propagation sont en partie réalisés à partir des lois d'évolutions phénoménologiques décrivant la vitesse d'avancée de la fissure dans un matériau en fonction de la contrainte appliquée. Dans l'objectif d'étudier ces données, qui sont susceptibles d'être modélisées de façon continue et qui laissent observer plusieurs régimes de propagation, nous proposons un modèle de régression polynomiale à plusieurs régimes, soumis à des hypothèses de régularité (continuité et/ou dérivabilité). Suite à cela, nous avons développé des méthodes d'inférence permettant d'estimer le nombre de régimes, les instants de transition et les paramètres de chaque régime. Ces résultats ne seront exploitables par le bureau d'études que s'ils sont obtenus en des temps de calculs raisonnables c'est-à-dire de l'ordre de quelques minutes. Chaque nouvelle méthode a donc été conçue dans l'objectif de réduire les temps de calculs nécessaires à l'estimation des paramètres du modèle. De plus, comme le nombre de régimes présents dans les données n'est pas connu a priori, les deux dernières méthodes que nous proposons n'utilisent aucun a priori sur ce nombre pour estimer les paramètres du modèle. Le travail présenté dans ce mémoire fait l'objet d'une collaboration entre l'équipe de Probabilités et Statistique de l'Institut Élie Cartan de l'université de Lorraine, l'équipe BIGS du centre Inria Nancy Grand Est et la société Safran Aircraft Engines.

Abstract

An aircraft engine is made up of several families of materials that undergo multiple degradation mechanisms from their manufacture but also during their flight cycle (take-off, landing, pressurization, pilot manoeuvring,...) or during its rest on the ground. One of these mechanisms is related to the phenomenon of fatigue which represents the degradation of a piece due to repeated cyclic stresses. This degradation results in the initiation of a crack and its propagation until the piece ruptures. The prediction of the propagation lifetime of parts is therefore a very sensitive point since it impacts both the dimensioning (step prior to the design of the pieces) and the maintenance procedures (repair or change of the piece). Propagation lifetime calculations are partly based on the laws of phenomenological evolution describing the rate of progress of the crack in a material as a function of the stress applied. In order to study these data, which are likely to be modelled continuously and which allow several propagation regimes to be observed, we propose a polynomial regression model with several regimes, subject to regularity assumptions (continuity and/or differentiability). Following this, we developed inference methods to estimate the number of regimes, transition times and parameters of each regime. These results will only be usable by the engineering office if they are obtained within reasonable calculation times, i.e. in the order of a few minutes. Each new method has therefore been designed to reduce the computation time required to estimate the model parameters. Moreover, since the number of regimes present in the data is not known a priori, the last two methods we propose do not use any a priori on this number to estimate the model parameters. The work presented in this thesis is the subject of a collaboration between the Probability and Statistics team of the Institut Elie Cartan of the University of Lorraine, the BIGS team of Inria Nancy Grand Est and Safran Aircraft Engines.

Chapitre 1

Contexte industriel et état de l'art

Sommaire

1.1 Contexte industriel	6
1.1.1 Phénomène de propagation de fissures de fatigue	6
1.1.2 Problématique industrielle	8
1.2 Modèles de propagation de fissure et source d'aléa	9
1.2.1 Paramètres intrinsèques et extrinsèques des matériaux	9
1.2.2 Modèles déterministes	11
1.2.3 Présence d'aléa	13
1.2.4 Modèles stochastiques	15
1.3 Processus de Markov déterministes par morceaux	18
1.3.1 Modèle multiplicatif	19
1.3.2 Modèles à un saut	20
1.4 Régression polynomiale par morceaux	20
1.4.1 Modèle de régression polynomiale par morceaux	21
1.4.2 Méthode d'inférence et algorithmes	22

1.1 Contexte industriel

Safran Aircraft Engines est une société française du groupe SAFRAN, spécialisée dans l'étude et la fabrication de moteurs pour l'industrie aéronautique et spatiale. Elle conçoit, développe, produit et commercialise, seule ou en coopération, des moteurs pour avions civils et militaires et pour satellites. En parallèle, la société propose aux compagnies aériennes, aux opérateurs d'avions et aux forces armées une gamme complète de services pour leurs moteurs. S'appuyant sur une expertise technologique unique et des capacités industrielles au meilleur niveau mondial, les nouvelles générations de moteurs proposés par Safran Aircraft Engines contribuent à rendre le transport aérien plus sûr et plus respectueux de l'environnement : ils sont plus silencieux, plus propres et consomment moins de carburant. Grâce aux nouvelles technologies de l'aéronautique, ce sont des milliards de voyageurs qui empruntent chaque année les routes aériennes du monde entier. Les attentes sont donc maximales en terme de fiabilité, disponibilité et performance. De ce fait, en aéronautique, la commercialisation des pièces moteur nécessite la caractérisation des propriétés de durée de vie de chaque pièce. Ces propriétés sont propres au matériau utilisé. La propriété qui nous intéresse dans ce mémoire est la propagation de fissures de fatigue.

1.1.1 Phénomène de propagation de fissures de fatigue

La fatigue des matériaux est un mécanisme de défaillance complexe qui, sous l'action de contraintes ou déformations répétées dans le temps, modifie les propriétés locales d'un matériau, ce qui peut entraîner la formation de défauts (appelés fissures) qui se propageront plus ou moins rapidement et entraîneront éventuellement la rupture de la pièce. En aéronautique, les risques de défaillance de certaines pièces moteur sont accentués par l'environnement extrême dans lequel ces pièces doivent fonctionner (cf figure 1.1). En effet, le cumul des sollicitations subies sur plusieurs dizaines de milliers d'heures de vols conduira à la fatigue des pièces qui composent le moteur.

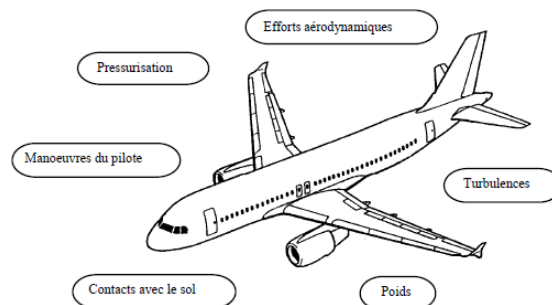


FIGURE 1.1 – Exemples de sollicitations subies par l'avion.

Le développement de techniques expérimentales sur l'étude de la rupture est lié à la première révolution industrielle (1780 – 1850) et à l'apparition des premiers grands accidents dans le domaine des transports de masse. La catastrophe ferroviaire de Meudon par exemple, est la première en France et l'une des premières dans le monde. Le 8 mai 1842, un train en provenance de Versailles et à destination de Paris a déraillé dans la tranchée de Bellevue à Meudon. C'est l'ingénieur écossais William Rankine qui, en étudiant les faciès de rupture d'essieux brisés lors de l'accident, a montré que le mode de rupture n'était pas lié à une fragilité des matériaux employés, mais à un mode de défaillance aujourd'hui connu comme «mode de rupture par fatigue». Les plus célèbres accidents dus au phénomène de fatigue ont eu lieu au XX^{ième} siècle. Nous pouvons par exemple citer : le crash du premier avion commercial propulsé par le turbo-réacteur Commet en 1954, du fait de la pressurisation du fuselage qui a provoqué l'apparition de fissures sur les coins des hublots rectangulaires; l'effondrement de la plateforme Alexandre Killian en 1980, suite à une fissure de fatigue présente dans l'une des six tubulures qui était reliée à une jambe de la plateforme. Depuis, les efforts de recherche sont largement portés sur l'étude des phénomènes d'amorçage et de

propagation des fissures de fatigue et les ingénieurs intègrent désormais ce phénomène dans la conception et les calculs de durées de vie des structures.

Chez Safran Aircraft Engines, les pièces les plus à risque d'un moteur d'avion sont dimensionnées (étape préalable à la conception des pièces) de façon à résister, pendant une période définie, aux sollicitations qui leur sont appliquées. La réglementation s'appuie sur le concept de la tolérance au dommage, présenté en figure 1.2, qui consiste à admettre l'apparition d'une ou plusieurs fissures durant la période d'exploitation, à condition de les détecter à l'aide de moyens non-destructifs et de les réparer. Les calculs de durée de vie en propagation des pièces sont en partie réalisés à

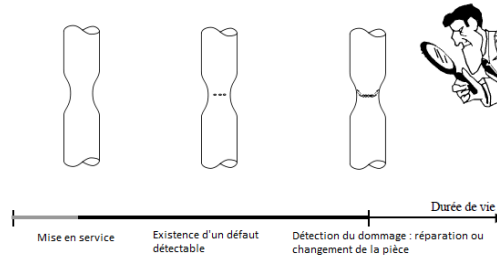


FIGURE 1.2 – Tolérance au dommage.

partir des lois d'évolutions phénoménologiques, qui décrivent la vitesse d'avancée de la fissure, notée da/dN en ($m/cycles$), en fonction du facteur d'intensité des contraintes ΔK , en $MPa\sqrt{m}$. Le facteur d'intensité des contraintes est une fonction de la longueur de fissure, de la géométrie de l'éprouvette testée pendant l'essai et de la contrainte appliquée. Ces lois d'évolutions sont estimées à partir des données expérimentales obtenues lors d'essais de propagation de fissure (détaillé en annexe A.1) durant lesquels la vitesse d'avancée de la fissure est mesurée pour différentes valeurs du facteur d'intensité des contraintes. L'allure empirique de la loi de la vitesse de propagation en fonction du facteur d'intensité des contraintes est représentée en échelle bilogarithmique dans la figure 1.3. Cette courbe implique trois régimes.

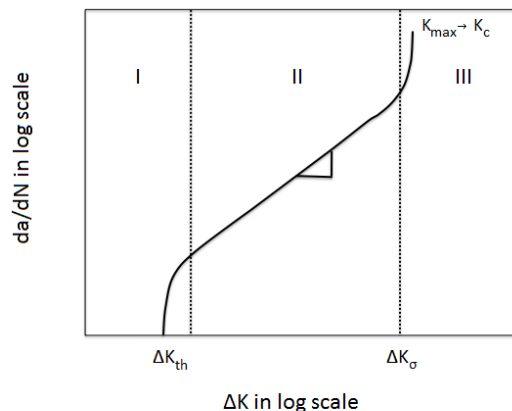


FIGURE 1.3 – Illustration schématique des différents régimes de propagation des fissures de fatigue. Les lignes verticales en pointillés indiquent les transitions entre les régimes.

- Régime 1 (amorçage et non propagation de la fissure) : lorsque la pièce sollicitée ne comporte aucune anomalie, des microfissures de surface se forment. Si le chargement est en dessous d'un certain seuil, la fissure ne se propage pas au niveau macroscopique. On réalise dans ce cas un essai de détermination du seuil de non propagation (d'amorçage) de la fissure.

- Régime 2 (évolution linéaire de la fissure) : les microfissures ont coalescé pour former une fissure de taille macroscopique. Au cours de cette étape, la fissure peut croître plus ou moins rapidement selon le matériau et l'intensité du chargement critique.
- Régime 3 (rupture finale) : il correspond à une croissance rapide et instable des fissures entraînant une rupture lorsque le facteur d'intensité de contrainte tend vers la valeur critique K_c .

1.1.2 Problématique industrielle

La particularité du schéma de la figure 1.3 est qu'il se compose d'une partie déterministe et d'une partie probabiliste. La partie déterministe est caractérisée par des modèles de forme connue pour chaque régime. La partie probabiliste est, comme le montre la figure 1.4, caractérisée par la variabilité entre différents essais des paramètres de chaque régime et des instants de transitions, notés ici T_1 et T_2 . En effet, de l'aléa peut être observé sur des données d'essais réalisés dans des conditions identiques. En conditions réelles, cet aléa peut être exacerbé par les phénomènes environnementaux tels que la vitesse ou la température. Dans le but de modéliser chaque courbe

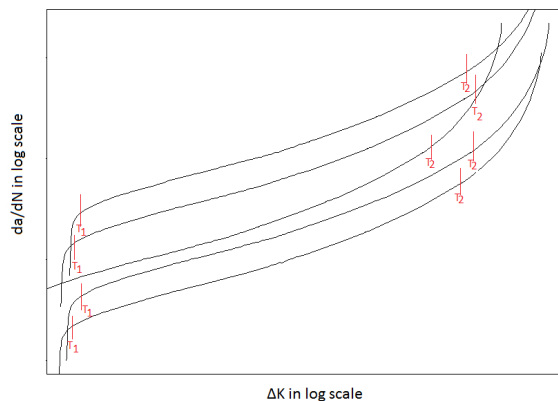


FIGURE 1.4 – Représentation schématique de plusieurs essais de vitesse de propagation de fissure.

de vitesse de propagation de fissure, nous proposons un modèle déterministe par morceaux caché. Ces modèles sont de bons candidats puisqu'ils sont stochastiques et qu'ils combinent à la fois trajectoires déterministes et transitions aléatoires. Nous parlons de «modèle caché» dans le sens où l'étude empirique des courbes de vitesse de propagation de fissure montre l'existence de trois régimes. Or, en pratique, les données d'essais ne permettent pas toujours d'observer cette allure de courbe. En effet, selon le matériau testé lors de l'essai, il est possible d'observer plus de trois régimes de propagation. De plus, la variabilité présente au sein d'un même essai (cf. figure 1.5), complexifie grandement la détection des instants de transition entre les régimes et donc a fortiori la modélisation des données présentes dans les différents régimes. L'objectif de la thèse consiste à caractériser la loi d'un essai de propagation de fissure, c'est-à-dire détecter le nombre de régimes, les instants de transition entre les régimes et la forme paramétrique des lois dans les différents régimes. Au vu de l'allure des courbes de vitesse de propagation de fissure et en s'inspirant des travaux de Chamroukhi et al. dans [14], nous proposons un modèle de régression polynomiale par morceaux. Ces modèles appartiennent à la famille des modèles déterministes par morceaux et semblent bien adaptés puisqu'ils permettent de prendre en compte l'existence de plusieurs régimes de propagation.

Le plan de thèse se déroulera de la façon suivante :

Dans les prochaines sections de ce chapitre, nous proposons un état de l'art (non exhaustif) des modèles déterministes, stochastiques et déterministes par morceaux utilisés dans la littérature pour modéliser le phénomène de propagation de fissure. Nous présentons également la régression polynomiale par morceaux que nous proposons d'utiliser dans ce mémoire. Dans le chapitre 2

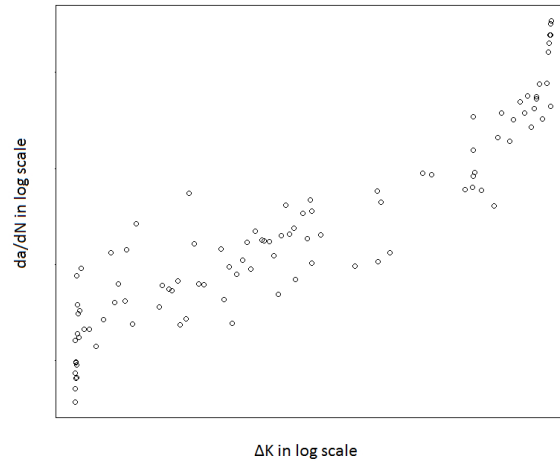


FIGURE 1.5 – Représentation bruitée d'un essai de vitesse de propagation de fissure.

nous présentons le modèle de régression polynomiale par morceaux proposé pour modéliser des phénomènes physiques comme la propagation de fissures, ainsi que des méthodes d'inférence permettant l'estimation de ce modèle. Les méthodes du chapitre 2 supposent connu le nombre de régimes, mais dans le chapitre 3 nous proposons deux nouvelles méthodes qui n'utilisent pas d'a priori sur ce nombre et proposent une estimation globale du nombre de régimes et de leurs paramètres. Le chapitre 4 présente le package R dans lequel ces méthodes ont été implémentées ainsi que l'ensemble des simulations réalisées pour montrer le bon comportement de l'ensemble des méthodes. Enfin, le chapitre 5 présente l'application de ces méthodes sur des données réelles de Safran Aircraft Engines ou des données de la littérature obtenues par Virkler et al. dans [59]. Nous terminons ce mémoire par une conclusion et des perspectives sur les travaux réalisés durant la thèse, dans le chapitre 6.

1.2 Modèles de propagation de fissure et source d'aléa

Dans cette section, nous présentons différents modèles qui ont été proposés dans la littérature pour décrire le phénomène de propagation de fissure, dans des matériaux métalliques non ferreux (alliages d'aluminium, de cuivre, de nickel, ...). Ces modèles peuvent être déterministes (modèles mécaniques qui utilisent les propriétés des matériaux) ou stochastiques (modèles statistiques permettant de caractériser le comportement aléatoire de la propagation de fissure). Dans le but de faciliter la compréhension des modèles déterministes, la section suivante présente une grande partie des paramètres intrinsèques et extrinsèques des matériaux.

1.2.1 Paramètres intrinsèques et extrinsèques des matériaux

Les modèles de propagation de fissure sont conçus à partir de plusieurs paramètres et conditions imposés en service ou en laboratoire qui agissent sur la propagation de fissure. Nous citons par exemple :

Les paramètres intrinsèques au matériau :

- La limite d'élasticité R_p : c'est la charge unitaire exprimée en MPa correspondant à la fin de la première période d'un essai de traction (ce type d'essai est présenté en annexe A.2).

$$R_p = \frac{F_e}{S_0},$$

où, F_e représente la force élastique et S_0 la section de l'éprouvette.

- Le module d'Young (ou module d'élasticité longitudinale) E : c'est une constante élastique qui, pour un matériau donné, lie la contrainte à la déformation et caractérise la raideur de la matière. À contrainte égale, un matériau ayant un module d'élasticité élevé subira une déformation plus faible qu'un matériau ayant un module d'élasticité faible. Il s'exprime en MPa.

$$E = \frac{R_p}{\frac{\Delta L}{l}}$$

où, R_p représente la limite d'élasticité et l correspond à la longueur de l'éprouvette.

- La résistance à la rupture R_m : elle représente la charge unitaire maximale que peut supporter l'éprouvette avant rupture et s'exprime en MPa.

$$R_m = \frac{F_r}{S_0}$$

où F_r représente la force à la rupture et S_0 la section de l'éprouvette.

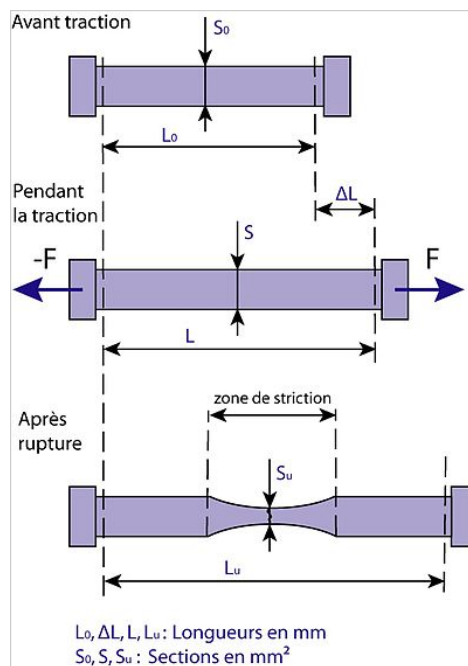


FIGURE 1.6 – Évolution d'une éprouvette soumise à un essai de traction.

Source web : *Allongement à la rupture.*

- L'allongement à la rupture $A\%$: illustré sur la figure 1.6, il définit la capacité d'un matériau à s'allonger avant de rompre lorsqu'il est sollicité en traction et s'exprime en pourcentage.

$$A\% = \frac{l_u - l_0}{l_0} \times 100,$$

où, l_u correspond à la longueur de l'éprouvette juste avant la rupture et l_0 est la longueur initiale (longueur de l'éprouvette avant le début de l'essai de traction).

- La ténacité du matériau K_c : c'est l'aptitude du matériau à résister à la déformation ou à la rupture sous un effort continu, en $MPa\sqrt{m}$.

Les paramètres des conditions d'essais :

- La micro-structure de l'éprouvette qui correspond à sa taille de grain.
- Le rapport de charge appliqué (rapport de contrainte) : $R = \frac{K_{\min,t}}{K_{\max,t}}$, où $K_{\max,t}$ et $K_{\min,t}$ sont respectivement les facteurs d'intensité de contraintes maximales et minimales en $MPa\sqrt{m}$. Ils varient avec le nombre de cycles t .

- La température : elle varie entre 20 et 720 degrés Celcius (selon le matériau testé). Des études (par exemple celle de Vautrot dans [58]) ont montré que la résistance de la plupart des matériaux diminue considérablement avec l'élévation de la température.
- Le chargement maximal appliqué à chaque éprouvette : σ_{\max} en MPa.
- Le niveau de contrainte appliquée $\Delta\sigma$: il se définit par $\Delta\sigma = \sigma_{\min} - \sigma_{\max}$, où σ_{\min} et σ_{\max} correspondent respectivement au chargement minimal et maximal.

1.2.2 Modèles déterministes

Les lois phénoménologiques utilisées par les ingénieurs mécaniciens pour modéliser la propagation de fissure sont généralement de la forme suivante, d'après Lin et Yang dans [33],

$$\frac{da_t}{dt} = L(t, \Delta K_t, K_{\max,t}, K_c, \sigma_{\max}, a_t, R), \quad (1.1)$$

avec $L(\cdot)$ une fonction positive et les différentes notations définies dans le tableau 1.1.

t	temps mesuré en nombre de cycles
a_t	longueur de fissure
da_t/dt	vitesse de propagation de la fissure
ΔK_t	facteur d'intensité des contraintes
K_c	ténacité ou seuil de rupture
$K_{\max,t}$	facteur d'intensité de contrainte maximale. Il varie avec t
R	rapport de contrainte
σ_{\max}	chargement maximal

TABEAU 1.1 – Notations.

Dans [41], Paris et Erdogan ont été les premiers à proposer une équation d'une telle forme. Cette équation, souvent appelée loi de Paris, ne décrit que la partie linéaire de la relation logarithmique entre la vitesse de propagation de fissure de fatigue et le facteur d'intensité des contraintes (régime 2) présentée dans la figure 1.3. Du fait de sa simplicité, c'est certainement le modèle le plus utilisé en propagation de fissure. Elle relie l'évolution de la longueur de fissure a_t au nombre de cycles de contrainte t à travers ΔK_t par l'équation suivante

$$\frac{da_t}{dt} = C(\Delta K_t)^m, \quad (1.2)$$

où C et m sont des paramètres constants caractéristiques du matériau. Dans la majorité des cas, ΔK_t est donné par la relation

$$\Delta K_t = Y(a_t)\Delta\sigma\sqrt{\pi a_t}, \quad (1.3)$$

où $Y(a_t)$ est un facteur sans dimension qui tient compte de la longueur de la fissure, de la géométrie de l'éprouvette et du niveau de contrainte appliqué. Malgré sa popularité et sa précision à décrire le second régime, le modèle donné par l'équation (1.2) n'est pas adapté pour décrire le comportement asymptotique des vitesses en début de propagation (régime 1) et à l'approche de la rupture (régime 3).

De nombreuses alternatives ont été proposées pour surmonter les limites de la loi de Paris. L'équation de Walker introduite dans [60] est l'une des premières équations simples qui tient compte du rapport de contrainte R , qui est un paramètre influent de la vitesse de propagation (un rapport de contrainte élevé se traduit par une vitesse de propagation plus élevée). Elle est définie par

$$\frac{da_t}{dt} = C[(1 - R)^m K_{\max,t}]^p, \quad (1.4)$$

où C , m et p sont les paramètres de cette loi. Forman et al dans [25] proposent un modèle appelé loi de Forman. Cette loi capture l'augmentation asymptotique de la vitesse de fissuration dans le régime 3 et intègre les effets du rapport de contrainte R et du seuil de rupture K_c . Cette loi a été obtenue en divisant la loi de Paris par un facteur qui devrait atteindre zéro quand le facteur d'intensité des contraintes atteindra un niveau critique. La forme générale de l'équation de Forman est la suivante

$$\frac{da_t}{dt} = \frac{C(\Delta K_t)^m}{(1-R)K_c - \Delta K_t}, \quad (1.5)$$

où K_c est le seuil de rupture qui correspond à la valeur du facteur d'intensité des contraintes nécessaire pour atteindre la rupture.

D'autres auteurs ont proposé des modèles plus complexes pour décrire l'ensemble des régimes de propagation. Ils tiennent généralement compte du rapport de contrainte R , du seuil d'amorçage ΔK_{th} et de la ténacité K_c . Nous pouvons par exemple citer le modèle BPQD, proposée par General Electric aviation et qui se définit par la formule suivante,

$$\frac{da_t}{dt} = e^B \left(\frac{\Delta K_t (1-R)^{(m\pm-1)}}{\Delta K_{th0}} \right)^P \left(\log \left(\frac{\Delta K_t (1-R)^{(m\pm-1)}}{\Delta K_{th0}} \right) \right)^Q \left(\log \left(\frac{K_c}{\Delta K_t (1-R)^{(m\pm-1)}} \right) \right)^D,$$

où,

- B , P , Q , D et $m\pm$ sont les paramètres du modèle.
- ΔK_{th0} est le facteur d'intensité de contrainte seuil pour un rapport de contrainte $R = 0$

Cette loi s'appuie sur une modélisation de la loi d'ouverture de la fissure de Walker. Dans le même ordre d'idée, nous pouvons citer le modèle utilisé dans le code NASGRO développé par la NASA et issu des travaux de Forman et Mettu dans [26]. Il se définit par l'équation suivante,

$$\frac{da_t}{dt} = c \left[\left(\frac{1-f}{1-R} \right) \Delta K_t \right]^n \frac{\left(1 - \frac{\Delta K_{th}}{\Delta K_t} \right)^p}{\left(1 - \frac{K_{max,t}}{K_{crit}} \right)^q},$$

où c , p , q et n sont des paramètres du modèle. K_{crit} est le facteur d'intensité de contrainte critique. f est une fonction d'ouverture de la fissure qui est définie par

$$f = \frac{K_{op}}{K_{max,t}} = \begin{cases} \max(R, A_0 + A_1 R + A_2 R^2 + A_3 R^3) & \text{si } R \geq 0, \\ A_0 + A_1 R & \text{si } -2 \leq R < 0, \\ A_0 - 2A_1 & \text{si } R < -2, \end{cases}$$

où K_{op} représente le facteur d'intensité de contrainte à l'ouverture. Il est supposé supérieur à $K_{min,t}$ au cours d'un cycle de charge. Les coefficients A_0 à A_3 sont donnés par

$$\begin{cases} A_0 = (0.825 - 0.34\alpha + 0.05\alpha^2) \left[\cos \left(\frac{\pi \sigma_{max}}{2\sigma_0} \right) \right]^{\frac{1}{\alpha}}, \\ A_1 = (0.415 - 0.071\alpha) \frac{\sigma_{max}}{\sigma_0}, \\ A_2 = 1 - A_0 - A_1 - A_3, \\ A_3 = 2A_0 + A_1 - 1, \end{cases}$$

où α représente le facteur de contrainte et $\frac{\sigma_{max}}{\sigma_0}$ correspond au rapport de contrainte maximale appliquée sur la contrainte d'écoulement. Le facteur d'intensité de contrainte seuil est donné par

$$\Delta K_{th} = \frac{\Delta K_{th0} \left(\frac{a}{a+a_0} \right)^{1/2}}{\left(\frac{1-f}{(1-A_0)(1-R)} \right)^{(1+c_{th}R)}},$$

où a_0 correspond à la longueur de fissure initiale et c_{th} est un coefficient de seuil. Les valeurs de ΔK_{th0} et c_{th} sont fournies par la base de données matériel. L'effet d'épaisseur de la fissure est également inclus dans le modèle par le rapport K_{crit}/K_c donné par

$$\frac{K_{crit}}{K_c} = 1 + B_k \exp \left[- \left(A_k \frac{e}{e_0} \right)^2 \right],$$

où A_k et B_k sont des paramètres d'ajustement, e représente l'épaisseur de la fissure et e_0 correspond à l'épaisseur de référence. Ce modèle permet de décrire l'ensemble des régimes de propagation en prenant en compte des caractéristiques du matériau : la ténacité, la limite élastique et des paramètres de chargements : rapport de contrainte et effet d'ouverture de la fissure.

La grande complexité de ces deux modèles vient du nombre de paramètres à estimer, ce qui nécessite des procédures numériques très lourdes. Du fait de sa simplicité (peu de paramètres à estimer) et de sa capacité à bien décrire le régime 2, la loi de Paris reste donc largement utilisée pour modéliser l'évolution de la vitesse de propagation de fissure.

1.2.3 Présence d'aléa

Données de Virkler

Les données expérimentales obtenues par Virkler et al. dans [59] sont une source d'information bien connue sur la fatigue des matériaux. Ces données, disponibles dans la littérature, sont probablement les plus célèbres et les plus utilisées pour modéliser la propagation de fissure de fatigue. Des études réalisées sur ces données seront présentées dans le chapitre 5.

Un total de 68 éprouvettes en aluminium 2024-T3 ont été pré-fissurées en surface de façon superficielle. Elles ont ensuite été soumises à un chargement d'amplitude constante $\Delta\sigma = 48.28$ MPa pour un rapport de contrainte $R = 0.2$. Pour réaliser ces tests, les ingénieurs ont utilisé des éprouvettes identiques provenant de la même pièce d'aluminium. Elles ont toutes été pré-fissurées de la même façon et la longueur de pré-fissuration de l'éprouvette est de $a_0 = 9$ mm, de telle sorte que l'étape d'initiation de la fissure soit considérée terminée. Les 68 courbes d'évolution de la longueur de fissure a_t en fonction du nombre de cycles t sont représentées sur la figure 1.7. Les mesures sont relevées jusqu'à ce que la fissure atteigne sa longueur critique $K_c = 49.8$ mm. Les relevés se font tous les 0.2 mm dans la plage $9.0 \leq a_t \leq 36.2$ mm, tous les 0.4 mm dans la plage $36.2 \leq a_t \leq 44.2$ mm et tous les 0.8 mm dans la dernière plage $44.2 \leq a_t \leq 49.8$ mm. Un total de 164 mesures ont été relevées pour chacun des 68 essais de propagation de fissure représentés dans la figure 1.7.

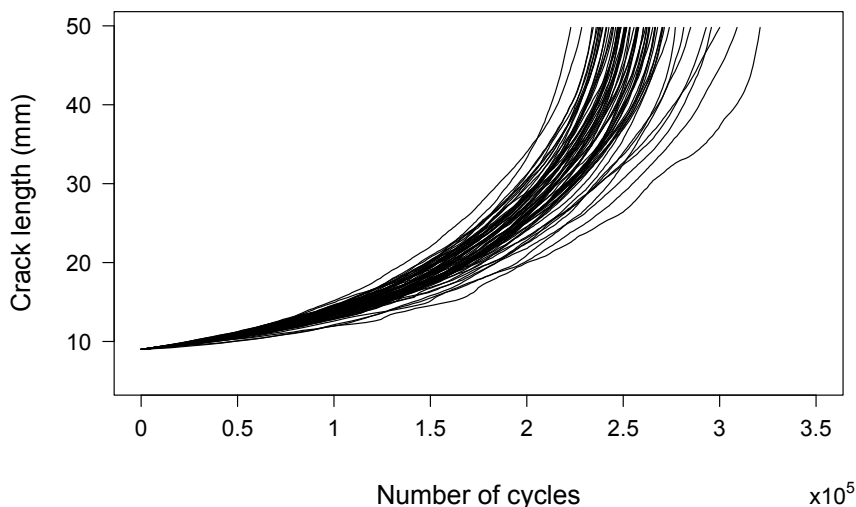


FIGURE 1.7 – 68 courbes expérimentales de longueur de fissure en fonction du nombre de cycles. Ces courbes sont issues des essais de propagation réalisés dans [59].

Variabilité dans les données

D'un point de vue macroscopique, les propriétés mécaniques des matériaux sont souvent considérées comme homogènes. Cependant, comme le montrent les données de Virkler présentées dans la figure 1.7, les données de propagation de fissure de fatigue présentent une variabilité statistique qui augmente avec le nombre de cycles. D'autres auteurs tels que Ghonem et Dore dans [27] ou Wu et Ni dans [63] montrent des résultats similaires. Le caractère aléatoire observé n'est pas un artefact de l'appareil de mesure, mais est bien dû au comportement du matériau, même lorsque les données ont été obtenues dans des conditions de charge identiques dans un environnement strictement contrôlé avec des éprouvettes découpées dans une même pièce. Ces effets aléatoires semblent varier non seulement d'un essai à l'autre, mais aussi pendant l'évolution de la fissure.

Nous pouvons supposer que les connaissances scientifiques de base sur la fatigue ne sont pas encore suffisantes pour décrire de manière strictement déterministe la propagation de fissure. Il n'est donc pas possible de proposer une loi déterministe qui serait valable pour tout individu provenant d'une même expérience, même lorsque l'expérience est répétée dans des conditions absolument identiques. Cette dispersion s'observe également dans la représentation logarithmique des résultats de vitesses de propagation de fissure de Virkler en fonction du facteur d'intensité des contraintes, présenté sur la figure 1.8. Comme le montre le schéma de la figure 1.3, une augmentation linéaire de da_t/dt est observée sur la plus grande partie de la propagation avec quelques changements à la fin et, dans une moindre mesure, au début.

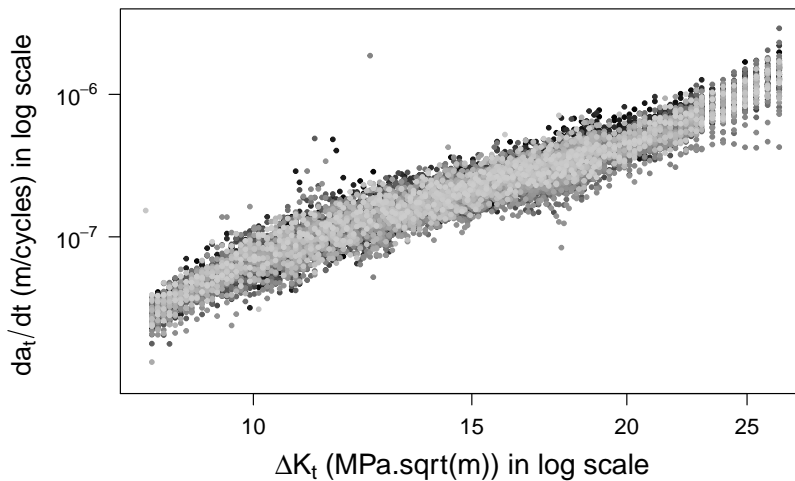


FIGURE 1.8 – 68 courbes de vitesse de propagation de fissure en fonction de ΔK_t . Ces courbes sont calculées à partir des essais de propagation réalisés dans [59].

En raison de la grande dispersion observée, les chercheurs ont commencé à utiliser des modèles statistiques pour caractériser le comportement de la propagation de fissure de fatigue. Il a été souligné que la dispersion restante est due à la nature essentiellement aléatoire de la croissance des fissures de fatigue qui est le résultat de l'inhomogénéité relative au matériau. Les modèles déterministes ne sont pas appropriés pour décrire le comportement de croissance des fissures et la variabilité présente dans les données doit être prise en compte de manière appropriée dans la conception et l'analyse de la structure critique pour la fatigue.

1.2.4 Modèles stochastiques

Afin de prendre en compte la dispersion observée dans les données, de nombreux auteurs se sont intéressés à des modèles stochastiques pour décrire l'évolution de la propagation de fissure de fatigue. Certains d'entre eux ont proposé des modèles pour décrire le processus d'initiation de la fissure, nous citons par exemple Kitamura et al. dans [32] et Sobczyk et Spencer dans [54]. Dans [54], les auteurs proposent des modèles empiriques. Ils proposent une caractérisation plus détaillée des effets aléatoires de la micro-structure, des propriétés mécaniques et de la fatigue des matériaux. Les résultats peuvent être utilisés pour le dimensionnement des pièces. L'avantage du modèle stochastique proposé par Kitamura et al. dans [32] est qu'il a été appliqué à deux jeux de données différents de Ohtani et al. dans [38, 39] dans lesquels des données expérimentales d'initiation de fissure sont disponibles. Ce modèle utilise les concepts d'accumulation des dommages et de dommage critique requis pour atteindre une défaillance, pour prédire la distribution de l'amorçage des fissures et leur croissance précoce. Il faut toutefois noter que le nombre de données basées sur l'amorçage des fissures est très faible et que cette partie de la propagation est difficile à modéliser. Ce n'est pas le cas pour les régimes de propagation et de rupture de la fissure pour lesquels il existe de nombreux modèles dans la littérature.

Les modèles stochastiques permettant de décrire les régimes de propagation et de rupture de la fissure peuvent être découpés en deux familles. La première correspond à une modélisation purement statistique des données de propagation de fissure alors que la seconde est basée sur des modèles stochastiques dérivés des lois phénoménologiques et qui sont des versions aléatoires des modèles déterministes présentés ci-dessus.

Modélisation statistique

Modèles empiriques. Un premier type de modèle consiste à déterminer la distribution statistique qui représente le mieux les données tout en prenant en compte la dispersion. Par exemple Wu et Ni dans [63] utilisent les statistiques descriptives pour étudier la distribution du nombre de cycles nécessaires pour atteindre une longueur de fissure spécifique ou sur la distribution des longueurs de fissure pour une valeur spécifique du nombre de cycles. Les lois de probabilité normale, log-normale et Weibull ont été testées pour ajuster 51 jeux de données différents. Les auteurs ont conclu que les cycles de chargement aléatoires sont mieux ajustés par la loi de Weibull (29/51). Ce type de modélisation semble adéquat pour décrire un jeu de données spécifique mais n'est pas assez flexible pour être étendu à plusieurs jeux de données ou pour faire de la prédiction. De plus, dans ce même article, les auteurs ont montré que les modèles qui prennent en compte les lois phénoménologiques (loi de Paris (1.2) par exemple) et l'aspect temporel sont mieux adaptés pour décrire le caractère aléatoire de la fatigue.

Modèles de sauts. Ces modèles ont été introduits par Bogdanoff et Kozin dans [9] et sont depuis utilisés par d'autres auteurs comme Kirkner et al. dans [31]. Dans les deux cas, les auteurs essaient de modéliser le processus de longueur de fissure a_t en temps continu, par un processus markovien de saut. Il n'y a pas d'aléa dans les transitions puisque quand le processus évolue il passe automatiquement à l'état suivant. En revanche, il y a de l'aléa sur le temps passé dans chacun des états, qui est modélisé par une loi exponentielle. Quand la longueur de fissure augmente, il est possible de modéliser l'accélération de la propagation en augmentant la valeur du paramètre de cette loi. Le modèle de McGill-Markov, introduit par Bogdanoff et Kozin dans [9] est un modèle de sauts inhomogène. Il modélise directement l'évolution temporelle de la longueur de fissure par un processus aléatoire qui se définit de la façon suivante,

- la longueur de fissure a_t au cycle t évolue comme une chaîne de Markov, c'est-à-dire que l'évolution future de la longueur de fissure ne dépend que de sa taille actuelle et pas de la suite des tailles passées;

- la loi du processus dynamique est donnée par la probabilité de transition suivante : soit deux nombres de cycles $t < t'$, cette probabilité est définie par,

$$\mathbf{P}(a_{t'} = j | a_t = i) = p_{ij}(t, t') = C_{j-1}^{j-i} q(t, t')^i (1 - q(t, t'))^{j-1},$$

avec

$$q(t, t') = \exp(-\Delta(t, t')) \quad \text{et} \quad \Delta(t, t') = \int_t^{t'} \frac{\lambda(1 + \lambda s)}{1 + \lambda s^\kappa} ds,$$

où λ et κ sont les paramètres du modèle.

Dans [61], Wu et Ni ont montré que ces modèles ajustent assez bien les données, néanmoins, ils n'ont pas d'interprétation physique. Ils peuvent cependant être généralisés aux processus de Markov déterministes par morceaux de façon à leur donner un sens.

Modèles stochastiques dérivés des lois déterministes

Lois déterministes à variables aléatoires. Des auteurs tels que Virkler et al. [59] ou Tanaka et al. [56] ont proposé les premiers modèles aléatoires de fatigue basés sur les lois phénoménologiques telle que la loi de Paris par exemple. En effet, les coefficients C et m de l'équation (1.2) peuvent être vus comme des constantes qui dépendent du matériau et de certaines conditions d'essai. C'est par l'intermédiaire de ces paramètres que les auteurs proposent de stochastiser le modèle de Paris en les rendant aléatoires. Dans un papier plus récent [52], Shen et al. considèrent également les paramètres C et m de la loi de Paris comme des variables aléatoires. À partir de l'analyse de données expérimentales permettant d'observer une relation linéaire entre m et C , ils ont décidé de modéliser soit C par une variable aléatoire de loi log-normale, soit m par une variable aléatoire de loi normale, mais ne modélisent pas le couple de paramètres de façon aléatoire. Il est alors facile de les estimer et de calculer les indicateurs de fiabilité du modèle. Cette approche fonctionne bien et donne une meilleure description du phénomène. Cependant, les auteurs ne semblent pas considérer la possibilité d'avoir conjointement C et m aléatoires, alors que c'est ce que tendent à montrer les données. Par ailleurs, ce modèle ignore les éventuels changements de dynamique dans la propagation de fissures puisque les lois de C et m ne dépendent pas du temps.

Lois déterministes introduisant un processus aléatoire. L'utilisation des systèmes différentiels permet de décrire les variations temporelles de la dispersion. Les modèles stochastiques prennent la forme générale des lois expérimentales (1.1) dans lesquelles est ajouté un facteur multiplicatif qui est un processus aléatoire.

- Processus log-normal. Dans [65, 64], Yang et Manning ont introduit ce modèle. Ils ont ajouté un facteur aléatoire multiplicatif X_t à l'équation générale des modèles déterministes (1.1),

$$\frac{da_t}{dt} = X_t L(t, \Delta K_t, K_{\max}, K_c, \sigma_{\max}, a_t, R) = X_t Q(a_t)^b,$$

où Q et b sont des constantes. Le coefficient multiplicatif X_t dépend du temps t (en nombre de cycles), c'est donc un processus stochastique. Les auteurs ont modélisé X_t comme un processus aléatoire log-normal stationnaire avec une valeur médiane de 1, un écart type σ_X et une fonction d'auto-covariance de la forme suivante,

$$\text{Cov}[X_t, X_{t'}] = \sigma_X^2 \exp(-\zeta|t' - t|).$$

Dans ce cas, la structure de probabilité du processus stochastique de croissance des fissures de fatigue peut être obtenue analytiquement. Le succès de ce modèle vient du fait qu'il y a seulement deux paramètres à estimer : l'écart-type σ_X et la covariance ζ . Wu et Ni dans [61, 62, 63] ont testé ce modèle sur des données expérimentales et ont conclu que le modèle log-normal le plus simple est suffisant pour décrire les données étudiées sur la croissance des fissures de fatigue. Cependant, l'hypothèse de stationnarité du processus peut être remise

en cause car la vitesse d'évolution des fissures suit différentes phases au cours du temps, avec en particulier la phase d'accélération lorsqu'on est proche de la rupture. Par ailleurs, pour les calculs de fiabilité il est nécessaire de connaître la loi de

$$W_T = \int_0^T X_t dt.$$

Sans aucune justification mathématique, les auteurs supposent que le processus W_T suit encore une loi log-normale dont on peut relier les paramètres à ceux du processus X_t . À notre connaissance aucune discussion ne permet de valider cette approximation. Dans [45], Ray et al. ont modélisé l'espérance de la longueur de la fissure $\mu_t = \mathbf{E}[a_t]$ par une loi déterministe qui ressemble à la loi de Paris,

$$\frac{d\mu_t}{dt} = C(\Delta K_t)^m.$$

En utilisant l'extension de Karhunen-Loève, ils ont proposé un modèle stochastique dans lequel ils considèrent que la longueur de fissure suit une loi log-normale à la place de l'hypothèse classique dans laquelle c'est le taux de croissance des fissures qui est considéré log-normal. Le modèle proposé a été vérifié sur les données expérimentales proposées par Virkler et al. dans [59] et Ghonem et Dore dans [27] à différent niveaux d'amplitude constante.

- Modèle polynomial. Dans [36], Ni s'est inspiré du modèle de Yang pour proposer un modèle stochastique polynomial de croissance des fissures de fatigue. Ce modèle est basé sur l'hypothèse que la croissance des fissures de fatigue est égale à une fonction polynomiale déterministe en terme de longueur de fissure de fatigue. Afin de prendre en compte la dispersion statistique de la croissance des fissures de fatigue, cette fonction est multipliée par un facteur aléatoire log-normal stationnaire.

$$\frac{da_t}{dt} = X_t(p + qa_t + ra_t^2).$$

Dans [37], Ni a vérifié la cohérence entre les résultats obtenus analytiquement à partir du modèle proposé et les données expérimentales. Ces résultats sont apparus satisfaisants. Pour le moment il n'existe aucune comparaison entre le modèle proposé et d'autres modèles à travers des données expérimentales

- Processus de diffusion. Les processus de diffusion associés à une équation différentielle stochastique pilotée par un mouvement brownien ont été largement étudiés dans la littérature pour modéliser la propagation de fissure. Dans [53], Sobczyk a été le premier auteur à essayer de modéliser la propagation de fissure par un processus de diffusion. Il propose une approximation asymptotique du processus de diffusion a_t qui devient valide quand le paramètre ϵ , qui apparaît dans la loi de propagation, tend vers 0,

$$\frac{da_t}{dt} = \epsilon \tilde{L}(a_t) X_t,$$

où \tilde{L} dépend de ΔK_t , K_{\max} , K_c , σ_{\max} et R qui sont des constantes connues ou des fonctions de a_t . Finalement \tilde{L} est seulement une fonction du temps et de la longueur de fissure. Les résultats asymptotiques et les techniques de moyennisation ont été initialement proposés par Stratonovich. Ils ont ensuite été consolidés par d'autres auteurs comme Khas'minskii dans [29, 30] ou Papanicolaou et Kohler dans [40], qui permettent de spécifier les conditions d'approximation de a_t par un processus de diffusion quand ϵ tend vers 0. Sur la base de ces constatations, Sobczyk dans [53] a été le premier à tenter de modéliser a_t par un processus de diffusion. Il a montré que dans un contexte homogène, le processus de diffusion a_t est gouverné par un système de la forme,

$$da_t = \mu(a_t)dt + \sigma(a_t)dB_t,$$

où μ et σ sont les fonctions de dérive et de diffusion et B_t est un mouvement brownien. Dans [33, 66] Lin et Yang ont complété ces résultats en les appliquant à des cas réels. Depuis, ces processus sont utilisés par de nombreux auteurs comme Ivanova et Naess dans [28], qui utilisent des méthodes d'approximation numérique des modèles de diffusion pour les calculs de fiabilité des structures. Ces modèles ont de bonnes justifications mathématiques et on peut en déduire des indicateurs de fiabilité. L'inconvénient de ce type de modèle vient de la dynamique du processus a_t qui est autorisé à croître et décroître continuellement au cours du temps par la modélisation, ce qui ne convient pas à la propagation de fissure pour laquelle la croissance est un caractère intrinsèque.

1.3 Processus de Markov déterministes par morceaux

Dans les années 1980, les processus markoviens déterministes par morceaux (PDMP pour l'anglais Piecewise Deterministic Markov Processes) ont été introduits par M.H.A. Davis comme une classe générale de modèles stochastiques non diffusifs. Ils servent notamment à décrire des systèmes physiques, qui évoluent selon des lois déterministes, et dont la dynamique peut être perturbée par deux types d'événements ponctuels et aléatoires :

- le premier type d'événement est lié à l'évolution déterministe des grandeurs physiques du système;
- le second type d'événement est purement stochastique et correspond à des défaillances de certains composants ou à des sollicitations extérieures.

Dans les deux cas, ces événements vont introduire des sauts dans le comportement du système. Entre deux événements, l'évolution du système est basée sur des lois physiques qui tiennent compte de la dépendance ou des interactions dynamiques entre les différents processus impliqués. Peu importe le phénomène modélisé, il est d'un grand intérêt d'utiliser des lois déterministes (qui viennent de la physique ou qui sont bien connues des praticiens) tout en autorisant des événements aléatoires. Ces systèmes sont décrits par deux variables : une variable discrète de mode qui traduit l'apparition des pannes ou des dysfonctionnements et caractérise ainsi le fonctionnement du système jusqu'à sa panne complète; une variable euclidienne d'état représentative des grandeurs physiques du système.

La trajectoire d'un PDMP (X_t) sur un espace d'état de la forme $\bigcup_{k \in K} \{k\} \times E_k$, où K est l'ensemble au plus dénombrable des régimes possibles et E_k l'espace d'état dans le régime k , est définie à partir de ses trois caractéristiques locales $(\lambda_k, Q_k, \Phi_k)_{k \in K}$:

- $\Phi_k : E_k \times \mathbb{R} \rightarrow E_k$ est le flot déterministe dans le régime k ;
- $\lambda_k : E_k \rightarrow \mathbb{R}_+$ est le taux de saut dans le régime k ;
- $Q_k : (E_k, \mathcal{B}(E_k)) \rightarrow [0, 1]$ est le noyau de transition dans le régime k .

Partant d'une condition initiale $X_0 = x$, la trajectoire de (X_t) est décrite comme suit. Le premier temps de saut est défini par sa fonction de survie,

$$\forall t \geq 0, \mathbf{P}(T_1 > t | X_0 = x) = \exp\left(-\int_0^t \lambda(\Phi(x, s)) ds\right).$$

La trajectoire entre 0 et le premier temps de saut T_1 est donnée par

$$\forall 0 \leq t \leq T_1, X_t = \begin{cases} \Phi(x, t) & \text{pour } t < T_1 \\ Z_1 & \text{sinon,} \end{cases}$$

où la variable Z_1 est définie à partir du noyau de transition de Q : pour toute fonction mesurable bornée ϕ ,

$$\mathbb{E}[\phi(Z_1) | T_1, X_0 = x] = \int \phi(u) Q(\Phi(x, T_1), du).$$

Partant de la position X_{T_1} , le temps de parcours $S_2 = T_2 - T_1$ et la future position Z_2 sont obtenus de la même manière et ainsi de suite. On obtient un processus à temps continu (X_t) ponctué par des sauts ayant lieu aux instants T_n . Les temps de parcours du flot sont les $S_n = T_n - T_{n-1}$. Enfin, les positions lors des sauts sont les $Z_n = X_{T_n}$. Au PDMP est donc associée une chaîne de Markov $(Z_n, S_n)_{n \geq 0}$. Une sélection appropriée de l'espace d'état et des principales caractéristiques du processus fournit une grande variété de modèles stochastiques couvrant de nombreuses applications comme la gestion de systèmes complexes étudiée par Dufour et Dutuit dans [22] et par Zhang et al. dans [67], la modélisation du phénomène de dégradation étudiée par de Saporta et al. dans [49] ou de la biologie par Beil et al. dans [5]. Les propriétés mathématiques et les implémentations numériques de ces processus ont été largement étudiés par M.H.A Davis dans [19, 18] mais également par Brandejsky et al. [11], de Saporta et Dufour dans [48] et de Saporta et al. dans [20], et les références qui y figurent.

Dans le cas de la propagation de fissure, nous savons d'une part que la longueur de fissure est croissante avec le temps et qu'elle ne permet pas l'utilisation de processus diffusifs pour la modélisation, car ils ne sont pas monotones. D'autre part la loi phénoménologique de Paris (1.2), ou d'autres lois comme celle de Walker (1.4) ou Forman (1.5), fonctionnent bien sur les régimes 2 et 3 et sont bien connues des mécaniciens. L'utilisation des PDMP pour modéliser la propagation de fissure permet donc

- d'utiliser des lois déterministes de propagation;
- d'introduire de l'aléa pour tenir compte de la variabilité entre les différentes fissures;
- d'introduire de l'aléa durant la propagation de fissure;
- de prendre en compte des évènements connus tels que le passage du régime 2 au régime 3.

Un des avantages de la modélisation par des PDMP est qu'elle permet de modéliser le changement de régime au niveau du temps de saut. Le caractère stochastique de ce changement est lié au caractère aléatoire du phénomène de propagation de différentes manières. Le changement de régime peut par exemple se produire à un moment prévisible lorsque certaines composantes du processus d'état atteignent un seuil, cela déclenche de façon déterministe un changement de mode dans le système. Le saut peut se produire à un temps aléatoire en raison d'évènements exogènes, comme un impact indépendant de la dynamique du système. Il peut également se produire de manière aléatoire en fonction de l'état du processus, typiquement à travers le taux de saut $\lambda(\Phi(x, t))$; ce qui semble intéressant pour modéliser le temps de saut entre les régimes 2 et 3. Le cadre des PDMP permet au temps de saut de dépendre de l'état du processus qui est, dans le cadre de la propagation, la longueur de fissure : plus la fissure est longue, plus le changement entre les régimes 2 et 3 est probable. Dans [1], Azaïs et Muller-Gueudin proposent une méthode pour inférer le taux de saut de ce dernier cas.

1.3.1 Modèle multiplicatif

L'utilisation des PDMP pour la modélisation de la propagation d'une fissure a été introduite par Chiquet et al. [15, 16]. Elle permet à la fois d'obtenir un modèle reposant sur des propriétés physiques (loi de Paris (1.2)), de rendre aléatoires les paramètres de cette loi et d'autoriser, comme pour les processus markoviens, des sauts de changement de régimes dans la propagation (accélération de la propagation au cours du temps). Dans cette approche les auteurs considèrent que la croissance des fissures de fatigue change en raison de petits chocs survenant à des moments aléatoires. Ils ont choisi d'utiliser les PDMP en ne leur attribuant aucune contrainte sur le nombre de sauts possibles du processus. Ils proposent un modèle dérivé de la loi de Paris dans lequel ils considèrent que le facteur $Y(a_t)$ de l'équation (1.3) est égal à 1. Ils rendent aléatoire cette loi en ajoutant un facteur multiplicatif donné par un processus de saut P_t , conduisant au système stochastique suivant,

$$\frac{da_t}{dt} = C(\Delta\sigma\sqrt{\pi})^m (a_t)^{m/2} P_t, \quad a_0 = a.$$

Le processus de Markov P_t est supposé avoir un espace d'état fini \mathcal{P} et une matrice génératrice constante $Q = (q_{ij})_{(i,j) \in \mathcal{P}^2}$. Finalement, $(X_t)_{t \geq 0} = (a_t, P_t)_{t \geq 0}$ est un PDMP d'espace d'état $\chi = \mathbb{R}_+ \times \mathcal{P}$, de taux de saut donné par $\lambda(x) = \lambda((a, p)) = \lambda(p) = q_{p,p}$ et de noyau de transition $Q(x, dy) = Q(p, p') = -q_{p,p'} / q_{p,p}$, donné par la matrice génératrice Q du processus de Markov. Les paramètres m et C sont identiques pour toutes les fissures, cependant chaque fissure est modélisée par un processus P_t différent. Ce terme permet à la fois de capturer l'aléa entre les fissures mais également l'aléa présent pendant la propagation. Les auteurs ont ajusté les données de propagation de Virkler et ont souligné une accélération des sauts du processus P_t au fur et à mesure que les longueurs de fissures tendent vers de grandes valeurs, annonçant l'arrivée dans le régime 3 qui correspond au régime de rupture.

1.3.2 Modèles à un saut

Dans [57, 2, 8, 3], Azaïs et al. ont proposé deux modèles de propagation de fissures utilisant les PDMP. Ces modèles sont plus simples que celui présenté dans la section précédente car ici les auteurs font l'hypothèse que la propagation de fissure peut s'exprimer sous la forme d'un PDMP avec un seul changement de régime. Cela permet de donner un sens physique au saut ou d'apporter de la flexibilité au modèle proposé. Les objectifs associés à ces deux modèles sont différents.

L'objectif du premier modèle est de caractériser la propagation à l'instant de saut entre les régimes 2 et 3. Le modèle proposé est une combinaison de deux lois déterministes utilisées de façon appropriée dans les différents régimes. Au début de la propagation, la longueur de fissure évolue selon la loi de Paris (donnée dans l'équation (1.2)) dans laquelle les paramètres C et m ont été choisis aléatoirement. Après un temps aléatoire, le modèle change de loi et passe de la loi de Paris à la loi de Forman (donnée dans l'équation (1.5)) avec de nouveaux paramètres pour m et C . Il n'y a qu'un seul changement dans la propagation. Dans ce cas, $X_t = (a_t, P_t)$, a_t représente la longueur de fissure; P_t désigne le mode de propagation à trois composants $P_t = (L_t, C_t, m_t)$, où $L_t \in \{1, 2\}$ indique la loi de propagation (Paris ou Forman) et (C_t, m_t) sont les paramètres constants de cette loi. La valeur initiale de L_t est déterministe et vaut 1 et la transition est également déterministe allant de 1 à 2, c'est-à-dire que $L_0 = 1$ et $Q((1, C_1, m_1), (2, \dots)) = 1$.

L'objectif du second modèle est de proposer une méthode de prédiction d'une fissure donnée lorsque l'on observe les premiers points de sa trajectoire. Cette méthode de prédiction est basée sur un modèle de PDMP plus simple que le précédent. En effet, en plus de ne tenir compte que d'un seul saut, les auteurs considèrent que dans ce modèle le flot déterministe est donné par la loi de Paris avant et après le saut. Lorsque le saut se produit, les paramètres de la loi de Paris changent de manière aléatoire. Dans ce cas le processus markovien de saut s'écrit, $X_t = (a_t, P_t)$, a_t est la longueur de fissure et P_t est un mode à deux composants (C_t, m_t) , où (C_t, m_t) sont les paramètres de la loi de Paris. La modélisation de la propagation, lorsque l'évolution de la fissure est donnée par deux équations différentielles successives de la loi de Paris, ajuste correctement les données. Le saut donne de la flexibilité au modèle pour l'ajustement et la seconde loi de Paris est naturelle car la méthode de prédiction est conçue pour prédire la propagation de fissure à partir des premiers points de sa trajectoire observés dans le régime de Paris.

Après avoir testé ces deux modèles sur les données de propagation de Virkler, les auteurs suggèrent que les modèles PDMP, même s'ils sont relativement simples, sont pertinents et peuvent être interprétés pour donner des informations utiles sur les processus mécaniques impliqués dans la propagation de fissure.

1.4 Régression polynomiale par morceaux

Lorsque les données semblent être issues d'un modèle déterministe par morceaux, la régression polynomiale par morceaux est intéressante car si le nombre de régimes et le degré des polynômes

sont choisis, alors le modèle devient paramétrique (coefficients des polynômes et instants de transition). La souplesse donnée par la diversité des polynômes et des régimes permet de modéliser beaucoup de types de courbes. De plus, l'estimation d'un tel modèle pour un degré de polynôme peu élevé permet de segmenter une courbe en un nombre fini de segments pour la recherche des différents régimes, ou de manière équivalente la recherche des ruptures dans la dynamique du modèle. On voit particulièrement son intérêt dans la modélisation d'une courbe de vitesse de propagation de fissure et la recherche des différents régimes qui la composent.

1.4.1 Modèle de régression polynomiale par morceaux

Afin de mieux décrire l'état de l'art au sujet de la régression polynomiale par morceaux, nous décrivons ici le modèle qui s'écrit,

$$Y = \varphi(X) + \varepsilon \Sigma(X), \quad (1.6)$$

où la fonction de lien φ est polynomiale par morceaux, c'est-à-dire

$$\varphi(x) = \sum_{k=1}^K \rho_{\alpha_k}(x) \mathbb{1}_{[\tau_{k-1}, \tau_k[}(x),$$

avec K correspondant au nombre de régimes. Les instants $(\tau_k)_{0 \leq k \leq K}$ sont supposés ordonnés de telle sorte que $\tau_i < \tau_j$ si et seulement si $i < j$. Pour tout k , l'intervalle $[\tau_{k-1}, \tau_k[$ correspond au régime k . Sur cet intervalle la fonction de lien est donnée par

$$\rho_{\alpha_k}(x) = \sum_{l=0}^{\delta} \alpha_k(l) x^l,$$

où le degré δ est supposé fixé. Le support de la fonction de lien est $[\tau_0, \tau_K[$, ce qui correspond à \mathbb{R} si $\tau_0 = -\infty$ et $\tau_K = +\infty$. Le modèle est gaussien et hétéroscédastique, c'est à dire que $\varepsilon \sim \mathcal{N}(0, 1)$ et la variance du signal est constante par morceaux et s'écrit,

$$\Sigma(x) = \sum_{k=1}^K \sigma_k \mathbb{1}_{[\tau_{k-1}, \tau_k[}(x),$$

où σ_k représente l'écart-type du signal dans le régime k .

Les premières publications de modèle de régression polynomiale par morceaux concernent des modèles à deux régimes uniquement. Quandt dans [44] estime un modèle linéaire (degré 1) par morceaux tandis que Robinson dans [46] estime un modèle polynomial par morceaux plus général. Ces deux auteurs ont pointé du doigt la difficulté de résoudre le cas où l'instant de transition entre les régimes n'est pas observé. McZgee et Carleton dans [34] sont les premiers auteurs à proposer une méthode de régression linéaire par morceaux qui s'appuie sur un modèle linéaire à plus de deux régimes et qu'ils proposent d'estimer en mariant la segmentation hiérarchique à la théorie de la régression standard. Depuis, d'autres auteurs comme Ferrari-Trecate et Muselli dans [24] ou encore Picard et al. dans [43] ont proposé des méthodes alternatives de régression polynomiale par morceaux pour estimer un modèle de régression linéaire par morceaux ou polynomiale par morceaux, c'est le cas de Brailovsky et Kempner dans [10] ou encore de Chamroukhi et al. dans [14, 13]. Ces modèles sont homoscedastiques car ils supposent que les différents régimes de régression polynomiale ont la même variance du bruit, c'est le cas dans [34, 10, 24]. Ce cadre peut être plus général en considérant un modèle hétéroscédastique qui autorise la variance du bruit à varier entre les différents régimes de régression polynomiale comme dans [43, 14, 13]. Cependant, l'ensemble des modèles proposés par ces auteurs ne supposent aucune contrainte de régularité.

Par ailleurs, le cadre de modélisation varie selon les auteurs et peut se scinder en deux catégories. On a d'une part la modélisation de données non temporelles : les auteurs souhaitent partitionner les données en attribuant chaque segment de courbe à un régime. C'est le cas de Ferrari et

Trecate dans [24] qui effectuent la recherche de régions homogènes dans l'espace, Picard et al. dans [43] recherchent des régions homogènes dans le génome et Chamroukhi dans [12] effectue de la classification de courbes. Dans ce contexte cela a évidemment du sens d'estimer les régimes sans aucune contrainte de régularité sur le modèle. D'autre part, l'objet peut être la modélisation de données temporelles : les auteurs souhaitent ajuster les données pour caractériser l'évolution successive d'un phénomène physique. Dans [10] Brailovsky et Kempner souhaitent restaurer la structure sous-jacente d'un signal, dans [14, 13] Chamroukhi et al. cherchent à caractériser le signal indiquant l'énergie électrique consommée pendant l'exploitation d'une voie d'aiguillage. Dans ce contexte, les auteurs n'imposent pas de contrainte de régularité au modèle, ce qui aurait pourtant du sens et que nous proposons dans cette thèse.

1.4.2 Méthode d'inférence et algorithmes

Pour estimer les paramètres du modèle (1.6), la pratique courante est d'utiliser la méthode du maximum de vraisemblance. Dans ce contexte on peut utiliser les techniques de programmation dynamique introduites par Bellman et Roth dans [7] et Stone dans [55] pour estimer les paramètres optimaux, ce qui est également fait dans [10, 43, 14]. En effet, lorsque la vraisemblance peut s'écrire sous la forme d'une somme de termes indépendants, l'usage de la programmation dynamique permet la maximisation indépendante et récursive de différents sous-intervalles du domaine considéré.

Dans le but d'ajouter de la souplesse à la segmentation, il est possible d'approcher le modèle (1.6) par un modèle à variable latente. Cette variable latente est de deux types : dans le cadre de modèle de mélanges polynomiaux par morceaux, la variable latente, notée Z_i , modélise le régime à l'instant i (voir par exemple [43, 12]) ; dans le cadre des modèles de Markov caché, Z_i est un processus de Markov de loi initiale π et de matrice de transition A qu'il faut estimer (voir par exemple [14, 13]). Le régime est attribué à la valeur k qui maximise $\mathbf{P}(Z_i = k|X_i)$. Dans ce contexte, la maximisation directe de la vraisemblance est trop complexe. Selon le type de variable latente, il est plutôt d'usage d'utiliser un algorithme EM proposé par Dempster et al. dans [21] (modèles de mélanges) ou un algorithme de Baum-Welch proposé par Baum et al. dans [4] (algorithme EM spécifique aux modèles de Markov cachés).

Enfin, nous souhaitons mentionner l'usage d'algorithmes hybrides comme c'est le cas dans [43], où les auteurs cherchent à déterminer des régions homogènes dans le génôme. L'algorithme proposé alterne algorithme EM (estimation des paramètres de chaque cluster lorsqu'ils sont prédéfinis) et programmation dynamique (estimation des clusters à partir des paramètres estimés par algorithme EM).

Chapitre 2

Régression polynomiale par morceaux sous contrainte de régularité

Sommaire

2.1 Introduction	24
2.2 Modèle de régression polynomiale par morceaux	24
2.2.1 Fonction de lien	24
2.2.2 Contrainte de régularité	25
2.2.3 Log-vraisemblance	25
2.2.4 Approximation par un modèle à variable latente	26
2.3 Maximum de vraisemblance à deux régimes	28
2.3.1 Modèle sans variable latente	29
2.3.2 Modèle à variable latente : algorithme EM	31
2.4 Généralisation à un nombre quelconque de régimes	33
2.4.1 Modèle sans variable latente	33
2.4.2 Modèle à variable latente et algorithme EM	35
2.5 Approche récursive : algorithme de type Bellman	36
2.5.1 Modèle sans variable latente	36
2.5.2 Lissage autour des instants de transition	39

2.1 Introduction

Les modèles de régression polynomiale par morceaux sont souples par le choix du degré des polynômes et des intervalles dans lesquels sont définis ces polynômes. Ils sont adaptés pour modéliser des phénomènes physiques comme la propagation de fissure, qui évolue selon plusieurs régimes successifs de propagation. Dans ce chapitre, nous proposons un modèle de ce type pour ajuster et segmenter des données dans le but de modéliser un phénomène temporel continu en identifiant les régimes successifs qui le composent. Malgré les généralités des nombreux modèles présentés dans l'état de l'art (cf. chapitre 1 section 1.4), le modèle proposé dans ce chapitre et les méthodes proposées pour estimer ses paramètres se distinguent sur plusieurs points. Premièrement, nous donnons la possibilité d'imposer des contraintes de régularité (continuité et/ou dérivabilité) sur la fonction de lien φ , ce qui se traduit en particulier par des contraintes aux instants de transition entre les régimes. En effet, comme nous avons en tête de modéliser un phénomène temporel continu, il nous paraissait intéressant d'ajouter cette possibilité. Ensuite, le statut de la variable latente utilisée pour modéliser l'appartenance au régime se différencie aussi des démarches précédentes. En effet, dans la littérature la variable latente modélise soit l'appartenance à une catégorie dans le cadre du clustering (c'est le cas dans [43, 12]) ou est un processus temporel de Markov caché dont il faut identifier les paramètres de la loi initiale et de la matrice de transition (c'est le cas dans [14, 13]). Dans notre cas elle correspond simplement à un lissage autour des instants de transition que l'on va chercher à estimer et n'est pas un processus de Markov. Enfin, nous proposons un algorithme de programmation dynamique pour l'optimisation d'une vraisemblance lissée. Cette méthode est inspirée des travaux de Bellman et Roth dans [7] qui ont étudié des modèles de régression linéaire par morceaux avec hypothèse de continuité. Elle a été conçue dans le but de limiter les temps de calcul nécessaires à l'estimation du modèle.

Avant de présenter l'algorithme de maximisation de la vraisemblance lissée, nous présentons différents modèles dont nous nous sommes inspirés pour l'élaborer : un modèle «standard» correspondant à l'équation (1.6) avec contraintes de régularité (estimation des régimes par optimisation de la vraisemblance sous contrainte). Nous affinons ensuite ce modèle en introduisant de l'aléa autour des instants de transition par une variable latente (estimation des régimes par algorithme EM). La segmentation en deux régimes étant centrale dans la programmation dynamique et les méthodes présentées dans le chapitre 3, nous commençons à chaque fois par l'étude du modèle à deux régimes, pour poursuivre sur leur généralisation à un nombre quelconque de régimes. Nous adoptons donc le plan suivant. En section 2.2, nous introduisons le modèle, les contraintes de régularité, la version du modèle avec variable latente et les différentes log-vraisemblances associées aux modèles. La section 2.3 propose une méthode de résolution pour le modèle à deux régimes, dans sa version standard (section 2.3.1) ou dans sa version à variable latente (section 2.3.2) par un algorithme EM. La section 2.4 généralise la section 2.3 dans le cas où le nombre de régimes est quelconque. Enfin, nous présentons l'algorithme de type Bellman pour l'approximation de la vraisemblance lissée en section 2.5. Les simulations relatives à ces méthodes sont reportées au chapitre 4.

2.2 Modèle de régression polynomiale par morceaux

2.2.1 Fonction de lien

Dans ce chapitre, nous nous intéressons à un modèle de régression avec un bruit additif

$$Y = \varphi(X) + \varepsilon \Sigma(X),$$

où la fonction de lien φ est polynomiale par morceaux, c'est-à-dire

$$\varphi(x) = \sum_{k=1}^K \rho_{\alpha_k}(x) \mathbb{1}_{[\tau_{k-1}, \tau_k[}(x), \quad (2.1)$$

avec K correspondant au nombre de régimes. Les instants $(\tau_k)_{0 \leq k \leq K}$ sont supposés ordonnés de telle sorte que $\tau_i < \tau_j$ si et seulement si $i < j$. Pour tout k , l'intervalle $[\tau_{k-1}, \tau_k[$ correspond au régime k . Sur cet intervalle la fonction de lien est donnée par

$$\rho_{\alpha_k}(x) = \sum_{l=0}^{\delta} \alpha_k(l) x^l,$$

où le degré δ est supposé fixé de telle sorte que $\alpha_k \in \mathbb{R}^{\delta+1}$. Le support de la fonction de lien est $[\tau_0, \tau_K[$, ce qui correspond à \mathbb{R} avec $\tau_0 = -\infty$ et $\tau_K = +\infty$. Enfin, nous supposons que le modèle est gaussien et homoscédastique par morceaux, c'est à dire que $\epsilon \sim \mathcal{N}(0, 1)$ et la variance du signal s'écrit de la façon suivante,

$$\Sigma(x) = \sum_{k=1}^K \sigma_k \mathbb{1}_{[\tau_{k-1}, \tau_k[}(x),$$

où σ_k représente l'écart-type dans le régime k .

2.2.2 Contrainte de régularité

Les données que nous étudions étant issues d'un phénomène physique susceptible d'être modélisé de façon continue ou dérivable, nous ajoutons une hypothèse de régularité à la fonction de lien φ . Cette hypothèse peut porter sur la continuité de la fonction de lien, elle est alors notée \mathcal{A}_c , ou sur sa dérivabilité, notée \mathcal{A}_d .

$\mathcal{A}_c(\tau_K)$: φ est continue sur l'intervalle $[\tau_0, \tau_K]$, c'est-à-dire,

$$\forall 1 < k \leq K, \rho_{\alpha_{k-1}}(\tau_{k-1}) = \rho_{\alpha_k}(\tau_{k-1}).$$

$\mathcal{A}_d(\tau_K)$: φ est continûment dérivable sur l'intervalle $[\tau_0, \tau_K]$, c'est-à-dire $\mathcal{A}_c(\tau_K)$ est vraie et,

$$\forall 1 < k \leq K, \dot{\rho}_{\alpha_{k-1}}(\tau_{k-1}) = \dot{\rho}_{\alpha_k}(\tau_{k-1}),$$

où pour tout α , $\dot{\rho}_{\alpha}(x)$ est la dérivée de $\rho_{\alpha}(x)$ par rapport à x .

Remarque 1. Si le polynôme est de degré $\delta = 0$ alors l'unique polynôme qui vérifie la contrainte de continuité est le polynôme constant. La contrainte de continuité n'est donc intéressante qu'à partir du degré 1. De même, lorsque le polynôme est de degré $\delta = 1$, alors l'unique solution qui vérifie la contrainte de dérivabilité est une droite. La contrainte de dérivabilité n'est donc intéressante qu'à partir du degré 2.

2.2.3 Log-vraisemblance

Nous supposons que nous observons des données $(X_i, Y_i)_{1 \leq i \leq n}$ issues de ce modèle de régression paramétrique, les instants d'observations X_i sont supposés déterministes et $Y_i = \varphi(X_i) + \epsilon_i \Sigma(X_i)$ avec les ϵ_i qui sont indépendants de loi $\mathcal{N}(0, 1)$. De cette manière, les Y_i sont des variables aléatoires gaussiennes indépendantes de moyenne $\varphi(X_i)$ et de variance $\Sigma(X_i)$.

Nous nous intéressons à l'estimation des quantités inconnues τ_{K-1} , α_K et σ_K^2 , où, par souci de lisibilité, \mathbf{A}_I désigne tout vecteur de la forme $(A_i)_{1 \leq i \leq I}$. Le support est supposé connu mais pourrait simplement être estimé par le minimum et le maximum des X_i . Le degré δ des polynômes est aussi supposé connu. Le nombre de régimes K est un paramètre crucial qui est supposé connu dans un premier temps.

La log-vraisemblance de ce modèle de régression est donnée (à une constante additive près) par

$$\mathcal{L}(\mathbf{T}_{K-1}, \mathbf{a}_K, \mathbf{S}_K^2) = -\frac{1}{2} \sum_{i=1}^n \sum_{k=1}^K \left[\log(S_k^2) + \frac{(Y_i - \rho_{\alpha_k}(X_i))^2}{S_k^2} \right] \mathbb{1}_{[\tau_{k-1}, \tau_k[}(X_i), \quad (2.2)$$

où $T_0 = \tau_0$ et $T_K = \tau_K$. Notre objectif vise à estimer l'ensemble des paramètres inconnus en maximisant la log-vraisemblance sous la contrainte de régularité $\mathcal{A}_c(\tau_K)$ ou $\mathcal{A}_d(\tau_K)$. Les paramètres \mathbf{T}_{K-1} , \mathbf{a}_K , \mathbf{S}_K^2 , qui réalisent ce maximum seront les estimateurs de τ_{K-1} , α_K , σ_K^2 .

2.2.4 Approximation par un modèle à variable latente

Modèle à variable latente

Dans l'objectif de rendre le modèle plus performant pour estimer à la fois les instants de transition entre les régimes et les paramètres de chaque régime, nous modifions le modèle précédent en rendant aléatoire l'appartenance de l'observation (X_i, Y_i) au régime k . En effet, cela permet d'apporter de la flexibilité au modèle, limitant ainsi les erreurs d'estimations liées à la présence d'observations dans un mauvais régime. Pour cela, nous introduisons les variables aléatoires $(Z_i)_{1 \leq i \leq n}$ avec Z_i discrète à valeurs dans $\{1, \dots, K\}$ qui représente le régime dans lequel on se trouve au temps X_i . Le modèle vérifie toujours l'équation $Y = \varphi(X) + \varepsilon \Sigma(X)$, en revanche les fonctions de lien entre X_i et Y_i et de variance s'écrivent

$$\varphi(X_i) = \sum_{k=1}^K \rho_{\alpha_k}(X_i) \mathbb{1}_{\{Z_i=k\}} \quad \text{et} \quad \Sigma(X_i) = \sum_{k=1}^K \sigma_k \mathbb{1}_{\{Z_i=k\}}.$$

En d'autres termes, cela signifie que la distribution de Y sachant X et Z est définie par

$$f_{Y|X,Z}(y|x, z, \tau_{K-1}, \alpha_K, \sigma_K^2) = \Phi_{\rho_{\alpha_z}(x), \sigma_z^2}(y),$$

où Φ_{μ, σ^2} désigne la densité de probabilité de la distribution gaussienne de moyenne μ et de variance σ^2 .

Nous modélisons la distribution de Z sachant X de la façon suivante. Si $2 \leq k \leq K-1$ alors

$$f_{Z|X}(k|x, \tau_{K-1}) = \begin{cases} 1 & \text{si} & x \in [\tau_{k-1} + e, \tau_k - e[\\ 0 & \text{si} & x < \tau_{k-1} - e \text{ ou } x > \tau_k + e \\ \kappa_+(x) & \text{si} & x \in [\tau_{k-1} - e, \tau_{k-1} + e[\\ \kappa_-(x) & \text{sinon,} \end{cases} \quad (2.3)$$

où $e \geq 0$ doit être suffisamment petit pour que toutes les probabilités conditionnelles soient bien définies, κ_+ est une fonction croissante et $\kappa_+(x) = 1 - \kappa_-(x)$. Aux bords, la définition est un peu différente. Si $k = 1$,

$$f_{Z|X}(1|x, \tau_{K-1}) = \begin{cases} 1 & \text{si} & x < \tau_1 - e \\ 0 & \text{si} & x \geq \tau_1 + e \\ \kappa_-(x) & \text{sinon.} \end{cases} \quad (2.4)$$

Si $k = K$,

$$f_{Z|X}(K|x, \tau_{K-1}) = \begin{cases} 0 & \text{si} & x < \tau_{K-1} - e \\ 1 & \text{si} & x \geq \tau_{K-1} + e \\ \kappa_+(x) & \text{sinon.} \end{cases} \quad (2.5)$$

La définition de la fonction $f_{Z|X}(z|x, \tau_{K-1})$ selon les équations (2.3), (2.4) et (2.5) assure sa continuité sur $[\tau_0, \tau_K]$. Dans la suite, nous proposons de prendre $\kappa_-(x) = \left(1 - \left[\frac{x - (\tau_k - e)}{2e}\right]^2\right)$. Pour cette valeur de $\kappa_-(x)$, la figure 2.1 représente la loi de Z sachant X autour de τ_1 . Cette définition de $f_{Z|X}(z|x, \tau_{K-1})$ permet d'apporter de la flexibilité au modèle en ce qui concerne les données proches des instants de transition τ_{K-1} , qui ont désormais une probabilité non-nulle d'appartenir à chacun des régimes $k-1$ et k autour de l'instant de transition τ_k . Cette flexibilité a pour but de limiter les erreurs d'estimations associées à la présence d'une ou plusieurs observations dans un mauvais régime, lorsque l'instant τ_k est mal estimé. Il est à noter que si $e = 0$, la variable Z devient déterministe et le modèle correspond à celui introduit précédemment.

Nous supposons que nous observons des données indépendantes $(X_i, Y_i)_{1 \leq i \leq n}$ de ce modèle de régression paramétrique et nous nous intéressons à l'estimation des quantités inconnues α_K , τ_{K-1} et σ_K^2 par maximum de vraisemblance sous contrainte de régularité $\mathcal{A}_c(\tau_K)$ ou $\mathcal{A}_d(\tau_K)$.

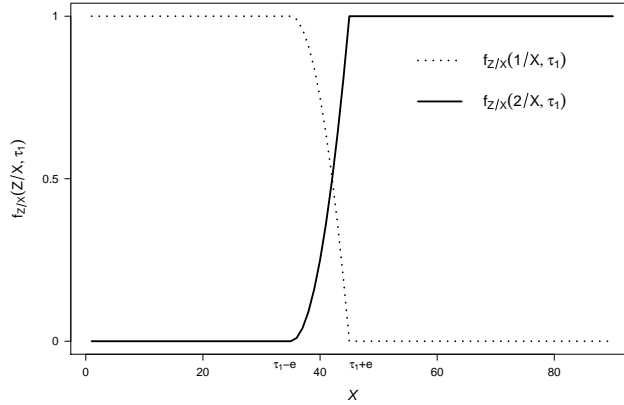


FIGURE 2.1 – Exemple de loi de Z sachant X autour de τ_1 .

Log-vraisemblances et maximisation

En présence de variables latentes, le problème de maximisation de la log-vraisemblance est complexe. Nous explicitons ici les différents objets mathématiques nécessaires pour l'estimation des paramètres. En effet, nous distinguons, la log-vraisemblance des données observées $(X_i, Y_i)_{1 \leq i \leq n}$ de celle des données complétées $(X_i, Y_i, Z_i)_{1 \leq i \leq n}$.

Par souci de lisibilité, nous posons $\boldsymbol{\theta}_K = (\mathbf{a}_K, \mathbf{S}_K^2)$ et nous notons $\tilde{\mathcal{L}}(\mathbf{T}_{K-1}, \boldsymbol{\theta}_K)$ la log-vraisemblance des observations $(X_i, Y_i)_{1 \leq i \leq n}$,

$$\tilde{\mathcal{L}}(\mathbf{T}_{K-1}, \boldsymbol{\theta}_K) = \sum_{i=1}^n \log [f_{Y|X}(Y_i|X_i, \mathbf{T}_{K-1}, \boldsymbol{\theta}_K)], \quad (2.6)$$

où $f_{Y|X}(Y_i|X_i, \mathbf{T}_{K-1}, \boldsymbol{\theta}_K)$ correspond à la vraisemblance de l'observation (X_i, Y_i) dans le modèle à variable latente,

$$f_{Y|X}(Y_i|X_i, \mathbf{T}_{K-1}, \boldsymbol{\theta}_K) = \sum_{z=1}^K f_{Y|X,Z}(Y_i|X_i, z, \mathbf{T}_{K-1}, \boldsymbol{\theta}_K) f_{Z|X}(z|X_i, \mathbf{T}_{K-1}).$$

Nous notons $\mathcal{L}_c(\mathbf{T}_{K-1}, \boldsymbol{\theta}_K)$ la log-vraisemblance complète, qui correspond à la log-vraisemblance des données observées et non observées $(X_i, Y_i, Z_i)_{1 \leq i \leq n}$,

$$\mathcal{L}_c(\mathbf{T}_{K-1}, \boldsymbol{\theta}_K) = \sum_{i=1}^n \log [f_{Y,Z|X}(Y_i, Z_i|X_i, \mathbf{T}_{K-1}, \boldsymbol{\theta}_K)], \quad (2.7)$$

où $f_{Y,Z|X}(Y_i, Z_i|X_i, \mathbf{T}_{K-1}, \boldsymbol{\theta}_K)$ se calcule ainsi,

$$f_{Y,Z|X}(Y_i, Z_i|X_i, \mathbf{T}_{K-1}, \boldsymbol{\theta}_K) = f_{Z|X,Y}(Z_i|X_i, Y_i, \mathbf{T}_{K-1}, \boldsymbol{\theta}_K) f_{Y|X}(Y_i|X_i, \mathbf{T}_{K-1}, \boldsymbol{\theta}_K).$$

L'espérance conditionnelle de la log-vraisemblance complète sachant les données observées Y s'écrit

$$\mathbf{E}[\mathcal{L}_c(\mathbf{T}_{K-1}, \boldsymbol{\theta}_K)|Y] = \sum_{i=1}^n \sum_{z=1}^K \log [f_{Y,Z|X}(Y_i, z|X_i, \mathbf{T}_{K-1}, \boldsymbol{\theta}_K)] f_{Z|X,Y}(z|X_i, Y_i, \mathbf{T}_{K-1}, \boldsymbol{\theta}_K),$$

ce qui revient à calculer la valeur moyenne de $\log [f_{Y,Z|X}(Y_i, z|X_i, \mathbf{T}_{K-1}, \boldsymbol{\theta}_K)]$ avec la pondération $f_{Z|X,Y}(z|X_i, Y_i, \mathbf{T}_{K-1}, \boldsymbol{\theta}_K)$. On peut prendre des paramètres différents pour ces deux termes : $\boldsymbol{\theta}_K$ pour le premier et $\boldsymbol{\theta}'_K$ pour le second. Alors, si on pose

$$\psi(\mathbf{T}_{K-1}, \boldsymbol{\theta}_K, \boldsymbol{\theta}'_K) = \sum_{i=1}^n \sum_{z=1}^K \log [f_{Y,Z|X}(Y_i, z|X_i, \mathbf{T}_{K-1}, \boldsymbol{\theta}_K)] f_{Z|X,Y}(z|X_i, Y_i, \mathbf{T}_{K-1}, \boldsymbol{\theta}'_K), \quad (2.8)$$

$\psi(\mathbf{T}_{K-1}, \boldsymbol{\theta}_K, \boldsymbol{\theta}'_K)$ s'interprète comme l'espérance conditionnelle de la log-vraisemblance $\mathcal{L}_c(\mathbf{T}_{K-1}, \boldsymbol{\theta}_K)$ sachant les données observées Y et la valeur du paramètre $\boldsymbol{\theta}'_K$, ce qu'on peut noter

$$\psi(\mathbf{T}_{K-1}, \boldsymbol{\theta}_K, \boldsymbol{\theta}'_K) = \mathbf{E}[\mathcal{L}_c(\mathbf{T}_{K-1}, \boldsymbol{\theta}_K) | Y, \boldsymbol{\theta}'_K].$$

On a le résultat important suivant.

Proposition 2. Si $\boldsymbol{\theta}_K^*$ maximise $\psi(\mathbf{T}_{K-1}, \cdot, \boldsymbol{\theta}'_K)$ alors,

$$\widetilde{\mathcal{L}}(\mathbf{T}_{K-1}, \boldsymbol{\theta}_K^*) \geq \widetilde{\mathcal{L}}(\mathbf{T}_{K-1}, \boldsymbol{\theta}_K).$$

Démonstration. En utilisant le fait que $f_{Y,Z|X}(Y, Z|X, \mathbf{T}_{K-1}, \boldsymbol{\theta}_K) = f_{Z|X,Y}(Z|X, Y, \mathbf{T}_{K-1}, \boldsymbol{\theta}_K) f_{Y|X}(Y|X, \mathbf{T}_{K-1}, \boldsymbol{\theta}_K)$, on peut réécrire $\widetilde{\mathcal{L}}(\mathbf{T}_{K-1}, \boldsymbol{\theta}_K)$ sous la forme suivante,

$$\widetilde{\mathcal{L}}(\mathbf{T}_{K-1}, \boldsymbol{\theta}_K) = \mathcal{L}_c(\mathbf{T}_{K-1}, \boldsymbol{\theta}_K) - \sum_{i=1}^n \log[f_{Z|X,Y}(Z_i|X_i, Y_i, \mathbf{T}_{K-1}, \boldsymbol{\theta}_K)].$$

En prenant l'espérance conditionnelle sachant Y et $\boldsymbol{\theta}'_K$, il est facile de montrer que

$$\widetilde{\mathcal{L}}(\mathbf{T}_{K-1}, \boldsymbol{\theta}_K) = \psi(\mathbf{T}_{K-1}, \boldsymbol{\theta}_K, \boldsymbol{\theta}'_K) - \mathbf{E} \left[\sum_{i=1}^n \log[f_{Z|X,Y}(Z_i|X_i, Y_i, \mathbf{T}_{K-1}, \boldsymbol{\theta}_K)] \middle| Y, \boldsymbol{\theta}'_K \right],$$

car $\widetilde{\mathcal{L}}(\mathbf{T}_{K-1}, \boldsymbol{\theta}_K)$ ne dépend pas de Z . Ainsi,

$$\begin{aligned} \widetilde{\mathcal{L}}(\mathbf{T}_{K-1}, \boldsymbol{\theta}_K) - \widetilde{\mathcal{L}}(\mathbf{T}_{K-1}, \boldsymbol{\theta}'_K) &= \psi(\mathbf{T}_{K-1}, \boldsymbol{\theta}_K, \boldsymbol{\theta}'_K) - \psi(\mathbf{T}_{K-1}, \boldsymbol{\theta}'_K, \boldsymbol{\theta}'_K) \\ &\quad + \mathbf{E} \left[\sum_{i=1}^n \log \left[\frac{f_{Z|X,Y}(Z_i|X_i, Y_i, \mathbf{T}_{K-1}, \boldsymbol{\theta}_K)}{f_{Z|X,Y}(Z_i|X_i, Y_i, \mathbf{T}_{K-1}, \boldsymbol{\theta}'_K)} \right] \middle| Y, \boldsymbol{\theta}'_K \right] \\ &\geq \psi(\mathbf{T}_{K-1}, \boldsymbol{\theta}_K, \boldsymbol{\theta}'_K) - \psi(\mathbf{T}_{K-1}, \boldsymbol{\theta}'_K, \boldsymbol{\theta}'_K), \end{aligned}$$

en vertu de l'inégalité de Jensen. Le résultat de la proposition est donc démontré. \square

Ce résultat fournit une stratégie de maximisation de la log-vraisemblance. Maximiser $\psi(\mathbf{T}_{K-1}, \boldsymbol{\theta}_K, \boldsymbol{\theta}'_K)$ par rapport à $\boldsymbol{\theta}_K$ permet de faire croître la log-vraisemblance.

2.3 Maximum de vraisemblance à deux régimes

Dans cette section, le nombre de régimes est fixé à $K = 2$ et nous souhaitons estimer les paramètres inconnus $\boldsymbol{\alpha}_2$, τ_1 et $\boldsymbol{\sigma}_2^2$. Pour ce faire, nous proposons deux méthodes. La méthode de la section 2.3.1 propose un algorithme pour déterminer l'estimateur de $(\tau_1, \boldsymbol{\alpha}_2, \boldsymbol{\sigma}_2^2)$, noté $(T_1, \mathbf{a}_2, \mathbf{S}_2^2)$, par maximum de vraisemblance sous la contrainte de régularité $\mathcal{A}_c(\tau_2)$ ou $\mathcal{A}_d(\tau_2)$. Celle de la section 2.3.2 estime les paramètres du modèle à variable latente à l'aide d'un algorithme EM. Dans la première méthode, le problème de maximisation de la log-vraisemblance se décompose de la façon suivante,

$$\max_{\substack{T_1, \mathbf{a}_2, \mathbf{S}_2^2 \\ \text{sous } \mathcal{A}_*(\tau_2)}} \mathcal{L}(T_1, \mathbf{a}_2, \mathbf{S}_2^2) = \max_{T_1} \max_{\substack{\mathbf{a}_2, \mathbf{S}_2^2 \\ \text{sous } \mathcal{A}_*(\tau_2)}} \mathcal{L}(T_1, \mathbf{a}_2, \mathbf{S}_2^2), \quad (2.9)$$

où $\mathcal{L}(T_1, \mathbf{a}_2, \mathbf{S}_2^2)$ est la log-vraisemblance du modèle sans variable latente de la section 2.2.1. Dans la seconde méthode nous résoudrons le problème de maximisation suivant,

$$\max_{\substack{T_1, \mathbf{a}_2, \mathbf{S}_2^2 \\ \text{sous } \mathcal{A}_*(\tau_K)}} \widetilde{\mathcal{L}}(T_1, \mathbf{a}_2, \mathbf{S}_2^2),$$

avec $\widetilde{\mathcal{L}}(T_1, \mathbf{a}_2, \mathbf{S}_2^2)$ la log-vraisemblance du modèle à variable latente de la section 2.2.4. Ce problème se décompose de la même façon que dans l'équation (2.9).

2.3.1 Modèle sans variable latente

Nous présentons une première méthode pour estimer les paramètres inconnus τ_1 , α_2 et σ_2^2 par maximum de vraisemblance sous l'hypothèse de régularité $\mathcal{A}_c(\tau_2)$ ou $\mathcal{A}_d(\tau_2)$. Comme détaillé dans l'équation (2.9), le problème de maximisation de la log-vraisemblance se décompose en deux sous-problèmes d'optimisation : la détermination de $(\mathbf{a}_2, \mathbf{S}_2^2)$ à T_1 fixé, qui fait l'objet des deux prochaines sections, puis la détermination de T_1 .

Propriété du maximum de vraisemblance en \mathbf{a}_2 et \mathbf{S}_2^2

Nous supposons que l'instant de transition T_1 est fixé (pas nécessairement à sa vraie valeur τ_1 qui est inconnue). Notre objectif est de maximiser la log-vraisemblance (2.2) en $(\mathbf{a}_2, \mathbf{S}_2^2)$ sous la contrainte de régularité $\mathcal{A}_c(\tau_2)$ ou $\mathcal{A}_d(\tau_2)$. Nous nous proposons de résoudre ce problème d'optimisation par la méthode des multiplicateurs de Lagrange. Dans la suite, nous introduisons des notations et objets mathématiques utiles pour expliciter nos algorithmes d'optimisation.

Le moment d'ordre r des observations X_i avant T_1 (après T_1 respectivement) s'écrit μ_1^r (μ_2^r respectivement),

$$\mu_1^r = \sum_{i=1}^n X_i^r \mathbb{1}_{\{X_i < T_1\}} \quad \text{et} \quad \mu_2^r = \sum_{i=1}^n X_i^r \mathbb{1}_{\{X_i \geq T_1\}}. \quad (2.10)$$

De même, nous définissons v_1^r et v_2^r de la façon suivante,

$$v_1^r = \sum_{i=1}^n X_i^r Y_i \mathbb{1}_{\{X_i < T_1\}} \quad \text{et} \quad v_2^r = \sum_{i=1}^n X_i^r Y_i \mathbb{1}_{\{X_i \geq T_1\}}. \quad (2.11)$$

À partir de ces notations, nous pouvons introduire la matrice carrée M de taille $2(\delta + 1)$ définie par

$$M = [w_1^0, w_2^0, \dots, w_1^\delta, w_2^\delta],$$

où pour $0 \leq b \leq \delta$,

$$\begin{aligned} w_1^b &= (\mu_1^{2\delta-b}, 0, \mu_1^{2\delta-b-1}, 0, \dots, \mu_1^{\delta-b}, 0)^t, \\ w_2^b &= (0, \mu_2^{2\delta-b}, 0, \mu_2^{2\delta-b-1}, \dots, 0, \mu_2^{\delta-b})^t, \end{aligned}$$

ainsi que le vecteur γ de taille $2(\delta + 1)$,

$$\gamma = (v_1^\delta, v_2^\delta, \dots, v_1^0, v_2^0)^t.$$

De plus, nous définissons les vecteurs $u_c(x, y)$, $x, y > 0$, et v_c , de dimension $2(\delta + 1)$, par

$$\begin{aligned} u_c(x, y) &= (xT_1^\delta, -yT_1^\delta, xT_1^{\delta-1}, -yT_1^{\delta-1}, \dots, x, y)^t, \\ v_c &= (T_1^\delta, -T_1^\delta, T_1^{\delta-1}, -T_1^{\delta-1}, \dots, 1, -1), \end{aligned}$$

où c représente la contrainte de continuité $\mathcal{A}_c(\tau_2)$. Nous définissons également

$$\begin{aligned} u_d(x, y) &= \frac{\partial u_c(x, y)}{\partial T_1} = (\delta x T_1^{\delta-1}, -\delta y T_1^{\delta-1}, (\delta-1)x T_1^{\delta-2}, -(\delta-1)y T_1^{\delta-2}, \dots, 0, 0)^t, \\ v_d &= \frac{\partial v_c}{\partial T_1} = (\delta T_1^{\delta-1}, -\delta T_1^{\delta-1}, (\delta-1)T_1^{\delta-2}, -(\delta-1)T_1^{\delta-2}, \dots, 0, 0), \end{aligned}$$

où d représente la contrainte de dérivabilité $\mathcal{A}_d(\tau_2)$. Enfin, nous définissons les matrices carrées $M_c(x, y)$ et $M_d(x, y)$ de tailles respectives $2(\delta + 1) + 1$ et $2(\delta + 1) + 2$ par

$$M_c(x, y) = \begin{pmatrix} v_c & 0 \\ M & u_c(x, y) \end{pmatrix} \quad \text{et} \quad M_d(x, y) = \begin{pmatrix} v_c & 0 & 0 \\ v_d & 0 & 0 \\ M & u_c(x, y) & u_d(x, y) \end{pmatrix}, \quad (2.12)$$

ainsi que les vecteurs γ_c et γ_d de tailles respectives $2(\delta + 1) + 1$ et $2(\delta + 1) + 2$ par

$$\gamma_c = \begin{pmatrix} 0 \\ \gamma \end{pmatrix} \quad \text{et} \quad \gamma_d = \begin{pmatrix} 0 \\ 0 \\ \gamma \end{pmatrix}. \quad (2.13)$$

Pour tout $a \in \mathbb{R}^{\delta+1}$, nous notons $V_{T_1}(1, a)$ ($V_{T_1}(2, a)$ respectivement) l'estimateur empirique de la variance dans le régime 1 (dans le régime 2 respectivement). Ils sont définis par

$$V_{T_1}(1, a) = \frac{\sum_{i=1}^n (Y_i - \rho_a(X_i))^2 \mathbb{1}_{\{X_i < T_1\}}}{\sum_{i=1}^n \mathbb{1}_{\{X_i < T_1\}}} \quad \text{et} \quad V_{T_1}(2, a) = \frac{\sum_{i=1}^n (Y_i - \rho_a(X_i))^2 \mathbb{1}_{\{X_i \geq T_1\}}}{\sum_{i=1}^n \mathbb{1}_{\{X_i \geq T_1\}}}. \quad (2.14)$$

La méthode des multiplicateurs de Lagrange conduit à la proposition suivante, qui est démontrée en annexe A.5.1.

Proposition 3. *En supposant l'instant de transition T_1 fixé (pas nécessairement à sa valeur optimale τ_1), l'optimum $(\mathbf{a}_2, \mathbf{S}_2^2)$ du problème d'optimisation (2.9) sous la contrainte de régularité $\mathcal{A}_*(\tau_2)$ avec $* \in \{c, d\}$, vérifie le système*

$$\begin{cases} M_*(\mathbf{S}_2^2)(\lambda, \mathbf{a}_2)^t & = & \gamma_*, \\ \mathbf{S}_2^2 & = & \begin{pmatrix} V_{T_1}(1, a_1) \\ V_{T_1}(2, a_2) \end{pmatrix}, \end{cases} \quad (2.15)$$

où $\lambda \in \mathbb{R}$ si $* = c$ ou $\lambda \in \mathbb{R}^2$ si $* = d$ est le multiplicateur de Lagrange.

Algorithme d'optimisation en \mathbf{a}_2 et \mathbf{S}_2^2

De la proposition 3, nous déduisons qu'à \mathbf{S}_2^2 fixé et sous la contrainte de régularité $\mathcal{A}_c(\tau_2)$ ou $\mathcal{A}_d(\tau_2)$, l'argument maximum sous contrainte \mathbf{a}_2 s'obtient en inversant la première ligne de l'équation (2.15) en \mathbf{a}_2 . De même, à \mathbf{a}_2 fixé et sous la contrainte de régularité $\mathcal{A}_c(\tau_2)$ ou $\mathcal{A}_d(\tau_2)$, l'argument maximum sous contrainte \mathbf{S}_2^2 s'obtient par les formules de l'équation (2.14). Nous proposons donc un algorithme de point fixe, qui fixera alternativement \mathbf{a}_2 et \mathbf{S}_2^2 , pour approcher l'optimum sous contrainte de régularité $\mathcal{A}_*(\tau_2)$ avec $* \in \{c, d\}$. Cet algorithme est itératif et le passage de l'étape (i) à l'étape $(i + 1)$, s'écrit de la façon suivante,

$$\begin{cases} \left(\lambda^{(i+1)}, \mathbf{a}_2^{(i+1)} \right)^t & = & M^* \left(\mathbf{S}_2^{2(i)} \right)^{-1} \gamma_*, \\ \mathbf{S}_2^{2(i+1)} & = & \begin{pmatrix} V_{T_1}(1, a_1^{(i+1)}) \\ V_{T_1}(2, a_2^{(i+1)}) \end{pmatrix}. \end{cases}$$

Nous avons choisi d'arrêter l'algorithme lorsque la valeur absolue des écarts entre les paramètres estimés à l'itération $(i + 1)$ et à l'itération (i) est inférieure à un seuil fixé. Ce critère d'arrêt sera toujours le même dans la suite du chapitre.

Résolution du problème global

L'instant de transition T_1 n'est plus considéré fixé et nous souhaitons résoudre le problème d'optimisation global (2.9). Pour ce faire, nous proposons de résoudre l'optimisation en T_1 à partir du problème de maximisation suivant,

$$\max_{\substack{T_1, \mathbf{a}_2, \mathbf{S}_2^2 \\ \text{sous } \mathcal{A}_*(\tau_2)}} \mathcal{L}(T_1, \mathbf{a}_2, \mathbf{S}_2^2) = \max_{T_1} \mathcal{L}(T_1, \mathbf{a}_2^*(T_1), \mathbf{S}_2^{2*}(T_1)), \quad (2.16)$$

où $(\mathbf{a}_2^*(T_1), \mathbf{S}_2^{2*}(T_1))$ est l'optimum de

$$\max_{\substack{\mathbf{a}_2, \mathbf{S}_2^2 \\ \text{sous } \mathcal{A}_*(\tau_2)}} \mathcal{L}(T_1, \mathbf{a}_2, \mathbf{S}_2^2),$$

obtenu par l'algorithme de point fixe de la section 2.3.1, pour une valeur de T_1 fixée.

Nous proposons de traiter cette optimisation en T_1 par une recherche trichotomique pour ses bonnes propriétés algorithmiques, mais une recherche exhaustive (plus coûteuse numériquement) peut être envisagée. Le détail de cette méthode se trouve en annexe A.3.1.

2.3.2 Modèle à variable latente : algorithme EM

Dans cette partie, notre objectif consiste à résoudre le problème suivant,

$$\max_{\substack{T_1, \mathbf{a}_2, \mathbf{S}_2^2 \\ \text{sous } \mathcal{A}_*(\tau_2)}} \widetilde{\mathcal{L}}(T_1, \mathbf{a}_2, \mathbf{S}_2^2) = \max_{T_1} \max_{\substack{\mathbf{a}_2, \mathbf{S}_2^2 \\ \text{sous } \mathcal{A}_*(\tau_2)}} \widetilde{\mathcal{L}}(T_1, \mathbf{a}_2, \mathbf{S}_2^2). \quad (2.17)$$

Propriété du maximum de vraisemblance en \mathbf{a}_2 et \mathbf{S}_2^2

Conditionnellement à l'instant de transition T_1 fixé (pas nécessairement à sa valeur optimale τ_1), le régime Z_i est une variable latente du sous-problème d'optimisation (2.17) en $\boldsymbol{\theta}_2 = (\mathbf{a}_2, \mathbf{S}_2^2)$. Or, en présence de variables latentes, la log-vraisemblance est difficile à calculer et il est d'usage d'utiliser un algorithme EM. Cet algorithme se base sur la proposition 2 et consiste à évaluer l'espérance conditionnelle de la log-vraisemblance des données complétées lorsque l'espérance est calculée par rapport à une autre valeur du paramètre $\boldsymbol{\theta}'_2$. Cette espérance conditionnelle, notée $\psi(T_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}'_2)$, est définie dans l'équation (2.8). Ainsi chaque itération de l'algorithme EM comporte deux phases : une phase de calcul de $\psi(T_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}'_2)$, pour une valeur courante $\boldsymbol{\theta}'_2$ du paramètre, notée E et une phase de maximisation de cette quantité par rapport à $\boldsymbol{\theta}_2$, notée M.

Le passage de l'algorithme de l'itération (l) à l'itération ($l+1$) est défini de la façon suivante :

- phase E : calcul de $\psi(T_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}'_2)$ pour la valeur courante $\boldsymbol{\theta}'_2$ du paramètre;
- phase M : estimation de $\boldsymbol{\theta}_2^{(l+1)} = (\mathbf{a}_2^{(l+1)}, \mathbf{S}_2^{2(l+1)})$ qui maximise $\psi(T_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}'_2)$ en $\boldsymbol{\theta}_2$ sous la contrainte de régularité $\mathcal{A}_c(\tau_2)$ ou $\mathcal{A}_d(\tau_2)$, c'est à dire,

$$\boldsymbol{\theta}_2^{(l+1)} = \underset{\substack{\boldsymbol{\theta}_2 \\ \text{sous } \mathcal{A}_*(\tau_2)}}{\operatorname{argmax}} \psi(T_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}'_2). \quad (2.18)$$

Grâce à la formule de $\psi(T_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}'_2)$ dans l'équation (2.8), nous avons

$$\begin{aligned} \psi(T_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}'_2) &\propto \sum_{z=1}^2 \sum_{i=1}^n f_{Z|X,Y}(z|X_i, Y_i, T_1, \boldsymbol{\theta}'_2) \log(f_{Z|X}(z|X_i, T_1)) \\ &\quad - \frac{1}{2} \sum_{i=1}^n \sum_{z=1}^2 \left[\log(\mathbf{S}_z^2) + \frac{(Y_i - \rho_{az}(X_i))^2}{\mathbf{S}_z^2} \right] f_{Z|X,Y}(z|X_i, Y_i, T_1, \boldsymbol{\theta}'_2). \end{aligned} \quad (2.19)$$

Le premier terme de l'équation (2.19) ne dépendant pas de $\boldsymbol{\theta}_2$, la maximisation de l'espérance conditionnelle se réduit à celle de son deuxième terme. Nous pouvons remarquer que ce terme ressemble fortement à la log-vraisemblance du modèle standard (sans variable latente) donné en (2.2) : les indicatrices ont été remplacées par les densités conditionnelles $f_{Z|X,Y}$. Les objets nécessaires à la compréhension de notre algorithme sont similaires à ceux de la section 2.3.1, à l'exception des quantités μ_1^r , μ_2^r , ν_1^r et ν_2^r . En effet, ces quantités définies par les équations (2.10) et (2.11) dépendent de la loi de Z sachant X et Y et se réécrivent sous la forme suivante,

$$\mu_{1,EM}^r = \sum_{i=1}^n X_i^r f_{Z|X,Y}(z|X_i, Y_i, T_1, \boldsymbol{\theta}_2^l) \quad \text{et} \quad \mu_{2,EM}^r = \sum_{i=1}^n X_i^r f_{Z|X,Y}(z|X_i, Y_i, T_1, \boldsymbol{\theta}_2^l), \quad (2.20)$$

$$\nu_{1,EM}^r = \sum_{i=1}^n X_i^r Y_i f_{Z|X,Y}(z|X_i, Y_i, T_1, \boldsymbol{\theta}_2^l) \quad \text{et} \quad \nu_{2,EM}^r = \sum_{i=1}^n X_i^r Y_i f_{Z|X,Y}(z|X_i, Y_i, T_1, \boldsymbol{\theta}_2^l), \quad (2.21)$$

où θ_2^l correspond à la valeur courante du paramètre à l'itération $(l) \rightarrow (l+1)$ de l'algorithme EM.

Dans la suite, nous notons $M_{*,EM}(x, y)$ et $\gamma_{*,EM}$, avec $*$ $\in \{c, d\}$, les matrices et vecteurs définis par les équations (2.12) et (2.13) en remplaçant les μ_i^r (respectivement v_i^r) par $\mu_{i,EM}^r$ (respectivement $v_{i,EM}^r$). Pour tout $a \in \mathbb{R}^{\delta+1}$, $V_{T_1,EM}(1, a)$ ($V_{T_1,EM}(2, a)$ respectivement) est l'estimateur empirique de la variance dans le régime 1 (dans le régime 2 respectivement) défini par,

$$V_{T_1,EM}(1, a) = \frac{\sum_{i=1}^n (Y_i - \rho_a(X_i))^2 f_{Z|X,Y}(z|X_i, Y_i, T_1, a_1^l, S_1^{2l})}{\sum_{i=1}^n f_{Z|X,Y}(z|X_i, Y_i, T_1, a_1^l, S_1^{2l})},$$

et

$$V_{T_1,EM}(2, a) = \frac{\sum_{i=1}^n (Y_i - \rho_a(X_i))^2 f_{Z|X,Y}(z|X_i, Y_i, T_1, a_2^l, S_2^{2l})}{\sum_{i=1}^n f_{Z|X,Y}(z|X_i, Y_i, T_1, a_2^l, S_2^{2l})}.$$

Comme précédemment, le problème d'optimisation sous la contrainte de régularité $\mathcal{A}_c(\tau_2)$ ou $\mathcal{A}_d(\tau_2)$ est en partie linéaire. Ce résultat est énoncé dans la proposition 4, démontrée en annexe A.6.

Proposition 4. *En supposant T_1 fixé (pas nécessairement à sa valeur optimale τ_1), l'optimum $(\mathbf{a}_2, \mathbf{S}_2^2)$ du problème d'optimisation (2.18) sous la contrainte de régularité $\mathcal{A}_*(\tau_2)$ avec $*$ $\in \{c, d\}$, vérifie le système*

$$\begin{cases} M_{*,EM}(\mathbf{S}_2^2)(\lambda, \mathbf{a}_2)^t & = \gamma_{*,EM}, \\ \mathbf{S}_2^2 & = \begin{pmatrix} V_{T_1,EM}(1, a_1) \\ V_{T_1,EM}(2, a_2) \end{pmatrix}, \end{cases} \quad (2.22)$$

où $\lambda \in \mathbb{R}$ si $*$ = c ou $\lambda \in \mathbb{R}^2$ si $*$ = d est le multiplicateur de Lagrange.

Algorithme d'optimisation en \mathbf{a}_2 et \mathbf{S}_2^2

La proposition 4 permet de définir un algorithme de point fixe d'approximation de l'optimum sous la contrainte de régularité $\mathcal{A}_c(\tau_2)$ ou $\mathcal{A}_d(\tau_2)$. Cet algorithme est appelé à chaque itération de l'algorithme EM. Au sein de l'itération $(l) \rightarrow (l+1)$ de l'algorithme EM, le passage de l'algorithme de point fixe de l'étape (i) à l'étape $(i+1)$ s'écrit de la façon suivante,

$$\begin{cases} \left(\lambda^{(i+1)}, \mathbf{a}_2^{(l+1),(i+1)} \right)^t & = M_{*,EM}(\mathbf{S}_2^{2(l+1),(i)})^{-1} \gamma_{*,EM}, \\ \mathbf{S}_2^{2(l+1),(i+1)} & = \begin{pmatrix} V_{T_1,EM}(1, a_1^{(l+1),(i+1)}) \\ V_{T_1,EM}(2, a_2^{(l+1),(i+1)}) \end{pmatrix}. \end{cases} \quad (2.23)$$

La proposition 2 assure que $\widetilde{\mathcal{L}}(T_1, \theta_2^{l+1}) \geq \widetilde{\mathcal{L}}(T_1, \theta_2^l)$ de sorte que la suite des log-vraisemblances $\widetilde{\mathcal{L}}(T_1, \theta_2^l)_{1 \leq l \leq n}$ est croissante et converge vers un maximum local de $\widetilde{\mathcal{L}}(T_1, \cdot)$.

Résolution du problème global

L'instant T_1 n'est plus considéré fixé et nous souhaitons résoudre le problème d'optimisation global défini dans l'équation (2.17). En présence de variables latentes, nous ne sommes pas en mesure de pouvoir résoudre ce problème d'optimisation. Nous pouvons toutefois l'approcher par le problème de maximisation suivant,

$$\max_{\substack{\mathbf{a}_2, T_1, \mathbf{S}_2^2 \\ \text{sous } \mathcal{A}_*(\tau_2)}} \widetilde{\mathcal{L}}(T_1, \mathbf{a}_2, \mathbf{S}_2^2) \approx \max_{T_1} \widetilde{\mathcal{L}}(T_1, \mathbf{a}_2^*(T_1), \mathbf{S}_2^{2*}(T_1)). \quad (2.24)$$

Il s'agit bien d'une approximation car $(\mathbf{a}_2^*(T_1), \mathbf{S}_2^{2*}(T_1))$ a été obtenu par un algorithme d'approximation d'optimisation de la log-vraisemblance de type EM (présenté ci-dessus) pour une valeur de T_1 fixée.

La résolution du problème d'optimisation (2.24) en T_1 est de nouveau faite par une recherche trichotomique. Dans la suite, nous généralisons ces deux méthodes d'inférence à un nombre quelconque de régimes.

2.4 Généralisation à un nombre quelconque de régimes

Dans cette section, nous nous plaçons dans le contexte où le nombre de régimes K est quelconque et nous souhaitons estimer les paramètres inconnus \mathbf{a}_K , τ_{K-1} et σ_K^2 par maximum de vraisemblance sous l'hypothèse de régularité $\mathcal{A}_c(\tau_K)$ ou $\mathcal{A}_d(\tau_K)$. Pour ce faire, nous proposons une méthode d'inférence avec ses versions standard (modèle sans variable latente) et lissée autour des instants de transition (modèle à variable latente et algorithme EM), qui sont une généralisation des méthodes à deux régimes présentées dans les sections 2.3.1 et 2.3.2. Dans la version standard, présentée à la section 2.4.1, le problème de maximisation de la log-vraisemblance se décompose ainsi,

$$\max_{\substack{\mathbf{T}_{K-1}, \mathbf{a}_K, \mathbf{S}_K^2 \\ \text{sous } \mathcal{A}_*(\tau_K)}} \mathcal{L}(\mathbf{T}_{K-1}, \mathbf{a}_K, \mathbf{S}_K^2) = \max_{\mathbf{T}_{K-1}} \max_{\substack{\mathbf{a}_K, \mathbf{S}_K^2 \\ \text{sous } \mathcal{A}_*(\tau_K)}} \mathcal{L}(\mathbf{T}_{K-1}, \mathbf{a}_K, \mathbf{S}_K^2), \quad (2.25)$$

où $\mathcal{L}(\mathbf{T}_{K-1}, \mathbf{a}_K, \mathbf{S}_K^2)$ est la log-vraisemblance du modèle sans variable latente de la section 2.2.1. Dans la version lissée autour des instants de transition, détaillée à la section 2.4.2, nous résoudrons le problème de maximisation suivant,

$$\max_{\substack{\mathbf{T}_{K-1}, \mathbf{a}_K, \mathbf{S}_K^2 \\ \text{sous } \mathcal{A}_*(\tau_K)}} \widetilde{\mathcal{L}}(\mathbf{T}_{K-1}, \mathbf{a}_K, \mathbf{S}_K^2),$$

avec $\widetilde{\mathcal{L}}(\mathbf{T}_{K-1}, \mathbf{a}_K, \mathbf{S}_K^2)$ la log-vraisemblance du modèle à variable latente de la section 2.2.4. Ce problème se décompose de la même façon que dans l'équation (2.25).

2.4.1 Modèle sans variable latente

Dans cette partie, nous généralisons la méthode à deux régimes de la section 2.3.1. Nous souhaitons résoudre le problème de maximisation de la log-vraisemblance (2.25) qui se décompose en deux sous-problèmes : l'estimation en $(\mathbf{a}_K, \mathbf{S}_K^2)$ à \mathbf{T}_{K-1} fixé, qui fait l'objet des deux prochaines sections, puis l'estimation en \mathbf{T}_{K-1} .

Propriété du maximum de vraisemblance en \mathbf{a}_K et \mathbf{S}_K^2

Lorsque le vecteur des instants de transition \mathbf{T}_{K-1} est fixé (pas nécessairement au vecteur optimal τ_{K-1} qui est inconnu), nous proposons de résoudre le sous-problème d'optimisation (2.25) en $(\mathbf{a}_K, \mathbf{S}_K^2)$ par la méthode des multiplicateurs de Lagrange. Les notations et objets mathématiques utiles pour expliciter nos algorithmes d'optimisation étant plus complexes que ceux proposés dans la section 2.3.1, nous les détaillons dans la suite.

Pour tout $1 \leq k \leq K$, nous notons μ_k^r le moment d'ordre r des observations X_i dans le régime k ,

$$\mu_k^r = \sum_{i=1}^n X_i^r \mathbb{1}_{\{T_{k-1} \leq X_i < T_k\}}, \quad (2.26)$$

avec $T_0 = \tau_0$ et $T_K = \tau_K$. De même, nous définissons v_k^r de la façon suivante,

$$v_k^r = \sum_{i=1}^n X_i^r Y_i \mathbb{1}_{\{T_{k-1} \leq X_i < T_1\}}. \quad (2.27)$$

À partir de ces notations, la matrice carrée M de taille $K(\delta + 1)$ s'écrit

$$M = [w_1^0, w_2^0, \dots, w_K^0, \dots, w_1^\delta, w_2^\delta, \dots, w_K^\delta],$$

où pour $0 \leq b \leq \delta$ et $0_j = (0, \dots, 0)^t$ de taille j ,

$$w_k^b = (0_{k-1}, \mu_k^{2\delta-b}, 0_{K-1}, \mu_k^{2\delta-b-1}, 0_{K-k}, \dots, 0_{k-1}, \mu_k^{\delta-b}, 0_{K-k})^t.$$

Le vecteur γ de taille $K(\delta + 1)$ s'écrit

$$\gamma = (v_1^\delta, v_2^\delta, \dots, v_K^\delta, \dots, v_1^0, \dots, v_K^0)^t.$$

De plus pour $1 \leq k \leq K-1$, les vecteurs $u_{c,k}(\mathbf{x})$, $\mathbf{x} = (x_1, \dots, x_K) \in \mathbb{R}_+^K$, et $v_{c,k}$, sont des vecteurs de dimension $K(\delta + 1)$, définis par

$$\begin{aligned} u_{c,k}(\mathbf{x}) &= (0_{k-1}, x_k T_k^\delta, -x_{k+1} T_k^\delta, 0_{K-2}, x_k T_k^{\delta-1}, -x_{k+1} T_k^{\delta-1}, 0_{(K-1)-k}, \dots, 0_{k-1}, x_k, -x_{k+1}, 0_{(K-1)-k})^t, \\ v_{c,k} &= (0_{k-1}, T_k^\delta, -T_k^\delta, 0_{K-2}, T_k^{\delta-1}, -T_k^{\delta-1}, 0_{(K-1)-k}, \dots, 0_{k-1}, 1, -1, 0_{(K-1)-k}), \end{aligned}$$

où c représente la contrainte de continuité $\mathcal{A}_c(\tau_K)$. Nous définissons également

$$\begin{aligned} u_{d,k}(\mathbf{x}) &= \frac{\partial u_{c,k}(\mathbf{x})}{\partial T_k}, \\ &= (0_{k-1}, \delta x_k T_k^{\delta-1}, -\delta x_{k+1} T_k^{\delta-1}, 0_{K-2}, (\delta-1)x_k T_k^{\delta-2}, -(\delta-1)x_{k+1} T_k^{\delta-2}, 0_{(K-1)-k}, \dots, 0_K)^t, \\ v_{d,k} &= \frac{\partial v_{c,k}}{\partial T_k}, \\ &= (0_{k-1}, \delta T_k^{\delta-1}, -\delta T_k^{\delta-1}, 0_{K-2}, (\delta-1)T_k^{\delta-2}, -(\delta-1)T_k^{\delta-2}, 0_{(K-1)-k}, \dots, 0_K), \end{aligned}$$

où d représente la contrainte de dérivabilité $\mathcal{A}_d(\tau_K)$. Les matrices carrées $M_c(\mathbf{x})$ et $M_d(\mathbf{x})$, de tailles respectives $K(\delta + 1) + (K-1)$ et $K(\delta + 1) + 2(K-1)$ sont définies par

$$M_c(\mathbf{x}) = \begin{pmatrix} v_{c,1} & 0 & 0 & 0 & 0 \\ v_{c,2} & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ v_{c,K-1} & 0 & 0 & 0 & 0 \\ M & u_{c,1}(\mathbf{x}) & u_{c,2}(\mathbf{x}) & \dots & u_{c,K-1}(\mathbf{x}) \end{pmatrix}, \quad (2.28)$$

et

$$M_d(\mathbf{x}) = \begin{pmatrix} v_{c,1} & 0 & 0 & 0 & 0 & 0 \\ v_{d,1} & 0 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ v_{c,K-1} & 0 & 0 & 0 & 0 & 0 \\ v_{d,K-1} & 0 & 0 & 0 & 0 & 0 \\ M & u_{c,1}(\mathbf{x}) & u_{d,1}(\mathbf{x}) & \dots & u_{c,K-1}(\mathbf{x}) & u_{d,K-1}(\mathbf{x}) \end{pmatrix}, \quad (2.29)$$

ainsi que les vecteurs γ_c et γ_d de tailles respectives $K(\delta + 1) + (K-1)$ et $K(\delta + 1) + 2(K-1)$,

$$\gamma_c = \begin{pmatrix} 0_{K-1} \\ \gamma \end{pmatrix} \quad \text{et} \quad \gamma_d = \begin{pmatrix} 0_{2(K-1)} \\ \gamma \end{pmatrix}. \quad (2.30)$$

De même que dans le cas à 2 régimes, nous définissons pour tout $a \in \mathbb{R}^{\delta+1}$ et $1 \leq k \leq K$, l'estimateur empirique de la variance dans le régime estimé k (régime défini sur $[T_{k-1}, T_k]$),

$$V_{T_k}(k, a) = \frac{\sum_{i=1}^n (Y_i - \rho_a(X_i))^2 \mathbb{1}_{\{T_{k-1} \leq X_i < T_k\}}}{\sum_{i=1}^n \mathbb{1}_{\{T_{k-1} \leq X_i < T_k\}}},$$

avec $T_0 = \tau_0$ et $T_K = \tau_K$, et nous notons $\mathbf{V}_{T_{K-1}}(\mathbf{a}_K)$ le vecteur des estimateurs empiriques de la variance de chaque régime estimé,

$$\mathbf{V}_{T_{K-1}}(\mathbf{a}_K) = (V_{T_{K-1}}(1, a_1), \dots, V_{T_{K-1}}(K, a_K))^t.$$

Le problème d'optimisation sous la contrainte de régularité $\mathcal{A}_c(\tau_K)$ ou $\mathcal{A}_d(\tau_K)$ est en partie linéaire. Ce résultat est énoncé dans la proposition 5, démontrée en annexe A.5.2.

Proposition 5. *En supposant que le vecteur d'instants de transition \mathbf{T}_{K-1} est fixé (pas nécessairement au vecteur optimal τ_{K-1}), l'optimum $(\mathbf{a}_K, \mathbf{S}_K^2)$ du problème d'optimisation (2.25) sous la contrainte $\mathcal{A}_*(\tau_K)$ avec $*$ $\in \{c, d\}$, vérifie le système*

$$\begin{cases} M_*(\mathbf{S}_K^2)(\lambda, \mathbf{a}_K)^t & = \gamma_*, \\ \mathbf{S}_K^2 & = \mathbf{V}_{T_{K-1}}(\mathbf{a}_K), \end{cases} \quad (2.31)$$

où $\lambda \in \mathbb{R}^{K-1}$ si $*$ = c ou $\lambda \in \mathbb{R}^{2(K-1)}$ si $*$ = d est le multiplicateur de Lagrange.

Algorithme d'optimisation en \mathbf{a}_K et \mathbf{S}_K^2

La proposition 5 permet de définir un algorithme de point fixe d'approximation de l'optimum sous contrainte $\mathcal{A}_*(\tau_K)$ avec $*$ $\in \{c, d\}$. Le passage de l'étape (i) à l'étape ($i+1$), s'écrit de la façon suivante,

$$\begin{cases} \left(\lambda^{(i+1)}, \mathbf{a}_K^{(i+1)} \right)^t & = M_* \left(\mathbf{S}_K^{2(i)} \right)^{-1} \gamma_*, \\ \mathbf{S}_K^{2(i+1)} & = \mathbf{V}_{T_{K-1}}(\mathbf{a}_K^{(i+1)}). \end{cases}$$

Résolution du problème global

Dans cette partie, le vecteur des instants de transition \mathbf{T}_{K-1} n'est plus considéré fixé. L'optimisation du problème global (2.25) en \mathbf{T}_{K-1} est un problème d'optimisation de dimension $K-1$ sous contraintes d'inégalités, $T_1 < T_2 < \dots < T_{K-1}$. Ce problème d'optimisation est résolu par une recherche exhaustive sur la grille des instants de mesure. Cette méthode a deux inconvénients majeurs : (i) la taille de la matrice M_c (ou M_d) à inverser pour résoudre le premier problème d'estimation est linéaire en le nombre de régimes (donc grande lorsque le nombre de régimes est grand), et (ii) cette inversion doit être réalisée à chaque appel de la log-vraisemblance à maximiser en les $K-1$ instants de transition.

2.4.2 Modèle à variable latente et algorithme EM

Dans cette section, nous souhaitons résoudre le problème d'optimisation suivant,

$$\max_{\substack{\mathbf{T}_{K-1}, \mathbf{a}_K, \mathbf{S}_K^2 \\ \text{sous } \mathcal{A}_*(\tau_K)}} \widetilde{\mathcal{L}}(\mathbf{T}_{K-1}, \mathbf{a}_K, \mathbf{S}_K^2) = \max_{\mathbf{T}_{K-1}} \max_{\substack{\mathbf{a}_K, \mathbf{S}_K^2 \\ \text{sous } \mathcal{A}_*(\tau_K)}} \widetilde{\mathcal{L}}(\mathbf{T}_{K-1}, \mathbf{a}_K, \mathbf{S}_K^2).$$

Comme détaillé dans la section 2.3.2, en présence de variable latente, nous ne sommes pas en mesure de résoudre ce problème d'optimisation. Nous pouvons cependant l'approcher par le problème de maximisation suivant,

$$\max_{\substack{\mathbf{T}_{K-1}, \mathbf{a}_K, \mathbf{S}_K^2 \\ \text{sous } \mathcal{A}_*(\tau_K)}} \widetilde{\mathcal{L}}(\mathbf{T}_{K-1}, \mathbf{a}_K, \mathbf{S}_K^2) \approx \max_{\mathbf{T}_{K-1}} \widetilde{\mathcal{L}}(\mathbf{T}_{K-1}, \mathbf{a}_K^*(\mathbf{T}_{K-1}), \mathbf{S}_K^{2*}(\mathbf{T}_{K-1})). \quad (2.32)$$

Il s'agit bien d'une approximation car $(\mathbf{a}_K^*(\mathbf{T}_{K-1}), \mathbf{S}_K^{2*}(\mathbf{T}_{K-1}))$ a été obtenu par un algorithme d'approximation d'optimisation de la log-vraisemblance de type EM à \mathbf{T}_{K-1} fixé. Nous ne détaillons pas l'optimisation en $(\mathbf{a}_K, \mathbf{S}_K^2)$ puisqu'il suffit de généraliser la section 2.3.2 à un nombre quelconque de régimes.

La résolution du problème d'optimisation (2.32) en \mathbf{T}_{K-1} est faite par une recherche exhaustive sur la grille des instants de mesure. Notons tout de même que cette méthode possède les deux mêmes inconvénients que la méthode précédente. La complexité de cet algorithme est rédhibitoire. C'est pourquoi nous avons développé une approche récursive qui fait l'objet de la section suivante.

2.5 Approche récursive : algorithme de type Bellman

Dans cette section, nous présentons une méthode d'inférence avec ses deux versions standard (modèle sans variable latente) et lissée¹. Cette méthode est inspirée de l'article [7] de Bellman et Roth, qui utilisent la programmation dynamique (méthode détaillée en annexe A.3.2) pour estimer un modèle linéaire par morceaux sous contrainte de continuité. Les différentes versions de cette méthode sont proposées dans les deux prochaines sections.

2.5.1 Modèle sans variable latente

Nous considérons que le nombre de régimes K est quelconque. Dans ce contexte, la stratégie de programmation dynamique appliquée à notre problème consiste à décomposer la log-vraisemblance (2.2) en la log-vraisemblance à droite (de T_{K-1}), notée \mathcal{L}^r et la log-vraisemblance à gauche (de T_{K-1}), notée \mathcal{L}^l ,

$$\begin{aligned} \mathcal{L}_{[\tau_0, \tau_K]}(\mathbf{T}_{K-1}, \mathbf{a}_K, \mathbf{S}_K^2) &= \mathcal{L}_{[\tau_0, T_{K-1}]}^l(\mathbf{T}_{K-1}, \mathbf{a}_K, \mathbf{S}_K^2) \\ &+ \mathcal{L}_{[T_{K-1}, \tau_K]}^r(\mathbf{T}_{K-1}, \mathbf{a}_K, \mathbf{S}_K^2). \end{aligned}$$

Nous pouvons remarquer que le premier terme \mathcal{L}^l ne dépend en réalité que de $(\mathbf{T}_{K-2}, \mathbf{a}_{K-1}, \mathbf{S}_{K-1}^2)$. De plus, le second terme \mathcal{L}^r ne dépend que de $(\mathbf{a}_K, \mathbf{S}_K^2)$. Finalement,

$$\begin{aligned} \mathcal{L}_{[\tau_0, \tau_K]}(\mathbf{T}_{K-1}, \mathbf{a}_K, \mathbf{S}_K^2) &= \mathcal{L}_{[\tau_0, T_{K-1}]}^l(\mathbf{T}_{K-2}, \mathbf{a}_{K-1}, \mathbf{S}_{K-1}^2) \\ &+ \mathcal{L}_{[T_{K-1}, \tau_K]}^r(\mathbf{a}_K, \mathbf{S}_K^2). \end{aligned} \quad (2.33)$$

Les instants de transition τ_{K-1} entre les régimes, les coefficients des polynômes \mathbf{a}_K et les variances σ_K^2 de chaque régime sont inconnus. Notre objectif consiste à les estimer en intégrant cette stratégie de programmation dynamique dans la résolution du problème de maximisation de la log-vraisemblance sous contrainte de régularité $\mathcal{A}_c(\tau_K)$ ou $\mathcal{A}_d(\tau_K)$. Pour ce faire nous commençons par décomposer ce problème de maximisation global en deux sous-problèmes d'optimisation,

$$\max_{\substack{\mathbf{T}_{K-1}, \mathbf{a}_K, \mathbf{S}_K^2 \\ \text{sous } \mathcal{A}_*(\tau_K)}} \mathcal{L}_{[\tau_0, \tau_K]}(\mathbf{T}_{K-1}, \mathbf{a}_K, \mathbf{S}_K^2) = \max_{\mathbf{T}_{K-1}} \max_{\substack{\mathbf{a}_K, \mathbf{S}_K^2 \\ \text{sous } \mathcal{A}_*(\tau_K)}} \mathcal{L}_{[\tau_0, \tau_K]}(\mathbf{T}_{K-1}, \mathbf{a}_K, \mathbf{S}_K^2). \quad (2.34)$$

Décomposition de la log-vraisemblance

Dans un premier temps, nous considérons que les instants de transition T_{K-1} entre les régimes sont fixés (pas nécessairement à τ_{K-1}). Nous allons résoudre le sous-problème de maximisation (2.34) en $(\mathbf{a}_K, \mathbf{S}_K^2)$ sous la contrainte de régularité $\mathcal{A}_c(\tau_K)$ ou $\mathcal{A}_d(\tau_K)$ en utilisant la stratégie de programmation dynamique définie dans l'équation (2.33). Cependant, sous cette contrainte, les log-vraisemblances à gauche et à droite de T_{K-1} sont liées, ce qui nous empêche de séparer le sous-problème de maximisation en $(\mathbf{a}_K, \mathbf{S}_K^2)$ en deux nouveaux sous-problèmes indépendants : la maximisation de la log-vraisemblance à gauche de T_{K-1} (qui dépend des paramètres des $K-1$ régimes à gauche de T_{K-1}) et la maximisation de la log-vraisemblance à droite de T_{K-1} (qui ne dépend que des paramètres du régime K situé à droite de T_{K-1}).

Nous nous intéressons d'abord au problème de maximisation sous contrainte de continuité $\mathcal{A}_c(\tau_K)$. Dans le but de décomposer le problème d'optimisation en $(\mathbf{a}_K, \mathbf{S}_K^2)$ sous cette contrainte, nous introduisons dans l'équation suivante une variable β correspondant à la valeur prise par la fonction

¹. Nous verrons que l'algorithme de type Bellman ne s'applique pas au modèle à variable latente. Nous proposerons donc une version lissée du modèle standard.

de lien φ de l'équation (2.1) en l'instant de transition T_{K-1} , le problème s'écrit alors,

$$\begin{aligned} \max_{\substack{\mathbf{a}_K, \mathbf{S}_K^2 \\ \text{sous } \mathcal{A}_c(\tau_K)}} \mathcal{L}_{[\tau_0, \tau_K]}(\mathbf{T}_{K-1}, \mathbf{a}_K, \mathbf{S}_K^2) &= \max_{\beta} \left(\underbrace{\max_{\substack{\mathbf{a}_{K-1}, \mathbf{S}_{K-1}^2 \\ \text{sous } \mathcal{A}_c(\mathbf{T}_{K-1}) \\ \text{et } \rho_{K-1}(\mathbf{T}_{K-1})=\beta}} \mathcal{L}_{[\tau_0, \mathbf{T}_{K-1}[}^l(\mathbf{T}_{K-2}, \mathbf{a}_{K-1}, \mathbf{S}_{K-1}^2)} \right. \\ &\quad \left. + \underbrace{\max_{\substack{a_K, S_K^2 \\ \text{sous } \rho_K(\mathbf{T}_{K-1})=\beta}} \mathcal{L}_{[\mathbf{T}_{K-1}, \tau_K]}^r(a_K, S_K^2)} \right). \end{aligned} \quad (2.35)$$

L'optimisation en β complexifie la résolution de ce problème d'optimisation en lui ajoutant une dimension. Cela sera d'autant plus vrai si l'on poursuit la décomposition de manière récursive, car dans ce cas, il faudra optimiser plusieurs variables β qui correspondront aux valeurs prises par la fonction de lien en les différents instants de transition \mathbf{T}_{K-1} . Pour diminuer la dimension de ce problème, nous proposons d'estimer l'optimum en β par un estimateur non-paramétrique de la fonction de lien φ de type Parzen-Rosenblatt (d'après les articles [42, 47] de Parzen et Rosenblatt), que nous notons \hat{g} . Cette idée vient de l'article [7] de Bellman et Roth dans lequel, pour la régression linéaire par morceaux sous contrainte de continuité, la fonction de lien φ est estimée par la moyenne des Y_i entre deux observations X_i .

$$\begin{aligned} \max_{\substack{\mathbf{a}_K, \mathbf{S}_K^2 \\ \text{sous } \mathcal{A}_c(\tau_K)}} \mathcal{L}_{[\tau_0, \tau_K]}(\mathbf{T}_{K-1}, \mathbf{a}_K, \mathbf{S}_K^2) &\simeq \max_{\substack{\mathbf{a}_{K-1}, \mathbf{S}_{K-1}^2 \\ \text{sous } \mathcal{A}_c(\mathbf{T}_{K-1}) \\ \text{et } \rho_{K-1}(\mathbf{T}_{K-1})=\hat{g}(\mathbf{T}_{K-1})}} \mathcal{L}_{[\tau_0, \mathbf{T}_{K-1}[}^l(\mathbf{T}_{K-2}, \mathbf{a}_{K-1}, \mathbf{S}_{K-1}^2) \\ &+ \max_{\substack{a_K, S_K^2 \\ \text{sous } \rho_K(\mathbf{T}_{K-1})=\hat{g}(\mathbf{T}_{K-1})}} \mathcal{L}_{[\mathbf{T}_{K-1}, \tau_K]}^r(a_K, S_K^2). \end{aligned}$$

La décomposition du problème de maximisation en $(\mathbf{a}_K, \mathbf{S}_K^2)$ est similaire sous contrainte de dérivabilité. Il suffit d'introduire la variable η correspondant à la valeur prise par la dérivée de la fonction de lien, notée $\dot{\varphi}(x)$, en l'instant de transition T_{K-1} . Comme pour β , nous estimons l'optimum en η par un estimateur non paramétrique de type Parzen-Rosenblatt, que l'on note $\hat{\ell}$, ce qui conduit au problème suivant,

$$\begin{aligned} \max_{\substack{\mathbf{a}_K, \mathbf{S}_K^2 \\ \text{sous } \mathcal{A}_d(\tau_K)}} \mathcal{L}_{[\tau_0, \tau_K]}(\mathbf{T}_{K-1}, \mathbf{a}_K, \mathbf{S}_K^2) &\simeq \max_{\substack{\mathbf{a}_{K-1}, \mathbf{S}_{K-1}^2 \\ \text{sous } \mathcal{A}_d(\mathbf{T}_{K-1}) \\ \text{et } \rho_{K-1}(\mathbf{T}_{K-1})=\hat{g}(\mathbf{T}_{K-1}) \\ \text{et } \dot{\rho}_{K-1}(\mathbf{T}_{K-1})=\hat{\ell}(\mathbf{T}_{K-1})}} \mathcal{L}_{[\tau_0, \mathbf{T}_{K-1}[}^l(\mathbf{T}_{K-2}, \mathbf{a}_{K-1}, \mathbf{S}_{K-1}^2) \\ &+ \max_{\substack{a_K, S_K^2 \\ \text{sous } \rho_K(\mathbf{T}_{K-1})=\hat{g}(\mathbf{T}_{K-1}) \\ \text{sous } \dot{\rho}_K(\mathbf{T}_{K-1})=\hat{\ell}(\mathbf{T}_{K-1})}} \mathcal{L}_{[\mathbf{T}_{K-1}, \tau_K]}^r(a_K, S_K^2). \end{aligned}$$

L'application récursive de la stratégie de programmation dynamique nous conduit à devoir résoudre le problème global suivant (sous contrainte de continuité),

$$\begin{aligned} \max_{\substack{\mathbf{a}_K, \mathbf{S}_K^2 \\ \text{sous } \mathcal{A}_c(\tau_K)}} \mathcal{L}_{[\tau_0, \tau_K]}(\mathbf{T}_{K-1}, \mathbf{a}_K, \mathbf{S}_K^2) &\simeq \max_{\substack{a_1, S_1^2 \\ \text{et } \rho_1(\mathbf{T}_1)=\hat{g}(\mathbf{T}_1)}} \mathcal{L}_{[\tau_0, \mathbf{T}_1[}^l(a_1, S_1^2) \\ &+ \sum_{k=2}^{K-1} \max_{\substack{a_k, S_k^2 \\ \text{sous } \rho_k(\mathbf{T}_{k-1})=\hat{g}(\mathbf{T}_{k-1}) \\ \text{et } \rho_k(\mathbf{T}_k)=\hat{g}(\mathbf{T}_k)}} \mathcal{L}_{[\mathbf{T}_{k-1}, \mathbf{T}_k[}^l(a_k, S_k^2) \\ &+ \max_{\substack{a_K, S_K^2 \\ \text{sous } \rho_K(\mathbf{T}_{K-1})=\hat{g}(\mathbf{T}_{K-1})}} \mathcal{L}_{[\mathbf{T}_{K-1}, \tau_K]}^r(\mathbf{T}_{K-1}, a_K, S_K^2). \end{aligned} \quad (2.36)$$

Algorithme d'optimisation en \mathbf{a}_K et \mathbf{S}_K^2

Nous remarquons que le problème d'optimisation global de l'équation (2.36) est décomposé en des sous-problèmes d'optimisation sur un régime, avec trois types de contraintes d'égalité : condition à droite uniquement (premier terme), à gauche uniquement (dernier terme) et aux deux bornes pour les autres termes. Comme précédemment, nous résolvons ces différents problèmes par la méthode des multiplicateurs de Lagrange. Nous donnons le détail de ces algorithmes en annexe A.4.

Résolution du problème global

Dans cette partie, le vecteur des instants de transition \mathbf{T}_{K-1} entre les régimes n'est plus considéré fixé et nous souhaitons résoudre le problème global (2.34) en l'approchant par le problème suivant (dans le cas de la contrainte de continuité),

$$\begin{aligned}
 & \max_{\substack{\mathbf{T}_{K-1}, \mathbf{a}_K, \mathbf{S}_K^2 \\ \text{sous } \mathcal{A}_c(\tau_K)}} \mathcal{L}_{[\tau_0, \tau_K]}(\mathbf{T}_{K-1}, \mathbf{a}_K, \mathbf{S}_K^2) \\
 \simeq & \max_{\mathbf{T}_{K-1}} \left[\max_{\mathbf{T}_{K-2}} \left[\underbrace{\max_{\substack{\mathbf{a}_{K-1}, \mathbf{S}_{K-1}^2 \\ \text{sous } \mathcal{A}_c(\mathbf{T}_{K-1}) \\ \text{et } \rho_{K-1}(\mathbf{T}_{K-1}) = \hat{g}(\mathbf{T}_{K-1})}} \mathcal{L}_{[\tau_0, \mathbf{T}_{K-1}]}^l(\mathbf{T}_{K-2}, \mathbf{a}_{K-1}, \mathbf{S}_{K-1}^2)} \right]}_{(K-1) \text{ régimes}} \right] \\
 & + \underbrace{\max_{\substack{\mathbf{a}_K, \mathbf{S}_K^2 \\ \text{sous } \rho_K(\mathbf{T}_{K-1}) = \hat{g}(\mathbf{T}_{K-1})}} \mathcal{L}_{[\mathbf{T}_{K-1}, \tau_K]}^r(\mathbf{a}_K, \mathbf{S}_K^2)} \Bigg]. \tag{2.37} \\
 & \hspace{10em} \text{1 régime}
 \end{aligned}$$

D'après l'équation (2.37), nous voyons que le problème d'estimation des paramètres \mathbf{a}_K , \mathbf{T}_{K-1} et \mathbf{S}_K^2 pour K régimes est équivalent à l'estimation des paramètres d'un modèle à $K-1$ régimes d'un côté et d'un modèle à 1 régime de l'autre. Ainsi, si nous avons résolu tous les problèmes à $K-1$ régimes sur les intervalles $[\tau_0, t]$ avec t défini sur la grille des instants de mesure et tous les problèmes à un régime sur $[t, \tau_K]$, la résolution du problème à K régimes sur $[\tau_0, \tau_K]$ se résume à une optimisation en t sur la grille des instants de mesure pour trouver l'instant de transition τ_{K-1} . En programmation dynamique, le stockage des résultats des sous-problèmes permet d'éviter les calculs redondants. De manière plus schématique, la résolution du problème d'estimation des paramètres pour K régimes peut être représentée selon la figure 2.2. La résolution du problème à K régimes consiste à optimiser t sur la grille des instants de mesures tout en utilisant directement les résultats obtenus lors de la résolution du problème à $K-1$ régimes sur $[\tau_0, t]$.

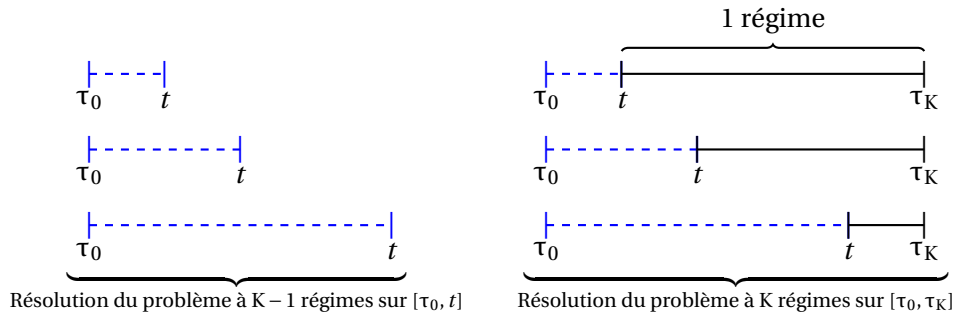


FIGURE 2.2 – Schéma de la résolution d'un problème à K régimes.

En appliquant récursivement la stratégie de programmation dynamique, l'équation (2.37) se ré-

écrit de la façon suivante,

$$\begin{aligned}
 \max_{\substack{\mathbf{a}_K, \mathbf{T}_{K-1}, \mathbf{S}_K^2 \\ \text{sous } \mathcal{A}_c(\tau_K)}} \mathcal{L}_{[\tau_0, \tau_K]}(\mathbf{T}_{K-1}, \mathbf{a}_K, \mathbf{S}_K^2) &\simeq \max_{\mathbf{T}_{K-1}} \left[\max_{\mathbf{T}_{K-2}} \left[\cdots \max_{\mathbf{T}_1} \left[\max_{\substack{a_1, S_1^2 \\ \text{et } \rho_1(\mathbf{T}_1) = \hat{g}(\mathbf{T}_1)}} \mathcal{L}_{[\tau_0, \mathbf{T}_1]}^l(a_1, S_1^2) \right. \right. \right. \\
 &+ \left. \left. \left. \max_{\substack{a_2, S_2^2 \\ \text{sous } \rho_2(\mathbf{T}_1) = \hat{g}(\mathbf{T}_1) \\ \text{et } \rho_2(\mathbf{T}_2) = \hat{g}(\mathbf{T}_2)}} \mathcal{L}_{[\mathbf{T}_1, \mathbf{T}_2]}^r(a_2, S_2^2) \right] \right. \\
 &\vdots \\
 &+ \left. \left. \left. \max_{\substack{a_{K-1}, S_{K-1}^2 \\ \text{sous } \rho_{K-1}(\mathbf{T}_{K-2}) = \hat{g}(\mathbf{T}_{K-2}) \\ \text{et } \rho_{K-1}(\mathbf{T}_{K-1}) = \hat{g}(\mathbf{T}_{K-1})}} \mathcal{L}_{[\mathbf{T}_{K-2}, \mathbf{T}_{K-1}]}^r(a_{K-1}, S_{K-1}^2) \right] \right. \\
 &+ \left. \left. \left. \max_{\substack{a_K, S_K^2 \\ \text{sous } \rho_K(\mathbf{T}_{K-1}) = \hat{g}(\mathbf{T}_{K-1})}} \mathcal{L}_{[\mathbf{T}_{K-1}, \tau_K]}^r(a_K, S_K^2) \right] \right].
 \end{aligned}$$

La maximisation indépendante des log-vraisemblances sur les différents intervalles permet de réduire considérablement sa complexité par rapport aux méthodes non récursives.

2.5.2 Lissage autour des instants de transition

L'algorithme de type Bellman ne s'applique pas au modèle à variable latente car l'algorithme EM utilise simultanément des données présentes dans plusieurs régimes, ce qui n'est pas compatible avec la décomposition droite-gauche de la log-vraisemblance. Afin d'approcher les résultats qui seraient obtenus par un algorithme EM, c'est-à-dire avec une certaine souplesse autour des instants de transition, nous introduisons une étape de lissage dans l'algorithme de type Bellman. Nous cherchons désormais $(\mathbf{a}_K, \mathbf{S}_K^2)$ qui maximise la log-vraisemblance lissée,

$$\mathcal{L}(\mathbf{T}_{K-1}, \mathbf{a}_K, \mathbf{S}_K^2) = -\frac{1}{2} \sum_{i=1}^n \sum_{z=1}^K \left[\log(S_z^2) + \frac{(Y_i - \rho_{a_z}(X_i))^2}{S_z^2} \right] f_{Z|X}(z|X_i, \mathbf{T}_{K-1}). \quad (2.38)$$

Pour tout $1 \leq k \leq K$, notons u_{k-1} et u_k les bornes du régime k tenant compte du lissage : si $2 \leq k \leq K-1$ alors $u_{k-1} = \mathbf{T}_{k-1} - e$ et $u_k = \mathbf{T}_k + e$; si $k = 1$, $u_0 = \tau_0$ et $u_1 = \mathbf{T}_1 + e$; si $k = K$, $u_{K-1} = \mathbf{T}_{K-1} - e$ et $u_K = \tau_K$. Alors la log-vraisemblance lissée du régime k , notée $\tilde{\mathcal{L}}_{[u_{k-1}, u_k]}(a_k, S_k^2)$ s'écrit

$$\tilde{\mathcal{L}}_{[u_{k-1}, u_k]}(a_k, S_k^2) = -\frac{1}{2} \sum_{\{i: X_i \in [u_{k-1}, u_k]\}} \left[\log(S_k^2) + \frac{(Y_i - \rho_{a_k}(X_i))^2}{S_k^2} \right] f_{Z|X}(k|X_i, \mathbf{T}_{K-1}),$$

où e correspond à l'écart fixé autour des instants de transition dans l'équation (2.3).

La stratégie de programmation dynamique revient à estimer les paramètres $(\mathbf{a}_K, \mathbf{S}_K^2)$ en résolvant le problème (2.39) qui scinde la log-vraisemblance lissée en K parties correspondant aux K régimes sous la contrainte de continuité $\mathcal{A}_c(\tau_K)$. Elle s'écrit

$$\begin{aligned}
 \max_{\substack{\mathbf{a}_K, \mathbf{S}_K^2 \\ \text{sous } \mathcal{A}_c(\tau_K)}} \mathcal{L}_{[\tau_0, \tau_K]}(\mathbf{T}_{K-1}, \mathbf{a}_K, \mathbf{S}_K^2) &\simeq \max_{\substack{a_1, S_1^2 \\ \text{sous } \rho_1(\mathbf{T}_1) = \hat{g}(\mathbf{T}_1)}} \tilde{\mathcal{L}}_{[\tau_0, \mathbf{T}_1 + e]}^l(a_1, S_1^2) \\
 &+ \sum_{k=2}^{K-1} \max_{\substack{a_k, S_k^2 \\ \text{sous } \rho_k(\mathbf{T}_{k-1}) = \hat{g}(\mathbf{T}_{k-1}) \\ \text{et } \rho_k(\mathbf{T}_k) = \hat{g}(\mathbf{T}_k)}} \tilde{\mathcal{L}}_{[\mathbf{T}_{k-1} - e, \mathbf{T}_k + e]}(a_k, S_k^2) \\
 &+ \max_{\substack{a_K, S_K^2 \\ \text{sous } \rho_K(\mathbf{T}_{K-1}) = \hat{g}(\mathbf{T}_{K-1})}} \tilde{\mathcal{L}}_{[\mathbf{T}_{K-1} - e, \tau_K]}^r(a_K, S_K^2). \quad (2.39)
 \end{aligned}$$

Comme détaillé dans la section précédente, le principe de décomposition est similaire sous contrainte de dérivabilité.

Résolution du problème global

Dans cette partie, le vecteur des instants de transition \mathbf{T}_{K-1} n'est plus considéré fixé et nous souhaitons résoudre le problème de maximisation global de la log-vraisemblance lissée (2.38) (toujours sous contrainte de continuité $\mathcal{A}_c(\tau_K)$) en l'approchant ainsi,

$$\begin{aligned}
 \max_{\substack{\mathbf{a}_K, \mathbf{S}_K^2 \\ \text{sous } \mathcal{A}_c(\tau_K)}} \mathcal{L}_{[\tau_0, \tau_K]}(\mathbf{T}_{K-1}, \mathbf{a}_K, \mathbf{S}_K^2) &\simeq \max_{\mathbf{T}_{K-1}} \left[\max_{\mathbf{T}_{K-2}} \left[\cdots \operatorname{argmax}_{\mathbf{T}_1} \left[\max_{\substack{a_1, S_1^2 \\ \text{sous } \rho_1(\mathbf{T}_1) = \hat{g}(\mathbf{T}_1)}} \widetilde{\mathcal{L}}_{[\tau_0, \mathbf{T}_1+e]}^l(a_1, S_1^2) \right. \right. \right. \\
 &+ \left. \left. \left. \max_{\substack{a_2, S_2^2 \\ \text{sous } \rho_2(\mathbf{T}_1) = \hat{g}(\mathbf{T}_1) \\ \text{et } \rho_2(\mathbf{T}_2) = \hat{g}(\mathbf{T}_2)}} \widetilde{\mathcal{L}}_{[\mathbf{T}_1-e, \mathbf{T}_2+e]}(a_2, S_2^2) \right] \right. \\
 &\vdots \\
 &+ \left. \left. \left. \max_{\substack{a_{K-1}, S_{K-1}^2 \\ \text{sous } \rho_{K-1}(\mathbf{T}_{K-2}) = \hat{g}(\mathbf{T}_{K-2}) \\ \text{et } \rho_{K-1}(\mathbf{T}_{K-1}) = \hat{g}(\mathbf{T}_{K-1})}} \widetilde{\mathcal{L}}_{[\mathbf{T}_{K-2}-e, \mathbf{T}_{K-1}+e]}(a_{K-1}, S_{K-1}^2) \right] \right. \\
 &+ \left. \left. \left. \max_{\substack{a_K, S_K^2 \\ \text{sous } \rho_K(\mathbf{T}_{K-1}) = \hat{g}(\mathbf{T}_{K-1})}} \mathcal{L}_{[\mathbf{T}_{K-1}-e, \tau_K]}^r(a_K, S_K^2) \right] \right.
 \end{aligned}$$

Comme précédemment, les sous-problèmes de maximisation sur les différents intervalles peuvent être résolus de façon indépendante et les instants de transition τ_{K-1} sont estimés sur la grille des instants de mesure. Le résultat est similaire pour la contrainte de dérivabilité.

Dans ce chapitre, nous avons présenté un modèle de régression polynomiale par morceaux contraint par des hypothèses de régularité, ainsi que des méthodes d'inférence permettant son estimation. Les simulations relatives à ces méthodes sont présentées dans le chapitre 4. Dans le prochain chapitre nous proposons d'autres méthodes d'inférence qui se distinguent de celles que nous venons de présenter, puisqu'elles n'utilisent aucun a priori sur le nombre de régimes K présents dans les données.

Chapitre 3

Méthodes d'inférence sans a priori sur le nombre de régimes

Sommaire

3.1 Introduction	42
3.2 Approche de type segmentation binaire	42
3.2.1 Modèle sans variable latente	42
3.2.2 Modèle à variable latente et algorithme EM	45
3.3 Approche de segmentation binaire pénalisée	45
3.3.1 Vraisemblance pénalisée	45
3.3.2 Vraisemblance pénalisée et algorithme EM	48
3.4 Tentative d'étude théorique de la pénalité	48
3.4.1 Expressions de \hat{a}_{n+m} , \hat{b}_{n+m} et $\hat{\sigma}_{n+m}^2$	49
3.4.2 Espérances et variances de \hat{a}_{n+m} , \hat{b}_{n+m} et $\hat{\sigma}_{n+m}^2$	50
3.4.3 Comportement asymptotique de $E[\hat{\sigma}_{n+m}^2]$	51

3.1 Introduction

Dans ce chapitre, nous souhaitons toujours estimer, par maximum de vraisemblance et sous contrainte de régularité $\mathcal{A}_c(\tau_K)$ ou $\mathcal{A}_d(\tau_K)$, le modèle de régression polynomiale par morceaux de la section 2.2. En revanche, les méthodes proposées ici nous permettent de nous démarquer de l'état de l'art (cf. chapitre 1 section 1.4) sur le fait que le nombre de régimes K n'est plus connu ou fixé a priori. En effet, dans la littérature, le nombre de régimes est souvent estimé a posteriori et en général par des techniques de sélection de modèle. On peut cependant noter que Chamroukhi dans [12] propose une méthode de classification de courbes lorsque le nombre de clusters est inconnu, mais cette problématique est différente de la notre qui consiste à identifier les différents régimes d'un signal temporel continu.

Les deux méthodes proposées dans ce chapitre (dans leurs versions «standard» et «lissée») sont récursives. La première est inspirée de l'algorithme de segmentation binaire introduit par Edwards et al. dans [23] et Scott et Knott dans [51] pour estimer un changement de moyenne ou de variance dans les données. Cette méthode consiste à comparer le signal sur un intervalle à sa décomposition en deux intervalles optimaux à l'aide, par exemple, du test de rapport de vraisemblance proposé par Scott et Knott dans [51]. Si cette décomposition en deux régimes est retenue, alors cette même procédure est appliquée à ces deux nouveaux intervalles. Cette procédure est répétée de telle sorte qu'un diagramme en arbre est exploré. Notre méthode, qui fait l'objet de la section 3.2, propose un algorithme de ce type pour estimer une ou plusieurs ruptures entre les régimes dans un modèle de régression polynomiale par morceaux sous contraintes de régularité. La seconde méthode, détaillée en section 3.3 utilise également cette notion de découpage en deux régimes, mais se veut évolutive puisqu'elle estime les régimes de la gauche vers la droite à l'aide d'une vraisemblance pénalisée. La pénalisation proposée dans cette méthode semble avoir un bon comportement face à l'introduction d'une ou plusieurs observations appartenant à un mauvais régime, ce que nous avons souhaité étudier dans la section 3.4.

3.2 Approche de type segmentation binaire

Dans cette section, nous proposons une méthode récursive avec ses deux versions «standard» (modèle sans variable latente) et «lissée» (modèle à variable latente et algorithme EM), qui est inspirée de l'algorithme de segmentation binaire proposé par Edwards et al. dans [23] et Scott et Knott dans [51].

3.2.1 Modèle sans variable latente

La méthode que nous présentons ici repose entièrement sur la méthode à deux régimes de la section 2.3.1, que nous proposons d'utiliser récursivement pour estimer un modèle à un nombre de régimes quelconque. Après avoir effectué un rappel sur la méthode à deux régimes de la section 2.3.1 dans la prochaine sous-partie, nous expliquons ensuite le fonctionnement de l'algorithme et la règle d'arrêt utilisée par la méthode. Cette règle est nécessaire car elle indique à l'algorithme s'il doit ou non poursuivre la segmentation pour l'estimation de régimes supplémentaires.

Rappel sur la méthode à deux régimes

La méthode à deux régimes de la section 2.3.1 consiste en la résolution du problème de maximisation suivant,

$$\max_{\substack{T_1, \mathbf{a}_2, \mathbf{S}_2^2 \\ \text{sous } \mathcal{A}_s(\tau_2)}} \mathcal{L}(T_1, \mathbf{a}_2, \mathbf{S}_2^2) = \max_{T_1} \max_{\substack{\mathbf{a}_2, \mathbf{S}_2^2 \\ \text{sous } \mathcal{A}_s(\tau_2)}} \mathcal{L}(T_1, \mathbf{a}_2, \mathbf{S}_2^2),$$

dans lequel la log-vraisemblance du modèle de régression à deux régimes s'écrit

$$\mathcal{L}(T_1, \mathbf{a}_2, \mathbf{S}_2^2) = -\frac{1}{2} \sum_{i=1}^n \sum_{k=1}^2 \left[\log(S_k^2) + \frac{(Y_i - \rho_{a_k}(X_i))^2}{S_k^2} \right] \mathbb{1}_{[T_{k-1}, T_k]}(X_i),$$

où $T_0 = \tau_0$ et $T_2 = \tau_2$. Pour résoudre ce problème d'optimisation, nous proposons de résoudre l'optimisation en T_1 à partir du problème de maximisation suivant,

$$\max_{\substack{T_1, \mathbf{a}_2, \mathbf{S}_2^2 \\ \text{sous } \mathcal{A}_*(\tau_2)}} \mathcal{L}(T_1, \mathbf{a}_2, \mathbf{S}_2^2) = \max_{T_1} \mathcal{L}(T_1, \mathbf{a}_2^*(T_1), \mathbf{S}_2^{2*}(T_1)),$$

où $(\mathbf{a}_2^*(T_1), \mathbf{S}_2^{2*}(T_1))$ est l'optimum de

$$\max_{\substack{\mathbf{a}_2, \mathbf{S}_2^2 \\ \text{sous } \mathcal{A}_*(\tau_2)}} \mathcal{L}(T_1, \mathbf{a}_2, \mathbf{S}_2^2),$$

obtenu par l'algorithme de point fixe décrit dans la section 2.3.1, pour une valeur de T_1 fixée.

Principe de la méthode

Le principe de récurrence est le suivant : on applique la méthode de segmentation en deux régimes 2.3.1 sur un intervalle I que l'on note $M_2(I)$ dans les figures 3.1 et 3.2. On teste ensuite si un seul régime sur I est meilleur que son découpage en deux intervalles I_1 et I_2 (selon un certain critère qui est détaillé plus loin). Cette étape est appelée règle d'arrêt dans les figures. Si cette règle répond «oui» I est conservé et la procédure s'arrête sans segmentation; sinon on applique cette même procédure sur I_1 et I_2 , c'est-à-dire $M_2(I_1)$ et $M_2(I_2)$. Cette méthode récursive est schématisée sur la figure 3.1. Les figures 3.2 et 3.3 montrent des exemples de découpages retenus par la méthode après 1 (figure 3.2) ou 2 (figure 3.3) étapes.

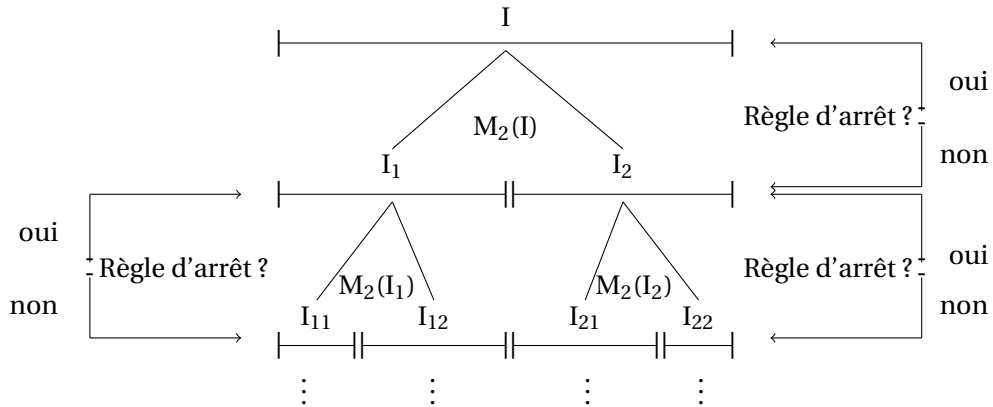


FIGURE 3.1 – Schéma du principe de segmentation de la méthode.

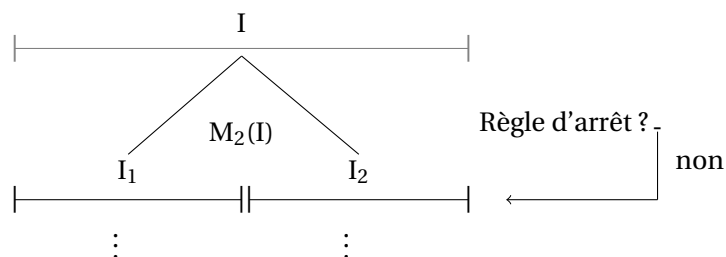


FIGURE 3.2 – Schéma de l'étape 1 de la méthode quand réponse «non» à la règle d'arrêt. L'intervalle en gris n'est pas retenu par la méthode.

Dans l'exemple de la figure 3.2, la règle d'arrêt répond «non» à la première étape de la méthode : les intervalles I_1 et I_2 sont retenus et la méthode poursuit la segmentation de ces intervalles pour l'estimation de régimes supplémentaires, ce qui nous donne le schéma de la figure 3.3. Dans cette figure la règle d'arrêt de gauche répond «oui» : l'intervalle I_1 est conservé et ne sera plus segmenté aux étapes suivantes. Celle de droite répond «non» : les intervalles I_{21} et I_{22} sont retenus et la méthode poursuit la segmentation de ces intervalles pour l'estimation de régimes supplémentaires. La méthode s'arrête lorsque toutes les règles d'arrêts répondent «oui». À ce moment là, nous obtenons le nombre de régimes estimé, noté \hat{K} et les estimateurs des paramètres $\mathbf{T}_{\hat{K}-1}$, $\mathbf{a}_{\hat{K}}$ et $\mathbf{S}_{\hat{K}}^2$.

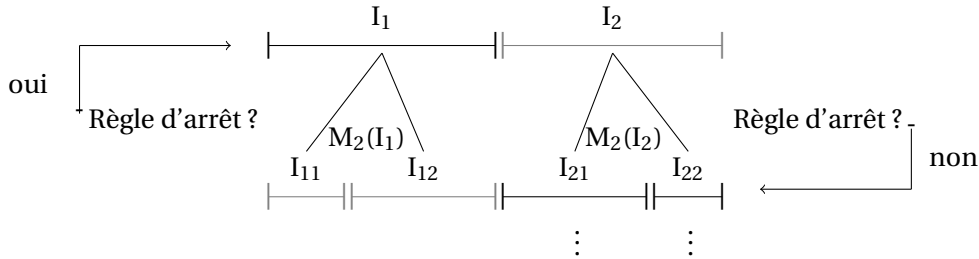


FIGURE 3.3 – Schéma de l'étape 2 de la méthode quand réponses «oui» (à gauche) et «non» (à droite) aux règles d'arrêts. Les intervalles en gris ne sont pas retenus par la méthode.

Règle d'arrêt

Le nombre de régimes étant inconnu a priori, il est nécessaire que la méthode proposée fonctionne selon une règle d'arrêt. Cette règle est appelée à chaque nouvelle application de la méthode à deux régimes sur un intervalle donné. Elle est basée sur un critère de sélection de modèle qui va nous permettre de comparer le modèle à un régime sur l'intervalle étudié, à celui obtenu par la méthode à deux régimes. Pour ce faire, nous proposons d'utiliser le BIC comme critère de comparaison de ces deux modèles. Ce critère a été proposé par Schwartz et al. dans [50] et se définit par la formule suivante,

$$\text{BIC} = -2\mathcal{L} + p \log(n),$$

avec \mathcal{L} la log-vraisemblance du modèle étudié, n le nombre d'observations de l'intervalle et p le nombre de paramètres libres. La règle standard consiste à conserver le modèle pour lequel ce critère a la plus faible valeur.

Dans notre cas, nous notons BIC_1 et BIC_2 les BIC des modèles à un et deux régimes et nous posons la règle suivante :

- si $\text{BIC}_1 < \text{BIC}_2$: nous conservons le modèle à un régime ;
- si $\text{BIC}_2 \leq \text{BIC}_1$: nous avons constaté empiriquement que dans ce cas la conservation du modèle à deux régimes était une règle trop stricte, puisqu'elle engendrait la conservation d'un nombre trop important de régimes. Nous avons donc choisi d'assouplir cette règle en calculant le pourcentage d'amélioration de BIC_2 par rapport à BIC_1 , c'est-à-dire,

$$\Omega = \frac{\text{BIC}_1 - \text{BIC}_2}{\text{BIC}_1} \times 100.$$

Pour une valeur seuil c choisie, on conservera le modèle à un régime si et seulement si $\Omega \leq c$.

Cette méthode n'estime que deux régimes à la fois, sur des intervalles dont le nombre d'observations diminue à chaque nouvelle étape de la méthode. Cela induit une faible complexité de la méthode. Elle a également l'avantage de ne tenir compte d'aucune information a priori sur le nombre de régimes.

3.2.2 Modèle à variable latente et algorithme EM

Dans cette section le principe de l'algorithme de découpage de l'intervalle I (maintenant appelé $M_{EM}(I)$) et de règle d'arrêt est identique à celui introduit à la section 3.2.1 sauf que $M_{EM}(I)$ correspond à la méthode de segmentation en deux régimes de la section 2.3.2 qui utilise un algorithme EM pour estimer les paramètres $(\mathbf{a}_2, \mathbf{S}_2^2)$ lorsque T_1 est fixé.

3.3 Approche de segmentation binaire pénalisée

Dans cette section, nous proposons une nouvelle méthode d'inférence avec ses deux versions «standard» (modèle sans variable latente) et «lissée» (modèle à variable latente et algorithme EM). Elle est issue de la segmentation binaire et consiste à chercher successivement les régimes d'un modèle de la gauche vers la droite. Elle repose sur une version pénalisée de la log-vraisemblance.

3.3.1 Vraisemblance pénalisée

La méthode que nous proposons repose sur la maximisation de la log-vraisemblance en $(\mathbf{a}_2, \mathbf{S}_2^2)$ sous contrainte de régularité $\mathcal{A}_c(\tau_2)$ ou $\mathcal{A}_d(\tau_2)$, pour une valeur de T_1 fixée, de la section 2.3.1. Nous notons $(\mathbf{a}_2^*(T_1), \mathbf{S}_2^{2*}(T_1))$ l'optimum de ce problème de maximisation obtenu par l'algorithme de point fixe 2.3.1.

$$(\mathbf{a}_2^*(T_1), \mathbf{S}_2^{2*}(T_1)) = \underset{\substack{\mathbf{a}_2, \mathbf{S}_2^2 \\ \text{sous } \mathcal{A}_s(\tau_2)}}{\arg \max} \mathcal{L}(T_1, \mathbf{a}_2, \mathbf{S}_2^2).$$

En revanche, nous avons choisi de pénaliser la maximisation en T_1 en utilisant le critère pénalisé

$$\mathcal{L}_p(T_1, \mathbf{a}_2, \mathbf{S}_2^2) = \mathcal{L}(T_1, \mathbf{a}_2, \mathbf{S}_2^2) - \sum_{i=1}^n \log(S_1^2) \mathbb{1}_{\{X_i < T_1\}},$$

de sorte que l'estimateur de τ_1 réalise

$$\max_{T_1} \mathcal{L}_p(T_1, \mathbf{a}_2^*(T_1), \mathbf{S}_2^{2*}(T_1)). \quad (3.1)$$

La pénalisation proposée dépend de S_1^2 qui correspond à la variance du régime 1 et du nombre d'observations dans ce même régime. Si l'intervalle $[\tau_0, T_1[$ contient des données du deuxième régime, c'est-à-dire si T_1 est plus grand que τ_1 , alors on s'attend à ce que la variance estimée S_1^2 augmente fortement. Par conséquent, la pénalité augmente dans ce cas. Ce phénomène devrait être accentué si l'intervalle $[\tau_0, T_1[$ contient des données des régimes suivants. Nous pouvons voir ce phénomène sur la figure 3.4 qui représente simultanément $\mathcal{L}(T, \mathbf{a}_2^*(T), \mathbf{S}_2^{2*}(T))$ et sa version pénalisée $\mathcal{L}_p(T, \mathbf{a}_2^*(T), \mathbf{S}_2^{2*}(T))$ en fonction de T pour un modèle à cinq régimes. On voit que ces deux log-vraisemblances sont très proches pour des valeurs de T inférieures à τ_1 , la pénalité a donc un faible impact sur la log-vraisemblance pour T_1 inférieur à τ_1 . En revanche, l'effet de la pénalité sur la log-vraisemblance est visible dès lors que T_1 est supérieur à τ_1 . On peut aussi remarquer que la log-vraisemblance non pénalisée présente un pic lors du 3^e et du 4^e instant de transition du processus. Ce pic est maximal au niveau du 4^e instants de transition. La log-vraisemblance pénalisée est quant à elle maximale au premier instant de transition, la pénalisation «privilégie» ce premier instant, ce qui est l'objectif recherché puisque la méthode proposée recherche successivement les régimes d'un modèle de la gauche vers la droite. Dans les prochaines parties, nous expliquons le fonctionnement de cette méthode ainsi que sa règle d'arrêt.

Principe de segmentation

Pour toutes observations $(X_i, Y_i)_{1 \leq i \leq n_1}$ définies sur un intervalle connu I , le principe de segmentation utilisé par la méthode est schématisé sur la figure 3.5. $M_2^p(I)$ signifie que l'on applique la méthode à deux régimes qui pénalise la maximisation en T_1 , aux données $(X_i, Y_i)_{1 \leq i \leq n_1}$ appartenant à l'intervalle I . Ceci, dans le but de le découper en deux régimes correspondant à I_1 et I_2 . Si

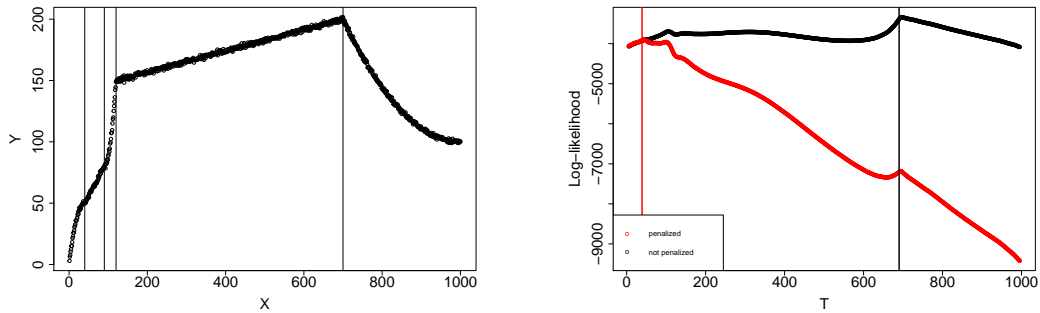


FIGURE 3.4 – Données simulées selon un modèle à cinq régimes où les droites verticales représentent les instants de transition (à gauche). Représentation simultanée de $\mathcal{L}(T, \mathbf{a}_2^*(T), \mathbf{S}_2^{2*}(T))$ et de sa version pénalisée $\mathcal{L}_p(T, \mathbf{a}_2^*(T), \mathbf{S}_2^{2*}(T))$ en fonction de T avec les droites verticales noir et rouge correspondant aux instants pour lesquels la log-vraisemblance et la log-vraisemblance pénalisée sont respectivement maximales (à droite).

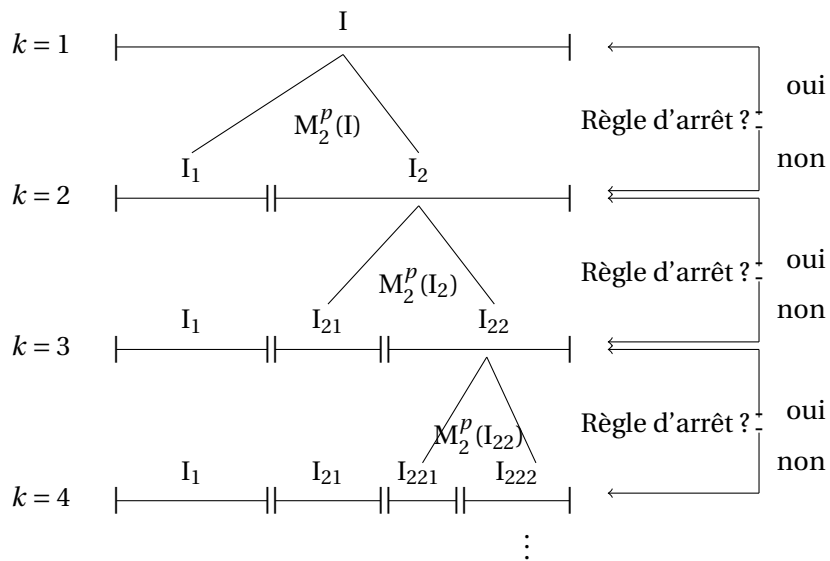


FIGURE 3.5 – Schéma du principe de segmentation de la méthode.

la règle d'arrêt répond «non», I_1 correspond au premier régime du modèle et la segmentation se poursuit sur I_2 pour la recherche de la deuxième rupture. Elle s'arrête dès que la règle d'arrêt répond «oui». Dans le cas où l'on obtiendrait les réponses des figures 3.6, 3.7 et 3.8 comme règles d'arrêts, les intervalles ciblés par la flèche correspondraient aux régimes retenus par la méthode à chaque étape. Dans l'exemple de la figure 3.6, la règle d'arrêt répond «non» à la première étape de la méthode : les intervalles I_1 et I_2 sont retenus et la méthode poursuit uniquement la segmentation de l'intervalle I_2 , ce qui nous donne le schéma de la figure 3.7. Dans cet exemple, la règle d'arrêt répond toujours «non» : En plus de I_1 , les intervalles I_{21} et I_{22} sont retenus et la méthode poursuit la segmentation de l'intervalle I_{22} , ce qui nous donne le schéma de la figure 3.8. Sur cette figure, la règle d'arrêt répond «oui» : la méthode s'arrête au bout de trois étapes et elle estime un modèle à $K = 3$ régimes, qui sont I_1 , I_{21} et I_{22} .

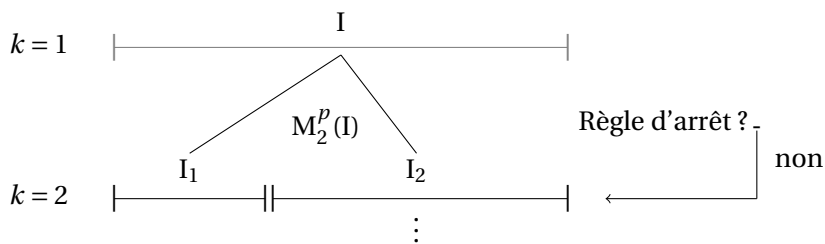


FIGURE 3.6 – Schéma de l'étape 1 de la méthode quand réponse «non» à la règle d'arrêt. L'intervalle en gris n'est pas retenu par la méthode.

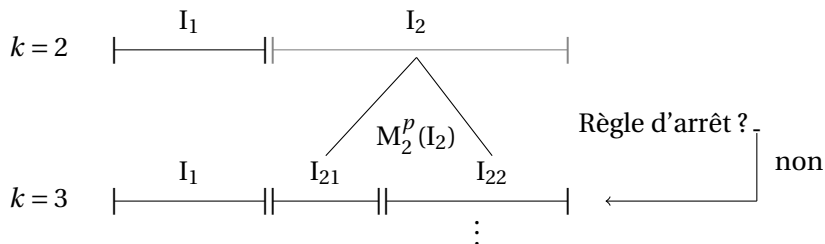


FIGURE 3.7 – Schéma de l'étape 2 de la méthode quand réponse «non» à la règle d'arrêt. L'intervalle en gris n'est pas retenu par la méthode.

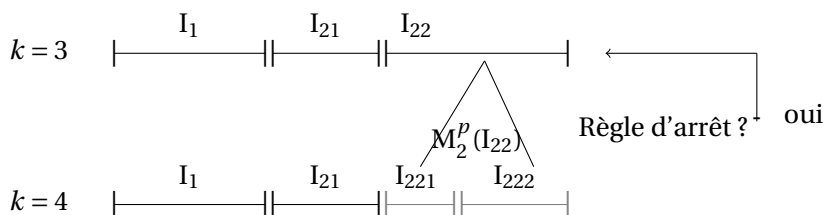


FIGURE 3.8 – Schéma de l'étape 3 de la méthode quand réponse «oui» à la règle d'arrêt. Les intervalles en gris ne sont pas retenus par la méthode.

Règle d'arrêt

Le nombre de régimes étant inconnu a priori, il est nécessaire que la méthode proposée fonctionne selon une règle d'arrêt. Cette règle est appelée à chaque nouvelle segmentation et est, comme précédemment, basée sur un critère de sélection de modèle de type BIC. En revanche nous comparons deux modèles sur l'intervalle I de départ, c'est-à-dire les modèles à k et à $k + 1$ régimes estimés aux étapes k et $k + 1$ de la méthode. Nous notons BIC_k et BIC_{k+1} , les BIC calculés a posteriori, à partir des log-vraisemblances non pénalisées de ces modèles. La règle de sélection de modèle est la suivante :

- si $\text{BIC}_k < \text{BIC}_{k+1}$, alors nous conservons le modèle à k régimes;
- si $\text{BIC}_k \geq \text{BIC}_{k+1}$, nous calculons le pourcentage d'amélioration de BIC_{k+1} par rapport à BIC_k , c'est-à-dire,

$$\Omega = \frac{\text{BIC}_k - \text{BIC}_{k+1}}{\text{BIC}_k} \times 100.$$

Pour une valeur seuil c choisie, on conservera le modèle à k régimes si et seulement si $\Omega \leq c$.

Comme précédemment, cette méthode a l'avantage de n'utiliser aucun a priori sur le nombre de régimes. De plus, elle n'estime que deux régimes à la fois, sur des intervalles dont le nombre d'observations diminue à chaque nouvelle étape de la méthode. Enfin, l'estimation des régimes s'effectue de manière ordonnée de la gauche vers la droite, c'est-à-dire que contrairement à la méthode précédente, celle-ci ne parcourt pas toutes les branches de l'arbre mais simplement la branche la plus à droite. Sa complexité est donc plus faible que celle de la méthode précédente.

3.3.2 Vraisemblance pénalisée et algorithme EM

Dans cette section, la méthode repose sur la maximisation en $(\mathbf{a}_2, \mathbf{S}_2^2)$ par algorithme EM, pour une valeur de T_1 fixée, de la méthode de la section 2.3.2,

$$(\mathbf{a}_2^{\text{EM}}(T_1), \mathbf{S}_2^{2\text{EM}}(T_1)) = \underset{\substack{\mathbf{a}_2, \mathbf{S}_2^2 \\ \text{sous } \mathcal{A}_*(T_2)}}{\text{argmax}} \widetilde{\mathcal{L}}(T_1, \mathbf{a}_2, \mathbf{S}_2^2).$$

En revanche, comme précédemment, nous pénalisons la maximisation en T_1 en utilisant

$$\widetilde{\mathcal{L}}_p(T_1, \mathbf{a}_2, \mathbf{S}_2^2) = \widetilde{\mathcal{L}}(T_1, \mathbf{a}_2, \mathbf{S}_2^2) - \sum_{i=1}^n \log(S_1^2) \mathbb{1}_{\{X_i < T_1\}},$$

de sorte que l'estimateur de τ_1 réalise

$$\max_{T_1} \widetilde{\mathcal{L}}_p(T_1, \mathbf{a}_2^{\text{EM}}(T_1), \mathbf{S}_2^{2\text{EM}}(T_1)).$$

Le principe de la méthode et la règle d'arrêt sont identiques à ceux présentés dans la section précédente.

3.4 Tentative d'étude théorique de la pénalité

Afin de mieux comprendre la pénalisation par $\sum_{i=1}^n \log(S_1^2) \mathbb{1}_{\{X_i < T_1\}}$ de la segmentation proposée dans la section 3.3, nous proposons ici d'étudier le comportement de S_1^2 dans le cadre suivant : nous estimons un unique modèle de régression linéaire sur des données issues de deux modèles différents.

Soit Y une variable aléatoire réelle à expliquer et X une variable explicative déterministe. Nous considérons que pour tout $1 \leq i \leq n$, le lien entre Y et X s'écrit sous la forme du modèle de régression linéaire donné par

$$Y_i = a + bX_i + \epsilon_i, \tag{3.2}$$

et que pour tout $n+1 \leq i \leq n+m$, le lien entre Y et X s'écrit sous la forme d'un autre modèle de régression linéaire donné par

$$Y_i = a^* + b^*X_i + \epsilon_i^*,$$

la distribution de l'erreur ϵ (ϵ^* respectivement) est indépendante de X et le modèle est supposé gaussien et homoscédastique, c'est-à-dire que $\epsilon \sim \mathcal{N}(0, \sigma^2)$ ($\epsilon^* \sim \mathcal{N}(0, \sigma^{2*})$ respectivement). a , b et σ^2 (a^* , b^* et σ^{2*} respectivement) sont des paramètres constants et inconnus de ce modèle.

Nous nous plaçons désormais dans le contexte où nous ne savons pas que ces $n + m$ observations sont issues de deux modèles linéaires différents et nous souhaitons estimer le modèle (3.2) sur l'ensemble des $n + m$ observations. Dans la suite nous noterons \hat{a}_{n+m} , \hat{b}_{n+m} et $\hat{\sigma}_{n+m}^2$ les estimateurs des paramètres du modèle (3.2) calculés sur les $n + m$ observations :

$$\begin{aligned}\hat{b}_{n+m} &= \frac{\sum_{i=1}^{n+m} (X_i - \bar{X}_{n+m})(Y_i - \bar{Y}_{n+m})}{\sum_{i=1}^{n+m} (X_i - \bar{X}_{n+m})^2}, \\ \hat{a}_{n+m} &= \bar{Y}_{n+m} - \hat{b}_{n+m}\bar{X}_{n+m}, \\ \hat{\sigma}_{n+m}^2 &= \frac{1}{n+m} \sum_{i=1}^{n+m} (Y_i - \hat{Y}_i)^2,\end{aligned}$$

où $\hat{Y}_i = \hat{a}_{n+m} + \hat{b}_{n+m}X_i$ et \bar{Y}_{n+m} (\bar{X}_{n+m} respectivement) désigne la moyenne des Y_i (des X_i respectivement) sur les $n + m$ observations. Nous définissons de même \bar{X}_n , \bar{Y}_n , etc.

Dans ce contexte, $\hat{\sigma}_{n+m}^2$ correspond au S_1^2 de la vraisemblance pénalisée, qui est notre estimateur d'intérêt. Dans la suite notre objectif est d'étudier le comportement moyen de $\hat{\sigma}_{n+m}^2$, notamment en fonction des nombres d'observations n et m .

3.4.1 Expressions de \hat{a}_{n+m} , \hat{b}_{n+m} et $\hat{\sigma}_{n+m}^2$

À partir des formules standard de la régression linéaire des estimateurs des moindres carrés, nous obtenons les formules suivantes pour \hat{a}_{n+m} , \hat{b}_{n+m} et $\hat{\sigma}_{n+m}^2$,

Lemme 6. Expression de \hat{b}_{n+m}

$$\begin{aligned}\hat{b}_{n+m} &= \frac{(a - a^*)n(\bar{X}_n - \bar{X}_{n+m})}{\sum_{i=1}^{n+m} (X_i - \bar{X}_{n+m})^2} + b \frac{\sum_{i=1}^n (X_i - \bar{X}_{n+m})^2}{\sum_{i=1}^{n+m} (X_i - \bar{X}_{n+m})^2} + b^* \frac{\sum_{i=n+1}^{n+m} (X_i - \bar{X}_{n+m})^2}{\sum_{i=1}^{n+m} (X_i - \bar{X}_{n+m})^2} \\ &+ \frac{\sum_{i=1}^n (X_i - \bar{X}_{n+m})\epsilon_i}{\sum_{i=1}^{n+m} (X_i - \bar{X}_{n+m})^2} + \frac{\sum_{i=n+1}^{n+m} (X_i - \bar{X}_{n+m})\epsilon_i^*}{\sum_{i=1}^{n+m} (X_i - \bar{X}_{n+m})^2}.\end{aligned}$$

Lemme 7. Expression de \hat{a}_{n+m}

$$\hat{a}_{n+m} = \frac{1}{n+m} \left((na + ma^*) + n(b - \hat{b}_{n+m})\bar{X}_n + m(b^* - \hat{b}_{n+m})\bar{X}_m + (n\bar{\epsilon}_n + m\bar{\epsilon}_m) \right).$$

Lemme 8. Expression de $\hat{\sigma}_{n+m}^2$

$$\begin{aligned}(n+m)\hat{\sigma}_{n+m}^2 &= n(a - \hat{a}_{n+m})^2 + 2n(a - \hat{a}_{n+m})(b - \hat{b}_{n+m})\bar{X}_n + 2n(a - \hat{a}_{n+m})\bar{\epsilon}_n \\ &+ (b - \hat{b}_{n+m})^2 \sum_{i=1}^n X_i^2 + 2(b - \hat{b}_{n+m}) \sum_{i=1}^n X_i \epsilon_i + \sum_{i=1}^n \epsilon_i^2 + m(a^* - \hat{a}_{n+m})^2 \\ &+ 2m(a^* - \hat{a}_{n+m})(b^* - \hat{b}_{n+m})\bar{X}_m + 2m(a^* - \hat{a}_{n+m})\bar{\epsilon}_m + (b^* - \hat{b}_{n+m})^2 \sum_{i=n+1}^{n+m} X_i^2 \\ &+ 2(b^* - \hat{b}_{n+m}) \sum_{i=n+1}^{n+m} X_i \epsilon_i^* + \sum_{i=n+1}^{n+m} \epsilon_i^{*2}.\end{aligned}$$

Ces différents résultats vont nous permettre de calculer les espérances et les variances de ces estimateurs.

3.4.2 Espérances et variances de \hat{a}_{n+m} , \hat{b}_{n+m} et $\hat{\sigma}_{n+m}^2$

Dans cette partie, nous nous appuyons sur les lemmes précédents pour donner des expressions explicites des espérances et variances des estimateurs.

Lemme 9. *Espérance de \hat{b}_{n+m}*

$$\mathbb{E}[\hat{b}_{n+m}] = b - \frac{(a^* - a)n(\bar{X}_n - \bar{X}_{n+m})}{\sum_{i=1}^{n+k} (X_i - \bar{X}_{n+m})^2} - \frac{(b - b^*) \sum_{i=n+1}^{n+m} (X_i - \bar{X}_{n+m})^2}{\sum_{i=1}^{n+m} (X_i - \bar{X}_{n+m})^2}.$$

Lemme 10. *Variance de \hat{b}_{n+m}*

$$\text{Var}[\hat{b}_{n+m}] = \frac{\sigma^2 \sum_{i=1}^n (X_i - \bar{X}_{n+m})^2 + \sigma^{2*} \sum_{i=n+1}^{n+m} (X_i - \bar{X}_{n+m})^2}{\left(\sum_{i=1}^{n+m} (X_i - \bar{X}_{n+m})^2 \right)^2}.$$

Lemme 11. *Espérance de \hat{a}_{n+m}*

$$\mathbb{E}[\hat{a}_{n+m}] = a - \frac{m(a - a^*)}{n + m} + \frac{n\bar{X}_n}{n + m} \mathbb{E}[b - \hat{b}_{n+m}] + \frac{m\bar{X}_m}{n + m} \mathbb{E}[b^* - \hat{b}_{n+m}].$$

Lemme 12. *Variance de \hat{a}_{n+m}*

$$\text{Var}[\hat{a}_{n+m}] = \text{Var}[\hat{b}_{n+m}] \frac{n^2\bar{X}_n^2 + m^2\bar{X}_m^2}{(n + m)^2} + \frac{n}{n + m} \sigma^2 + \frac{m}{n + m} \sigma^{2*}.$$

D'après les lemmes 9 et 11, \hat{b}_{n+m} et \hat{a}_{n+m} sont des estimateurs biaisés de b et a . Pour $\hat{\sigma}_{n+m}^2$ nous donnons seulement l'expression de son espérance, qui nécessite la définition des termes N_1 à N_5 ,

$$N_1 = (a - a^*)^2 \left(nm \sum_{i=1}^{n+m} (X_i - \bar{X}_{n+m})^2 - (n + m)(n\bar{X}_n - n\bar{X}_{n+m})^2 \right),$$

$$N_2 = (a - a^*)(b - b^*) \left(2nm \left(\bar{X}_n \sum_{i=n+1}^{n+m} (X_i - \bar{X}_{n+m})^2 + \bar{X}_m \sum_{i=1}^n (X_i - \bar{X}_{n+m})^2 \right) - 2(n + m)\bar{X}_{n+m}(n\bar{X}_n - n\bar{X}_{n+m}) \right),$$

$$N_3 = (b - b^*)^2 \left((n + m) \sum_{i=n+1}^{n+m} (X_i - \bar{X}_{n+m})^2 \sum_{i=1}^n (X_i - \bar{X}_{n+m})^2 + (n + m)\bar{X}_{n+m}(n\bar{X}_n - n\bar{X}_{n+m}) \times N_4 + N_5 \right),$$

$$N_4 = \sum_{i=n+1}^{n+m} (X_i - \bar{X}_{n+m})^2 - \sum_{i=1}^n (X_i - \bar{X}_{n+m})^2 - \bar{X}_{n+m}(n\bar{X}_n - n\bar{X}_{n+m}),$$

$$N_5 = nm\bar{X}_n\bar{X}_m \sum_{i=1}^{n+m} (X_i - \bar{X}_{n+m})^2.$$

Proposition 13. *Espérance de $\hat{\sigma}_{n+m}^2$*

$$\begin{aligned} \mathbb{E}[\hat{\sigma}_{n+m}^2] &= \sigma^2 \left(\frac{-\sum_{i=1}^n (X_i - \bar{X}_{n+m})^2}{(n + m) \sum_{i=1}^{n+m} (X_i - \bar{X}_{n+m})^2} + \frac{n}{n + m} - \frac{n}{(n + m)^2} \right) \\ &+ \sigma^{2*} \left(\frac{-\sum_{i=n+1}^{n+m} (X_i - \bar{X}_{n+m})^2}{(n + m) \sum_{i=1}^{n+m} (X_i - \bar{X}_{n+m})^2} + \frac{m}{n + m} - \frac{m}{(n + m)^2} \right) \\ &+ \frac{1}{(n + m)^2 \sum_{i=1}^{n+m} (X_i - \bar{X}_{n+m})^2} (N_1 + N_2 + N_3). \end{aligned}$$

Cette formule est compliquée et donc difficilement interprétable. Dans la suite, nous supposons que les m données aberrantes sont toutes situées à droite des n premières données, afin de poursuivre l'analyse.

3.4.3 Comportement asymptotique de $\mathbb{E}[\hat{\sigma}_{n+m}^2]$

Dans cette partie on suppose que les X_i sont définis par $X_i = \frac{iT}{n}$, où T est strictement positif. Par conséquent si $i \leq n$, $X_i \leq T$ et sinon $X_i > T$. Autrement dit, les données aberrantes sont situées à droite des n premières données. De plus, nous faisons l'hypothèse que les deux modèles linéaires sont continus au niveau de l'instant de transition T , c'est-à-dire $a + bT = a^* + b^*T$. Cette hypothèse nous permet de réécrire l'équation précédente de la façon suivante,

Proposition 14. *Espérance de $\hat{\sigma}_{n+m}^2$ dans un modèle continu à deux régimes*

$$\begin{aligned} \mathbb{E}[\hat{\sigma}_{n+m}^2] &= \sigma^2 \left(\frac{-n(3m^2 + n^2 - 1) + n(n+m+1)(n+m-1)^2}{(n+m)^2(n+m+1)(n+m-1)} \right) \\ &+ \sigma^{2*} \left(\frac{-m(m^2 + 3n^2 - 1) + m(n+m+1)(n+m-1)^2}{(n+m)^2(n+m+1)(n+m-1)} \right) \\ &+ \frac{(b-b^*)^2 T^2 (m^3(4n^2 - 6n + 2) + m^2(6n^2 - 6n) + m(2n^2 - 2))}{12n(n+m)^2(n+m+1)(n+m-1)}. \end{aligned}$$

Démonstration. D'après la proposition 13 et l'hypothèse $X_i = \frac{iT}{n}$ on a

$$\begin{aligned} (n+m)^2(n+m+1)(n+m-1)\mathbb{E}[\hat{\sigma}_{n+m}^2] &= \sigma^2(-n(3m^2 + n^2 - 1) + n(n+m+1)(n+m-1)^2) \\ &+ \sigma^{2*}(-m(m^2 + 3n^2 - 1) + m(n+m+1)(n+m-1)^2) \\ &+ (a-a^*)^2(nm(-3nm + (n+m+1)(n+m-1))) \\ &+ (a-a^*)(b-b^*)Tk(n+1)(m^2 - nm + 2n^2 - n - 1) \\ &+ \frac{(b-b^*)^2 nm T^2}{12n} (D_1 + D_2), \end{aligned}$$

où

$$\begin{aligned} D_1 &= (m^2 + 3n^2 - 1)(3m^2 + n^2 - 1) \\ &- 3m(n+m+1)(m^2 + 3n^2 - 1), \end{aligned}$$

et

$$\begin{aligned} D_2 &= 3n(n+m+1)(3m^2 + n^2 - 1) \\ &- 9(n+m+1)^2 nm \\ &+ 3(n+m+1)(n+m-1)(n+1)(m+2n+1). \end{aligned}$$

On obtient le résultat en ajoutant l'hypothèse $a + bT = a^* + b^*T$. \square

Lorsque m devient très grand devant n , $\mathbb{E}[\hat{\sigma}_{n+m}^2]$ converge vers σ^{2*} . Désormais nous posons $m = \alpha n$, avec $\alpha \geq 0$ et nous faisons tendre n vers l'infini. α représente la proportion de données du deuxième régime qui est donc fixé. On a le résultat suivant,

$$\mathbb{E}[\hat{\sigma}_{n+m}^2] \rightarrow \frac{3(\sigma^2 + \alpha\sigma^{2*})(\alpha+1)^3 + \alpha^3(b-b^*)^2 T^2}{3(\alpha+1)^4} = f(\alpha).$$

Nous pouvons maintenant étudier le comportement asymptotique de $f(\alpha)$ et cela passe par l'étude du signe de sa dérivée notée $\frac{\partial f(\alpha)}{\partial \alpha}$ et définie par

$$\frac{\partial f(\alpha)}{\partial \alpha} = \frac{3(\sigma^{2*} - \sigma^2)(\alpha+1)^3 - (b-b^*)^2 T^2 \alpha^2 (\alpha-3)}{3(\alpha+1)^5}.$$

Le dénominateur $3(\alpha+1)^5$ est toujours positif, de ce fait le signe de $\frac{\partial f(\alpha)}{\partial \alpha}$ ne dépend que du signe du numérateur. Dans la suite, nous étudions les variations de $f(\alpha)$ dans le cas où $(b-b^*)^2 T^2 > 0$, qui correspond à un changement de régime. Les calculs réalisés pour l'obtention de ces variations sont présentés en annexe A.7. L'expression de la dérivée de $f(\alpha)$ permet de donner un premier résultat dans le cas de l'égalité des variances dans les deux régimes.

Proposition 15. *Égalité des variances*

On suppose : $\sigma^{2*} = \sigma^2$. Dans ce cas, la fonction f est croissante de 0 à 3 puis décroissante, avec $f(0) = \sigma^2$, $f(3) = \frac{256\sigma^2 + 9(b-b')^2T^2}{256}$ et sa limite en l'infini est σ^2 .

Afin d'étudier les cas où σ^{2*} est différent de σ^2 nous posons $\gamma = \frac{(\sigma^{2*} - \sigma^2)}{(b-b')^2T^2}$. Selon la valeur de γ le comportement de f n'est pas le même.

Proposition 16. $\sigma^{2*} > \sigma^2$

On suppose $\gamma > 0$. Nous pouvons distinguer les cas suivants :

- $\gamma \leq \frac{1}{3}$: dans ce cas f est croissante avec $f(0) = \sigma^2$ et sa limite en l'infini est σ^{2*} ;
- $\gamma > \frac{1}{3}$: dans ce cas f est croissante jusqu'en

$$\alpha_1 = \frac{1}{3\left(\gamma - \frac{1}{3}\right)} \left(\sqrt[3]{-(18\gamma^2 + 15\gamma + 1) + 9\left(\gamma - \frac{1}{3}\right)\sqrt{\gamma\left(\gamma + \frac{1}{12}\right)}} \right. \\ \left. + \sqrt[3]{-(18\gamma^2 + 15\gamma + 1) - 9\left(\gamma - \frac{1}{3}\right)\sqrt{\gamma\left(\gamma + \frac{1}{12}\right)} - (3\gamma + 1)} \right) > 3,$$

puis décroissante, avec $f(0) = \sigma^2$ et sa limite en l'infini est σ^{2*} .

Proposition 17. $\sigma^2 > \sigma^{2*}$

On suppose $\gamma \leq 0$. Nous pouvons distinguer les cas suivants :

- $\gamma \geq \frac{-1}{12}$: dans ce cas f est décroissante jusqu'en

$$\alpha_2 = -\frac{\sqrt{9\gamma+1}}{\sqrt{9\left(\gamma - \frac{1}{3}\right)^2}} \cos\left(\frac{\phi}{3}\right) + \frac{\sqrt{9\gamma+1}}{\sqrt{3\left(\gamma - \frac{1}{3}\right)^2}} \sin\left(\frac{\phi}{3}\right) - \frac{3\gamma+1}{3\gamma-1} < 1,$$

avec

$$\phi = \arccos\left(-\frac{(18\gamma^2 + 15\gamma + 1)}{(9\gamma + 1)\sqrt{9\gamma + 1}}\right),$$

puis croissante jusqu'en

$$\alpha_3 = \frac{2\sqrt{9\gamma+1}}{\sqrt{9\left(\gamma - \frac{1}{3}\right)^2}} \cos\left(\frac{\phi}{3}\right) - \frac{3\gamma+1}{3\gamma-1} < 3,$$

et enfin décroissante, avec $f(0) = \sigma^2$ et sa limite en l'infini est σ^{2*} ;

- $\gamma \leq \frac{-1}{12}$: dans ce cas f est décroissante, avec $f(0) = \sigma^2$ et sa limite en l'infini est σ^{2*} .

La fonction f définie en α_2 s'écrit de la façon suivante,

$$f(\alpha_2) = 3 \left(\sigma^2 + \left(-\frac{\sqrt{9\gamma+1}}{\sqrt{9\left(\gamma - \frac{1}{3}\right)^2}} \cos\left(\frac{\phi}{3}\right) + \frac{\sqrt{9\gamma+1}}{\sqrt{3\left(\gamma - \frac{1}{3}\right)^2}} \sin\left(\frac{\phi}{3}\right) - \frac{3\gamma+1}{3\gamma-1} \right) \sigma^{2*} \right) \\ \times \left(-\frac{\sqrt{9\gamma+1}}{\sqrt{9\left(\gamma - \frac{1}{3}\right)^2}} \cos\left(\frac{\phi}{3}\right) + \frac{\sqrt{9\gamma+1}}{\sqrt{3\left(\gamma - \frac{1}{3}\right)^2}} \sin\left(\frac{\phi}{3}\right) - \frac{2}{3\gamma-1} \right)^3 \\ + \frac{\left(-\frac{\sqrt{9\gamma+1}}{\sqrt{9\left(\gamma - \frac{1}{3}\right)^2}} \cos\left(\frac{\phi}{3}\right) + \frac{\sqrt{9\gamma+1}}{\sqrt{3\left(\gamma - \frac{1}{3}\right)^2}} \sin\left(\frac{\phi}{3}\right) - \frac{3\gamma+1}{3\gamma-1} \right)^3 (b-b')^2 T^2}{\left(-\frac{\sqrt{9\gamma+1}}{\sqrt{9\left(\gamma - \frac{1}{3}\right)^2}} \cos\left(\frac{\phi}{3}\right) + \frac{\sqrt{9\gamma+1}}{\sqrt{3\left(\gamma - \frac{1}{3}\right)^2}} \sin\left(\frac{\phi}{3}\right) - \frac{2}{3\gamma-1} \right)^4}.$$

et on a les bornes suivantes,

$$\frac{3}{4}\sigma^2 + \frac{1}{4}\sigma^{2*} < f(\alpha_2) < \sigma^2.$$

La fonction f définie en α_3 s'écrit

$$f(\alpha_3) = \frac{3 \left(\sigma^2 + \left(\frac{2\sqrt{9\gamma+1}}{\sqrt{9(\gamma-\frac{1}{3})^2}} \cos\left(\frac{\phi}{3}\right) - \frac{3\gamma+1}{3\gamma-1} \right) \sigma^{2*} \right) \times \left(\frac{2\sqrt{9\gamma+1}}{\sqrt{9(\gamma-\frac{1}{3})^2}} \cos\left(\frac{\phi}{3}\right) - \frac{2}{3\gamma-1} \right)^3 + \frac{\left(\frac{2\sqrt{9\gamma+1}}{\sqrt{9(\gamma-\frac{1}{3})^2}} \cos\left(\frac{\phi}{3}\right) - \frac{3\gamma+1}{3\gamma-1} \right)^3 (b-b')^2 T^2}{3 \left(\frac{2\sqrt{9\gamma+1}}{\sqrt{9(\gamma-\frac{1}{3})^2}} \cos\left(\frac{\phi}{3}\right) - \frac{2}{3\gamma-1} \right)^4}.$$

et on a les bornes suivantes,

$$\frac{3}{4}\sigma^2 + \frac{1}{4}\sigma^{2*} < f(\alpha_3) < \sigma^2 + \frac{9}{256}(b-b^*)^2 T^2$$

Les propositions 15, 16 et 17 donnent le comportement de l'espérance de $\hat{\sigma}_{n+m}^2$ dans le cas asymptotique d'un grand nombre d'observations n dans le premier régime et m dans le second en fonction de $\alpha = \frac{m}{n}$. Bien sur, ces résultats asymptotiques ne donnent pas de résultat formel sur le comportement de la variance $\hat{\sigma}_{n+m}^2$ estimée sur deux régimes pour des valeurs fixes de m et n mais si nous faisons malgré tout cette extrapolation alors nous voyons que dans le cas d'égalité des variances, la pénalité $\sum_{i=1}^{n+m} \log(\hat{\sigma}_{n+m}^2)$, proposée dans la section 3.3 sera maximale en $\alpha = 3$. Dans le cas où $\sigma^{2*} > \sigma^2$, la pénalité sera d'autant plus grande que la différence entre les variances sera grande, en particulier en α_1 dans le cas où $\gamma > \frac{1}{3}$. Pour que la pénalisation appliquée à la vraisemblance soit efficace, il faut qu'elle soit la plus grande possible lorsqu'on intègre des données d'un mauvais régime. Ces deux cas sont donc en faveur de cela. Ce n'est pas le cas lorsque l'on considère $\sigma^2 > \sigma^{2*}$, qui jouera le rôle inverse, c'est-à-dire que la pénalité diminuera avec l'ajout de données appartenant à un mauvais régime. Dans ce cas, la méthode de la section 3.3 ne sera pas efficace.

Dans ce chapitre nous avons présenté deux nouvelles méthodes récursives qui n'utilisent aucun a priori sur le nombre de régimes pour estimer un modèle de régression polynomiale par morceaux, ainsi qu'une tentative d'étude théorique de la pénalité proposée dans la seconde méthode. Les simulations relatives à ces méthodes sont présentées dans le chapitre 4.

Chapitre 4

Package R et simulations

Sommaire

4.1 Package R : HSPOR	56
4.1.1 Fonction H2SPOR	56
4.1.2 Fonction HKSPOR	57
4.1.3 Fonction HKSPOR_DynProg	57
4.1.4 Fonction H2SPOR_DynProg	59
4.2 Simulations	59
4.2.1 Analyse de l'algorithme H2SPOR pour des modèles à deux régimes	59
4.2.2 Analyse des algorithmes HKSPOR, HKSPOR_DynProg, H2SPOR_DynProg et Bin_Pen	66
4.2.3 Exemple de cas difficiles pour H2SPOR_DynProg et Bin_Pen	76
4.2.4 Estimation a posteriori du nombre de régimes pour HKSPOR_DynProg	78
4.3 Discussion	78

Les méthodes présentées dans les chapitres 2 et 3 ont été implémentées dans un package R qui fait l'objet de la section 4.1. À partir de ces implémentations et dans le but d'étudier le comportement des différentes méthodes, nous avons réalisé des simulations qui sont présentées dans la section 4.2. Les résultats obtenus dans ces simulations font l'objet d'une discussion, présentée dans la section 4.3.

4.1 Package R : HSPOR

Cette section a pour but de présenter le package R que nous avons conçu et qui contient l'ensemble des méthodes qui ont été présentées dans les chapitres 2 et 3, à l'exception de la méthode de la section 3.3, que l'on nommera `Bin_Pen` dans la suite, qui est issue d'une approche plus empirique pour laquelle nous n'avons pas obtenu de garanties théoriques ou numériques suffisantes. Le package se nomme HSPOR pour Hidden Smooth POLynomial Regression et est disponible sur le CRAN. L'objectif de ce package est de proposer des fonctions facilement utilisables par les utilisateurs (User Friendly). Il contient quatre fonctions principales nommées H2SPOR, HKSPOR, HKSPOR_DynProg et H2SPOR_DynProg que nous présentons rapidement dans les prochaines sections. Pour un détail plus complet de ces fonctions, l'aide du package HSPOR se trouve en annexe A.8. Dans la suite, l'ensemble des temps de calculs indiqués dans la suite ont été obtenus à partir d'un ordinateur portable équipé d'un processeur intel Core i7.

4.1.1 Fonction H2SPOR

La fonction H2SPOR permet d'estimer un modèle de régression polynomiale à deux régimes à l'aide des méthodes présentées dans les sections 2.3.1 et 2.3.2. Cette fonction prend seulement en entrée les données X et Y ainsi que le degré `deg` du modèle que l'on souhaite estimer. Les autres paramètres sont définis par défaut. Parmi ces paramètres, `EM` est un paramètre de choix de la méthode qui sera utilisée pour estimer le modèle. Plus précisément, si `EM=False`, la fonction estime le modèle à deux régimes par la méthode de la section 2.3.1. C'est celle de la section 2.3.2 qui est utilisée quand `EM=True` (c'est le cas par défaut).

Les résultats obtenus en sortie de la fonction sont présentés dans la figure 4.1 soit : un tableau contenant les paramètres du modèle estimé par la fonction H2SPOR, qui s'affiche dans la console ; le graphique des résultats associés qui s'affiche dans l'onglet plot de l'environnement Rstudio, lorsque l'option `plotG=True` (c'est le cas par défaut). Le tableau obtenu en sortie se lit de la façon

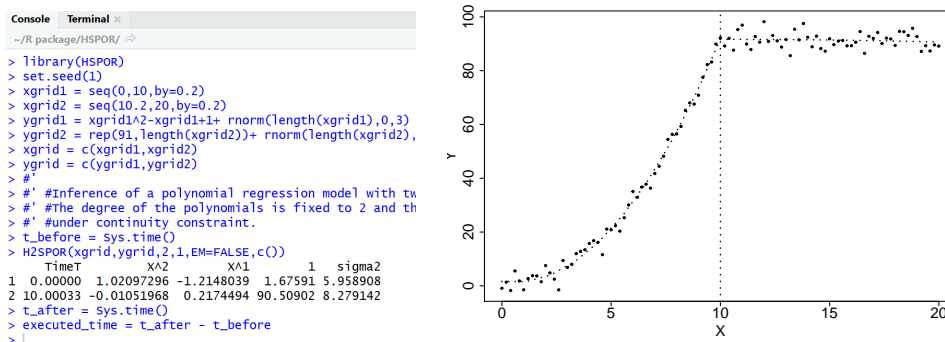


FIGURE 4.1 – Sortie obtenue en 2.7 secondes par la fonction H2SPOR dans la console de Rstudio (à gauche) et dans l'onglet plot de Rstudio (à droite), lorsqu'elle estime un modèle de degré 2 à deux régimes.

suivante : la première colonne du tableau contient la valeur de la première observation X_1 (une estimation de τ_0), suivie en deuxième ligne de l'instant de transition estimé T_1 ; les colonnes 2 à $2 + deg$ contiennent les estimations des coefficients des polynômes des régimes 1 et 2 ; la dernière

colonne contient les variances estimées des régimes 1 et 2. Le graphique proposé en sortie affiche les données X et Y (points) le modèle estimé (en pointillés) et l’instant de transition estimé (droite verticale en pointillés).

4.1.2 Fonction HKSPOR

La fonction HKSPOR est une généralisation de la fonction H2SPOR et permet d’estimer un modèle de régression polynomiale à un nombre de régimes K quelconque, à l’aide des méthodes présentées dans les sections 2.4.1 et 2.4.2. Elle prend simplement en entrée des données X et Y, le degré deg et le nombre de régimes K du modèle que l’on souhaite estimer. Les autres paramètres sont définis par défaut. Comme précédemment, EM est un paramètre de choix de la méthode qui sera utilisée pour estimer le modèle. Plus précisément, si EM=False, la fonction estime le modèle par la méthode de la section 2.4.1. C’est celle de la section 2.4.2 qui est utilisée quand EM=True (c’est le cas par défaut).

Comme le montre la figure 4.2, les résultats obtenus en sortie de la fonction sont similaires à ceux de la fonction H2SPOR (un tableau et un graphique). Le nombre de lignes du tableau dépend du nombre de régimes que contient le modèle estimé.

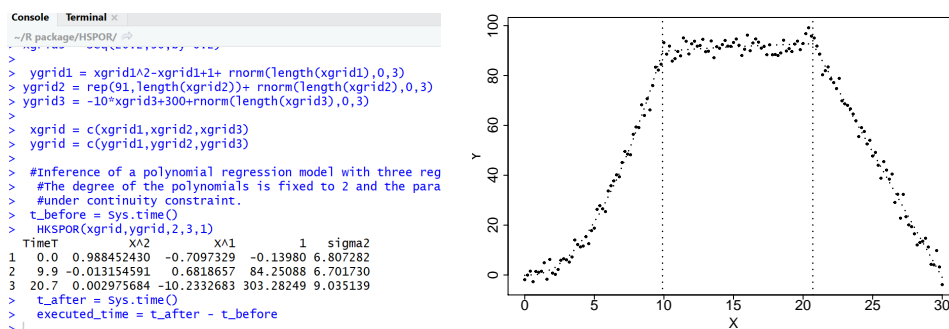


FIGURE 4.2 – Sortie obtenue en 47.6 minutes par la fonction HKSPOR dans la console de Rstudio (à gauche) et dans l’onglet plot de Rstudio (à droite), lorsqu’elle estime un modèle de degré 2 à trois régimes.

4.1.3 Fonction HKSPOR_DynProg

La fonction HKSPOR_DynProg utilise la stratégie de programmation dynamique pour estimer un modèle de régression polynomiale à un nombre de régimes K quelconque, par les méthodes présentées dans les sections 2.5.1 et 2.5.2. Lorsque EM=False, le modèle est estimé par la méthode de la section 2.5.1, sinon il est estimé par celle de la section 2.5.2 (c’est le cas par défaut).

Comme le montre la figure 4.3, les résultats obtenus en sortie sont similaires à ceux des fonctions précédentes (un tableau et un graphique). Par défaut, l’option verbose=False. Lorsque verbose=True alors les résultats obtenus en sortie sont ceux de la figure 4.4. En effet, la fonction HKSPOR_DynProg utilise le principe de programmation dynamique qui consiste à dire qu’estimer un modèle à K régimes c’est équivalent à estimer un modèle à K – 1 régimes d’un côté et un modèle à 1 régime de l’autre. De ce fait, lorsque verbose=True, en plus du tableau d’estimation des paramètres du modèle à K régimes, la fonction affiche les tableaux intermédiaires des estimations des paramètres des modèles de 1 à K – 1 régimes. Il en est de même avec les graphiques affichés en sortie.

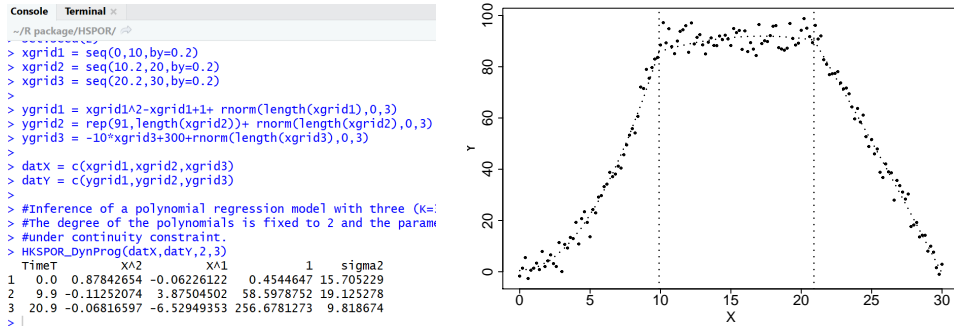


FIGURE 4.3 – Sortie obtenue en 3.6 minutes par la fonction HKSPOR_DynProg dans la console de Rstudio (à gauche) et dans l’onglet plot de Rstudio (à droite), lorsqu’elle estime un modèle de degré 2 à trois régimes.

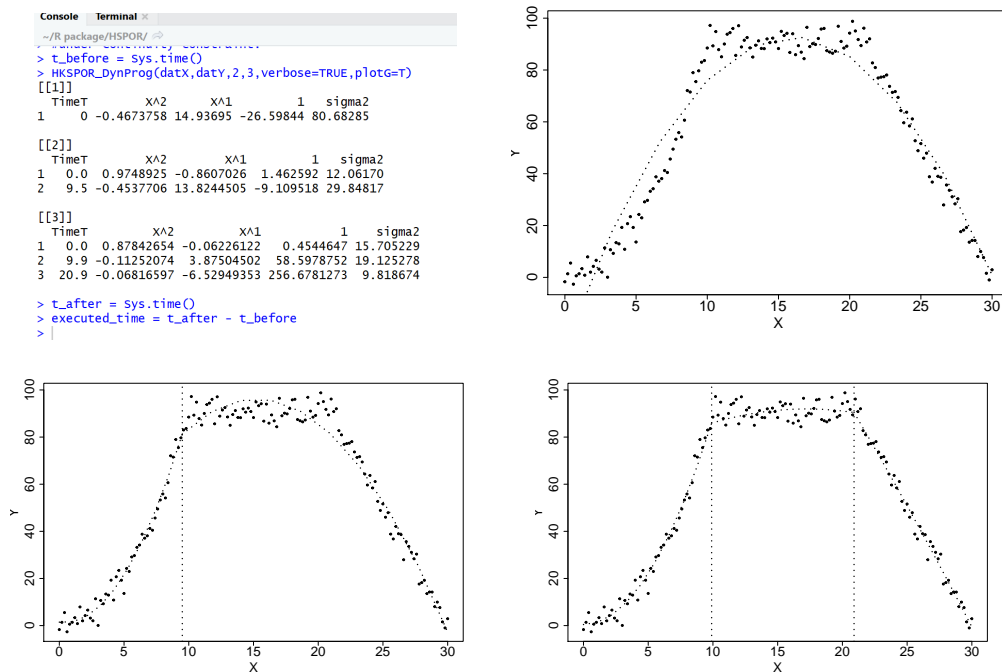


FIGURE 4.4 – Sortie obtenue en 3.6 minutes par la fonction HKSPOR_DynProg dans la console de Rstudio (en haut à gauche) et dans l’onglet plot de Rstudio (les trois autres graphiques), lorsqu’elle estime un modèle de degré 2 à trois régimes, quand l’option verbose=True.

4.1.4 Fonction H2SPOR_DynProg

La fonction H2SPOR_DynProg permet d'estimer un modèle de régression polynomiale par morceaux lorsque le nombre de régimes K est a priori inconnu à l'aide des méthodes présentées dans les sections 3.2.1 et 3.2.2. Lorsque `EM=False`, le modèle est estimé par la méthode de la section 3.2.1, sinon c'est par celle de la section 3.2.2 (c'est le cas par défaut).

Comme le montre la figure 4.5, les résultats obtenus en sortie sont similaires à ceux des fonctions précédentes (un tableau et un graphique). Le nombre de régimes du modèle à estimer étant a

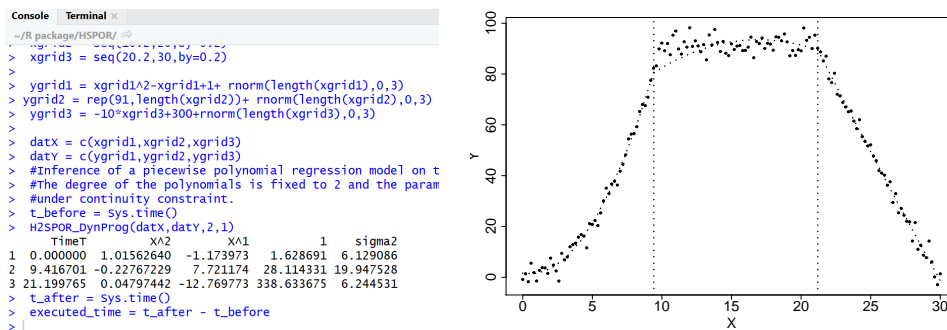


FIGURE 4.5 – Sortie obtenue en 2.1 minutes par la fonction H2SPOR_DynProg dans la console de Rstudio (à gauche) et dans l’onglet plot de Rstudio (à droite), lorsqu’elle estime un modèle de degré 2.

priori inconnu, le nombre de lignes du tableau affiché en sortie correspondra au nombre de régimes \hat{K} du modèle estimé par la fonction H2SPOR_DynProg.

4.2 Simulations

Dans cette section, nous proposons d’analyser l’algorithme H2SPOR pour des modèles à deux régimes, cette étude fait l’objet de la section 4.2.1. Nous étudions ensuite les algorithmes HKSPOR, HKSPOR_DynProg, H2SPOR_DynProg et Bin_Pen pour des modèles à un nombre quelconque de régimes dans la section 4.2.2. Nous poursuivons ensuite sur l’analyse de cas difficiles pour les algorithmes H2SPOR_DynProg et Bin_Pen dans la section 4.2.3. Nous terminons enfin par l’estimation a posteriori du nombre de régimes pour HKSPOR_DynProg dans la section 4.2.4. Pour rappel, les caractéristiques des différentes méthodes sont présentées dans le tableau 4.1.

	H2SPOR	HKSPOR	HKSPOR_DynProg	H2SPOR_DynProg	Bin_Pen
Nombre quelconque de régimes	non seulement 2	oui	oui	oui	oui
Estimation du nombre de régimes	non	possible a posteriori	possible a posteriori	oui	oui

TABLEAU 4.1 – Tableau des caractéristiques des différentes méthodes.

4.2.1 Analyse de l’algorithme H2SPOR pour des modèles à deux régimes

Dans cette section, nous illustrons le bon comportement des deux méthodes que nous avons présentées dans les sections 2.3.1 et 2.3.2 et qui sont implémentées dans l’algorithme H2SPOR avec `EM=False` ou `EM=True` respectivement.

Nous avons quatre objectifs correspondant à chacune des quatre prochaines sections qui sont (i) étudier l'impact du nombre d'observations sur l'estimation des paramètres, (ii) comparer les méthodes H2SPOR(EM=False) et H2SPOR(EM=True) lorsque les données étudiées présentent un ou plusieurs points leviers, (iii) vérifier si la méthode détecte un changement de variance lorsque la dynamique du régime n'est pas modifiée et enfin (iv) étudier l'impact de la contrainte de dérivabilité imposée dans l'estimation des paramètres lorsque celle-ci n'est pas vérifiée par le modèle.

Impact du nombre d'observations

L'objectif de cette section est d'étudier l'impact du nombre n d'observations (X_i, Y_i) sur la qualité des estimations de α_2 , τ_1 et σ_2^2 , obtenus par H2SPOR(EM=False). Pour cela, nous testons différentes valeurs de n : $n = 100$, $n = 200$ et $n = 500$.

Modèle de simulation A Nous considérons le modèle donné par

$$Y_i = \sum_{k=1}^2 \left(\rho_{\alpha_k}(X_i) + \sigma_k^2 \right) \mathbb{1}_{[\tau_{k-1}, \tau_k]}(X_i),$$

où le degré des polynômes $(\rho_{\alpha_k})_{1 \leq k \leq 2}$ est fixé à $\delta = 2$ et les coefficients sont $\alpha_2^1 = (1, 0.5)$, $\alpha_2^1 = (-2, 5)$ et $\alpha_2^0 = (10, -6.5)$. La variance du bruit est identique dans les deux régimes et vaut $\sigma_1^2 = \sigma_2^2 = 2.25$. Les observations $(X_i)_{1 \leq i \leq n}$ sont définies sur $[\tau_0 = 0, \tau_2 = 9]$ et sont réparties de façon régulière avec un tiers des observations qui se trouvent dans le premier régime et deux tiers dans le second ce qui correspond à un instant de transition entre les régimes fixé à $\tau_1 = 3$.

Les graphiques de la figure 4.6 représentent des données simulées suivant le modèle A pour différentes valeurs de n . Pour chaque valeur de n , nous réalisons 100 simulations indépendantes du

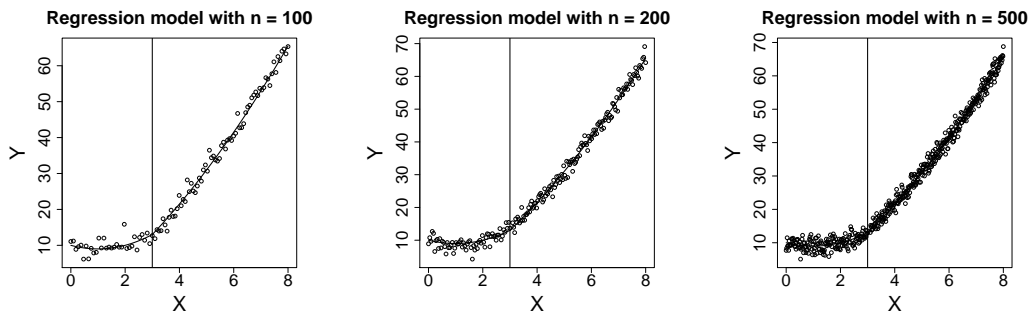


FIGURE 4.6 – Données simulées selon le modèle A pour $n = 100$, $n = 200$ et $n = 500$ observations. Le vrai modèle est en trait plein et l'instant de transition τ_1 est représenté par la droite verticale.

modèle. À chaque simulation, les estimations $\hat{\alpha}_2$, \hat{T}_1 et \hat{S}_2^2 des paramètres α_2 , τ_1 et σ_2^2 sont déterminés par la méthode H2SPOR(EM=False). Ces estimations sont comparées aux vraies valeurs des paramètres dans la figure 4.7.

Les résultats observés sur la figure 4.7 montrent que la méthode d'estimation fonctionne bien. Les estimateurs des coefficients des polynômes ne sont pas biaisés et comme attendu la précision de l'estimation augmente avec n . De plus, nous remarquons que la précision des estimations est meilleure dans le régime 2 ce qui s'explique par le nombre de points deux fois plus nombreux dans ce régime que dans le premier. Nous remarquons également que les variances S_1^2 et S_2^2 sont biaisées vers le bas lorsque le nombre d'observations est modeste. En effet, l'estimation \hat{S}_2^2 de σ_2^2 à $\hat{\alpha}_2$ connus, donnée dans les équations (A.11), correspond à l'estimation biaisée de la variance (division par le nombre n_i d'observations dans le régime i au lieu de $n_i - 3$ en tenant compte du nombre de coefficients estimés dans le régime). Nous obtenons des résultats similaires lorsque

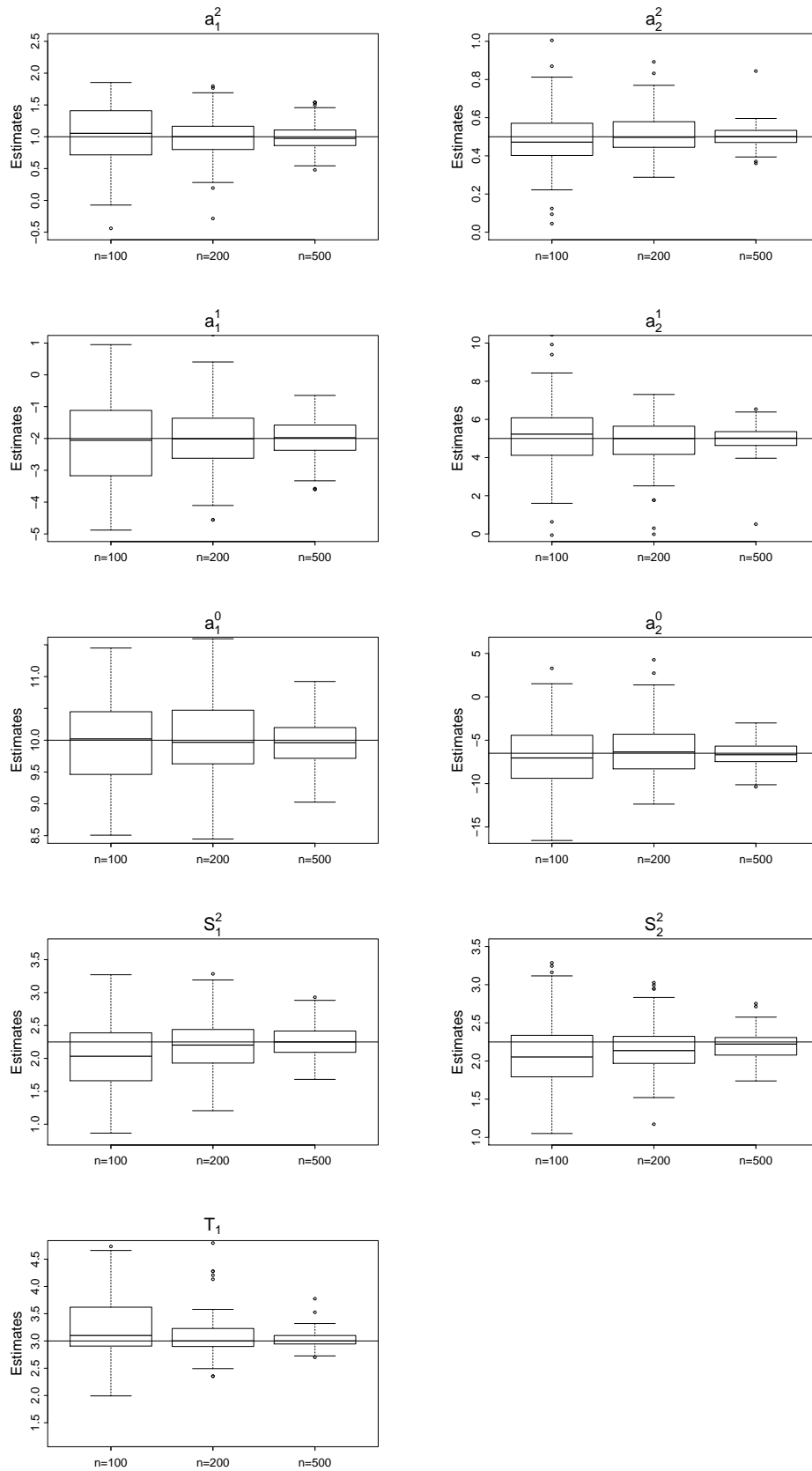


FIGURE 4.7 – Boxplots des estimateurs des coefficients des polynômes \mathbf{a}_2 , des variances S_2^2 et de l'instant de transition T_1 obtenus par H2SPOR (EM=False), en fonction du nombre d'observations n pour $B = 100$ répétitions indépendantes du modèle A. La droite horizontale représente la vraie valeur du paramètre.

l'on étudie l'impact du nombre n d'observations (X_i, Y_i) , sur la qualité des estimateurs \mathbf{a}_2 , T_1 et \mathbf{S}_2^2 obtenus par H2SPOR (EM=True). C'est pourquoi, nous avons choisi de ne pas montrer les résultats ici.

Avantage de l'algorithme EM en présence de données aberrantes

L'objectif de cette section est de comparer les méthodes H2SPOR (EM=False) et H2SPOR (EM=True) lorsque les données étudiées présentent un ou plusieurs points leviers.

Modèle de simulation B Nous considérons le modèle donné par

$$Y_i = \sum_{k=1}^2 \left(\rho_{\alpha_k}(X_i) + \sigma_k^2 \right) \mathbb{1}_{[\tau_{k-1}, \tau_k[}(X_i),$$

où le degré des polynômes $(\rho_{\alpha_k})_{1 \leq k \leq 2}$ est fixé à $\delta = 1$ et les coefficients sont $\alpha_1^1 = (1, 0)$ et $\alpha_2^0 = (0, 3)$. La variance du bruit est identique dans les deux régimes et vaut $\sigma_1^2 = \sigma_2^2 = 0.04$. Les observations $(X_i)_{1 \leq i \leq n}$ avec $n = 69$, sont définies sur $[\tau_0 = 0, \tau_2 = 8]$ et sont réparties de façon régulière. L'instant de transition entre les régimes est fixé à $\tau_1 = 3$.

La figure 4.8 présente des données simulées selon le modèle B dans lesquelles nous avons sciemment ajouté deux points leviers, c'est-à-dire que deux observations issues du second régime ont été placées dans le premier régime. Nous simulons 100 réalisations indépendantes de ce modèle.

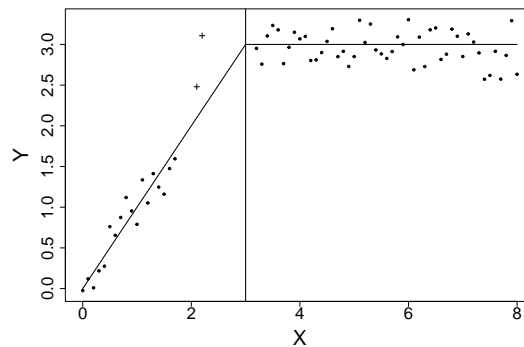


FIGURE 4.8 – Modèle de régression polynomiale à deux régimes contenant 2 points levier (+). Le vrai modèle est en trait plein et l'instant de transition τ_1 est représenté par la droite verticale.

Pour chaque simulation, les estimations \mathbf{a}_2 , T_1 et \mathbf{S}_2^2 des paramètres α_2 , τ_1 et σ_2^2 sont obtenus par les deux méthodes. Dans la figure 4.9, nous ne présentons que les résultats obtenus pour les paramètres du premier régime et de l'instant de transition T_1 . Nous avons en effet constaté que les paramètres du second régime sont bien estimés par les deux méthodes.

L'examen des résultats observés sur la figure 4.9 nous permet de dire que le modèle à variable latente est plus souple quant au traitement des points leviers. En effet, H2SPOR (EM=True) donne de meilleures estimations des coefficients des polynômes et de la variance du premier régime. Vraisemblablement, ceci est dû au fait que les deux points leviers sont attribués au régime 2. En revanche, le modèle standard (sans variable latente) n'a pas cette souplesse et la pente α_1^1 de la droite est sur-estimée par H2SPOR (EM=False) à cause de ces deux points. Fort de ces résultats, dans la suite nous ne considérons plus que la méthode d'estimation associée au modèle à variable latente dans les simulations.

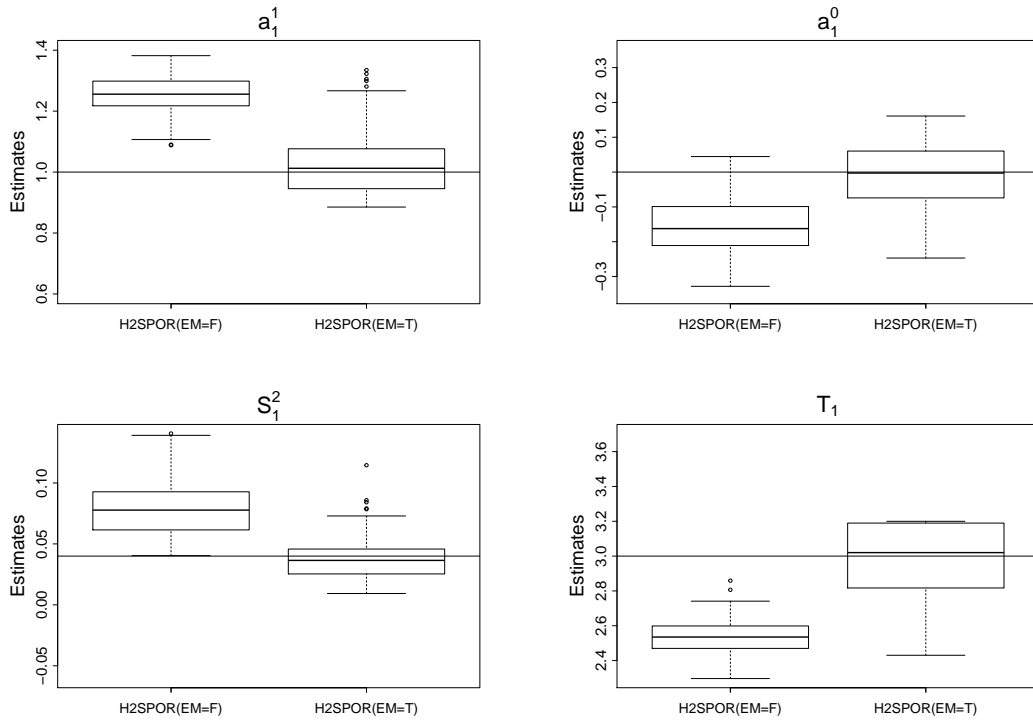


FIGURE 4.9 – Boxplots des valeurs estimées a_1 , T_1 et S_1^2 par H2SPOR (EM=False) et H2SPOR (EM=True) pour $B = 100$ répétitions indépendantes du modèle **B**. La droite horizontale représente la vraie valeur du paramètre.

Détection d'un changement de variance

L'objectif de cette section est de vérifier que la méthode détecte un changement de variance lorsque la dynamique du régime n'est pas modifiée. Pour ce faire, nous proposons d'étudier le modèle suivant, pour lequel nous faisons varier la variance du second régime en lui faisant prendre des valeurs de plus en plus grandes.

Modèle de simulation C (σ_2^2) Nous considérons un modèle de régression polynomiale constant par morceaux donné par,

$$Y_i = \sum_{k=1}^2 \left(3 + \sigma_k^2 \right) \mathbb{1}_{[\tau_{k-1}, \tau_k]}(X_i).$$

La variance du bruit du premier régime vaut $\sigma_1^2 = 0.04$ tandis que celle du second régime est une variable du modèle et vaudra dans nos tests $\sigma_2^2 = 0.09$, $\sigma_2^2 = 0.25$, $\sigma_2^2 = 0.64$ et $\sigma_2^2 = 1$. Les observations $(X_i)_{1 \leq i \leq n}$ avec $n = 100$, sont définies sur $[\tau_0 = 0, \tau_2 = 10]$. Elles sont réparties de façon régulière avec un quart des observations qui se trouvent dans le premier régime et trois quarts dans le second ce qui correspond à un instant de transition entre les régimes fixé à $\tau_1 = 2.5$.

Les graphiques de la figure 4.10 présentent des réalisations du modèle **C** (σ_2^2) en fonction des valeurs prises par σ_2^2 . Nous réalisons 100 répétitions indépendantes de ce modèle pour les différentes valeurs de σ_2^2 . À chaque simulation les estimations \mathbf{a}_2 , T_1 et \mathbf{S}_2^2 des paramètres $\boldsymbol{\alpha}_2$, τ_1 et $\boldsymbol{\sigma}_2^2$ sont obtenus par la méthode H2SPOR (EM=True). Les estimations ou les erreurs relatives absolues des paramètres sont représentées dans la figure 4.11.

Les résultats obtenus sur la figure 4.11 montrent que la méthode détecte un changement de variance au bon moment et l'estimation du temps de transition est d'autant meilleure que la différence entre les variances est grande. Nous pouvons de plus observer que l'estimation des pa-

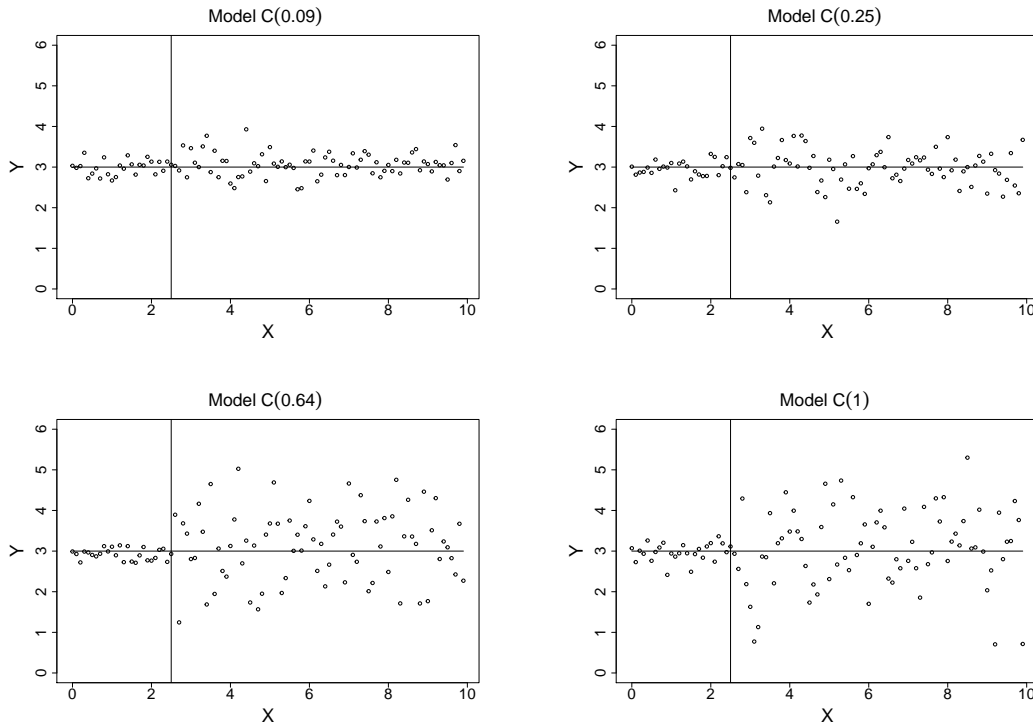


FIGURE 4.10 – Données simulées selon le modèle $C(\sigma_2^2)$ lorsque $\sigma_2^2 = 0.09$ (en haut à gauche), $\sigma_2^2 = 0.25$ (en haut à droite), $\sigma_2^2 = 0.64$ (en bas à gauche) et $\sigma_2^2 = 1$ (en bas à droite). Le vrai modèle est en trait plein (droite horizontale) et l'instant de transition τ_1 est représenté par la droite verticale.

ramètres est correcte dans les deux régimes. Les variances estimées dans le second régime sont biaisées mais restent proches des vraies valeurs.

Effet de la contrainte de dérivabilité lorsqu'elle n'est pas vérifiée par le modèle

L'objectif de cette section consiste à étudier l'effet de la contrainte de dérivabilité sur l'estimation des paramètres α_2 , τ_1 et σ_2^2 lorsque celle-ci n'est pas vérifiée par le modèle considéré.

Modèle de simulation D Nous considérons le modèle donné par

$$Y_i = \sum_{k=1}^2 \left(\rho_{\alpha_k}(X_i) + \sigma_k^2 \right) \mathbb{1}_{[\tau_{k-1}, \tau_k[}(X_i)$$

où le degré des polynômes $(\rho_{\alpha_k})_{1 \leq k \leq 2}$ est fixé à $\delta = 1$ et les coefficients sont $\alpha_1^1 = (-1, 2)$ et $\alpha_1^0 = (5, -4)$. La variance du bruit est identique dans les deux régimes et vaut $\sigma_1^2 = \sigma_2^2 = 0.25$. Les observations $(X_i)_{1 \leq i \leq n}$ avec $n = 61$, sont définies sur $[\tau_0 = 0, \tau_2 = 6]$. Elles sont réparties de façon régulière et la première moitié des observations se trouve dans le premier régime, l'autre moitié se situe dans le second ce qui correspond à un instant de transition entre les régimes fixé à $\tau_1 = 3$. Il faut remarquer que le modèle proposé est continu en tout point de $[\tau_0, \tau_2]$ mais la contrainte de dérivabilité en τ_1 n'est pas vérifiée par ce modèle.

Nous réalisons plusieurs répétitions indépendantes de ce modèle. Pour chaque simulation les estimations $\hat{\alpha}_2$, \hat{T}_1 et \hat{S}_2^2 des paramètres α_2 , τ_1 et σ_2^2 sont obtenus par H2SPOR (EM=True) sous la contrainte de continuité seule d'une part et sous la contrainte de dérivabilité d'autre part. Les résultats observés étant toujours du même type, nous présentons les ajustements d'une seule simulation dans la figure 4.12. Comme le montre la figure 4.12, l'ajustement des données est mauvais

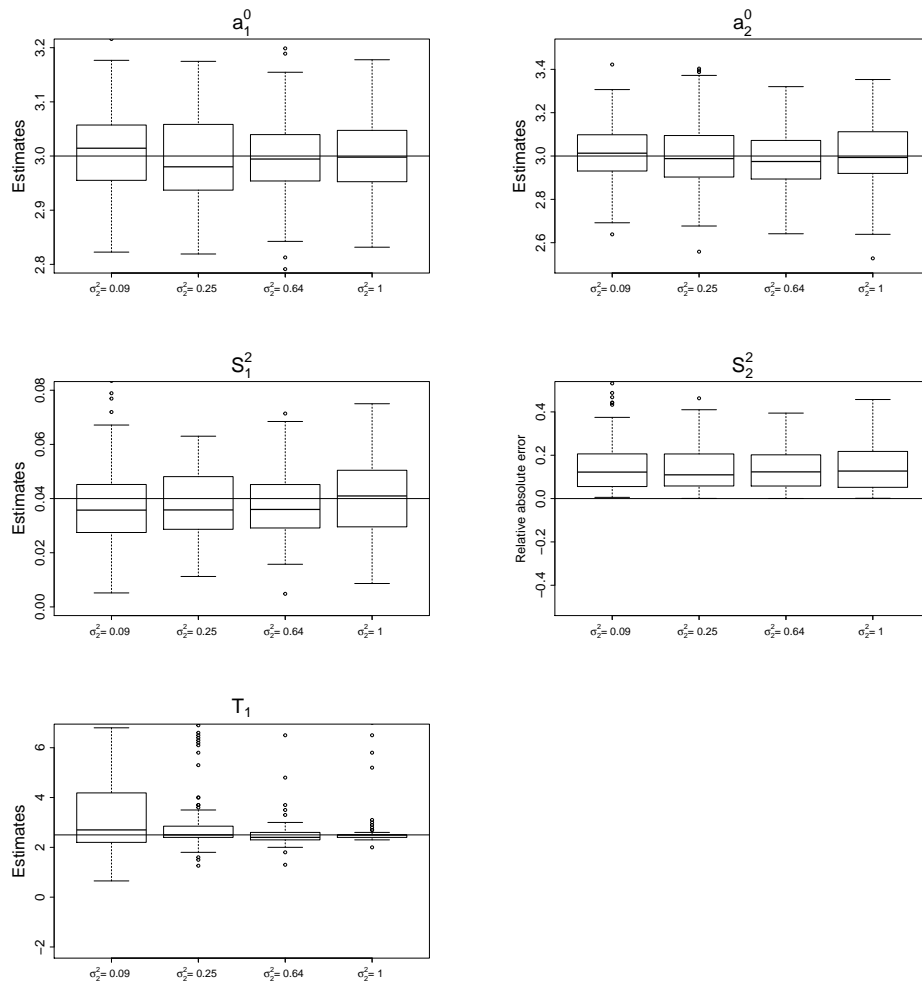


FIGURE 4.11 – Boxplots des estimations des coefficients constants des polynômes a_1^0 et a_2^0 , de l’instant de transition T_1 et de la variance du bruit du premier régime S_1^2 , ainsi que les erreurs d’estimations relatives absolues faites sur la variance du second régime par H2SPOR(EM=True), pour $B = 100$ répétitions indépendantes du modèle $C(\sigma_2^2)$ lorsque $\sigma_2^2 = 0.09$, $\sigma_2^2 = 0.25$, $\sigma_2^2 = 0.64$ et $\sigma_2^2 = 1$. La droite horizontale représente soit la vraie valeur du paramètre, soit une erreur nulle d’estimation.

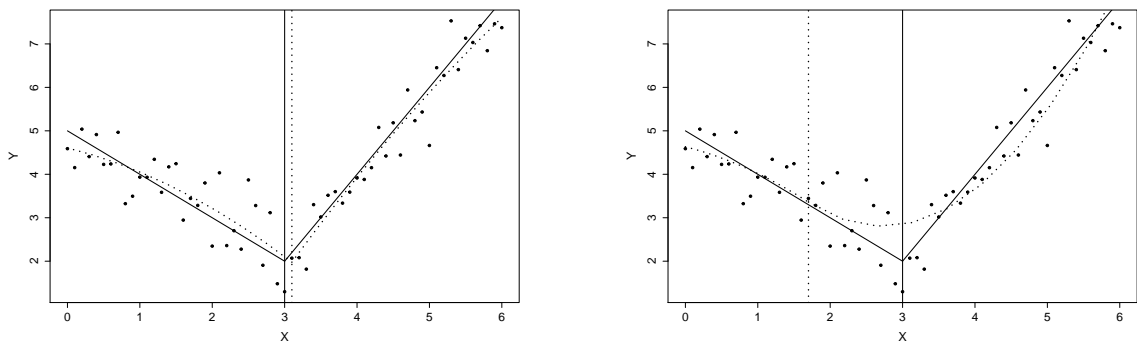


FIGURE 4.12 – Réalisations du modèle **D** (points) et ajustement des données par H2SPOR(EM=True) (trait en pointillés) sous contrainte de continuité (à gauche) ou sous contrainte de dérivabilité (à droite). Le vrai modèle est en trait plein et l’instant de transition τ_1 est représenté par la droite verticale en trait plein alors que son estimation T_1 est en pointillé.

lorsque l'estimation des paramètres se fait sous contrainte de dérivabilité alors que cette hypothèse n'est pas vérifiée par le modèle. Lors de l'estimation de modèles présentés dans cet article, il est donc judicieux de bien réfléchir au type de contrainte en fonction des connaissances a priori sur le modèle.

4.2.2 Analyse des algorithmes HKSPOR, HKSPOR_DynProg, H2SPOR_DynProg et Bin_Pen

Dans cette section, nous souhaitons illustrer le bon comportement des méthodes que nous avons présentées dans les sections 2.4.2, 2.5.2, 3.2.2 et 3.3.2 respectivement et qui sont implémentées dans les fonctions HKSPOR, HKSPOR_DynProg et H2SPOR_DynProg du package HSPOR présenté dans la section 4.1. Pour faciliter la compréhension, dans la suite nous utilisons ces termes pour parler de ces méthodes et nous appelons Bin_Pen la méthode de la section 3.3.2.

Nous avons deux objectifs correspondant aux deux prochaines sections qui sont (i) étudier l'intérêt de la programmation dynamique en comparant HKSPOR et HKSPOR_DynProg en terme de qualité d'estimation des paramètres mais également de temps de calcul, (ii) comparer les méthodes récursives c'est-à-dire HKSPOR_DynProg, H2SPOR_DynProg et Bin_Pen sur la détection des instants de transition lorsque H2SPOR_DynProg et Bin_Pen estiment le bon nombre de régimes d'une part, sur la qualité d'ajustement des données d'autre part et enfin sur les temps de calcul.

Intérêt de la programmation dynamique : HKSPOR VS HKSPOR_DynProg

Dans cette section, nous souhaitons comparer les algorithmes HKSPOR et HKSPOR_DynProg aussi bien sur la qualité d'estimation des paramètres que sur le temps de calcul nécessaire à l'obtention de ces estimations. Ces deux méthodes prennent en entrée le nombre de régimes, il n'y a pas d'enjeu sur ce point.

Modèle de simulation E Nous considérons le modèle donné par

$$Y_i = \sum_{k=1}^3 \left(\rho_{\alpha_k}(X_i) + \sigma_k^2 \right) \mathbb{1}_{[\tau_{k-1}, \tau_k]}(X_i),$$

où le nombre de régimes est fixé à $K = 3$, le degré des polynômes $(\rho_{\alpha_k})_{1 \leq k \leq 3}$ vaut $\delta = 2$ et les coefficients sont $\alpha_1^2 = (-1.22, 0, 1.25)$, $\alpha_1^1 = (7.33, -0.29, -25)$ et $\alpha_1^0 = (1, 12.86, 135)$. La variance du bruit est identique dans les trois régimes et vaut $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 0.16$. Les observations $(X_i)_{1 \leq i \leq n}$ avec $n = 190$, sont définies sur $[\tau_0 = 1, \tau_3 = 12]$ et sont réparties de façon régulière. Les instants de transition entre les régimes sont fixés à $\tau_1 = 3$ et $\tau_2 = 10$.

Le graphique de la figure 4.13 représente des données simulées suivant le modèle E.

Nous simulons 100 réalisations indépendantes de ce modèle. Pour chaque simulation, les estimations \mathbf{a}_3 , \mathbf{T}_2 et \mathbf{S}_3^2 sont obtenues par les deux méthodes et les temps de calcul nécessaires à ces estimations ont été enregistrés. La somme des erreurs relatives absolues faites sur les coefficients α_1 , α_2 et α_3 des polynômes de chaque régime et les estimations des paramètres τ_2 et σ_3^2 sont représentées dans la figure 4.14. Dans la figure 4.15 se trouvent les distributions des temps de calcul nécessaires aux deux méthodes pour estimer le modèle.

Les résultats observés sur les figures 4.14 montrent que les deux méthodes d'estimation fonctionnent tout aussi en ce qui concerne l'estimation des instants de transition que celle des coefficients des polynômes. On peut observer un léger biais dans l'estimation des coefficients des polynômes. Les biais relatifs sont assez proches et peuvent s'expliquer par la présence d'un faible nombre d'observations dans les différents régimes. Finalement, le résultat qui permet de vraiment distinguer ces deux méthodes est le temps de calcul (figure 4.15). En effet, quand HKSPOR met plus d'une heure trente à estimer les paramètres du modèle à trois régimes, HKSPOR_DynProg ne met que 7 minutes.

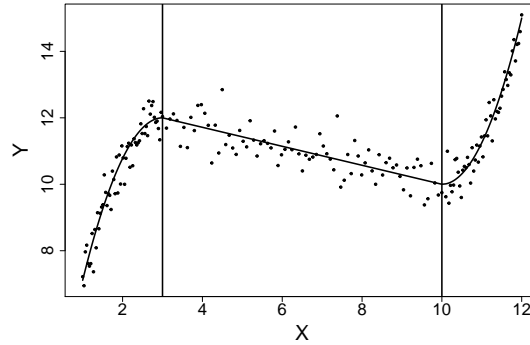


FIGURE 4.13 – Données simulées selon le modèle E. Le vrai modèle est en trait plein et les instants de transition τ_1 et τ_2 sont représentés par les droites verticales.

Comparaison des algorithmes HKSPOR_DynProg, H2SPOR_DynProg et Bin_Pen

Dans cette section, nous allons comparer les algorithmes HKSPOR_DynProg, H2SPOR_DynProg et Bin_Pen. Contrairement à la section précédente, les méthodes H2SPOR_DynProg et Bin_Pen estiment le nombre de régimes. La comparaison des trois méthodes se fera désormais sur la détection des instants de transition lorsque H2SPOR_DynProg et Bin_Pen estiment le bon nombre de régimes. Nous étudierons également la qualité des ajustements aux données et les temps de calcul nécessaires à l'estimation du modèle. Pour réaliser cette étude, nous proposons de comparer ces trois méthodes à partir de quatre modèles différents comportant respectivement deux, trois, quatre et six régimes.

Modèle de simulation F Nous considérons le modèle suivant,

$$Y_i = \sum_{k=1}^2 \left(\rho_{\alpha_k}(X_i) + \sigma_k^2 \right) \mathbb{1}_{[\tau_{k-1}, \tau_k]}(X_i),$$

où le nombre de régimes est fixé à $K = 2$, le degré des polynômes $(\rho_{\alpha_k})_{1 \leq k \leq 2}$ vaut $\delta = 2$ et les coefficients sont $\alpha_2^2 = (-7.76, 2.09)$, $\alpha_1^1 = (53.45, -41.86)$ et $\alpha_0^0 = (0, 211.28)$. La variance du bruit est identique dans les deux régimes et vaut $\sigma_1^2 = \sigma_2^2 = 4$. Les observations $(X_i)_{1 \leq i \leq n}$ avec $n = 200$, sont définies sur $[\tau_0 = 1, \tau_2 = 10]$ et sont réparties de façon régulière. L'instant de transition entre les régimes est fixé à $\tau_1 = 3.44$. Ce modèle a été généré aléatoirement et assure la continuité du modèle au niveau de l'instant τ_1 .

Le graphique de la figure 4.16 représente des données simulées suivant le modèle F.

Nous simulons 100 réalisations indépendantes de ce modèle et pour chaque simulation, nous estimons ce modèle avec les trois méthodes. H2SPOR_DynProg et Bin_Pen estiment dans 100% des cas le bon nombre de régimes qui est de 2. Dans le but de comparer l'ajustement du modèle F par les modèles estimés par les méthodes HKSPOR_DynProg, H2SPOR_DynProg et Bin_Pen, nous proposons d'utiliser le RMSE (Root Mean Squared Error) qui correspond à l'écart-type des résidus, c'est-à-dire,

$$\text{RMSE} = \sqrt{\sum_{i=1}^n \frac{(Y_i - \hat{Y}_i)^2}{n}},$$

et qui indique l'ajustement absolu du modèle aux données (différence entre les valeurs observées Y et les valeurs prédites par le modèle \hat{Y}). Plus le RMSE est proche de 0, meilleur est l'ajustement. La figure 4.17 représente les distributions des estimations T_1 et des RMSE obtenues par les trois

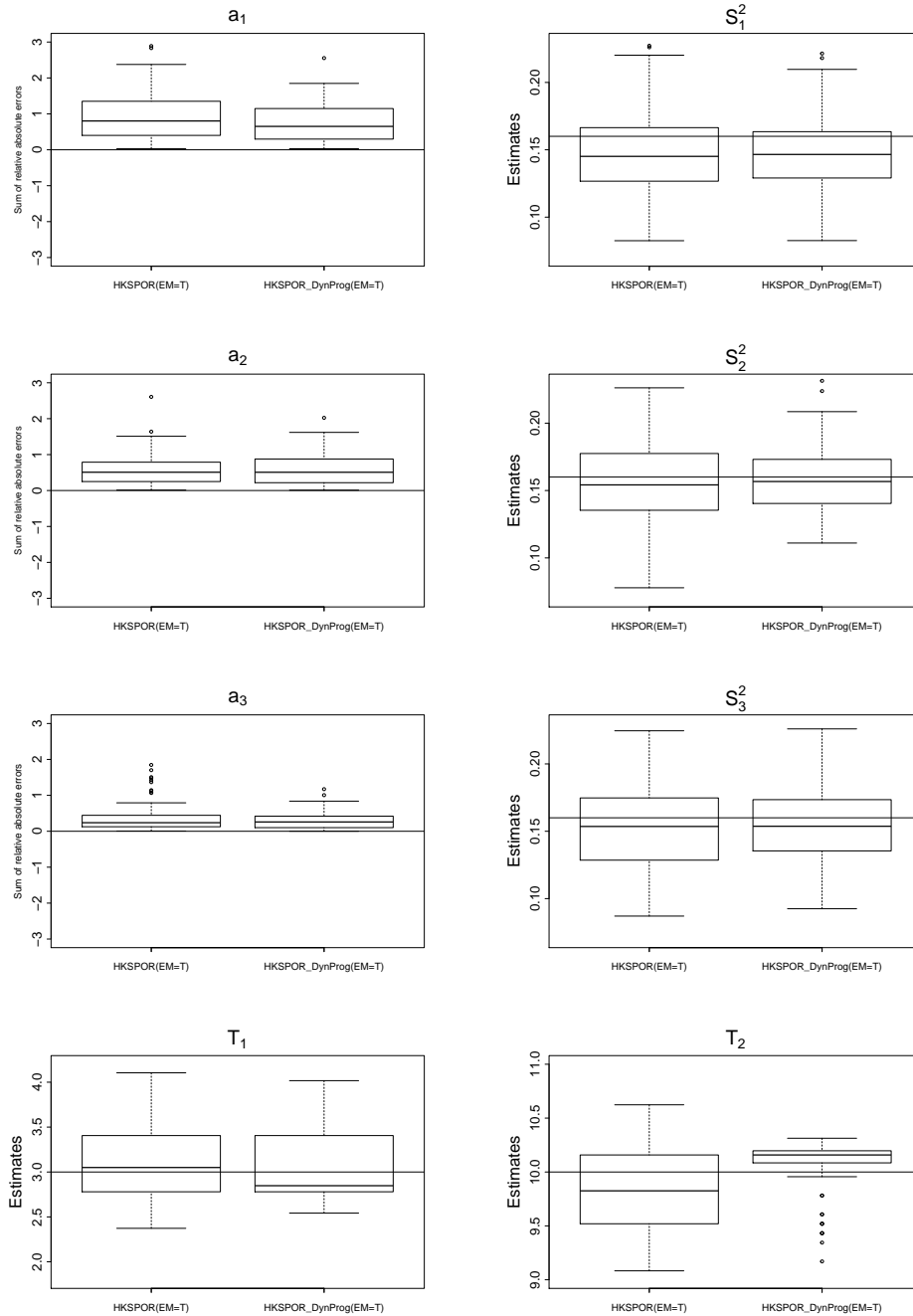


FIGURE 4.14 – Boxplots des erreurs relatives absolues des coefficients a_1 , a_2 et a_3 des polynômes de chaque régime estimés par les méthodes HKSPOR et HKSPOR_DynProg, ainsi que les estimations des variances S_3^2 et des instants de transition T_2 obtenues par ces deux mêmes méthodes pour $B = 100$ répétitions indépendantes du modèle E. La droite horizontale représente soit une erreur nulle, soit la vraie valeur du paramètre.

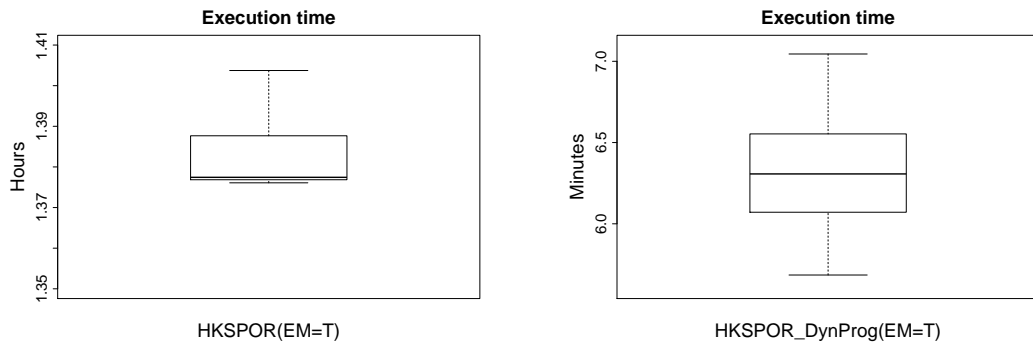


FIGURE 4.15 – Boxplots des temps de calculs nécessaires à HKSPOR (en heures) et à HKSPOR_DynProg (en minutes) pour estimer les paramètres α_3 , τ_2 et σ_3^2 , pour $B = 100$ répétitions indépendantes du modèle **E**.

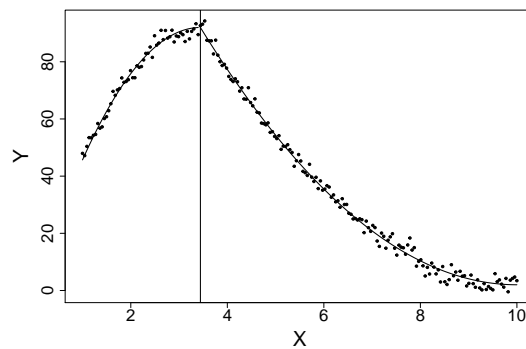


FIGURE 4.16 – Données simulées selon le modèle **F**. Le vrai modèle est en trait plein et l’instant de transition τ_1 est représenté par la droite verticale.

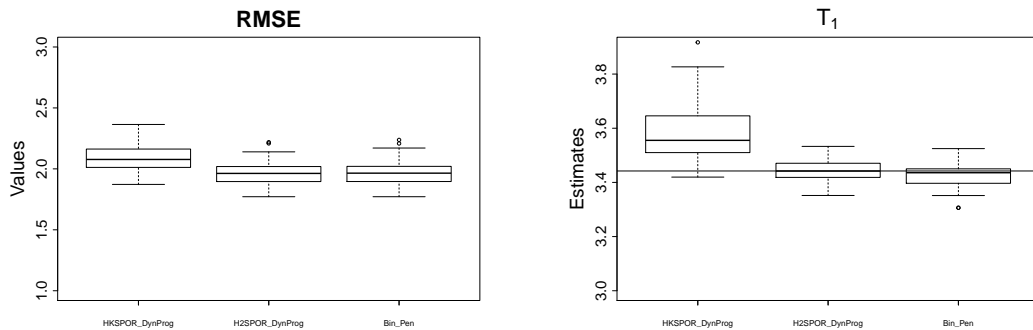


FIGURE 4.17 – Boxplots des estimations de l’instant de transition τ_1 obtenues par HKSPOR_DynProg, H2SPOR_DynProg et Bin_Pen dans le cas où les méthodes trouvent le bon nombre de régimes du modèle **F** (soit $B = 100$) et distribution des RMSE pour $B = 100$ répétitions indépendantes du modèle **F**. La droite horizontale représente la vraie valeur de l’instant de transition τ_1 .

méthodes. Les estimations T_1 sont comparées à la vraie valeur de l'instant de transition τ_1 . La figure 4.18 représente la distribution des temps de calcul nécessaires à chacune des méthodes pour estimer le modèle.

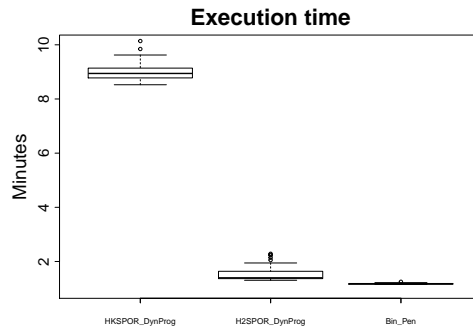


FIGURE 4.18 – Boxplots des temps de calculs nécessaires à HKSPOR_DynProg, H2SPOR_DynProg et Bin_Pen pour estimer les paramètres du modèle, pour $B = 100$ répétitions indépendantes du modèle F.

On voit dans la figure 4.17 que H2SPOR_DynProg et Bin_Pen obtiennent de meilleurs résultats d'estimation de τ_1 que HKSPOR_DynProg. Malgré ces résultats, les RMSE présentés sur cette même figure indiquent que les trois méthodes sont plutôt équivalentes en terme d'ajustement. Pour ce qui est des temps de calcul, la figure 4.18 indique que Bin_Pen est la plus rapide, elle est suivie de très près par H2SPOR_DynProg. La plus lente est HKSPOR_DynProg qui met plus de 8 minutes à estimer le modèle quand les deux autres en mettent moins de 2.

Modèle de simulation E Le modèle E est un modèle de régression polynomiale à trois régimes qui a été défini dans la section 4.2.2. Nous simulons 100 réalisations indépendantes de ce modèle et pour chaque simulation, nous estimons ce modèle avec les trois méthodes. H2SPOR_DynProg et Bin_Pen estiment le bon nombre de régimes dans 76 et 100 simulations respectivement. Dans les 24 autres cas, les modèles estimés par H2SPOR_DynProg contiennent de 2 à 5 régimes, alors que le vrai modèle en possède 3.

Dans la figure 4.19 nous présentons les distributions des RMSE calculés à partir des 100 modèles prédits par les trois méthodes ainsi que les estimations des instants de transitions T_1 et T_2 obtenues par la méthode HKSPOR_DynProg dans tous les cas (pas d'estimation du nombre de régimes) et par les méthodes H2SPOR_DynProg et Bin_Pen lorsqu'elles estiment le bon nombre de régimes. Les distributions des temps de calcul nécessaires à chacune des méthodes pour estimer le modèle sont représentés sur la figure 4.20.

Les résultats observés sur la figure 4.19 montrent que HKSPOR_DynProg est cette fois-ci légèrement meilleure que H2SPOR_DynProg et Bin_Pen, puisque les instants de transition τ_1 et τ_2 sont sur ou sous-estimés avec ces méthodes lorsqu'elles estiment le bon nombre de régimes du modèle E. Les distributions des RMSE présentées dans cette même figure montrent que les méthodes restent équivalentes en terme d'ajustement. Pour ce qui est du temps de calcul, lorsque HKSPOR_DynProg met en moyenne 10 minutes pour estimer le modèle, H2SPOR_DynProg ne met que 2 minutes et Bin_Pen 1 minute pour estimer le modèle.

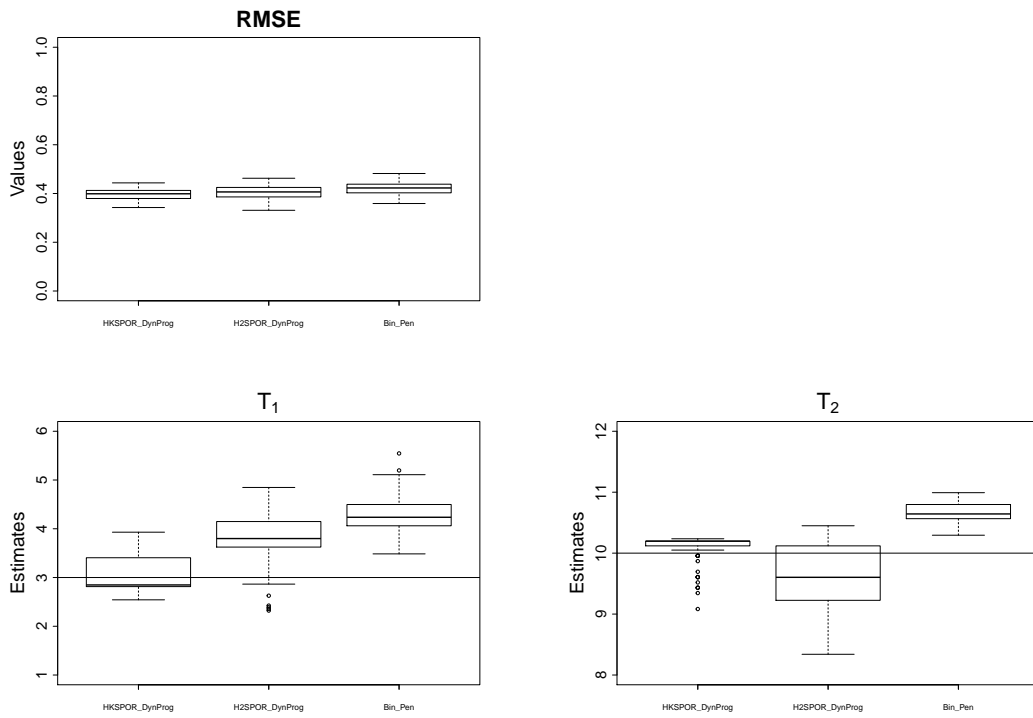


FIGURE 4.19 – Distributions des RMSE calculés pour les trois méthodes pour $B = 100$ répétitions indépendantes du modèle **E** et boxplots des estimations des instants de transition τ_1 et τ_2 obtenues par HKSPOR_DynProg pour $B = 100$ répétitions indépendantes du modèle **E** et par H2SPOR_DynProg et Bin_Pen dans le cas où elles trouvent le bon nombre de régimes, c'est-à-dire pour $B = 76$ et $B = 100$ simulations du modèle **E** respectivement. La droite horizontale représente la vraie valeur du paramètre.

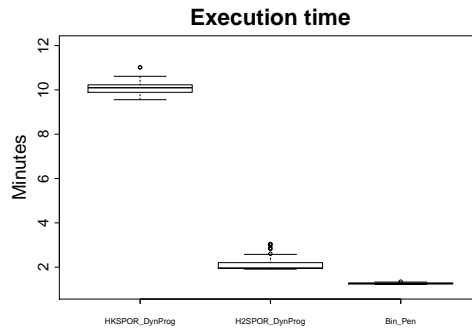


FIGURE 4.20 – Boxplots des temps de calculs nécessaires à HKSPOR_DynProg, H2SPOR_DynProg et Bin_Pen pour estimer les paramètres du modèle, pour $B = 100$ répétitions indépendantes du modèle **E**.

Modèle de simulation G Nous considérons le modèle suivant,

$$Y_i = \sum_{k=1}^4 \left(\rho_{\alpha_k}(X_i) + \sigma_k^2 \right) \mathbb{1}_{[\tau_{k-1}, \tau_k[}(X_i),$$

où le nombre de régimes est fixé à $K = 4$, le degré des polynômes $(\rho_{\alpha_k})_{1 \leq k \leq 4}$ vaut $\delta = 2$ et les coefficients sont $\alpha_2^2 = (-4.98, 3.87, -32.55, 2.07)$, $\alpha_1^1 = (34.28, -43.78, 453.80, -41.39)$ et $\alpha_0^0 = (0, 163.87, -1485.47, 283.93)$. La variance du bruit est identique dans les quatre régimes et vaut $\sigma_1^2 = \dots = \sigma_4^2 = 4$. Les observations $(X_i)_{1 \leq i \leq n}$ avec $n = 200$, sont définies sur $[\tau_0 = 1, \tau_4 = 10]$ et sont réparties de façon régulière. Les instants de transition entre les régimes sont fixés à $\tau_1 = 3.44$, $\tau_2 = 5.66$ et $\tau_3 = 6.97$. Ce modèle a été généré aléatoirement et assure la continuité du modèle au niveau des instants τ_1 , τ_2 et τ_3 .

Le graphique de la figure 4.21 représente des données simulées suivant le modèle G. Nous simu-

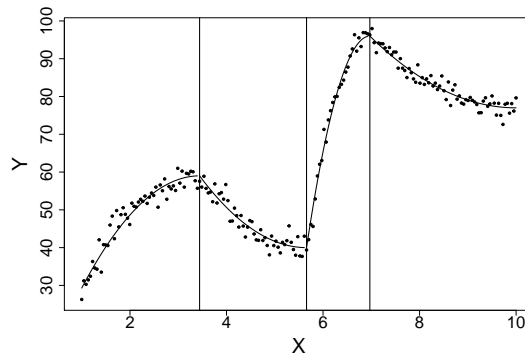


FIGURE 4.21 – Données simulées selon le modèle G. Le vrai modèle est en trait plein et les instants de transition τ_1 , τ_2 et τ_3 sont représentés par les droites verticales.

lons 100 réalisations indépendantes de ce modèle et pour chaque simulation, nous estimons ce modèle avec les trois méthodes. La méthode H2SPOR_DynProg (Bin_Pen respectivement) estime dans 76 cas (1 cas respectivement) le bon nombre de régimes qui est de 4. Bin_Pen estime dans 90% des cas des modèles à 3 régimes (oubli du premier instant de transition), dans les autres cas elle estime des modèles à 1 ou 2 régimes. Les modèles estimés par H2SPOR_DynProg dans les 24 autres cas contiennent de 2 à 4 régimes. La figure 4.22 représente les estimations des instants de transitions τ_1 , τ_2 et τ_3 obtenues par HKSPOR_DynProg pour les 100 répétitions indépendantes du modèle G (cette méthode n'estime pas le nombre de régimes) et par H2SPOR_DynProg et Bin_Pen lorsqu'elles estiment le bon nombre de régimes du modèle G. Cette figure contient également les distributions des RMSE calculés à partir des 100 modèles estimés par les trois méthodes. Enfin, la figure 4.23 correspond aux distributions des temps de calcul nécessaires à chacune des méthodes pour estimer le modèle.

Les résultats de la figure 4.22 montrent encore une fois que HKSPOR_DynProg est légèrement meilleure que H2SPOR_DynProg et Bin_Pen qui vont avoir tendance à sur ou sous-estimer davantage les instants de transition. D'après les distributions des RMSE présentés sur cette même figure, nous pouvons remarquer que les méthodes HKSPOR_DynProg et H2SPOR_DynProg restent équivalentes en terme d'ajustement, ce n'est pas le cas de Bin_Pen qui, en n'estimant pas le bon nombre de régimes, est moins bonne en terme d'ajustement. Pour ce qui est du temps de calcul, nous observons les mêmes ordres de temps de calcul que pour les modèles précédents.

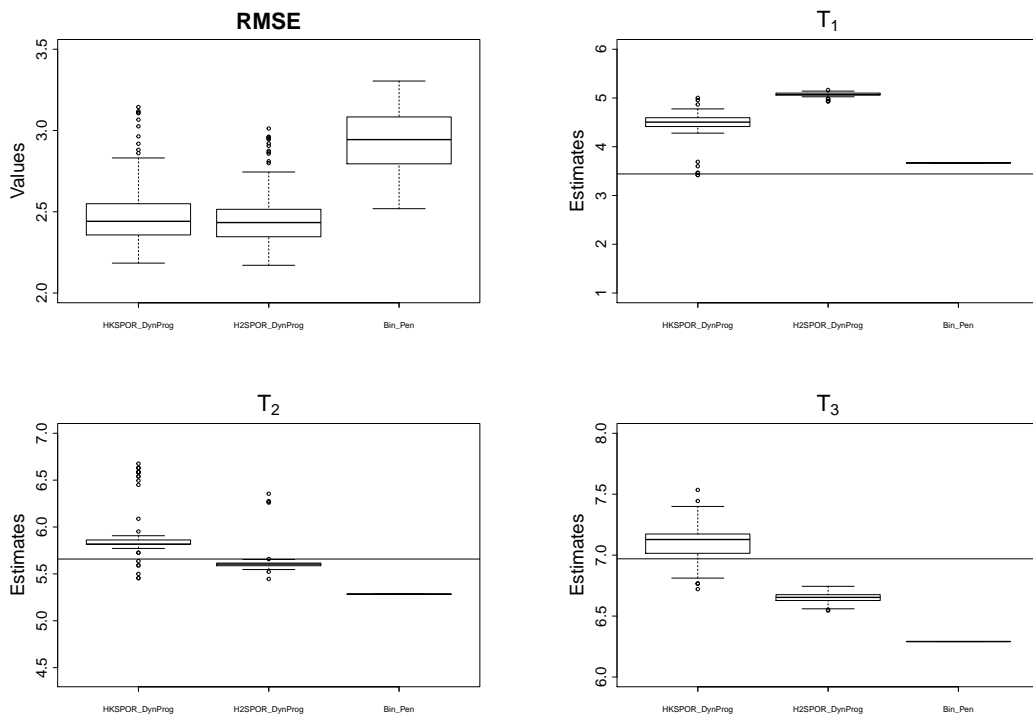


FIGURE 4.22 – Distributions des RMSE calculés pour les trois méthodes pour $B = 100$ répétitions indépendantes du modèle \mathbf{G} et boxplots des estimations des instants de transition T_1 , T_2 et T_3 obtenues par HKSPOR_DynProg pour $B = 100$ répétitions indépendantes du modèle \mathbf{G} et par H2SPOR_DynProg et Bin_Pen lorsqu'elles estiment le bon nombre de régimes, c'est-à-dire pour $B = 76$ et $B = 1$ réalisations du modèle \mathbf{G} respectivement. La droite horizontale représente la vraie valeur du paramètre.

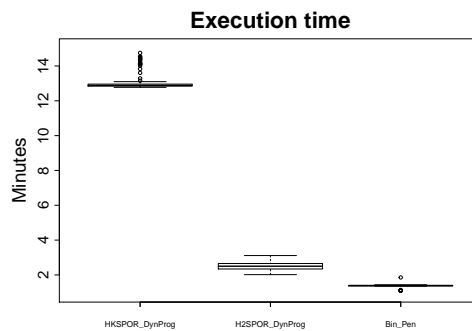


FIGURE 4.23 – Boxplots des temps de calculs nécessaires à HKSPOR_DynProg, H2SPOR_DynProg et Bin_Pen pour estimer les paramètres du modèle, pour $B = 100$ répétitions indépendantes du modèle \mathbf{G} .

Modèle de simulation H Nous considérons le modèle suivant,

$$Y_i = \sum_{k=1}^6 \left(\rho_{\alpha_k}(X_i) + \sigma_k^2 \right) \mathbb{1}_{[\tau_{k-1}, \tau_k[}(X_i),$$

où le nombre de régimes est fixé à $K = 6$, le degré des polynômes $(\rho_{\alpha_k})_{1 \leq k \leq 6}$ vaut $\delta = 2$ et les coefficients sont $\alpha_1^2 = (1.72, -0.41, 1.21, -0.22, 0.55, -1.93)$, $\alpha_1^1 = (-2.11, 1.51, -1.31, -0.16, -0.89, -0.12)$ et $\alpha_1^0 = (-0.69, 9.71, 0.09, 15.96, 5.92, 54.89)$. La variance du bruit est identique dans les six régimes et vaut $\sigma_1^2 = \dots = \sigma_6^2 = 0.04$. Les observations $(X_i)_{1 \leq i \leq n}$ avec $n = 275$, sont définies sur $[\tau_0 = 3, \tau_6 = 5.04]$ et sont réparties de façon régulière. Les instants de transition entre les régimes sont fixés à $\tau_1 = 3.22$, $\tau_2 = 3.45$, $\tau_3 = 3.76$, $\tau_4 = 4.14$ et $\tau_5 = 4.61$. Ce modèle a été généré aléatoirement et assure la continuité du modèle au niveau des instants de transition τ_1 à τ_5 .

Le graphique de la figure 4.24 représente des données simulées suivant le modèle H. Nous simu-

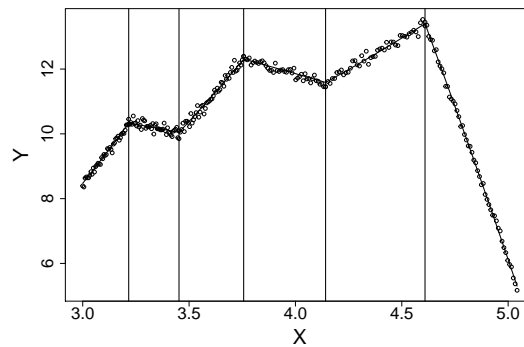


FIGURE 4.24 – Données simulées selon le modèle H. Le vrai modèle est en trait plein et les instants de transition τ_1 à τ_5 sont représentés par les droites verticales.

lons 100 réalisations indépendantes du modèle H. L'allure des données s'apparentant plus à un modèle de degré 1, pour chaque simulation, nous estimons un modèle de degré 1 avec les trois méthodes. H2SPOR_DynProg estime le bon nombre de régimes dans 89% des cas, dans les autres cas le modèle estimé a 7 régimes, alors que le vrai modèle n'en contient que 6. À contrario, Bin_Pen n'estime jamais le bon nombre de régimes puisque dans 100% des cas le modèle estimé ne possède que 2 régimes. La figure 4.25 représente les distributions des RMSE calculés à partir des 100 modèles estimés par les trois méthodes et les estimations des instants de transition τ_1 à τ_5 , obtenues par HKSPOR_DynProg dans les 100 simulations (pas d'estimation du nombre de régimes) et par H2SPOR_DynProg et Bin_Pen lorsqu'elles estiment le bon nombre de régimes. Puisque Bin_Pen estime dans 100% des cas un modèle à deux régimes, nous présentons dans la figure 4.26 l'instant de transition estimé par cette méthode. Enfin, la figure 4.27 représente les distributions des temps de calcul nécessaires à chacune des méthodes pour estimer le modèle.

Les résultats observés sur la figure 4.25 montrent que les estimations des instants de transition sont meilleures avec H2SPOR_DynProg qu'avec HKSPOR_DynProg. Nous remarquons tout de même que la méthode HKSPOR_DynProg ne détecte pas l'instant de transition τ_2 , les estimations T_2 , T_3 et T_4 sont en fait celles de τ_3 , τ_4 et τ_5 et la méthode ajoute un instant de transition à la fin. D'après la figure 4.26, Bin_Pen estime correctement l'instant de transition τ_5 mais passe à côté de tous les autres. Les distributions des RMSE sur la figure 4.25 vont dans le même sens et montrent que H2SPOR_DynProg ajuste mieux ces données. Enfin, en ce qui concerne les temps de calcul, on voit dans la figure 4.27 que Bin_Pen met en moyenne moins d'une minute pour estimer ce modèle. Ce résultat n'est pas surprenant dans le sens où le modèle estimé ne contient que deux régimes. En

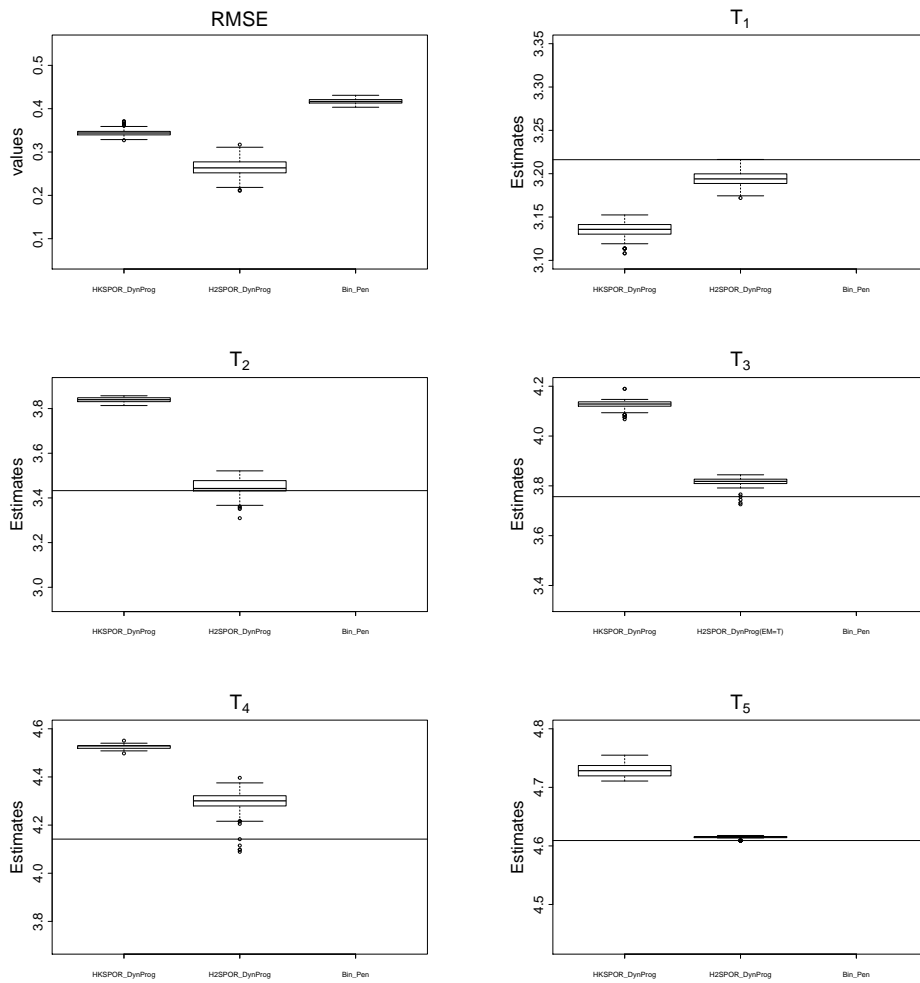


FIGURE 4.25 – Boxplots des RMSE calculés pour les trois méthodes pour $B = 100$ répétitions indépendantes du modèle \mathbf{H} et distributions des estimations des instants de transition T_1 à T_5 obtenues par la méthode HK-SPOR_DynProg, pour $B = 100$ répétitions indépendantes du modèle \mathbf{H} et par les méthodes H2SPOR_DynProg et Bin_Pen lorsqu’elles estiment le bon nombre de régimes du modèle \mathbf{H} , c’est-à-dire pour $B = 89$ et $B = 0$ simulations respectivement. La droite horizontale représente la vraie valeur du paramètre.

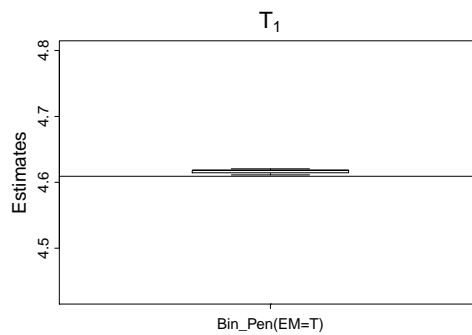


FIGURE 4.26 – Boxplot des estimations de l’instant de transition obtenues par Bin_Pen lorsqu’elle estime un modèle à deux régimes au lieu de six ($B = 100$). La droite horizontale représente la valeur du vrai saut le plus proche soit τ_5 .

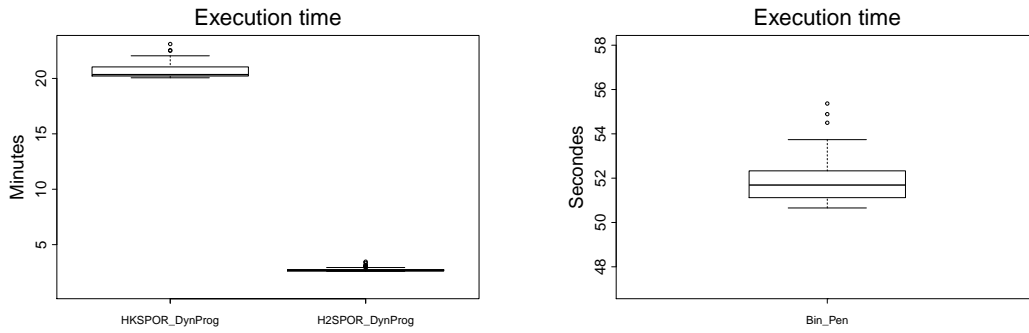


FIGURE 4.27 – Boxplots des temps de calculs nécessaires à HKSPOR_DynProg, H2SPOR_DynProg (en minutes, à gauche) et Bin_Pen (en secondes, à droite) pour estimer les paramètres du modèle, pour $B = 100$ répétitions indépendantes du modèle H .

ce qui concerne les autres méthodes, comme précédemment, H2SPOR_DynProg est plus rapide et met moins de 3 minutes à estimer le modèle quand HKSPOR_DynProg en met un peu plus de 20.

4.2.3 Exemple de cas difficiles pour H2SPOR_DynProg et Bin_Pen

Dans cette section, nous souhaitons étudier les cas où H2SPOR_DynProg et Bin_Pen n'estiment pas correctement le nombre de régimes. Pour ce faire, nous proposons d'étudier 100 modèles différents de la forme suivante,

Modèle de simulation I Nous considérons le modèle suivant,

$$Y_i = \sum_{k=1}^3 \left(\rho_{\alpha_k}(X_i) + \sigma_k^2 \right) \mathbb{1}_{[\tau_{k-1}, \tau_k[}(X_i),$$

où le nombre de régimes est fixé à $K = 3$ et le degré des polynômes $(\rho_{\alpha_k})_{1 \leq k \leq 3}$ vaut $\delta = 2$. Les coefficients α_3 sont générés aléatoirement, mais assurent la continuité de la fonction de lien entre X et Y . La variance du bruit est identique dans les trois régimes et vaut $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 1$. Les observations $(X_i)_{1 \leq i \leq n}$ avec $n = 500$, sont définies sur $[\tau_0 = 1, \tau_3 = 100]$. Les instants de transition τ_1 et τ_2 entre les régimes sont générés aléatoirement, mais vérifient $\tau_1 < \tau_2 < \tau_3$.

Nous simulons 100 modèles de ce type et pour chaque simulation, nous estimons le modèle avec les méthodes H2SPOR_DynProg et Bin_Pen. Elles estiment correctement le nombre de régimes dans 66% et 53% des cas. Nous nous intéressons désormais aux simulations pour lesquelles l'estimation du modèle est incorrecte, c'est-à-dire pour lesquelles le nombre de régimes n'a pas été correctement estimé. Dans la figure 4.28 (figure 4.29 respectivement), 3 (4 respectivement) exemples types de données simulées pour lesquels H2SPOR_DynProg (Bin_Pen respectivement) n'estime pas correctement le nombre de régimes sont présentés. Les graphiques de la figure 4.28 montrent que cette méthode n'estime pas d'instant de transition lorsque les polynômes de deux régimes successifs sont trop proches (graphique en haut à gauche), où lorsqu'un régime ne contient pas assez d'observations (graphique en haut à droite). Au contraire, la méthode va estimer trop de régimes quand les données présentent des ruptures trop abruptes (graphique en bas à gauche). Ce phénomène peut s'expliquer par le fait que lorsqu'un instant de transition est estimé de façon légèrement biaisée, la rupture entre les régimes est toujours observable ce qui entrainera l'estimation d'un nouvel instant de transition qui sera moins biaisé. Ceux de la figure 4.29 montrent que Bin_Pen n'estime pas d'instant de transition lorsque les polynômes de deux régimes sont trop proches (graphiques en haut à gauche et en bas à droite), mais également lorsque les deux modèles aux extrémités sont trop proches ou trop abrupts (graphiques en haut à droite et en bas à gauche).

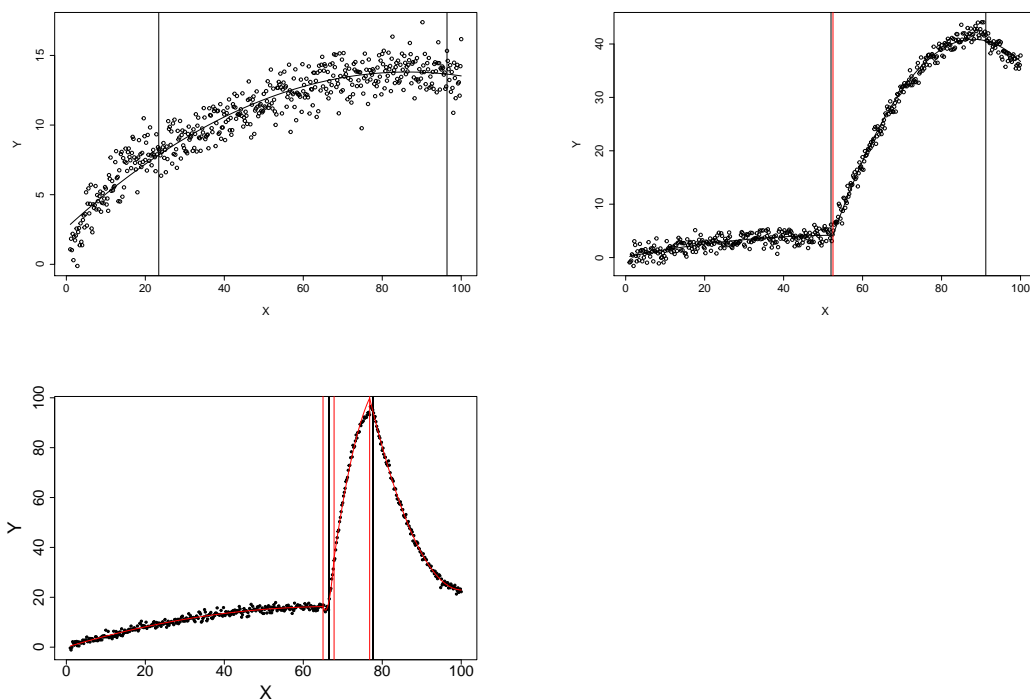


FIGURE 4.28 – Exemples de données simulées (points) pour lesquels H2SPOR_DynProg n'estime pas le bon nombre de régimes. Le modèle estimé par H2SPOR_DynProg est en trait plein. Les droites verticales noires correspondent aux vrais instants de transition, les rouges aux instants estimés par H2SPOR_DynProg.

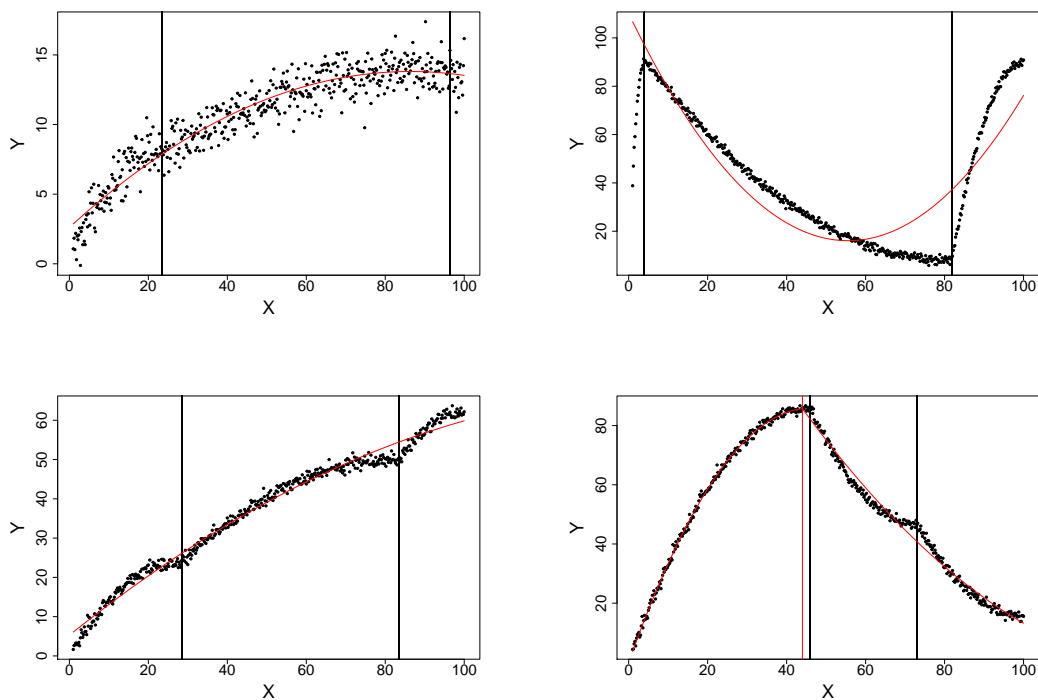


FIGURE 4.29 – Exemples de données simulées (points) pour lesquels Bin_Pen n'estime pas le bon nombre de régimes. Le modèle estimé par Bin_Pen est en trait plein. Les droites verticales noires correspondent aux vrais instants de transition, les rouges aux instants estimés par Bin_Pen.

4.2.4 Estimation a posteriori du nombre de régimes pour HKSPOR_DynProg

Dans cette section, nous proposons une estimation a posteriori du nombre de régimes optimal qui serait retenu par la méthode HKSPOR_DynProg. Pour ce faire, nous réalisons cette estimation pour les modèles **E**, **F** et **G** présentés dans la section précédente. Le nombre de régimes optimal est obtenu à l'aide du BIC, qui nous permet de comparer les modèles estimés par HKSPOR_DynProg avec un nombre de régimes différent.

Modèle E

Nous simulons 100 réalisations indépendantes du modèle **E** qui est un modèle à trois régimes. Pour chaque simulation on estime les paramètres du modèle et les instants de transition avec HKSPOR_DynProg pour K allant de 1 à 5 régimes. On choisit le modèle a posteriori en conservant celui qui minimise le BIC et cette méthode estime le bon nombre de régimes (3) dans tous les cas. Ces résultats sont similaires à ceux obtenus par la méthode Bin_Pen dans les simulations précédentes du modèle **E**, durant lesquelles nous avons déjà montré la qualité d'ajustement aux données des modèles à trois régimes estimés par HKSPOR_DynProg.

Modèle F

Nous simulons 100 réalisations indépendantes du modèle **F** qui est un modèle à deux régimes. Pour chaque simulation on estime les paramètres du modèle et les instants de transition avec HKSPOR_DynProg pour K allant de 1 à 4 régimes. On choisit le modèle a posteriori en conservant celui qui minimise le BIC, nous pouvons voir les résultats dans le tableau 4.2. Ceux-ci indiquent que le nombre de régimes optimal (selon le critère BIC) est de 2 dans 61% des cas. Ces résultats sont moins bons que ceux obtenus par H2SPOR_DynProg et Bin_Pen qui estiment le bon nombre de régimes dans 100% des cas dans les simulations précédentes du modèle **F**. La qualité d'ajustement aux données du modèle à deux régimes estimé par HKSPOR_DynProg a été donné à la section 4.2.2, figure 4.17.

Nombre de régimes	1	2	3	4
Pourcentage de modèles estimés par HKSPOR_DynProg	0	61	38	1

TABLEAU 4.2 – Distribution du nombre de régimes estimés par optimisation du critère BIC pour $B = 100$ réalisations indépendantes du modèle **F** à deux régimes.

Modèle G

Nous simulons 100 réalisations indépendantes du modèle **G** qui est un modèle à quatre régimes. Pour chaque simulation on estime les paramètres du modèle et les instants de transition avec HKSPOR_DynProg pour K allant de 1 à 6 régimes. On choisit le modèle a posteriori en conservant celui qui minimise le BIC et cette méthode estime six régimes (au lieu de 4) dans tous les cas. À l'inverse de Bin_Pen qui sous-estimait constamment le nombre de régimes dans les simulations précédentes du modèle **G**, les modèles retenus selon le critère BIC sur-estiment le nombre de régimes. Selon la forme des données, le critère BIC n'est pas toujours adéquat pour assurer un bon ajustement et une bonne segmentation des données.

4.3 Discussion

H2SPOR Les simulations réalisées sur les modèles de régression polynomiale à deux régimes, dans la section 4.2.1, ont permis de montrer le bon comportement des méthodes à deux régimes présentées dans les sections 2.3.1 et 2.3.2, en particulier lorsque le nombre de données est grand.

Nous avons montré l'avantage d'utiliser un modèle à variables latentes et un algorithme EM pour l'estimation lorsque les données présentent des points levier proches des instants de transitions. De plus, ces méthodes sont capables de détecter un changement de variance entre deux régimes. Enfin, nous avons montré l'intérêt d'avoir un bon a priori sur le modèle afin de ne pas utiliser une contrainte de régularité qui serait non appropriée aux données.

HKSPOR L'intérêt de cette méthode est surtout théorique car trop lente, en particulier à cause de l'optimisation des instants de transition. Les résultats montrent toutefois une certaine efficacité dans l'estimation des paramètres et des instants de transition.

HKSPOR_DynProg Contrairement à **HKSPOR**, **HKSPOR_DynProg** permet un calcul plus rapide des estimateurs. Malgré une étape d'approximation de la vraisemblance, cette méthode fournit des résultats satisfaisants, même si sa robustesse n'a pas été examinée en profondeur dans ce chapitre. Notons que le schéma de contrainte impose de choisir un modèle de degré 3 pour estimer une trajectoire dérivable, alors que toutes les autres méthodes nécessitent seulement un modèle de degré 2. Enfin le critère BIC, utilisé pour sélectionner le nombre de régimes a posteriori, a tendance à sur-estimer le nombre de transitions.

H2SPOR_DynProg Malgré une approximation moins précise de la vraisemblance, **H2SPOR_DynProg** fournit des résultats satisfaisants et cela en un temps 5 fois plus court que **HKSPOR_DynProg** pour $n = 200$. De plus cette méthode permet d'estimer le bon nombre de régimes dans au moins 2/3 des cas dans les simulations effectuées. **H2SPOR_DynProg** est donc sans doute l'algorithme à privilégier dans le contexte de notre étude.

Bin_Pen Les résultats numériques obtenus sont plutôt hétérogènes : excellents sur les modèles **E** et **F** ou très mauvais sur les modèles **G** et **H**. L'intérêt principal de **Bin_Pen** réside dans son caractère évolutif qui pourrait permettre une application à des données on-line.

Chapitre 5

Application sur données réelles

Sommaire

5.1 Données de fluage et de traction	82
5.2 Données de propagation de fissures	86
5.2.1 Données Safran Aircraft Engines	86
5.2.2 Données de Virkler	87
5.2.3 Données de recherche du seuil d'amorçage	89

Dans ce chapitre nous proposons d'appliquer les méthodes HKSPOR_DynProg, H2SPOR_DynProg et Bin_Pen sur des données réelles de la littérature obtenues par Virkler et al. dans [59] et de Safran Aircraft Engines. En ce qui concerne les données de Safran, nous proposons, en plus des données de propagation, d'étudier des données de fluage et de traction qui évoluent selon trois régimes et qui peuvent être modélisées par des polynômes. Sachant que les segmentations des données réelles sont inconnues, ce chapitre n'a pas pour objectif de valider nos méthodes mais simplement de présenter les résultats de modélisation et de segmentation obtenus par nos méthodes sur les différents jeux de données étudiés. Par souci de confidentialité, dans la suite du chapitre l'échelle des graphiques obtenus sur les données de Safran a été volontairement modifiée.

5.1 Données de fluage et de traction

Les performances mécaniques des matériaux mises en jeu à haute température sont dépendantes du temps, il est donc nécessaire de s'y intéresser à haute température, sous sollicitation constante sur des durées importantes. Dans de telles conditions, la déformation progressive du matériau est appelée fluage. Un essai de fluage consiste à mesurer la déformation en fonction du temps. En pratique un essai de fluage consiste à maintenir une éprouvette à haute température sous chargement constant (et non sous contrainte constante) et à mesurer la déformation en fonction du temps. L'allure classique d'une courbe de fluage est représentée dans la figure 5.1. Cette courbe permet de distinguer trois stades de fluage avant la rupture :

- le stade de fluage primaire : décroissance de la vitesse de déformation avec le temps;
- le stade de fluage secondaire : évolution constante de la vitesse de déformation;
- le stade de fluage tertiaire : apparition de striction ou formation de vide interne à l'éprouvette.

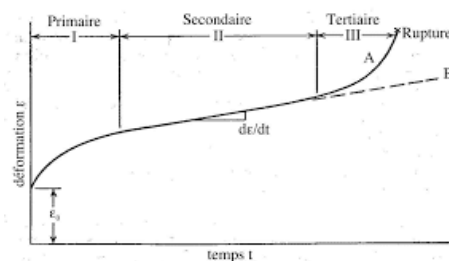


FIGURE 5.1 – Illustration schématique des différents stades du fluage. La courbe A correspond à un essai sous chargement constant; la courbe B correspond à un essai sous contrainte constante.

Source web : *Essai de fluage*.

Un essai de traction permet de déterminer le comportement d'un matériau face au différentes sollicitations qu'il peut rencontrer en période d'utilisation. Cet essai, détaillé en annexe A.2, fournit en sortie des données d'allongement à la rupture en fonction de la force appliquée. L'allure classique d'une courbe de traction représentée sur la figure A.4 laisse observer trois zones de traction. Les données obtenues en sortie de ces deux types d'essais (fluage et traction) semblent pouvoir être modélisées par un modèle de régression polynomiale par morceaux, c'est pourquoi nous proposons de tester nos méthodes sur les données de ces essais fournies par Safran Aircraft Engines.

Les figures 5.2 et 5.4 représentent les segmentations et ajustements obtenus par les trois méthodes lorsqu'elles estiment un modèle de degré 2 sur quatre essais de fluage et trois essais de traction respectivement. Pour HKSPOR_DynProg, le nombre de régimes est a priori fixé à $K = 3$ (résultat attendu). Nous proposons de comparer la qualité des ajustements obtenus à partir des distributions des RMSE représentées dans les figures 5.2 et 5.4.

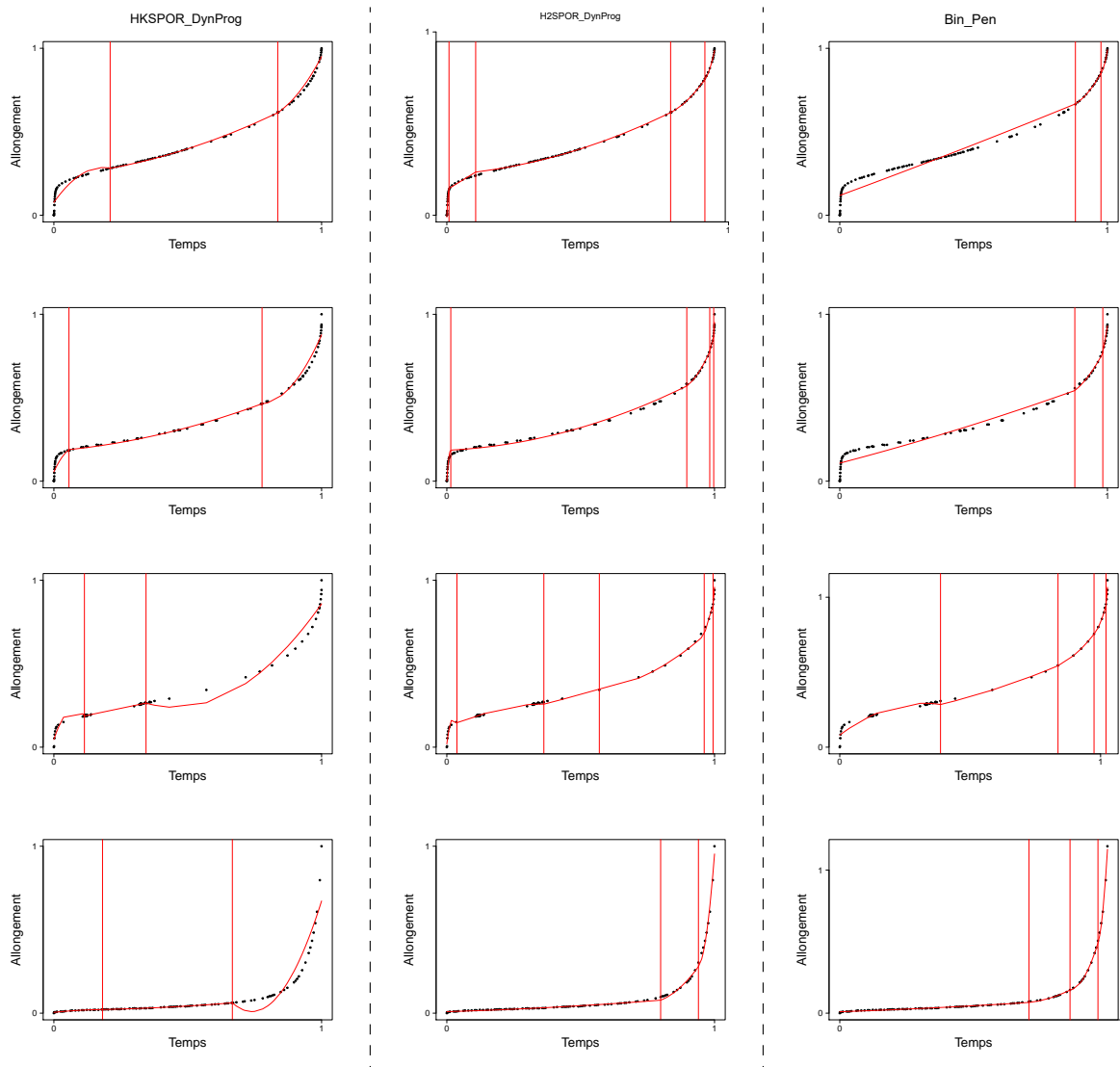


FIGURE 5.2 – Ajustements et segmentations de quatre essais de fluage (un essai par ligne) par les méthodes HKSPOR_DynProg (à gauche), H2SPOR_DynProg (au milieu) et Bin_Pen (à droite) lorsqu’elles estiment un modèle de degré 2. Les courbes d’ajustements et les instants de transition estimés par les méthodes sont représentés en rouge.

Source des données : Safran Aircraft Engines.

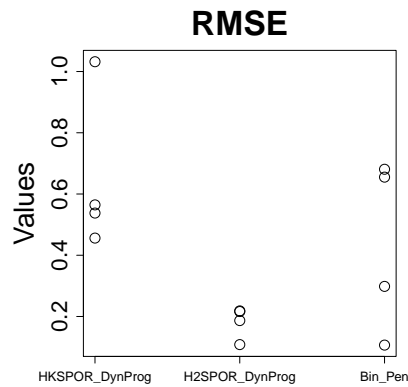


FIGURE 5.3 – Distributions des RMSE calculés à partir des valeurs prédites par les modèles estimés par HKSPOR_DynProg, H2SPOR_DynProg et Bin_Pen, pour B = 4 essais de fluage.

La figure 5.2 montre des résultats intéressants avec HKSPOR_DynProg (colonne de gauche), à l'exception des essais trois (ligne 3) et quatre (ligne 4) pour lesquels la méthode estime un peu trop tôt le deuxième instant de transition par rapport aux essais précédents. Nous pouvons d'ailleurs noter que, contrairement aux autres essais, le premier régime n'est pas vraiment présent dans le quatrième essai. La méthode H2SPOR_DynProg sur-estime dans la majorité des cas le nombre de régimes indiqué par Safran (soit $K = 3$). C'est également le cas de Bin_Pen pour les troisième et quatrième essais de fluage. Dans les deux premiers cas, Bin_Pen estime le bon nombre de régimes, nous pouvons tout de même considérer que la segmentation obtenue n'est pas correcte puisque l'asymptote verticale observée au début de ces données n'est pas correctement ajustée par la méthode. Ce résultat n'est d'ailleurs pas surprenant car les polynômes de faible degré ne permettent pas de capturer cette asymptote. Enfin, la répartition des données du troisième essai est un peu différente de celle des deux autres et semble avoir une influence sur les ajustements et segmentations obtenus par les différentes méthodes. La figure 5.3 indique une meilleure qualité d'ajustement pour H2SPOR_DynProg, ce résultat est lié au nombre plus grand de régimes estimés (en particulier pour le troisième essai).

La figure 5.4 montre des résultats intéressants avec la méthode HKSPOR_DynProg (colonne de gauche), en particulier le troisième essai de traction pour lequel la méthode estime des régimes similaires à ceux de la courbe de traction de la figure A.4. Dans les deux autres essais la décroissance du troisième régime n'est pas suffisamment importante pour engendrer l'estimation d'un saut. H2SPOR_DynProg sur-estime très fortement le nombre de régimes (24 régimes pour le premier essai et 9 pour les autres) indiqué par Safran (soit $K = 3$). La méthode Bin_Pen estime le premier régime mais pas le second. Comme précédemment, la figure 5.5 indique une meilleure qualité d'ajustement pour H2SPOR_DynProg, ce résultat est lié à l'estimation d'un nombre beaucoup plus important de régimes.

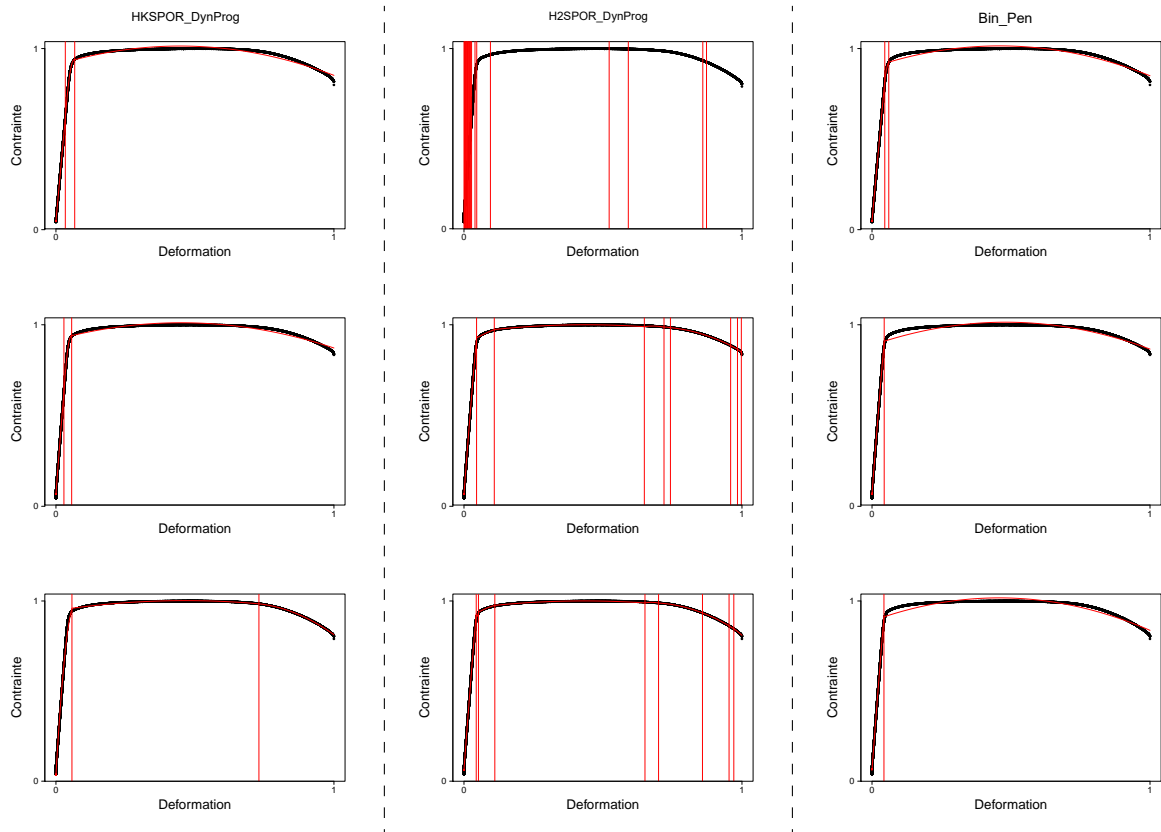


FIGURE 5.4 – Ajustements et segmentations de trois essais de traction (un essai par ligne) par HKSPOR_DynProg (à gauche), H2SPOR_DynProg (au milieu) et Bin_Pen (à droite) lorsqu’elles estiment un modèle de degré 2. Les courbes d’ajustements et les instants de transition estimés par les méthodes sont représentés en rouge.

Source des données : Safran Aircraft Engines.

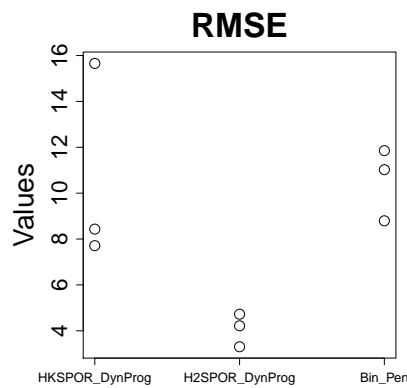


FIGURE 5.5 – Distributions des RMSE calculés à partir des valeurs prédites par les modèles estimés par HKSPOR_DynProg, H2SPOR_DynProg et Bin_Pen, pour B = 3 essais de traction.

5.2 Données de propagation de fissures

Dans cette section nous appliquons nos méthodes sur des données réelles de propagation de fissures fournies par Safran Aircraft Engines, cette étude fait l'objet de la prochaine section, mais également sur des données de la littérature, qui font l'objet de la section 5.2.2.

5.2.1 Données Safran Aircraft Engines

Dans cette section nous appliquons les méthodes HKSPOR_DynProg, H2SPOR_DynProg et Bin_Pen sur des données réelles de propagation de fissures fournies par Safran. Nous disposons de 19 essais de propagation de fissures courtes réalisés sur des éprouvettes d'inconel 718, dans des conditions identiques de température $T = 450$ degrés celsius et de rapport de charge $R = 0.05$. Dans la figure 5.6, nous présentons trois de ces essais de propagation et les segmentations et ajustements associés. Ces essais sont représentés en vitesse $\frac{da_t}{dt}$ en fonction du facteur d'intensité des contraintes ΔK_t en échelle bi-logarithmique. La figure 5.7 présente les distributions des RMSE des trois méthodes calculés pour les 19 essais.

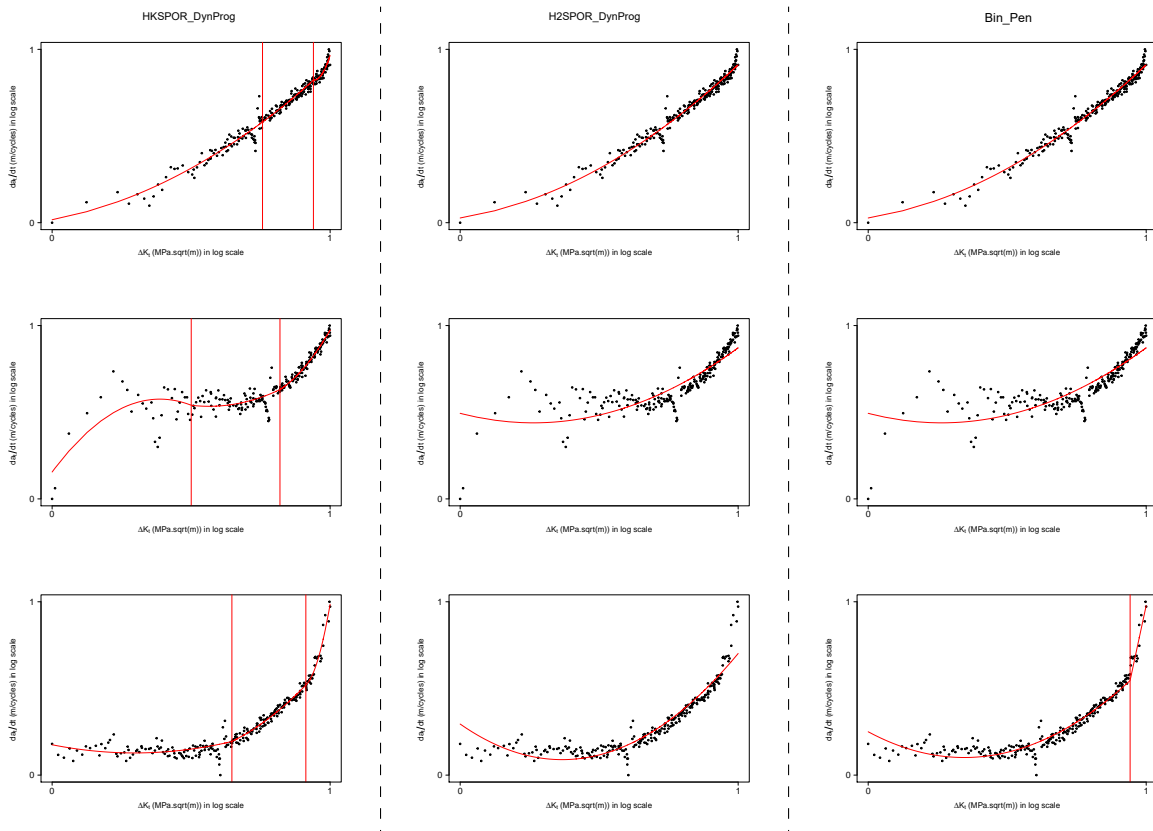


FIGURE 5.6 – Ajustements et segmentations de trois essais de propagation (un essai par ligne) en vitesse en fonction du facteur d'intensité des contraintes par les méthodes HKSPOR_DynProg (à gauche), H2SPOR_DynProg (au milieu) et Bin_Pen (à droite), lorsqu'elles estiment un modèle de degré 2. Ces essais sont représentés en échelle bi-logarithmique. Les courbes d'ajustements et les instants de transition estimés par les méthodes sont représentés en rouge.

Source des données : Safran Aircraft Engines.

À l'aide du critère BIC, HKSPOR_DynProg estime 3 régimes dans 100% des cas. H2SPOR_DynProg et Bin_Pen estiment majoritairement un seul régime. En effet, H2SPOR_DynProg estime un seul régime dans 100% des cas et Bin_Pen estime un seul régime dans 14/19 essais et deux régimes dans 5/19 essais. Les résultats de la figure 5.6 montrent que les instants de transitions estimés par HKSPOR_DynProg sont souvent représentatifs d'une rupture dans les données, cette rupture

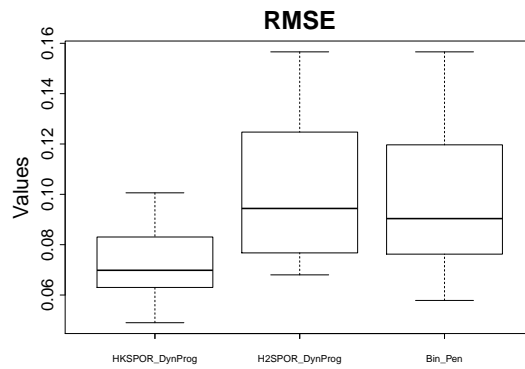


FIGURE 5.7 – Distributions des RMSE calculés à partir des valeurs prédites par les modèles estimés par les méthodes HKSPOR_DynProg, H2SPOR_DynProg et Bin_Pen, pour B = 19 essais de propagation de fissures de fatigue.

n'est pas captée par les autres méthodes. Les distributions des RMSE représentés dans la figure 5.7 indiquent que les trois méthodes sont bonnes en terme d'ajustement. HKSPOR_DynProg est légèrement meilleure car elle estime plus de régimes.

5.2.2 Données de Virkler

Dans cette section, nous proposons de tester nos méthodes sur les données de propagation de fissures de la littérature (essais de propagation proposés par Virkler et al. dans [59]) que nous avons présentées dans la section 1.2.3. Nous disposons de 68 essais de propagation de fissures longues réalisés sur des éprouvettes en aluminium dans des conditions d'essais identiques. Dans la figure 5.8 nous présentons 3 des 68 essais de propagation et les segmentations et ajustements associés (estimés en degré 1). Comme précédemment, ces essais sont représentés en vitesse $\frac{da_t}{dt}$ en fonction du facteur d'intensité des contraintes ΔK_t en échelle bi-logarithmique. La figure 5.9 montre les RMSE des trois méthodes calculés pour les 68 essais.

H2SPOR_DynProg estime dans 100% des cas un seul régime dans les données de Virkler. Bin_Pen estime dans 67/68 cas un seul régime et une fois deux régimes. Enfin, HKSPOR_DynProg estime a posteriori deux régimes dans 100% des cas. Les données de Virkler ayant pour objectif de valider le modèle de Paris sur le régime 2, ces données appartiennent majoritairement à ce régime, les résultats obtenus par les méthodes H2SPOR_DynProg et Bin_Pen ne sont pas vraiment surprenants.

HKSPOR_DynProg estime toujours deux régimes sur les données de Virkler nous proposons donc d'analyser l'instant de transition et les caractéristiques de la fissure au niveau de la transition. Ces statistiques sont données dans le tableau 5.1. Cette étude statistique a déjà été réalisée par Azaïs

	Moyenne	Ecart-type	Min	Max
Longueurs de fissure (mm)	25.34	11.31	11.2	42.2
Instants de transition (nombre de cycles)	174552	65584	55221	308158
ΔK_t (MPa \sqrt{m})	14.88	4.55	9.22	21.9
K_C (MPa \sqrt{m})	18.59	5.69	11.53	27.50

TABEAU 5.1 – Statistiques concernant les longueurs de fissure au instants de transition, les instants de transition en termes de nombre de cycles et le facteur d'intensité de contrainte correspondant.

et al. dans [3] après avoir estimé le modèle PDMP qui autorise un saut pour passer du modèle de Paris à celui de Forman (cf. chapitre 1, section 1.3.2) mais également après estimation du modèle

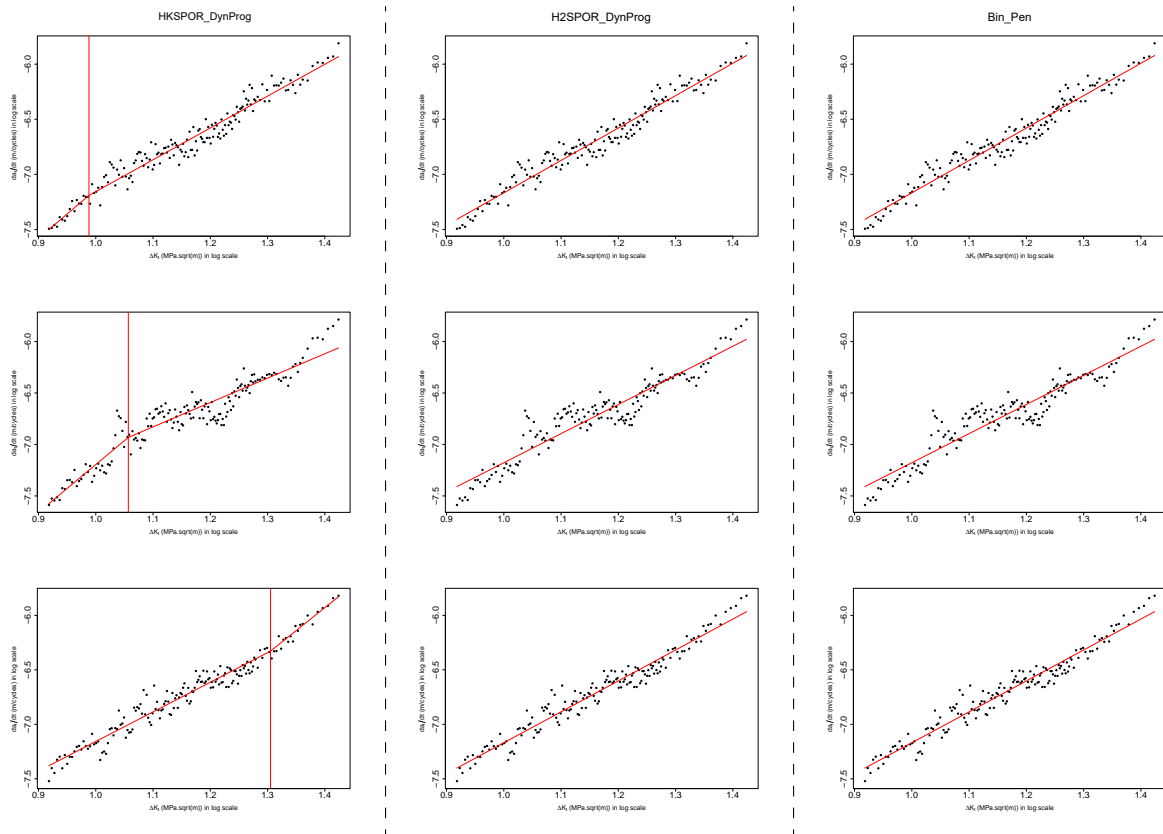


FIGURE 5.8 – Ajustements et segmentations de trois essais de propagation (un essai par ligne) en vitesse en fonction du facteur d'intensité des contraintes par les méthodes HKSPOR_DynProg (à gauche), H2SPOR_DynProg (au milieu) et Bin_Pen (à droite), lorsqu'elles estiment un modèle de degré 1. Ces essais sont représentés en échelle bi-logarithmique. Les courbes d'ajustements et les instants de transition estimés par les méthodes sont représentés en rouge.

Source des données : Virkler et al [59].

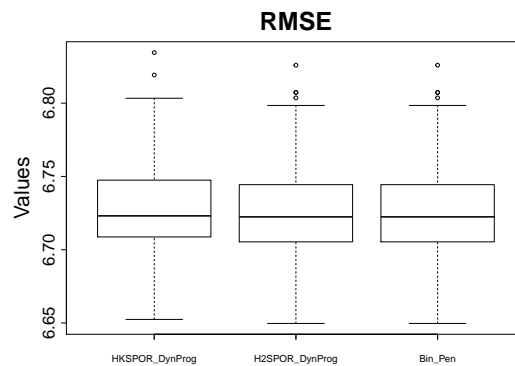


FIGURE 5.9 – Boxplots des RMSE calculés à partir des valeurs prédites par les modèles estimés par les méthodes HKSPOR_DynProg, H2SPOR_DynProg et Bin_Pen, pour B = 68 essais de propagation de fissures de fatigue.

à deux régimes (cf. chapitre 2, section 2.3.1), sur ces mêmes données. Contrairement aux résultats obtenus dans [3], qui allaient en faveur d’une transition tardive dans la vitesse de propagation, ce tableau montre que la méthode HKSPOR_DynProg estime dans 2/3 des cas une transition en début de courbe. Ce résultat est confirmé par la figure 5.10 qui représente les instants de transition estimés par Azaïs et al. dans [3] et celles obtenues par la méthode HKSPOR_DynProg. Nous pouvons tout de même noter que Azaïs et al. dans [3] ajustent les courbes de longueur de fissures en fonction du nombre de cycles et utilisent des lois empiriques précises (Paris puis Forman) pour modéliser ces données. Pour rappel, ces lois empiriques sont conçues pour modéliser l’évolution linéaire de la vitesse dans le régime 2 (Paris) et son augmentation asymptotique dans le régime 3 (Forman), ce qui explique un instant de transition tardif. Cependant, la modélisation des 68 courbes de propagation a nécessité plusieurs semaines de temps de calcul sur serveur. Dans notre cas, HKSPOR_DynProg ajuste les courbes de vitesse de propagation de fissure en fonction du facteur d’intensité des contraintes et l’espace de recherche du modèle de régression polynomiale par morceaux considéré est naïf par rapport au comportement que peut avoir la vitesse de propagation de fissure. Cependant, la modélisation des 68 courbes s’obtient en un temps beaucoup plus raisonnable, c’est-à-dire en moins de trois heures.

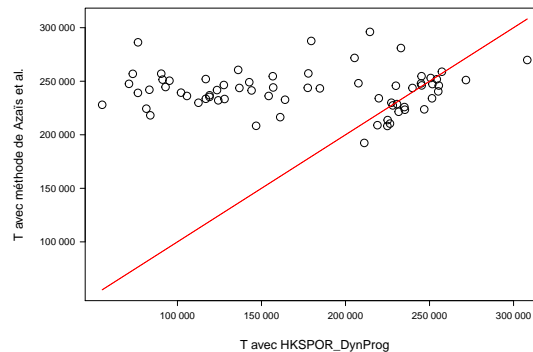


FIGURE 5.10 – Nuage de points des instants de transition estimés par HKSPOR_DynProg (en abscisse) ou par Azaïs et al. dans [3] (en ordonnée).

5.2.3 Données de recherche du seuil d’amorçage

Pour obtenir les données du premier régime de propagation il faut réaliser des essais de seuils d’amorçage de la fissure. Durant ces essais l’objectif est de déterminer le seuil à partir duquel la fissure ne propage pas. Ces essais sont souvent peu nombreux et pas toujours disponibles. Dans la suite, nous présentons deux essais de seuil sur lesquels nous avons appliqué nos méthodes. Ces essais ne sont pas réalisés dans des conditions identiques de température ($T = 350$ degré Celsius pour le premier et $T = 400$ degré Celsius pour le second). La figure 5.11 représente ces deux essais de seuils et les segmentations et ajustements associés obtenus en degré 2. Comme précédemment, ces essais sont représentés en vitesse $\frac{da_f}{dt}$ en fonction du facteur d’intensité des contraintes ΔK_I en échelle bi-logarithmique. Les distributions des RMSE sont représentés sur la figure 5.12.

Les résultats présentés dans la figure 5.11 montrent que selon l’essai, les méthodes n’estiment pas toujours le même nombre de régimes. Les méthodes HKSPOR_DynProg et Bin_Pen n’estiment pas correctement la première partie des données. Ce n’est pas le cas de H2SPOR_DynProg qui estime des transitions en début de propagation. Comme pour les données de fluage, ces résultats s’expliquent par le fait que les polynômes de faible degré ne permettent pas de capturer l’asymptote verticale que nous observons au début des données. Les distributions des RMSE présentés dans la figure 5.12 vont en faveur de H2SPOR_DynProg qui estime un plus grand nombre de régimes.

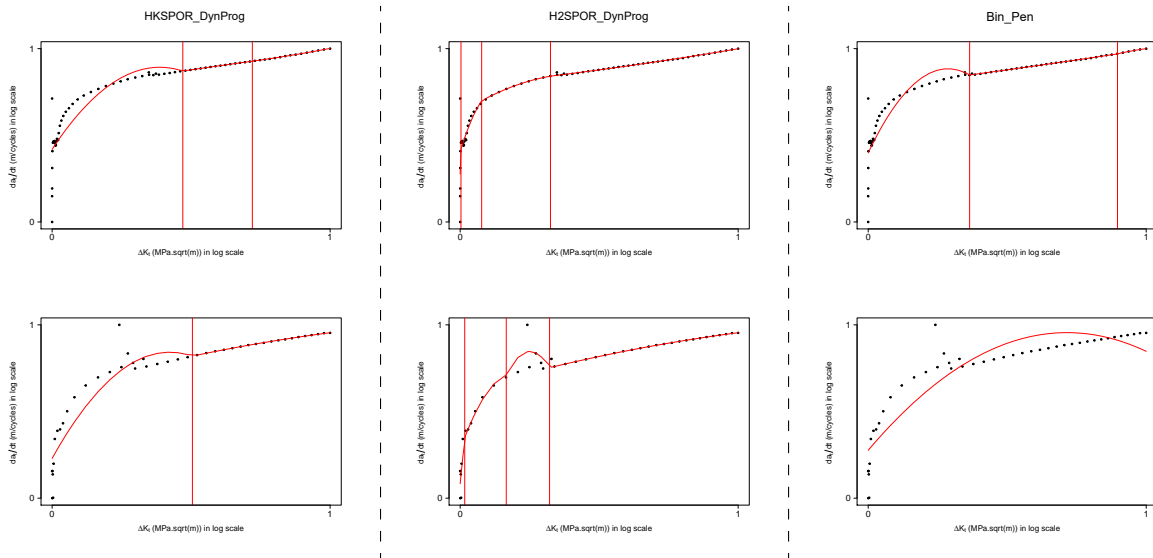


FIGURE 5.11 – Ajustements et segmentations de deux essais de seuil de non propagation (un essai par ligne) en vitesse en fonction du facteur d'intensité des contraintes par les méthodes HKSPOR_DynProg (à gauche), H2SPOR_DynProg (au milieu) et Bin_Pen (à droite), lorsqu'elles estiment un modèle de degré 2. Ces essais sont représentés en échelle bi-logarithmique. Les courbes d'ajustements et les instants de transition estimés par les méthodes sont représentés en rouge.

Source des données : Safran Aircraft Engines.

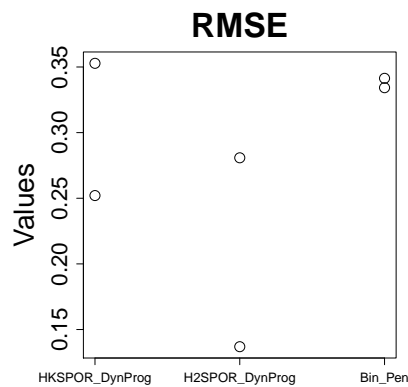


FIGURE 5.12 – Distributions des RMSE calculés à partir des valeurs prédites par les modèles estimés par les méthodes HKSPOR_DynProg, H2SPOR_DynProg et Bin_Pen pour B = 2 essais de seuil de non propagation.

Chapitre 6

Conclusion et perspectives

Sommaire

6.1 Conclusion	92
6.2 Perspectives	93

6.1 Conclusion

Dans le but de caractériser la loi d'un essai de propagation de fissure, c'est-à-dire détecter le nombre de régimes, les instants de transition entre les régimes et la forme paramétrique des différents régimes, nous avons proposé un modèle de régression polynomiale par morceaux soumis à des contraintes de régularité. Ce choix vient de la souplesse de ces modèles qui intervient dans le choix du degré des polynômes mais également des intervalles dans lesquels sont définis ces polynômes. Suite à cela, nous avons proposé plusieurs méthodes qui permettent l'estimation de ce modèle. Chaque nouvelle méthode étant toujours conçue pour réduire les temps de calcul de la méthode précédente et pouvoir être exploitable par Safran Aircraft Engines.

Dans cette thèse, nous nous démarquons de l'état de l'art présenté dans le chapitre 1 par plusieurs points. Nous proposons un modèle de régression polynomiale par morceaux pour décrire les courbes de vitesse de propagation de fissure. Dans l'idée de modéliser un phénomène temporel et continu, nous avons donné la possibilité d'imposer des contraintes de régularité (continuité et/ou dérivabilité) sur la fonction de lien φ . Le statut de la variable latente utilisée pour modéliser l'appartenance au régime n'est pas un processus de Markov mais correspond à un lissage autour des instants de transition que l'on cherche à estimer. Afin d'allier la rapidité d'un algorithme récursif et la souplesse des transitions donnée par un modèle à variable latente, nous avons proposé un algorithme de programmation dynamique pour l'optimisation d'une vraisemblance lissée. Enfin, nous avons proposé deux méthodes d'inférence récursives qui n'utilisent aucun a priori sur le nombre de régimes présents dans les données.

Les comportements des différentes méthodes présentées dans les chapitres 2 et 3 pour estimer un modèle de régression polynomiale par morceaux sous contrainte de régularité, ont été étudiés au travers de plusieurs simulations dans le chapitre 4. Des simulations ont été réalisées sur des modèles à deux régimes. Elles ont permis de montrer le bon comportement de la méthode à deux régimes avec ou sans variable latente, en particulier lorsque le nombre d'observations est grand. Nous avons montré l'intérêt d'utiliser un modèle à variable latente et un algorithme EM pour l'estimation lorsque les données présentent des points aberrants, en l'occurrence les points leviers. Nous avons constaté qu'il est possible de détecter un changement de variance entre deux régimes avec cette méthode H2SPOR. Enfin, nous avons vu qu'il est utile d'avoir un bon a priori sur le modèle afin de ne pas utiliser de contrainte de régularité non appropriée aux données.

Les simulations réalisées sur des modèles à un nombre quelconque de régimes nous ont permis de comparer les différentes méthodes adaptées à ce contexte. Les résultats obtenus indiquent que malgré une certaine efficacité dans l'estimation des paramètres et des instants de transition, la méthode HKSPOR est trop lente pour être utilisée et n'est vraiment intéressante que d'un point de vue théorique. Contrairement à HKSPOR, la méthode HKSPOR_DynProg réalise une étape d'approximation de la log-vraisemblance. Malgré cela, cette méthode fournit des résultats satisfaisants et permet un calcul plus rapide des estimateurs. Nous pouvons tout de même noter que sa robustesse n'a pas été examinée en profondeur dans les simulations. De plus, le critère BIC, utilisé pour sélectionner le nombre de régimes a posteriori a tendance à sur-estimer le nombre de transitions. Comme HKSPOR_DynProg, la méthode H2SPOR_DynProg approche également la vraisemblance mais de manière moins précise. Malgré cela, les résultats obtenus sont satisfaisants, d'autant plus qu'ils sont obtenus en un temps 5 fois plus court qu'avec HKSPOR_DynProg, lorsque $n = 200$. De plus, contrairement à HKSPOR_DynProg qui utilise un a priori sur le nombre de régimes, cette méthode estime ce nombre et obtient de bons résultats dans au moins 2/3 des cas dans les simulations effectuées. Les cas limites pour lesquels la méthode n'estime pas le bon nombre de régimes sont lorsque les polynômes sont trop proches dans les différents régimes (estimation d'un modèle à un régime) ou lorsque les ruptures sont trop abruptes, ce qui entraîne une sur-estimation du nombre de régimes. Enfin, les résultats obtenus pour Bin_Pen sont plutôt hétérogènes. La méthode peut être excellente comme très mauvaise. L'intérêt de cette méthode réside simple-

ment dans son caractère évolutif qui pourrait permettre une application à des données on-line. H2SPOR_DynProg est sans doute l'algorithme à privilégier dans le contexte de notre étude.

Dans le chapitre 5 nous avons testé nos méthodes récursives, c'est-à-dire HKSPOR_DynProg, H2SPOR_DynProg et Bin_Pen sur des données réelles de la littérature ou de Safran Aircraft Engines. Cette étude était simplement illustrative dans le sens où les segmentations réelles de ces données sont inconnues. Dans ce chapitre nous avons présenté les ajustements aux données obtenus par ces trois méthodes. Les différents résultats obtenus nous permettent tout de même de conclure que les polynômes de degré faible ne parviennent pas à saisir des variations trop brutales et ne sont donc pas adaptés pour modéliser les asymptotes observées sur les données de fluage et de seuil de non propagation.

Enfin, à l'exception de Bin_Pen qui est issue d'une approche plus empirique pour laquelle nous n'avons pas obtenu de garanties théoriques ou numériques suffisantes, l'ensemble de ces méthodes ont été implémentées dans un package R qui se nomme HSPOR et qui est disponible sur le CRAN.

6.2 Perspectives

Perspectives scientifiques Le choix du degré des polynômes reste une perspective. En effet, celui-ci était fixé dans les différentes analyses que nous avons effectuées, mais il pourrait tout aussi bien être estimé a posteriori à l'aide d'un critère de sélection de modèle tel que le BIC. Dans le contexte de notre étude, il est tout de même important de ne pas choisir un degré trop élevé car dans ces cas là, un seul polynôme suffirait à décrire l'ensemble de la courbe, ce qui ne correspond à notre objectif de détection des différents régimes.

La fonction de lien φ est soumise à des contraintes de régularité (continuité et/ou dérivabilité). Cependant, il n'est pas toujours évident de savoir a priori si la fonction étudiée est dérivable ou non. Il serait donc intéressant de donner la possibilité à l'algorithme de tester les deux contraintes, de les comparer en terme de RMSE et de ne donner en sortie que le meilleur résultat. De plus, nous pourrions imposer de nouvelles contraintes, par exemple que la fonction soit doublement différentiable. L'ajout d'une contrainte de ce type aura un impact sur la complexité des méthodes (la matrice à inverser sera de taille plus grande), mais également sur le degré des polynômes à partir duquel elle sera applicable.

Nous avons proposé de traiter l'optimisation des instants de transition par une recherche trichotomique pour ces bonnes propriétés d'optimisation. Nous pourrions effectuer une recherche exhaustive mais cela serait plus coûteux numériquement ou alors tester une autre méthode d'optimisation. C'est une piste à explorer.

La robustesse de la méthode HKSPOR_DynProg n'a pas été testée durant les simulations. Il serait pertinent de le faire en réalisant davantage de simulations. De plus, la sélection du nombre de régimes se fait a posteriori par le critère BIC. Comme nous l'avons énoncé dans la conclusion, ce critère a tendance à sur-estimer le nombre de régimes. Il serait donc intéressant d'intégrer le calcul du BIC dans l'algorithme pour pouvoir étudier le pourcentage d'amélioration du BIC entre un modèle à K régimes et un modèle à $K + 1$ régimes. Ce pourcentage d'amélioration est déjà intégré dans les règles d'arrêts des méthodes H2SPOR_DynProg et Bin_Pen et a effectivement permis de diminuer le nombre de régimes retenus. Notons que ce pourcentage est aujourd'hui fixé au sein de ces algorithmes. Il faudrait étudier l'influence du nombre d'observations et/ou du nombre de régimes (réels et/ou estimés) du modèle sur ce pourcentage pour pouvoir, à terme, le rendre dépendant de ces paramètres.

En ce qui concerne la méthode H2SPOR_DynProg, lorsqu'elle sur-estime le nombre de régimes en présence d'une rupture abrupte de la pente entre deux régimes, c'est souvent lié à une première estimation biaisée d'un instant de transition que la méthode corrige en ajoutant une nouvelle transition. Pour pallier ce problème, il serait intéressant de valider a posteriori le ou les premiers instants estimés par la méthode. Comme les instants de transitions ne sont pas estimés dans l'ordre, pour pouvoir faire cette étape de validation a posteriori il faudrait ajouter dans le tableau des résultats obtenu en sortie, l'étape à laquelle chaque instant a été estimé. La grande difficulté de cette étude est de décider du nombre d'instant à valider a posteriori.

Enfin, la méthode Bin_Pen a l'avantage d'être évolutive. Il serait donc intéressant de trouver une nouvelle pénalité qui aurait le comportement asymptotique voulu et qui permettrait d'améliorer les résultats d'estimation. À terme, nous pourrions implémenter cette méthode dans le package HSPOR.

Perspectives industrielles Dans l'objectif d'étudier l'impact de la température et de la taille de grain du matériau sur la dispersion observée dans les données de vitesse de propagation de fissure, il est nécessaire de disposer de plus de données réelles de propagation mais également de détection du seuil de non propagation. À partir de ces données, nous pourrions calculer des modèles moyens par température et par taille de grain. Un modèle moyen se définit par la loi des instants de transition et la loi des paramètres de chaque régime qui ont été calculés sur des essais réalisés dans des conditions identiques. L'impact de la température ou de la taille grain s'observerait au travers de l'évolution de ces lois.

En ce qui concerne la modélisation des données de propagation, le choix des polynômes est peut-être à revoir, dans le sens où ils ne sont pas adaptés pour saisir les variations brutales comme celles observées dans les données de seuil de non propagation, mais également car ils sont naïfs par rapport au comportement de la vitesse de propagation de fissure. Il serait intéressant de remplacer ces polynômes par un ou plusieurs modèles déterministes de type Paris, Forman ou encore Walker (cf. chapitre 1, section 1.2.2) qui sont bien connus des mécaniciens. Ce remplacement entraînera la réécriture de la vraisemblance et l'estimation des coefficients des polynômes sera remplacée par celle des coefficients de la loi de Paris (Forman ou Walker). Pour utiliser ces méthodes il faudra faire le calcul théorique des paramètres qui maximisent cette vraisemblance.

Bibliographie

- [1] R. AZAIS et A. MULLER-GUEUDIN. « Optimal choice among a class of nonparametric estimators of the jump rate for piecewise-deterministic Markov processes ». In : *Electronic Journal of Statistics* 10.2 (2016), p. 3648-3692.
- [2] R. AZAIS et al. « Estimation, simulation et prévision d'un modèle de propagation de fissures par des processus markoviens déterministes par morceaux ». In : *Actes du congrès lambda-mu 17 17e Congrès de Maîtrise des Risques et de Sécurité de Fonctionnement 5-7 octobre 2010 La Rochelle*. France, 2010, H5-C3.
- [3] Romain AZAIS, Anne GÉGOUT-PETIT et Florine GRECIET. « Rupture Detection in Fatigue Crack Propagation ». In : *Statistical Inference for Piecewise-deterministic Markov Processes* (2018), p. 173-207.
- [4] Leonard E BAUM et al. « A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains ». In : *The annals of mathematical statistics* 41.1 (1970), p. 164-171.
- [5] Michael BEIL et al. « Simulating the formation of keratin filament networks by a piecewise-deterministic Markov process ». In : *Journal of theoretical biology* 256.4 (2009), p. 518-532.
- [6] Richard BELLMAN. « On the theory of dynamic programming ». In : *Proceedings of the National Academy of Sciences of the United States of America* 38.8 (1952), p. 716.
- [7] Richard BELLMAN et Robert ROTH. « Curve fitting by segmented straight lines ». In : *Journal of the American Statistical Association* 64.327 (1969), p. 1079-1084.
- [8] A. BEN ABDESSALEM et al. « Stochastic modelling and prediction of fatigue crack propagation using piecewise-deterministic Markov processes ». In : *Journal of Risk and Reliability* 230.4 (2016), p. 405-416.
- [9] J.L. BOGDANOFF et F. KOZIN. *Probabilistic models of cumulative damage*. A Wiley-Interscience publication. Etats-Unis : John Wiley & Sons, 1985, p. 341. ISBN : 978-0-471-88180-3.
- [10] Victor L BRAILOVSKY et Yulia KEMPNER. « Application of piece-wise regression to detecting internal structure of signal ». In : *Pattern recognition* 25.11 (1992), p. 1361-1370.
- [11] A. BRANDEJSKY, B. de SAPORTA et F. DUFOUR. « Optimal stopping for partially observed piecewise-deterministic Markov processes ». In : *Stochastic Processes and their Applications* 123.8 (2013), p. 3201-3238.
- [12] Faïcel CHAMROUKHI. « Unsupervised learning of regression mixture models with unknown number of components ». In : *Journal of Statistical Computation and Simulation* 86.12 (2016), p. 2308-2334.
- [13] Faïcel CHAMROUKHI et al. « A regression model with a hidden logistic process for signal parametrization ». In : *arXiv preprint arXiv:1312.6994* (2013).
- [14] Faïcel CHAMROUKHI et al. « Time series modeling by a regression approach based on a latent process ». In : *Neural Networks* 22.5-6 (2009), p. 593-602.

- [15] J. CHIQUET et N. LIMNIOS. « Estimating stochastic dynamical systems driven by a continuous-time jump Markov process ». In : *Methodology and Computing in Applied Probability* 8.4 (2006), p. 431-447. ISSN : 1573-7713. DOI : [10.1007/s11009-006-0423-z](https://doi.org/10.1007/s11009-006-0423-z). URL : <https://doi.org/10.1007/s11009-006-0423-z>.
- [16] J. CHIQUET, N. LIMNIOS et M. EID. « Piecewise deterministic Markov processes applied to fatigue crack growth modelling ». In : *Journal of Statistical Planning and Inference* 139.5 (2009), p. 1657-1667. ISSN : 0378-3758. DOI : <http://dx.doi.org/10.1016/j.jspi.2008.05.034>.
- [17] Thomas CORMEN. « Introduction à l'algorithmique ». In : (1997).
- [18] M.H.A. DAVIS. *Markov models and optimization*. T. 49. Monographs on Statistics and Applied Probability. London : Chapman & Hall, 1993, p. xiv+295. ISBN : 0-412-31410-X.
- [19] M.H.A. DAVIS. « Piecewise-Deterministic Markov-Processes - A General-Class of Non-Diffusion Stochastic-Models ». In : *Journal Of The Royal Statistical Society. Series B-Methodological* 46.3 (1984), p. 353-388.
- [20] B. DE SAPORTA, F. DUFOUR et K. GONZALEZ. « Numerical method for optimal stopping of piecewise deterministic Markov processes ». In : *Ann. Appl. Probab.* 20 (2010), p. 1607-1637. DOI : [10.1214/09-AAP667](https://doi.org/10.1214/09-AAP667).
- [21] Arthur P DEMPSTER, Nan M LAIRD et Donald B RUBIN. « Maximum likelihood from incomplete data via the EM algorithm ». In : *Journal of the Royal Statistical Society: Series B (Methodological)* 39.1 (1977), p. 1-22.
- [22] F. DUFOUR et Y. DUTUIT. « Dynamic Reliability: A new model ». In : *Proceedings of $\lambda\mu$ 13-ESREL02, Lyon, France*. 2002.
- [23] Anthony WF EDWARDS et Luigi Luca CAVALLI-SFORZA. « A method for cluster analysis ». In : *Biometrics* (1965), p. 362-375.
- [24] Giancarlo FERRARI-TRECATE et Marco MUSELLI. « A new learning method for piecewise linear regression ». In : *International conference on artificial neural networks*. Springer. 2002, p. 444-449.
- [25] R. FORMAN, V. KEARNEY et R. ENGLE. « Numerical analysis of crack propagation in cyclic-loaded structures ». In : *Journal of basic Engineering* 89.3 (1967), p. 459-463. DOI : [10.1115/1.3609637](https://doi.org/10.1115/1.3609637). URL : <http://dx.doi.org/10.1115/1.3609637>.
- [26] R.G. FORMAN et S.R. METTU. *Behavior of surface and corner cracks subjected to tensile and bending loads in Ti-6Al-4V alloy*. Rapp. tech. National Aeronautics et Space Administration, 1990, p. 59.
- [27] H. GHONEM et S. DORE. « Experimental study of the constant-probability crack growth curves under constant amplitude loading ». In : *Engineering Fracture Mechanics* 27.1 (1987), p. 1-25. DOI : [10.1016/0013-7944\(87\)90002-6](https://doi.org/10.1016/0013-7944(87)90002-6). URL : [https://doi.org/10.1016/0013-7944\(87\)90002-6](https://doi.org/10.1016/0013-7944(87)90002-6).
- [28] A. IVANOVA et A. NAESS. « Importance sampling for dynamic systems by approximate calculation of the optimal control function ». In : *Proc. 4th Int. Conf. on Mathematical Methods in Reliability*. 2004.
- [29] R.Z. KHAS'MINSKII. « A limit theorem for the solutions of differential equations with random right-hand sides ». In : *Theory of Probability & Its Applications* 11.3 (1966), p. 390-406.
- [30] R.Z. KHAS'MINSKII. « On stochastic processes defined by differential equations with a small parameter ». In : *Theory of Probability & Its Applications* 11.2 (1966), p. 211-228.
- [31] D.J. KIRKNER, K. SOBCZYK et B.F. SPENCER. « On the relationship of the cumulative jump model for random fatigue to empirical data ». In : *Probabilistic engineering mechanics* 14.3 (1999), p. 257-267. DOI : [10.1016/S0266-8920\(98\)00004-6](https://doi.org/10.1016/S0266-8920(98)00004-6). URL : [https://doi.org/10.1016/S0266-8920\(98\)00004-6](https://doi.org/10.1016/S0266-8920(98)00004-6).

- [32] T. KITAMURA, L.J. GHOSN et R. OHTANI. *Stochastic modeling of crack initiation and short-crack growth under creep and creep-fatigue conditions*. Rapp. tech. National Aeronautics et Space Administration, 1989, p. 20.
- [33] Y.K. LIN et Yang J.N. « On statistical moments of fatigue crack propagation ». In : *Engineering Fracture Mechanics* 18.2 (1983), p. 243-256. DOI : [10.1016/0013-7944\(83\)90136-4](https://doi.org/10.1016/0013-7944(83)90136-4). URL : [https://doi.org/10.1016/0013-7944\(83\)90136-4](https://doi.org/10.1016/0013-7944(83)90136-4).
- [34] Victor E MCZGEE et Willard T CARLETON. « Piecewise regression ». In : *Journal of the American Statistical Association* 65.331 (1970), p. 1109-1124.
- [35] John A NELDER et Roger MEAD. « A simplex method for function minimization ». In : *The computer journal* 7.4 (1965), p. 308-313.
- [36] C.C. NI. « Formulation of a Polynomial Stochastic Fatigue Crack Growth Model ». In : *Advanced Materials Research*. T. 909. Trans Tech Publ. 2014, p. 467-471.
- [37] C.C. NI. « Verification of a Polynomial Stochastic Fatigue Crack Growth Model ». In : *2015 International Conference on Intelligent Systems Research and Mechatronics Engineering*. Atlantis Press. 2015.
- [38] R. OHTANI. « Growth and distribution of microcracks at the surface of smooth specimen of 304 stainless steel in creep and effect of high temperature oxidation ». In : *J. Soc. Mater. Sci., Jpn.* 32.357 (1983), p. 635-639.
- [39] R. OHTANI, T. KINAMI et H. SAKAMOTO. « Small crack propagation in high temperature creep-fatigue of 304 stainless steel. » In : *Trans. Japan soc. Mech. Eng.(Ser. A)*. 52.480 (1986), p. 1824-1830.
- [40] G.C. PAPANICOLAOU et W. KOHLER. « Asymptotic theory of mixing stochastic ordinary differential equations ». In : *Communications on Pure and Applied Mathematics* 27.5 (1974), p. 641-668.
- [41] P.C. PARIS et F. ERDOGAN. « A critical Analysis of Crack Propagation Laws ». In : *Journal of Basic Engineering* 85.4 (1963), p. 528-533. DOI : [10.1115/1.3656900](https://doi.org/10.1115/1.3656900). URL : <http://dx.doi.org/10.1115/1.3656900>.
- [42] Emanuel PARZEN. « On estimation of a probability density function and mode ». In : *The annals of mathematical statistics* 33.3 (1962), p. 1065-1076.
- [43] Franck PICARD et al. « A segmentation/clustering model for the analysis of array CGH data ». In : *Biometrics* 63.3 (2007), p. 758-766.
- [44] Richard E QUANDT. « The estimation of the parameters of a linear regression system obeying two separate regimes ». In : *Journal of the American statistical association* 53.284 (1958), p. 873-880.
- [45] A. RAY, S. TANGIRALA et S. PHOHA. « Stochastic modeling of fatigue crack propagation ». In : *Applied Mathematical Modelling* 22.3 (1998), p. 197-204. DOI : [10.1016/S0307-904X\(98\)00013-4](https://doi.org/10.1016/S0307-904X(98)00013-4). URL : [https://doi.org/10.1016/S0307-904X\(98\)00013-4](https://doi.org/10.1016/S0307-904X(98)00013-4).
- [46] Donald Edward ROBISON. « Estimates for the points of intersection of two polynomial regressions ». In : *Journal of the American Statistical Association* 59.305 (1964), p. 214-224.
- [47] Murray ROSENBLATT. « Remarks on some nonparametric estimates of a density function ». In : *The Annals of Mathematical Statistics* (1956), p. 832-837.
- [48] B. de SAPORTA et F. DUFOUR. « Numerical method for impulse control of piecewise deterministic Markov processes. » In : *Automatica* 48.5 (2012), p. 779-793.
- [49] B. de SAPORTA et al. « Optimal stopping for the predictive maintenance of a structure subject to corrosion ». In : *Journal of Risk and Reliability* 226 (2) (2012), p. 169-181.
- [50] Gideon SCHWARZ et al. « Estimating the dimension of a model ». In : *The annals of statistics* 6.2 (1978), p. 461-464.

- [51] Andrew Jhon SCOTT et M KNOTT. « A cluster analysis method for grouping means in the analysis of variance ». In : *Biometrics* (1974), p. 507-512.
- [52] W. SHEN, A.B.O. SOBOYEJO et W.O. SOBOYEJO. « Probabilistic modeling of fatigue crack growth in Ti-6Al-4V ». In : *International journal of fatigue* 23.10 (2001), p. 917-925. DOI : [10.1016/S0142-1123\(01\)00045-7](https://doi.org/10.1016/S0142-1123(01)00045-7). URL : [https://doi.org/10.1016/S0142-1123\(01\)00045-7](https://doi.org/10.1016/S0142-1123(01)00045-7).
- [53] K. SOBCZYK. « On the Markovian models for fatigue accumulation ». In : *Journal de Mécanique Théorique et Appliquée* (1982), p. 147-160.
- [54] K. SOBCZYK et B.F. SPENCER. « Random microstructural effects on fatigue accumulation ». In : *International journal of fatigue* 17.8 (1995), p. 521-530. DOI : [10.1016/0142-1123\(95\)00053-4](https://doi.org/10.1016/0142-1123(95)00053-4). URL : [https://doi.org/10.1016/0142-1123\(95\)00053-4](https://doi.org/10.1016/0142-1123(95)00053-4).
- [55] Henry STONE. « Approximation of curves by line segments ». In : *Mathematics of Computation* (1961), p. 40-47.
- [56] S. TANAKA, M. ICHIKAWA et S. AKITA. « Variability of m and C in the fatigue crack propagation law ». In : *International Journal of Fracture* 17.5 (1981), R121-R124.
- [57] M. TOUZET et al. « Modèle probabiliste de propagation de fissures de fatigue ». In : *20ème Congrès Français de Mécanique, 28 août/2 sept. 2011*. Besançon, France : AFM, Maison de la Mécanique, 39/41 rue Louis Blanc, 92400 Courbevoie, France (FR), 2011.
- [58] Mathieu VAUTROT. « Étude du comportement mécanique des matériaux dans des conditions étendues de vitesses et de températures: application à l'acier C68 dans le cas d'une opération de formage incrémental ». Thèse de doct. 2012.
- [59] D. VIRKLER, B. HILLBERRY et P. GOEL. « The statistical nature of fatigue crack propagation ». In : *J Engng Mater Technol* 101.2 (1979), p. 148-153.
- [60] K. WALKER. « The Effect of Stress Ratio during Crack Propagation and Fatigue for 2024-T3 and 7075-T6 Aluminum ». In : *Effects of environment and complex load history on fatigue life*. ASTM International, 1970.
- [61] W.F. WU et C.C. NI. « A study of stochastic fatigue crack growth modeling through experimental data ». In : *Probabilistic Engineering Mechanics* 18.2 (2003), p. 107-118. DOI : [10.1016/S0266-8920\(02\)00053-X](https://doi.org/10.1016/S0266-8920(02)00053-X). URL : [https://doi.org/10.1016/S0266-8920\(02\)00053-X](https://doi.org/10.1016/S0266-8920(02)00053-X).
- [62] W.F. WU et C.C. NI. « Probabilistic models of fatigue crack propagation and their experimental verification ». In : *Probabilistic Engineering Mechanics* 19.3 (2004), p. 247-257.
- [63] W.F. WU et C.C. NI. « Statistical aspects of some fatigue crack growth data ». In : *Engineering Fracture Mechanics* 74.18 (2007), p. 2952-2963. DOI : [10.1016/j.engfracmech.2006.08.019](https://doi.org/10.1016/j.engfracmech.2006.08.019). URL : <https://doi.org/10.1016/j.engfracmech.2006.08.019>.
- [64] J.N. YANG et S.D. MANNING. « A simple second order approximation for stochastic crack growth analysis ». In : *Engineering Fracture Mechanics* 53.5 (1996), p. 677-686. DOI : [10.1016/0013-7944\(95\)00130-1](https://doi.org/10.1016/0013-7944(95)00130-1). URL : [https://doi.org/10.1016/0013-7944\(95\)00130-1](https://doi.org/10.1016/0013-7944(95)00130-1).
- [65] J.N. YANG et S.D. MANNING. « Stochastic crack growth analysis methodologies for metallic structures ». In : *Engineering Fracture Mechanics* 37.5 (1990), p. 1105-1124. DOI : [10.1016/0013-7944\(90\)90032-C](https://doi.org/10.1016/0013-7944(90)90032-C). URL : [https://doi.org/10.1016/0013-7944\(90\)90032-C](https://doi.org/10.1016/0013-7944(90)90032-C).
- [66] J.N. YANGT. « A stochastic theory of fatigue crack propagation ». In : (1985).
- [67] H. ZHANG et al. « Piecewise deterministic Markov processes and dynamic reliability ». In : *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability* 222.4 (2008), p. 545-551.

Annexe A

Complément d'information

Sommaire

A.1 Essai de propagation de fissure de fatigue	II
A.2 Essai de traction sur un matériau	III
A.3 Deux méthodes d'optimisation du chapitre 2	IV
A.3.1 Recherche trichotomique	IV
A.3.2 Programmation dynamique	V
A.4 Estimation d'un régime sous contraintes	VI
A.4.1 Modèle de régression polynomiale par morceaux sur un régime contraint .	VI
A.4.2 Maximum de vraisemblance à un régime contraint	VII
A.5 Résolution des problèmes d'optimisation sous contrainte de régularité sur le modèle sans variable latente.	IX
A.5.1 Optimisation à deux régimes	IX
A.5.2 Optimisation à un nombre de régimes K quelconque	XI
A.5.3 Optimisation à un régime	XII
A.6 Résolution des problèmes d'optimisation sous contrainte de régularité pour le modèle à variable latente.	XIV
A.7 Complément d'étude du comportement asymptotique de $\mathbb{E}[\hat{\sigma}_{n+m}^2]$	XV
A.8 Documentation package HSPOR	XXIII

A.1 Essai de propagation de fissure de fatigue

Ce type d'essai consiste à soumettre une éprouvette pré-fissurée à un chargement cyclique, afin d'établir certaines propriétés fondamentales des matériaux. Les éprouvettes utilisées pour ces essais de propagation de fissure en fatigue sont définies comme sur le plan de la figure A.1. Les

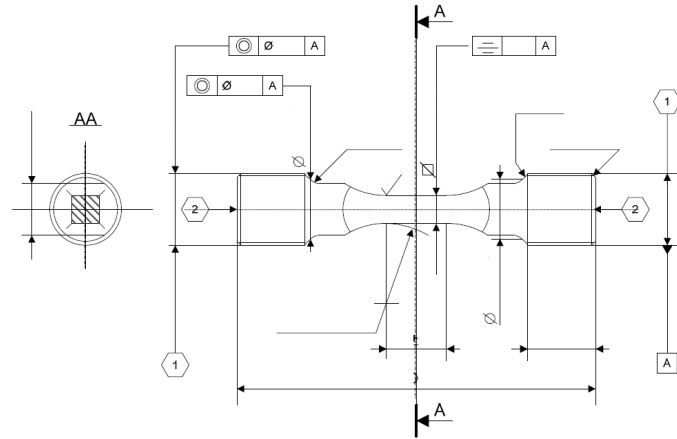


FIGURE A.1 – Plan de référence d'une éprouvette de propagation de fissures.

Source : Safran Aircraft Engines.

essais sont conduits sur des machines asservies, qui doivent pouvoir réaliser plusieurs fonctions comme : maintenir des cycles de charge constants en amplitude et en valeur moyenne pendant toute la durée d'un essai; travailler sur une gamme de fréquence; imposer à l'éprouvette différentes formes de cycles de charge.

Pour préparer l'éprouvette, des fils lui sont fixés pour l'arrivée du courant et la mesure de la différence de potentiel. Le processus de mise en charge doit être identique pour toutes les éprouvettes testées dans les mêmes conditions. La fréquence de l'essai peut varier selon que l'on veut mesurer la vitesse de fissuration à l'air, en milieu corrosif, ou le seuil de non propagation, mais doit cependant rester constante au cours de l'essai. Celui-ci peut être mené soit jusqu'à la rupture de l'éprouvette, soit jusqu'à ce que la fissure atteigne une longueur maximale autorisée par les normes d'applications des concepts de la mécanique de la rupture. Il doit être réalisé de bout en bout sans interruption.

Pour amorcer la fissure de fatigue à partir d'une entaille, les mêmes conditions de charge et de fréquence que celles utilisées pendant l'essai doivent être appliquées. À la fin de l'essai, nous obtenons les relevés de la longueur de fissure, notée a , en fonction du nombre de cycles testés t . La vitesse de propagation de la fissure $\frac{da}{dt}$ est calculée par la méthode de la sécante : à partir de 2 couples (a_i, t_i) et (a_{i+1}, t_{i+1}) on assimile $\frac{da}{dt}$ à la quantité $\frac{\Delta a}{\Delta t} = \frac{a_{i+1} - a_i}{t_{i+1} - t_i}$. Le facteur d'intensité des contraintes ΔK_t est calculé à partir de sa formule (1.3), en considérant une longueur moyenne de fissure $\bar{a}_i = \frac{a_i + a_{i+1}}{2}$.

$$\Delta K_i = Y(\bar{a}_i) \Delta \sigma \sqrt{\pi \bar{a}_i},$$

où $Y(\bar{a}_i)$ est un facteur sans dimension qui tient compte de la longueur de fissure, de la géométrie de l'éprouvette et du niveau de contrainte appliqué $\Delta \sigma$.

Les relevés de la longueur de fissure en fonction du nombre de cycles doivent satisfaire à deux exigences contradictoires : ne pas être trop rapprochés pour que l'erreur due à la mesure ne soit pas du même ordre de grandeur que la propagation de fissure entre deux relevés et ne pas être trop éloignés pour retenir un maximum d'informations sur la propagation réelle de la fissure. Ce phénomène est illustré sur la figure A.2. Une perte d'information est visible sur la courbe rouge,

tandis qu'un surplus d'information est repérable sur la courbe grise. Un compromis entre ces deux extrêmes revient à tracer la courbe verte, qui est plus satisfaisante.

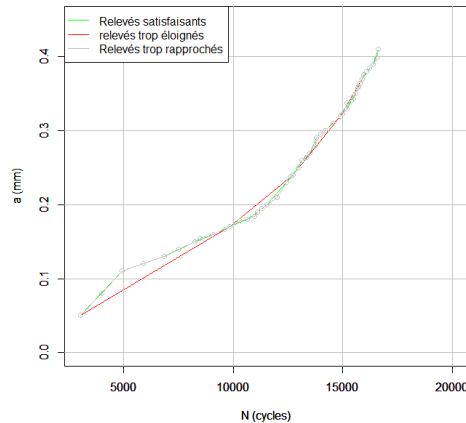


FIGURE A.2 – Illustration des exigences à satisfaire.

A.2 Essai de traction sur un matériau

Pour déterminer le comportement d'un matériau face aux différentes sollicitations qu'il peut rencontrer en période d'utilisation, ces sollicitations sont reproduites à l'aide d'essais statiques ou dynamiques, généralement effectués sur des éprouvettes normalisées afin de connaître les caractéristiques chiffrées du matériau (module de Young, allongement, ...). Le schéma de la figure A.3 représente deux types d'éprouvettes normalisées qui sont utilisées pour réaliser un essai de traction. La longueur calibrée L_0 (en mm) de l'éprouvette est liée à la section S_0 (en mm^2) par la relation $L_0 = 5,65\sqrt{S_0}$.

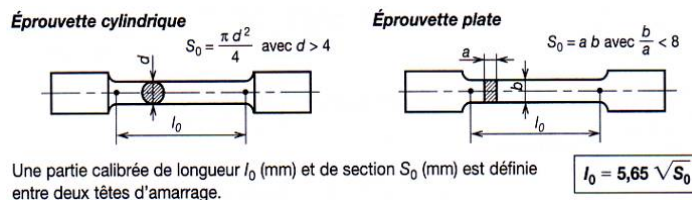


FIGURE A.3 – Éprouvettes normalisées.

Source web : [Les essais sur le matériau acier.](#)

Durant un essai de traction les têtes de l'éprouvette sont fixées dans la machine de traction et on va solliciter cette éprouvette en traction uni-axiale jusqu'à la rupture pour déterminer ses caractéristiques mécaniques. Le déroulement de l'essai de traction est le suivant (il fait référence à la courbe de traction présentée dans la figure A.4) :

1. l'éprouvette est au repos, la force de traction $F = 0$ et l'allongement vaut 0;
2. on tire sur l'éprouvette avec une force F_1 qui entraîne un allongement e_1 de l'éprouvette;
3. on supprime l'effort et l'allongement disparaît. On est dans la zone élastique du matériau;
4. on sollicite l'éprouvette avec une force plus importante F_2 , cela entraîne un allongement e_3 de l'éprouvette;
5. on supprime la force F_2 , mais l'éprouvette ne retrouve pas son état initial. Nous sommes dans la zone plastique du matériau. Il reste un allongement e_2 ;
6. on exerce un effort de traction supplémentaire plus important que l'effort F_2 . Après avoir atteint un maximum, la force décroît et l'éprouvette s'amincit en un endroit appelé zone de striction, puis il y a rupture. Après cette rupture on mesure l'allongement $e_4 = L_u - L_0$.

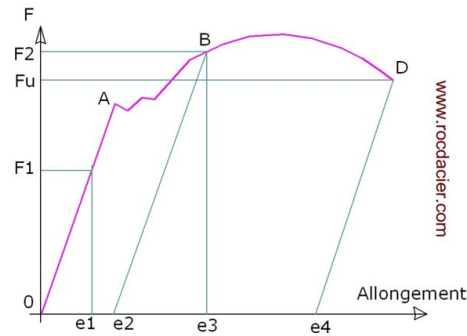


FIGURE A.4 – Diagramme conventionnel d'un essai de traction.
Source web : *Cours essai de traction.*

L'évolution de l'éprouvette durant l'essai est représentée dans la figure A.5.

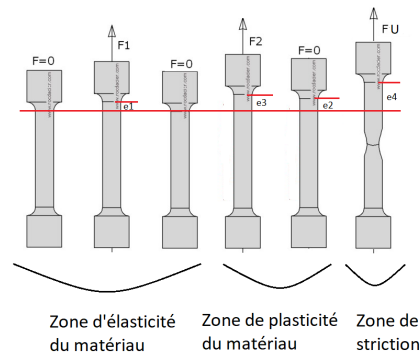


FIGURE A.5 – Évolution schématisée de l'éprouvette durant l'essai de traction.
Source web : schéma inspiré du *cours essai de traction.*

A.3 Deux méthodes d'optimisation du chapitre 2

Cette annexe a pour but de présenter deux méthodes d'optimisation abordées, mais non détaillées dans le chapitre 2.

A.3.1 Recherche trichotomique

La recherche trichotomique consiste à déterminer le maximum d'une fonction concave W définie sur un intervalle de temps connu $[m, n]$. Cet intervalle est ensuite divisé en trois parts égales de longueur $h = \frac{n-m}{3}$ afin de proposer deux valeurs $t_1 = m + h$ et $t_2 = n - h$ à l'étape suivante.

- Si $W(t_1) \leq W(t_2)$, on recommence la procédure sur $[t_1, n]$;
- Si $W(t_1) > W(t_2)$, on recommence la procédure sur $[m, t_2]$.

Dans notre cas,

- l'intervalle $[m, n]$ correspond à l'intervalle des observations X_i , c'est à dire $[\tau_0, \tau_K]$;
- t_1 et t_2 sont deux instants de transition que nous comparons;
- dans le cas de la méthode standard (sans variable latente) à deux régimes,

$$W(t) = \mathcal{L}(t, \mathbf{a}_2^*(t), \mathbf{S}_2^{2*}(t)),$$

définie dans l'équation (2.16).

Nous arrêtons l'algorithme trichotomique lorsque la différence entre les deux bornes de l'intervalle est plus petite qu'une valeur seuil fixée. D'autres méthodes numériques sont possibles, nous pouvons par exemple citer l'algorithme de Nelder-Mead, proposé par Nelder et Mead dans [35].

A.3.2 Programmation dynamique

La programmation dynamique est une méthode algorithmique utilisée pour résoudre des problèmes d'optimisation, dont le concept a été introduit au début des années 1950 par Richard Bellman dans [6]. Elle consiste à résoudre un problème en le décomposant en plusieurs sous-problèmes, puis à résoudre les sous-problèmes en sauvegardant les résultats intermédiaires. C'est un paradigme alternatif à celui de «diviser pour mieux régner» qui se décompose en les trois étapes suivantes :

- diviser : on divise les données initiales en plusieurs sous parties;
- régner : on résout récursivement chacun des sous problèmes associés;
- combiner : on combine les différents résultats pour obtenir une solution au problème initial.

Contrairement à ce paradigme, la programmation dynamique tient compte du fait que les sous-problèmes sont liés entre eux et permet, à l'aide de ces deux formes récursive ou itérative, d'éviter les appels récursifs redondants en stockant les valeurs intermédiaires.

Les deux formes que peut prendre la programmation dynamique sont les suivantes :

1. Une forme récursive appelée «Top Down» : on utilise directement la formule de récurrence et lors d'un appel récursif on regarde dans le tableau de mémoire cache si le travail n'a pas déjà été effectué avant d'effectuer un calcul;
2. Une forme itérative appelée «Bottom Up» : on résout d'abord les sous problèmes de la plus petite taille, puis ceux de la taille d'au dessus. Au fur et à mesure on stocke les résultats obtenus dans un tableau de mémoire cache. On continue jusqu'à la taille voulue.

Un des principaux champs d'application de la programmation dynamique est la résolution de problèmes d'optimisation. Il s'agit de problèmes dont chaque solution possède une valeur. On cherche alors une solution de valeur optimale (minimale ou maximale). Cormen et co ont exposé dans [17] quatre étapes dans la conception d'une solution optimale :

1. caractériser la structure d'une solution optimale;
2. définir par récurrence la valeur d'une solution optimale;
3. calculer la valeur d'une solution optimale par une méthode Top Down ou Bottom Up;
4. construire la solution optimale.

Un exemple élémentaire de programmation dynamique est la pyramide de nombres. Dans cet exemple, l'objectif est de déterminer le chemin, partant du sommet pour aller vers le bas, qui maximise le total des nombres traversés. Dans la figure A.6, le total maximal des nombres traversés est de 23 et le chemin des nombres traversés est représenté en rouge. Réaliser cette étude au travers d'un algorithme naïf consisterait à étudier l'ensemble des 8 chemins possibles puis à choisir celui qui a le plus grand total. Sachant qu'une pyramide à n niveaux contient 2^{n-1} chemins, cela reviendrait à effectuer $2^n - 2$ calculs, ce qui serait très rapidement impraticable pour une valeur de n grande. Le paradigme de la programmation dynamique permet donc d'obtenir un algorithme efficace en définissant des sous problèmes et en écrivant une relation de récurrence.

Nous notons,

- x la position dans la pyramide;
- $v(x)$ le nombre écrit à cette position;
- $c(x)$ le total des nombres traversés dans un chemin maximal issu de x .

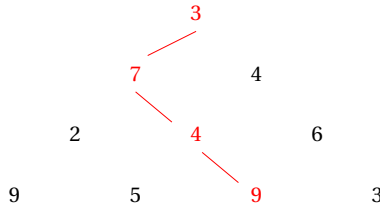


FIGURE A.6 – Pyramide de nombres avec son chemin optimal en rouge.

Les sous problèmes consistent à calculer les valeurs de $c(x)$ pour tout x . La valeur de la solution optimale est $c(x)$ et la solution en elle-même est la liste des x qui composent le chemin maximal. La relation de récursivité peut s'écrire de la façon suivante,

- $c(x) = v(x)$ pour tout x situé à la base de la pyramide;
- $c(x) = v(x) + \max(c(x_g), c(x_d))$ pour tout autre position x où x_g et x_d sont les enfants de x .

Le calcul de $c(x)$ maximal peut se faire à l'aide d'un algorithme itératif ascendant en stockant les valeurs déjà calculées dans un tableau, soit à l'aide d'un algorithme récursif descendant avec mémorisation (terme informatique désignant la mémorisation des valeurs retournées par une fonction). L'important est de conserver dans un tableau les valeurs intermédiaires. La séquence des calculs effectués par la relation de récursivité définie ci-dessus est indiquée en gras :

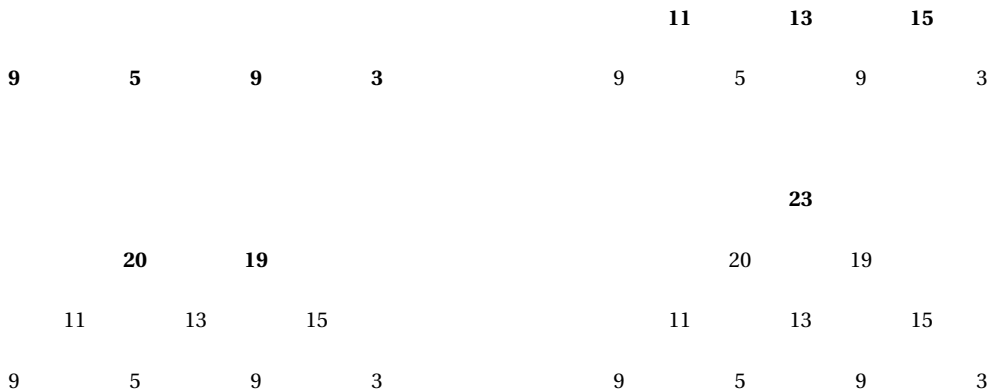


FIGURE A.7 – Séquence des calculs effectués par la relation de récursivité.

A.4 Estimation d'un régime sous contraintes

Nous présentons ici la méthode d'estimation d'un régime (ie \tilde{a} et \tilde{S}^2) sur un intervalle donné avec des contraintes aux bornes pour résoudre le problème d'optimisation de l'équation (2.36).

A.4.1 Modèle de régression polynomiale par morceaux sur un régime constraint

Fonction de lien

Soit $\tilde{\tau}_1$ et $\tilde{\tau}_2$ fixés et connus. Soit x , y , \dot{x} et \dot{y} fixés. Dans cette partie, nous travaillons dans un régime quelconque borné par $\tilde{\tau}_1$ et $\tilde{\tau}_2$. On note,

$$\rho_{\tilde{\alpha}}(x) = \sum_{l=0}^{\delta} \tilde{\alpha}(l) x^l,$$

la fonction de lien du régime défini sur l'intervalle $[\tilde{\tau}_1, \tilde{\tau}_2]$ et $\tilde{\sigma}^2$ la variance dans ce régime.

Contrainte de régularité

Les hypothèses de régularité soumises à la fonction de lien deviennent une contrainte de valeur aux bornes :

$$\mathcal{A}_c(\tilde{\tau}_1, x) \leftrightarrow \tilde{\rho}_{\tilde{\alpha}}(\tilde{\tau}_1) = x.$$

$$\mathcal{A}_d(\tilde{\tau}_1, x) \leftrightarrow \mathcal{A}_c(\tilde{\tau}_1, x) \text{ vraie et } \dot{\rho}_{\tilde{\alpha}}(\tilde{\tau}_1) = \dot{x}.$$

Log-vraisemblance

Notre objectif est d'estimer les paramètres inconnus $\tilde{\alpha}$ et $\tilde{\sigma}^2$ en maximisant la log-vraisemblance suivante qui s'écrit (à une constante additive près),

$$\mathcal{L}_{[\tilde{\tau}_1, \tilde{\tau}_2]}(\tilde{\alpha}, \tilde{S}^2) = -\frac{1}{2} \sum_{i=1}^n \left[\log(\tilde{S}^2) + \frac{(Y_i - \tilde{\rho}_{\tilde{\alpha}}(X_i))^2}{\tilde{S}^2} \right] \mathbb{1}_{[\tilde{\tau}_1, \tilde{\tau}_2]}(X_i).$$

A.4.2 Maximum de vraisemblance à un régime contraint

Propriété du maximum de vraisemblance en $\tilde{\alpha}$ et \tilde{S}^2

Le problème de maximisation de la log-vraisemblance en $(\tilde{\alpha}, \tilde{S}^2)$ sous contrainte de régularité $\mathcal{A}_*(\tilde{\tau}_1, x) \vee \mathcal{A}_*(\tilde{\tau}_2, y)$ ou $\mathcal{A}_*(\tilde{\tau}_1, x) \wedge \mathcal{A}_*(\tilde{\tau}_2, y)$ avec $*$ $\in \{c, d\}$ s'écrit

$$(\hat{\tilde{\alpha}}, \hat{\tilde{\sigma}}^2) = \underset{\tilde{\alpha}, \tilde{S}^2}{\operatorname{argmax}} \mathcal{L}_{[\tilde{\tau}_1, \tilde{\tau}_2]}(\tilde{\alpha}, \tilde{S}^2). \quad (\text{A.1})$$

sous $\begin{cases} \mathcal{A}_*(\tilde{\tau}_1, x) \vee \mathcal{A}_*(\tilde{\tau}_2, y) \\ \text{ou} \\ \mathcal{A}_*(\tilde{\tau}_1, x) \wedge \mathcal{A}_*(\tilde{\tau}_2, y) \end{cases}$

Nous nous proposons de résoudre ce problème par la méthode des multiplicateurs de Lagrange. Dans la suite, nous introduisons des notations et objets mathématiques utiles pour exploiter nos algorithmes d'optimisation.

Le moment d'ordre r des observations X_i sur l'intervalle $[\tilde{\tau}_1, \tilde{\tau}_2]$ s'écrit $\tilde{\mu}^r$,

$$\tilde{\mu}^r = \sum_{i=1}^n X_i^r \mathbb{1}_{\{\tilde{\tau}_1 \leq X_i < \tilde{\tau}_2\}}. \quad (\text{A.2})$$

De même, nous définissons $\tilde{\nu}^r$ de la façon suivante,

$$\tilde{\nu}^r = \sum_{i=1}^n X_i^r Y_i \mathbb{1}_{\{\tilde{\tau}_1 \leq X_i < \tilde{\tau}_2\}}. \quad (\text{A.3})$$

À partir de ces notations, nous introduisons la matrice carrée \tilde{M} de taille $(\delta + 1)$,

$$\tilde{M} = [\tilde{w}^0, \tilde{w}^1, \dots, \tilde{w}^\delta],$$

où

$$\tilde{w}^b = (\mu^{2\delta-b}, \mu^{2\delta-b-1}, \dots, \mu^{\delta-b})^t,$$

ainsi que le vecteur $\tilde{\gamma}$ de taille $(\delta + 1)$,

$$\tilde{\gamma} = (\tilde{\nu}^\delta, \tilde{\nu}^{\delta-1}, \dots, \tilde{\nu}^0)^t.$$

De plus, nous définissons les vecteurs $u_{c,j}(\omega)$, $\omega > 0$, et $v_{c,j}$ tels que,

$$\begin{aligned}\tilde{u}_{c,j}(\omega) &= (\omega \tilde{\tau}_j^\delta, \omega \tilde{\tau}_j^{\delta-1}, \dots, \omega)^t, \\ \tilde{v}_{c,j} &= (\tilde{\tau}_j^\delta, \tilde{\tau}_j^{\delta-1}, \dots, 1),\end{aligned}$$

où c représente la contrainte de continuité $\mathcal{A}_c(\tilde{\tau}_1, x)$, si $j = 1$ et $\mathcal{A}_c(\tilde{\tau}_2, y)$, si $j = 2$. Nous définissons également

$$\begin{aligned}\tilde{u}_{d,j}(\omega) &= \partial_{\tilde{\tau}_j} u_{c,j}(\omega) = (\delta \omega \tilde{\tau}_j^{\delta-1}, (\delta-1) \omega \tilde{\tau}_j^{\delta-2}, \dots, 0)^t, \\ \tilde{v}_{d,j} &= \partial_{\tilde{\tau}_j} v_{c,j} = (\delta \tilde{\tau}_j^{\delta-1}, (\delta-1) \tilde{\tau}_j^{\delta-2}, \dots, 0),\end{aligned}$$

où d représente la contrainte de dérivabilité $\mathcal{A}_d(\tilde{\tau}_1, x)$, si $j = 1$ et $\mathcal{A}_d(\tilde{\tau}_2, y)$, si $j = 2$. Enfin, nous définissons les matrices carrées $M_c(\omega)$ et $M_d(\omega)$ ainsi que les vecteurs γ_c et γ_d tels que :

- Sous la contrainte $\mathcal{A}_*(\tilde{\tau}_1, x)$, avec $*$ $\in \{c, d\}$, ces matrices sont de tailles $(\delta+1)+1$ et $(\delta+1)+2$ et s'écrivent,

$$\tilde{M}_c(\omega) = \begin{pmatrix} \tilde{v}_{c,1} & 0 \\ \tilde{M} & \tilde{u}_{c,1}(\omega) \end{pmatrix}, \quad \tilde{M}_d(\omega) = \begin{pmatrix} \tilde{v}_{c,1} & 0 & 0 \\ \tilde{v}_{d,1} & 0 & 0 \\ \tilde{M} & \tilde{u}_{c,1}(\omega) & \tilde{u}_{d,1}(\omega) \end{pmatrix}, \quad (\text{A.4})$$

et les vecteurs γ_c et γ_d sont définis par,

$$\tilde{\gamma}_c = \begin{pmatrix} x \\ \tilde{\gamma} \end{pmatrix}, \quad \tilde{\gamma}_d = \begin{pmatrix} x \\ \dot{x} \\ \tilde{\gamma} \end{pmatrix}. \quad (\text{A.5})$$

- Sous la contrainte $\mathcal{A}_*(\tilde{\tau}_1, x) \wedge \mathcal{A}_*(\tilde{\tau}_2, y)$, avec $*$ $\in \{c, d\}$, ces matrices sont de tailles $(\delta+1)+2$ et $(\delta+1)+4$ et s'écrivent,

$$\tilde{M}_c(\omega) = \begin{pmatrix} \tilde{v}_{c,1} & 0 & 0 \\ \tilde{v}_{c,2} & 0 & 0 \\ \tilde{M} & \tilde{u}_{c,1}(\omega) & \tilde{u}_{c,2}(\omega) \end{pmatrix}, \quad \tilde{M}_d(\omega) = \begin{pmatrix} \tilde{v}_{c,1} & 0 & 0 & 0 & 0 \\ \tilde{v}_{d,1} & 0 & 0 & 0 & 0 \\ \tilde{v}_{c,2} & 0 & 0 & 0 & 0 \\ \tilde{v}_{d,2} & 0 & 0 & 0 & 0 \\ \tilde{M} & \tilde{u}_{c,1}(\omega) & \tilde{u}_{d,1}(\omega) & \tilde{u}_{c,2}(\omega) & \tilde{u}_{d,2}(\omega) \end{pmatrix}, \quad (\text{A.6})$$

et les vecteurs γ_c et γ_d sont définis par,

$$\tilde{\gamma}_c = \begin{pmatrix} x \\ y \\ \tilde{\gamma} \end{pmatrix}, \quad \tilde{\gamma}_d = \begin{pmatrix} x \\ \dot{x} \\ y \\ \dot{y} \\ \tilde{\gamma} \end{pmatrix}. \quad (\text{A.7})$$

Nous notons $V(\tilde{a})$, l'estimateur empirique de la variance dans le régime défini sur l'intervalle $[\tilde{\tau}_1, \tilde{\tau}_2]$ tel que pour tout $\tilde{a} \in \mathbb{R}^{\delta+1}$,

$$V(\tilde{a}) = \frac{\sum_{i=1}^n (Y_i - \rho_{\tilde{a}}(X_i))^2 \mathbb{1}_{\{\tilde{\tau}_1 \leq X_i < \tilde{\tau}_2\}}}{\sum_{i=1}^n \mathbb{1}_{\{\tilde{\tau}_1 \leq X_i < \tilde{\tau}_2\}}}.$$

Le problème d'optimisation de l'équation (A.1) est partiellement linéaire. Ce résultat est énoncé dans la proposition 18 et sera démontré en annexe A.5.3.

Proposition 18. *L'optimum (\tilde{a}, \tilde{S}^2) du problème d'optimisation (A.1) sous la contrainte de régularité $\mathcal{A}_*(\tilde{\tau}_1, x) \vee \mathcal{A}_*(\tilde{\tau}_2, y)$ ou $\mathcal{A}_*(\tilde{\tau}_1, x) \wedge \mathcal{A}_*(\tilde{\tau}_2, y)$ avec $*$ $\in \{c, d\}$, vérifie ce système*

$$\begin{cases} \tilde{M}_*(\tilde{S}^2)(\lambda, \tilde{a})^t &= \tilde{\gamma}_*, \\ \tilde{S}^2 &= V(\tilde{a}). \end{cases} \quad (\text{A.8})$$

où $\lambda \in \mathbb{R}$ sous $\mathcal{A}_c(\tilde{\tau}_1, x) \vee \mathcal{A}_c(\tilde{\tau}_2, y)$, $\lambda \in \mathbb{R}^2$ sous $\mathcal{A}_d(\tilde{\tau}_1, x) \vee \mathcal{A}_d(\tilde{\tau}_2, y)$ ou $\mathcal{A}_c(\tilde{\tau}_1, x) \wedge \mathcal{A}_c(\tilde{\tau}_2, y)$, $\lambda \in \mathbb{R}^4$ sous $\mathcal{A}_d(\tilde{\tau}_1, x) \wedge \mathcal{A}_d(\tilde{\tau}_2, y)$ est le multiplicateur de Lagrange.

Algorithme d'optimisation en $\tilde{\mathbf{a}}$ et $\tilde{\mathbf{S}}^2$

La proposition 18 permet de définir un algorithme de point fixe d'approximation de l'optimum de (A.1) sous la contrainte de régularité $\mathcal{A}_*(\tilde{\tau}_1, x) \vee \mathcal{A}_*(\tilde{\tau}_2, y)$ ou $\mathcal{A}_*(\tilde{\tau}_1, x) \wedge \mathcal{A}_*(\tilde{\tau}_2, y)$ avec $*$ $\in \{c, d\}$. Le passage de l'étape (i) à l'étape $(i+1)$ s'écrit de la façon suivante,

$$\begin{cases} (\lambda^{(i+1)}, \tilde{\mathbf{a}}^{(i+1)})^t &= \tilde{\mathbf{M}}_* \left(\tilde{\mathbf{S}}^{2(i)} \right)^{-1} \tilde{\mathbf{Y}}_*, \\ \tilde{\mathbf{S}}^{2(i+1)} &= \mathbf{V}(\tilde{\mathbf{a}}^{(i+1)}). \end{cases}$$

A.5 Résolution des problèmes d'optimisation sous contrainte de régularité sur le modèle sans variable latente.

Dans les problèmes de maximisation de cette annexe nous souhaitons estimer les paramètres α_K et σ_K^2 par maximum de vraisemblance sous contrainte de continuité ou dérivabilité. Nous sommes dans le contexte d'un problème d'optimisation sous contraintes d'égalités dans lequel la fonction à optimiser et les fonctions de contraintes sont différentiables par rapport aux paramètres. Le nombre de contraintes peut varier en fonction du type de contrainte (continuité ou dérivabilité) et du problème étudié. Pour résoudre ces problèmes d'optimisation nous proposons d'utiliser la méthode des multiplicateurs de Lagrange. Dans la suite, nous introduisons des notations qui permettront d'explicitier le système d'équations nécessaire à la recherche d'un point critique. Soit δ le degré fixé des polynômes, alors pour $1 \leq k \leq K$, on note a_k^0, \dots, a_k^δ les coefficients du polynôme ρ_{a_k} du régime k , avec a_k^δ le coefficient devant le terme de plus haut degré et a_k^0 le terme constant. À partir de ces notations, \mathbf{a}_K peut se réécrire sous la forme suivante,

$$\mathbf{a}_K = (a_1^\delta \quad \dots \quad a_1^0 \quad a_2^\delta \quad \dots \quad a_2^0 \quad \dots \quad a_K^\delta \quad \dots \quad a_K^0)^t.$$

A.5.1 Optimisation à deux régimes

Dans cette partie nous nous rapportons à la section 2.3.1 qui nous place dans le contexte où le nombre de régimes est fixé à $K = 2$. En travaillant à T_1 fixé, notre objectif est de résoudre le problème de maximisation suivant,

$$(\hat{\alpha}_2, \hat{\sigma}_2^2) = \underset{\substack{\mathbf{a}_2, \mathbf{S}_2^2 \\ \text{sous } \mathcal{A}_*(\tau_2)}}{\arg \max} \mathcal{L}(T_1, \mathbf{a}_2, \mathbf{S}_2^2),$$

où $\mathcal{L}(T_1, \mathbf{a}_2, \mathbf{S}_2^2)$ est la log-vraisemblance du modèle sans variable latente de la section 2.2.1 qui s'écrit (à une constante additive près),

$$\mathcal{L}(T_1, \mathbf{a}_2, \mathbf{S}_2^2) = -\frac{1}{2} \sum_{i=1}^n \sum_{k=1}^2 \left[\log(S_k^2) + \frac{(Y_i - \rho_{a_k}(X_i))^2}{S_k^2} \right] \mathbb{1}_{\{T_{k-1}, T_k\}}(X_i). \quad (\text{A.9})$$

Sous l'hypothèse de continuité $\mathcal{A}_c(\tau_2)$, la fonction de contrainte $q(\mathbf{a}_2)$ s'écrit

$$q(\mathbf{a}_2) = \rho_{a_1}(T_1) - \rho_{a_2}(T_1) = 0.$$

Dans la résolution du problème de maximisation sous contrainte de continuité, il est facile de vérifier que la condition de surjectivité de la différentielle de $q(\mathbf{a}_2)$ est vérifiée. De ce fait, le lagrangien $\mathbb{L}_{\mathcal{L}, q}(\mathbf{a}_2, \mathbf{S}_2^2, \lambda) : \mathbb{R}^3 \rightarrow \mathbb{R}$ peut s'écrire,

$$\mathbb{L}_{\mathcal{L}, q}(\mathbf{a}_2, \mathbf{S}_2^2, \lambda) = \mathcal{L}(T_1, \mathbf{a}_2, \mathbf{S}_2^2) - \lambda q(\mathbf{a}_2), \quad (\text{A.10})$$

où $\lambda \in \mathbb{R}$ représente le multiplicateur de Lagrange associé à la contrainte $q(\mathbf{a}_2)$. Nous allons donc chercher les paramètres qui réalisent le maximum sous contrainte de (A.9) parmi les points critiques de (A.10). En reprenant les notations des équations (2.10) et (2.11), la recherche d'un point

critique se fait par la résolution du système de dimension $2(\delta + 1) + 3$ suivant,

$$\left\{ \begin{array}{l} \rho_{a_1}(T_1) - \rho_{a_2}(T_1) = 0 \\ v_1^\delta = a_1^\delta \mu_1^{2\delta} + \dots + a_1^0 \mu_1^\delta + \lambda S_1^2 T_1^\delta \\ v_1^{\delta-1} = a_1^\delta \mu_1^{2\delta-1} + \dots + a_1^0 \mu_1^{\delta-1} + \lambda S_1^2 T_1^{\delta-1} \\ \vdots \\ v_1^0 = a_1^\delta \mu_1^\delta + \dots + a_1^0 \mu_1^0 + \lambda S_1^2 \\ v_2^\delta = a_2^\delta \mu_2^{2\delta} + \dots + a_2^0 \mu_2^\delta - \lambda S_2^2 T_1^\delta \\ v_2^{\delta-1} = a_2^\delta \mu_2^{2\delta-1} + \dots + a_2^0 \mu_2^{\delta-1} - \lambda S_2^2 T_1^{\delta-1} \\ \vdots \\ v_2^0 = a_2^\delta \mu_2^\delta + \dots + a_2^0 \mu_2^0 - \lambda S_2^2 \\ S_1^2 = \frac{\sum_{i=1}^n (Y_i - \rho_{a_1}(X_i))^2 \mathbb{1}_{\{X_i < T_1\}}}{\sum_{i=1}^n \mathbb{1}_{\{X_i < T_1\}}} \\ S_2^2 = \frac{\sum_{i=1}^n (Y_i - \rho_{a_2}(X_i))^2 \mathbb{1}_{\{X_i \geq T_1\}}}{\sum_{i=1}^n \mathbb{1}_{\{X_i \geq T_1\}}} \end{array} \right. \quad (\text{A.11})$$

Nous remarquons que ce système d'équations n'est pas linéaire en les coefficients. En revanche, si S_2^2 est considéré connu dans les $2(\delta + 1) + 1$ premières équations, alors celles-ci sont linéaires en \mathbf{a}_2 et λ . En effet, à S_2^2 connu, le vecteur des paramètres (λ, \mathbf{a}_2) vérifie

$$M_c(S_2^2)(\lambda, \mathbf{a}_2)^t = \gamma_c,$$

avec M_c et γ_c définis dans les équations (2.12) et (2.13) de la section 2.3.1. La résolution de ce système linéaire nécessite que M_c soit inversible. D'autre part, si \mathbf{a}_2 est connu alors S_2^2 s'obtient par la formule de l'estimateur empirique de la variance détaillée dans les deux dernières équations du système (A.11). Ce résultat entraîne la proposition 3 et l'algorithme itératif 2.3.1 qui alterne entre les résolutions à S_2^2 fixé et à \mathbf{a}_2 fixé. La démarche est identique sous contrainte de dérivabilité qui s'écrit

$$\left\{ \begin{array}{l} q_1(\mathbf{a}_2) = \rho_{a_1}(T_1) - \rho_{a_2}(T_1) = 0 \\ q_2(\mathbf{a}_2) = \dot{\rho}_{a_1}(T_1) - \dot{\rho}_{a_2}(T_1) = 0. \end{array} \right.$$

La recherche du point critique se fait par la résolution du système de dimension $2(\delta + 1) + 4$ et qui justifie les équations (2.12), (2.13) et (2.15) de la section 2.3.1,

$$\left\{ \begin{array}{l} \rho_{a_1}(T_1) - \rho_{a_2}(T_1) = 0 \\ \dot{\rho}_{a_1}(T_1) - \dot{\rho}_{a_2}(T_1) = 0 \\ v_1^\delta = a_1^\delta \mu_1^{2\delta} + \dots + a_1^0 \mu_1^\delta + \lambda_1 S_1^2 T_1^\delta + \delta \lambda_2 S_1^2 T_1^{\delta-1} \\ v_1^{\delta-1} = a_1^\delta \mu_1^{2\delta-1} + \dots + a_1^0 \mu_1^{\delta-1} + \lambda_1 S_1^2 T_1^{\delta-1} + (\delta - 1) \lambda_2 S_1^2 T_1^{\delta-2} \\ \vdots \\ v_1^0 = a_1^\delta \mu_1^\delta + \dots + a_1^0 \mu_1^0 + \lambda_1 S_1^2 \\ v_2^\delta = a_2^\delta \mu_2^{2\delta} + \dots + a_2^0 \mu_2^\delta - \lambda_1 S_2^2 T_1^\delta - \delta \lambda_2 S_2^2 T_1^{\delta-1} \\ v_2^{\delta-1} = a_2^\delta \mu_2^{2\delta-1} + \dots + a_2^0 \mu_2^{\delta-1} - \lambda_1 S_2^2 T_1^{\delta-1} - (\delta - 1) \lambda_2 S_2^2 T_1^{\delta-2} \\ \vdots \\ v_2^0 = a_2^\delta \mu_2^\delta + \dots + a_2^0 \mu_2^0 - \lambda_1 S_2^2 \\ S_1^2 = \frac{\sum_{i=1}^n (Y_i - \rho_{a_1}(X_i))^2 \mathbb{1}_{\{X_i < T_1\}}}{\sum_{i=1}^n \mathbb{1}_{\{X_i < T_1\}}} \\ S_2^2 = \frac{\sum_{i=1}^n (Y_i - \rho_{a_2}(X_i))^2 \mathbb{1}_{\{X_i \geq T_1\}}}{\sum_{i=1}^n \mathbb{1}_{\{X_i \geq T_1\}}} \end{array} \right.$$

A.5.2 Optimisation à un nombre de régimes K quelconque

Dans cette partie, nous nous rapportons à la section 2.4.1 qui nous place dans le contexte où le nombre de régimes K est considéré quelconque. À \mathbf{T}_{K-1} fixé, nous souhaitons résoudre le problème de maximisation suivant,

$$(\hat{\mathbf{a}}_K, \hat{\mathbf{S}}_K^2) = \underset{\substack{\mathbf{a}_K, \mathbf{S}_K^2 \\ \text{sous } \mathcal{A}_*(\tau_K)}}{\operatorname{argmax}} \mathcal{L}(\mathbf{T}_{K-1}, \mathbf{a}_K, \mathbf{S}_K^2),$$

où $\mathcal{L}(\mathbf{T}_{K-1}, \mathbf{a}_K, \mathbf{S}_K^2)$ est la log-vraisemblance du modèle sans variable latente de la section 2.2.1 qui s'écrit (à une constante additive près),

$$\mathcal{L}(\mathbf{T}_{K-1}, \mathbf{a}_K, \mathbf{S}_K^2) = -\frac{1}{2} \sum_{i=1}^n \sum_{k=1}^K \left[\log(S_k^2) + \frac{(Y_i - \rho_{a_k}(X_i))^2}{S_k^2} \right] \mathbb{1}_{[\mathbf{T}_{k-1}, \mathbf{T}_k]}(X_i).$$

En suivant le raisonnement de la section précédente et en reprenant les notations des équations (2.26) et (2.27), la recherche du point critique sous contrainte de continuité se fait par la résolution du système de dimension $K(\delta + 1) + 2K - 1$ suivant, qui correspond aux équations (2.28), (2.30) et (2.31) de la section 2.4.1,

$$\left\{ \begin{array}{l} \rho_{a_1}(\mathbf{T}_1) - \rho_{a_2}(\mathbf{T}_1) = 0 \\ \rho_{a_2}(\mathbf{T}_2) - \rho_{a_3}(\mathbf{T}_3) = 0 \\ \vdots \\ \rho_{a_{K-1}}(\mathbf{T}_{K-1}) - \rho_{a_K}(\mathbf{T}_{K-1}) = 0 \\ v_1^\delta = a_1^\delta \mu_1^{2\delta} + \dots + a_1^0 \mu_1^\delta + \lambda_1 S_1^2 T_1^\delta \\ v_1^{\delta-1} = a_1^\delta \mu_1^{2\delta-1} + \dots + a_1^0 \mu_1^{\delta-1} + \lambda_1 S_1^2 T_1^{\delta-1} \\ \vdots \\ v_2^0 = a_2^\delta \mu_2^\delta + \dots + a_2^0 \mu_2^0 + \lambda_1 S_1^2 \\ v_2^\delta = a_2^\delta \mu_2^{2\delta} + \dots + a_2^0 \mu_2^\delta - \lambda_1 S_2^2 T_1^\delta + \lambda_2 S_2^2 T_2^\delta \\ v_2^{\delta-1} = a_2^\delta \mu_2^{2\delta-1} + \dots + a_2^0 \mu_2^{\delta-1} - \lambda_1 S_2^2 T_1^{\delta-1} + \lambda_2 S_2^2 T_2^{\delta-1} \\ \vdots \\ v_2^0 = a_2^\delta \mu_2^\delta + \dots + a_2^0 \mu_2^0 - \lambda_1 S_2^2 + \lambda_2 S_2^2 \\ \vdots \\ v_K^\delta = a_K^\delta \mu_K^{2\delta} + \dots + a_K^0 \mu_K^\delta - \lambda_{K-1} S_K^2 T_{K-1}^\delta \\ v_K^{\delta-1} = a_K^\delta \mu_K^{2\delta-1} + \dots + a_K^0 \mu_K^{\delta-1} - \lambda_{K-1} S_K^2 T_{K-1}^{\delta-1} \\ \vdots \\ v_K^0 = a_K^\delta \mu_K^\delta + \dots + a_K^0 \mu_K^0 - \lambda_{K-1} S_K^2 \\ S_1^2 = \frac{\sum_{i=1}^n (Y_i - \rho_{a_1}(X_i))^2 \mathbb{1}_{\{X_i < T_1\}}}{\sum_{i=1}^n \mathbb{1}_{\{X_i < T_1\}}} \\ S_2^2 = \frac{\sum_{i=1}^n (Y_i - \rho_{a_2}(X_i))^2 \mathbb{1}_{\{T_1 \leq X_i < T_2\}}}{\sum_{i=1}^n \mathbb{1}_{\{T_1 \leq X_i < T_2\}}} \\ \vdots \\ S_K^2 = \frac{\sum_{i=1}^n (Y_i - \rho_{a_K}(X_i))^2 \mathbb{1}_{\{X_i \geq T_{K-1}\}}}{\sum_{i=1}^n \mathbb{1}_{\{X_i \geq T_{K-1}\}}}. \end{array} \right.$$

La démarche est identique sous contrainte de dérivabilité. La recherche du point critique se fait par la résolution du système de dimension $K(\delta + 1) + 3K - 2$ suivant, qui correspond aux équations

(2.29), (2.30) et (2.31) de la section 2.4.1,

$$\left\{ \begin{array}{l}
 \rho_{a_1}(T_1) - \rho_{a_2}(T_1) = 0 \\
 \dot{\rho}_{a_1}(T_1) - \dot{\rho}_{a_2}(T_1) = 0 \\
 \vdots \\
 \rho_{a_{K-1}}(T_{K-1}) - \rho_{a_K}(T_{K-1}) = 0 \\
 \dot{\rho}_{a_{K-1}}(T_{K-1}) - \dot{\rho}_{a_K}(T_{K-1}) = 0 \\
 v_1^\delta = a_1^\delta \mu_1^{2\delta} + \dots + a_1^0 \mu_1^\delta + \lambda_1 S_1^2 T_1^\delta + \delta \lambda_2 S_1^2 T_1^{\delta-1} \\
 v_1^{\delta-1} = a_1^\delta \mu_1^{2\delta-1} + \dots + a_1^0 \mu_1^{\delta-1} + \lambda_1 S_1^2 T_1^{\delta-1} + (\delta-1) \lambda_2 S_1^2 T_1^{\delta-2} \\
 \vdots \\
 v_1^0 = a_1^\delta \mu_1^\delta + \dots + a_1^0 \mu_1^0 + \lambda_1 S_1^2 \\
 v_2^\delta = a_2^\delta \mu_2^{2\delta} + \dots + a_2^0 \mu_2^\delta - \lambda_1 S_2^2 T_1^\delta - \delta \lambda_2 S_2^2 T_1^{\delta-1} + \lambda_3 S_2^2 T_2^\delta + \delta \lambda_4 S_2^2 T_2^{\delta-1} \\
 v_2^{\delta-1} = a_2^\delta \mu_2^{2\delta-1} + \dots + a_2^0 \mu_2^{\delta-1} - \lambda_1 S_2^2 T_1^{\delta-1} - (\delta-1) \lambda_2 S_2^2 T_1^{\delta-2} + \lambda_3 S_2^2 T_2^{\delta-1} + (\delta-1) \lambda_4 S_2^2 T_2^{\delta-2} \\
 \vdots \\
 v_2^0 = a_2^\delta \mu_2^\delta + \dots + a_2^0 \mu_2^0 - \lambda_1 S_2^2 + \lambda_3 S_2^2 \\
 \vdots \\
 v_K^\delta = a_K^\delta \mu_K^{2\delta} + \dots + a_K^0 \mu_K^\delta - \lambda_{(2(K-1)-1)} S_K^2 T_{K-1}^\delta - \delta \lambda_{2(K-1)} S_K^2 T_{K-1}^{\delta-1} \\
 v_K^{\delta-1} = a_K^\delta \mu_K^{2\delta-1} + \dots + a_K^0 \mu_K^{\delta-1} - \lambda_{(2(K-1)-1)} S_K^2 T_{K-1}^{\delta-1} - (\delta-1) \lambda_{2(K-1)} S_K^2 T_{K-1}^{\delta-2} \\
 \vdots \\
 v_K^0 = a_K^\delta \mu_K^\delta + \dots + a_K^0 \mu_K^0 - \lambda_{(2(K-1)-1)} S_K^2 \\
 S_1^2 = \frac{\sum_{i=1}^n (Y_i - \rho_{a_1}(X_i))^2 \mathbb{1}_{\{X_i < T_1\}}}{\sum_{i=1}^n \mathbb{1}_{\{X_i < T_1\}}} \\
 S_2^2 = \frac{\sum_{i=1}^n (Y_i - \rho_{a_2}(X_i))^2 \mathbb{1}_{\{T_1 \leq X_i < T_2\}}}{\sum_{i=1}^n \mathbb{1}_{\{T_1 \leq X_i < T_2\}}} \\
 \vdots \\
 S_K^2 = \frac{\sum_{i=1}^n (Y_i - \rho_{a_K}(X_i))^2 \mathbb{1}_{\{X_i \geq T_{K-1}\}}}{\sum_{i=1}^n \mathbb{1}_{\{X_i \geq T_{K-1}\}}}.
 \end{array} \right.$$

Ces deux systèmes justifient la proposition 5.

A.5.3 Optimisation à un régime

Dans cette partie, nous nous rapportons au problème présenté en annexe A.4 dans lequel nous travaillons sur un régime unique défini sur $[\tilde{\tau}_1, \tilde{\tau}_2]$. x , y , \dot{x} et \dot{y} sont des valeurs fixées et nous souhaitons résoudre le problème de maximisation donné par

$$(\hat{\alpha}, \hat{\sigma}^2) = \underset{\tilde{a}, \tilde{S}^2}{\operatorname{arg\,max}} \quad \mathcal{L}_{[\tilde{\tau}_1, \tilde{\tau}_2]}(\tilde{a}, \tilde{S}^2),$$

sous $\left\{ \begin{array}{l} \mathcal{A}_*(\tilde{\tau}_1, x) \vee \mathcal{A}_*(\tilde{\tau}_2, y) \\ \text{ou} \\ \mathcal{A}_*(\tilde{\tau}_1, x) \wedge \mathcal{A}_*(\tilde{\tau}_2, y) \end{array} \right.$

où $\mathcal{L}_{[\tilde{\tau}_1, \tilde{\tau}_2]}(\tilde{a}, \tilde{S}^2)$ est la log-vraisemblance définie (à une constante additive près) par

$$\mathcal{L}_{[\tilde{\tau}_1, \tilde{\tau}_2]}(\tilde{a}, \tilde{S}^2) = -\frac{1}{2} \sum_{i=1}^n \left[\log(\tilde{S}^2) + \frac{(Y_i - \rho_{\tilde{a}}(X_i))^2}{\tilde{S}^2} \right] \mathbb{1}_{[\tilde{\tau}_1, \tilde{\tau}_2]}(X_i).$$

On note $\tilde{a}^0, \dots, \tilde{a}^\delta$ les coefficients du polynôme $\rho_{\tilde{a}}$, avec \tilde{a}^δ le coefficient devant le terme de plus haut degré et \tilde{a}^0 le terme constant.

En suivant le raisonnement de l'annexe (A.5.1) et en reprenant les notations des équations (A.2) et (A.3), nous détaillons dans les deux prochaines sections les différents systèmes d'équations à résoudre pour la recherche d'un point critique, en fonction de la contrainte de régularité appliquée.

Résolution du problème d'optimisation sous contrainte de continuité

Sous la contrainte $A_c(\tilde{\tau}_1, x)$, la recherche du point critique se fait par la résolution du système de dimension $(\delta + 1) + 2$ suivant, qui correspond aux équations (A.4), (A.5) et (A.8) de l'annexe A.4,

$$\left\{ \begin{array}{l} \rho_{\tilde{a}}(\tilde{\tau}_1) - x = 0 \\ \tilde{v}^\delta = \tilde{a}^\delta \tilde{\mu}^{2\delta} + \dots + \tilde{a}^0 \tilde{\mu}^\delta + \lambda \tilde{S}^2 \tilde{\tau}_1^\delta \\ \tilde{v}^{\delta-1} = \tilde{a}^\delta \tilde{\mu}^{2\delta-1} + \dots + \tilde{a}^0 \tilde{\mu}^{\delta-1} + \lambda \tilde{S}^2 \tilde{\tau}_1^{\delta-1} \\ \vdots \\ \tilde{v}^0 = \tilde{a}^\delta \tilde{\mu}^\delta + \dots + \tilde{a}^0 \tilde{\mu}^0 + \lambda \tilde{S}^2 \\ \tilde{S}^2 = \frac{\sum_{i=1}^n (Y_i - \rho_{\tilde{a}}(X_i))^2 \mathbb{1}_{\{\tilde{\tau}_1 \leq X_i < \tilde{\tau}_2\}}}{\sum_{i=1}^n \mathbb{1}_{\{\tilde{\tau}_1 \leq X_i < \tilde{\tau}_2\}}} \end{array} \right.$$

Sous la contrainte $A_c(\tilde{\tau}_2, y)$, la recherche du point critique se fait par la résolution du système précédent dans lequel il faut remplacer $\tilde{\tau}_1$ par $\tilde{\tau}_2$ et x par y .

Sous la contrainte $A_c(\tilde{\tau}_1, x) \wedge \mathcal{A}_c(\tilde{\tau}_2, y)$, la recherche du point critique se fait par la résolution du système de dimension $(\delta + 1) + 3$ suivant, qui correspond aux équations (A.6), (A.7) et (A.8) de l'annexe A.4,

$$\left\{ \begin{array}{l} \rho_{\tilde{a}}(\tilde{\tau}_1) - x = 0 \\ \rho_{\tilde{a}}(\tilde{\tau}_2) - y = 0 \\ \tilde{v}^\delta = \tilde{a}^\delta \tilde{\mu}^{2\delta} + \dots + \tilde{a}^0 \tilde{\mu}^\delta + \lambda_1 \tilde{S}^2 \tilde{\tau}_1^\delta + \lambda_2 \tilde{S}^2 \tilde{\tau}_2^\delta \\ \tilde{v}^{\delta-1} = \tilde{a}^\delta \tilde{\mu}^{2\delta-1} + \dots + \tilde{a}^0 \tilde{\mu}^{\delta-1} + \lambda_1 \tilde{S}^2 \tilde{\tau}_1^{\delta-1} + \lambda_2 \tilde{S}^2 \tilde{\tau}_2^{\delta-1} \\ \vdots \\ \tilde{v}^0 = \tilde{a}^\delta \tilde{\mu}^\delta + \dots + \tilde{a}^0 \tilde{\mu}^0 + \lambda_1 \tilde{S}^2 + \lambda_2 \tilde{S}^2 \\ \tilde{S}^2 = \frac{\sum_{i=1}^n (Y_i - \rho_{\tilde{a}}(X_i))^2 \mathbb{1}_{\{\tilde{\tau}_1 \leq X_i < \tilde{\tau}_2\}}}{\sum_{i=1}^n \mathbb{1}_{\{\tilde{\tau}_1 \leq X_i < \tilde{\tau}_2\}}} \end{array} \right.$$

Résolution du problème d'optimisation sous contrainte de dérivabilité

Sous la contrainte $A_d(\tilde{\tau}_1, x)$, la recherche du point critique se fait par la résolution du système de dimension $(\delta + 1) + 3$ suivant, qui correspond aux équations (A.4), (A.5) et (A.8) de l'annexe A.4,

$$\left\{ \begin{array}{l} \rho_{\tilde{a}}(\tilde{\tau}_1) - x = 0 \\ \dot{\rho}_{\tilde{a}}(\tilde{\tau}_1) - \dot{x} = 0 \\ \tilde{v}^\delta = \tilde{a}^\delta \tilde{\mu}^{2\delta} + \dots + \tilde{a}^0 \tilde{\mu}^\delta + \lambda_1 \tilde{S}^2 \tilde{\tau}_1^\delta + \delta \lambda_2 \tilde{S}^2 \tilde{\tau}_1^{\delta-1} \\ \tilde{v}^{\delta-1} = \tilde{a}^\delta \tilde{\mu}^{2\delta-1} + \dots + \tilde{a}^0 \tilde{\mu}^{\delta-1} + \lambda_1 \tilde{S}^2 \tilde{\tau}_1^{\delta-1} + (\delta - 1) \lambda_2 \tilde{S}^2 \tilde{\tau}_1^{\delta-2} \\ \vdots \\ \tilde{v}^0 = \tilde{a}^\delta \tilde{\mu}^\delta + \dots + \tilde{a}^0 \tilde{\mu}^0 + \lambda_1 \tilde{S}^2 \\ \tilde{S}^2 = \frac{\sum_{i=1}^n (Y_i - \rho_{\tilde{a}}(X_i))^2 \mathbb{1}_{\{\tilde{\tau}_1 \leq X_i < \tilde{\tau}_2\}}}{\sum_{i=1}^n \mathbb{1}_{\{\tilde{\tau}_1 \leq X_i < \tilde{\tau}_2\}}} \end{array} \right.$$

Sous la contrainte $A_d(\tilde{\tau}_2, y)$, la recherche du point critique se fait par la résolution du système précédent dans lequel il faut remplacer $\tilde{\tau}_1$ par $\tilde{\tau}_2$ et x par y .

Sous la contrainte $A_d(\tilde{\tau}_1, x) \wedge \mathcal{A}_d(\tilde{\tau}_2, y)$, la recherche du point critique se fait par la résolution du système de dimension $(\tilde{\delta} + 1) + 5$ suivant et qui correspond aux équations (A.6), (A.7) et (A.8) de l'annexe A.4,

$$\left\{ \begin{array}{l} \rho_{\tilde{a}}(\tilde{\tau}_1) - x = 0 \\ \dot{\rho}_{\tilde{a}}(\tilde{\tau}_1) - \dot{x} = 0 \\ \rho_{\tilde{a}}(\tilde{\tau}_2) - y = 0 \\ \dot{\rho}_{\tilde{a}}(\tilde{\tau}_2) - \dot{y} = 0 \\ \tilde{v}^\delta = \tilde{a}^\delta \tilde{\mu}^{2\delta} + \dots + \tilde{a}^0 \tilde{\mu}^\delta + \lambda_1 \tilde{S}^2 \tilde{\tau}_1^\delta + \delta \lambda_2 \tilde{S}^2 \tilde{\tau}_1^{\delta-1} + \lambda_3 \tilde{S}^2 \tilde{\tau}_2^\delta + \delta \lambda_4 \tilde{S}^2 \tilde{\tau}_2^{\delta-1} \\ \tilde{v}^{\delta-1} = \tilde{a}^\delta \tilde{\mu}^{2\delta-1} + \dots + \tilde{a}^0 \tilde{\mu}^{\delta-1} + \lambda_1 \tilde{S}^2 \tilde{\tau}_1^{\delta-1} + (\delta-1) \lambda_2 \tilde{S}^2 \tilde{\tau}_1^{\delta-2} + \lambda_3 \tilde{S}^2 \tilde{\tau}_2^{\delta-1} + (\delta-1) \lambda_4 \tilde{S}^2 \tilde{\tau}_2^{\delta-2} \\ \vdots \\ \tilde{v}^0 = \tilde{a}^\delta \tilde{\mu}^\delta + \dots + \tilde{a}^0 \tilde{\mu}^0 + \lambda_1 \tilde{S}^2 + \lambda_3 \tilde{S}^2 \\ \tilde{S}^2 = \frac{\sum_{i=1}^n (Y_i - \rho_{\tilde{a}}(X_i))^2 \mathbb{1}_{\{\tilde{\tau}_1 \leq X_i < \tilde{\tau}_2\}}}{\sum_{i=1}^n \mathbb{1}_{\{\tilde{\tau}_1 \leq X_i < \tilde{\tau}_2\}}} \end{array} \right.$$

La proposition 18 et l'algorithme A.4.2 sont donc justifiés.

A.6 Résolution des problèmes d'optimisation sous contrainte de régularité pour le modèle à variable latente.

Dans cette annexe, nous nous rapportons à la section 2.3.2 qui nous place dans le contexte où le nombre de régimes est fixé à $K = 2$. À T_1 fixé, nous proposons de résoudre le problème de maximisation suivant, qui doit être résolu à chaque itération de l'algorithme EM, par la méthode des multiplicateurs de Lagrange.

$$\begin{aligned} (\mathbf{a}_2^{(l+1)}, \mathbf{S}_2^{2(l+1)}) = \underset{\substack{\mathbf{a}_2, \mathbf{S}_2^2 \\ \text{sous } \mathcal{A}_*(\tau_2)}}{\operatorname{argmax}} \psi(T_1, (\mathbf{a}_2, \mathbf{S}_2^2), (\mathbf{a}_2^l, \mathbf{S}_2^{2l})). \end{aligned}$$

Nous avons vu à la section 2.3.2 que la maximisation de $\psi(T_1, (\mathbf{a}_2, \mathbf{S}_2^2), (\mathbf{a}_2^l, \mathbf{S}_2^{2l}))$ en $(\mathbf{a}_2, \mathbf{S}_2^2)$ était équivalente à celle de

$$-\frac{1}{2} \sum_{i=1}^n \sum_{z=1}^2 \left[\log(S_z^2) + \frac{(Y_i - \rho_{a_z}(X_i))^2}{S_z^2} \right] f_{Z|X,Y}(z|X_i, Y_i, T_1, \mathbf{a}_2^l, \mathbf{S}_2^{2l})$$

En suivant le raisonnement de l'annexe A.5.1 et en reprenant les notations des équations (2.20) et (2.21), la recherche du point critique se fait par la résolution des systèmes suivants, de dimensions $2(\delta + 1) + 2$ et $2(\delta + 1) + 3$, sous contrainte de continuité ou de dérivabilité. Ces systèmes corres-

pondent à la proposition 4 et l'algorithme 2.3.2 de la section 2.3.2,

$$\left\{ \begin{array}{l} \rho_{a_1}(T_1) - \rho_{a_2}(T_1) = 0 \\ v_{1,EM}^\delta = a_1^\delta \mu_{1,EM}^{2\delta} + \dots + a_1^0 \mu_{1,EM}^\delta + \lambda S_1^2 T_1^\delta \\ v_{1,EM}^{\delta-1} = a_1^\delta \mu_{1,EM}^{2\delta-1} + \dots + a_1^0 \mu_{1,EM}^{\delta-1} + \lambda S_1^2 T_1^{\delta-1} \\ \vdots \\ v_{1,EM}^0 = a_1^\delta \mu_{1,EM}^\delta + \dots + a_1^0 \mu_{1,EM}^0 + \lambda S_1^2 \\ v_{2,EM}^\delta = a_2^\delta \mu_{2,EM}^{2\delta} + \dots + a_2^0 \mu_{2,EM}^\delta - \lambda S_2^2 T_1^\delta \\ v_{2,EM}^{\delta-1} = a_2^\delta \mu_{2,EM}^{2\delta-1} + \dots + a_2^0 \mu_{2,EM}^{\delta-1} - \lambda S_2^2 T_1^{\delta-1} \\ \vdots \\ v_{2,EM}^0 = a_2^\delta \mu_{2,EM}^\delta + \dots + a_2^0 \mu_{2,EM}^0 - \lambda S_2^2 \\ S_1^2 = \frac{\sum_{i=1}^n (Y_i - \rho_{a_1}(X_i))^2 f_{Z|X,Y}(z|X_i, Y_i, (\mathbf{a}'_2, \mathbf{S}'_2))}{\sum_{i=1}^n f_{Z|X,Y}(z|X_i, Y_i, (\mathbf{a}'_2, \mathbf{S}'_2))} \\ S_2^2 = \frac{\sum_{i=1}^n (Y_i - \rho_{a_2}(X_i))^2 f_{Z|X,Y}(z|X_i, Y_i, (\mathbf{a}'_2, \mathbf{S}'_2))}{\sum_{i=1}^n f_{Z|X,Y}(z|X_i, Y_i, (\mathbf{a}'_2, \mathbf{S}'_2))}. \end{array} \right.$$

$$\left\{ \begin{array}{l} \rho_{a_1}(T_1) - \rho_{a_2}(T_1) = 0 \\ \dot{\rho}_{a_1}(T_1) - \dot{\rho}_{a_2}(T_1) = 0 \\ v_{1,EM}^\delta = a_1^\delta \mu_{1,EM}^{2\delta} + \dots + a_1^0 \mu_{1,EM}^\delta + \lambda_1 S_1^2 T_1^\delta + \delta \lambda_2 S_1^2 T_1^{\delta-1} \\ v_{1,EM}^{\delta-1} = a_1^\delta \mu_{1,EM}^{2\delta-1} + \dots + a_1^0 \mu_{1,EM}^{\delta-1} + \lambda_1 S_1^2 T_1^{\delta-1} + (\delta-1) \lambda_2 S_1^2 T_1^{\delta-2} \\ \vdots \\ v_{1,EM}^0 = a_1^\delta \mu_{1,EM}^\delta + \dots + a_1^0 \mu_{1,EM}^0 + \lambda_1 S_1^2 \\ v_{2,EM}^\delta = a_2^\delta \mu_{2,EM}^{2\delta} + \dots + a_2^0 \mu_{2,EM}^\delta - \lambda_1 S_2^2 T_1^\delta - \delta \lambda_2 S_2^2 T_1^{\delta-1} \\ v_{2,EM}^{\delta-1} = a_2^\delta \mu_{2,EM}^{2\delta-1} + \dots + a_2^0 \mu_{2,EM}^{\delta-1} - \lambda_1 S_2^2 T_1^{\delta-1} - (\delta-1) \lambda_2 S_2^2 T_1^{\delta-2} \\ \vdots \\ v_{2,EM}^0 = a_2^\delta \mu_{2,EM}^\delta + \dots + a_2^0 \mu_{2,EM}^0 - \lambda_1 S_2^2 \\ S_1^2 = \frac{\sum_{i=1}^n (Y_i - \rho_{a_1}(X_i))^2 f_{Z|X}(1|X_i, T_1)}{\sum_{i=1}^n f_{Z|X}(1|X_i, T_1)} \\ S_2^2 = \frac{\sum_{i=1}^n (Y_i - \rho_{a_2}(X_i))^2 f_{Z|X}(2|X_i, T_1)}{\sum_{i=1}^n f_{Z|X}(2|X_i, T_1)}. \end{array} \right.$$

A.7 Complément d'étude du comportement asymptotique de $\mathbb{E}[\hat{\sigma}_{n+m}^2]$

Cette annexe fait référence à la section 3.4.3 dans laquelle nous étudions le comportement asymptotique de $\mathbb{E}[\hat{\sigma}_{n+m}^2]$. Nous avons

$$\mathbb{E}[\hat{\sigma}_{n+m}^2] \rightarrow \frac{3(\sigma^2 + \alpha\sigma^{2*})(\alpha+1)^3 + \alpha^3(b-b^*)^2 T^2}{3(\alpha+1)^4} = f(\alpha).$$

L'étude des variations de $f(\alpha)$ passe par l'étude du signe de sa dérivée notée $\frac{\partial f(\alpha)}{\partial \alpha}$. Dans cette annexe, nous proposons de montrer les différents calculs réalisés pour obtenir les propositions 15, 16 et 17.

Dans le but de déterminer les valeurs de α qui annulent $\frac{\partial f(\alpha)}{\partial \alpha}$, nous posons $\gamma = \frac{(\sigma^{2*} - \sigma^2)}{(b-b^*)^2 T^2}$ et nous considérons le cas $b \neq b^*$ pour résoudre l'équation suivante,

$$\frac{\alpha^2(\alpha-3)}{3(\alpha+1)^3} = \gamma,$$

ce qui revient à résoudre l'équation de degré 3 suivante,

$$a\alpha^3 + b\alpha^2 + c\alpha + d = 0, \quad (\text{A.12})$$

avec,

$$a = (1 - 3\gamma),$$

$$b = (-3 - 9\gamma),$$

$$c = -9\gamma,$$

$$d = -3\gamma.$$

Nous proposons de résoudre cette équation de degré 3 par la méthode de Cardan. Cette méthode consiste à diviser l'équation A.12 par a et à poser

$$\begin{aligned} \alpha &= X - \frac{b}{3a} \\ &= X + \frac{(1 + 3\gamma)}{(1 - 3\gamma)}, \end{aligned}$$

ce qui permet d'étudier l'équation de forme simplifiée (qui n'a pas de terme en X^2) suivante,

$$X^3 + pX + q = 0, \quad (\text{A.13})$$

avec,

$$\begin{aligned} p &= \frac{-b^2}{3a^2} + \frac{c}{a} \\ &= \frac{-(9\gamma + 1)}{3(\gamma - \frac{1}{3})^2}, \end{aligned}$$

$$\begin{aligned} q &= \frac{2b^3}{27a^3} - \frac{bc}{3a^2} + \frac{d}{a} \\ &= \frac{2(18\gamma^2 + 15\gamma + 1)}{27(\gamma - \frac{1}{3})^3}. \end{aligned}$$

En trouvant les racines de l'équation A.13, nous serons en mesure de trouver celle de l'équation A.12. Pour ce faire, l'astuce de Cardan consiste à poser $X = U + V$ et à introduire deux nouvelles inconnues au lieu d'une. L'intérêt étant d'avoir un degré de liberté supplémentaire dans l'équation A.13 qui devient alors,

$$(U + V)^3 + p(U + V) + q = 0,$$

qui s'écrit de manière équivalente,

$$U^3 + V^3 + (U + V)(3UV + p) + q = 0$$

L'avantage d'avoir une équation à deux inconnus est que nous pouvons leur imposer une relation supplémentaire, comme $3UV + p = 0$. En faisant cela l'intérêt de l'astuce de Cardan se révèle, car grâce à cette relation, nous sommes désormais à la recherche de U et V qui sont solutions du système suivant,

$$\begin{cases} U^3 + V^3 = -q, \\ U^3V^3 = \left(\frac{-p}{3}\right)^3, \end{cases}$$

ce qui est équivalent à chercher les valeurs de U^3 et V^3 qui sont solutions de l'équation

$$X^2 + qX + \left(\frac{-p}{3}\right)^3 = 0.$$

Le discriminant de ce polynôme, noté Δ , s'écrit

$$\begin{aligned}\Delta &= \frac{q^2}{4} + \frac{p^3}{27} \\ &= \frac{4\gamma\left(\gamma + \frac{1}{12}\right)}{9\left(\gamma - \frac{1}{3}\right)^4}.\end{aligned}$$

Il est défini $\forall \gamma \neq \frac{1}{3}$. Dans le cas où $\gamma = \frac{1}{3}$, l'équation (A.12) s'écrit $-6\alpha^2 - 3\alpha - 1 = 0$ et ce polynôme ne possède pas de racines réelles. Le signe de ce déterminant donne les indications suivantes sur les racines de l'équation (A.12) :

- si $\Delta > 0$: l'équation $X^3 + pX + q = 0$ possède une racine réelle et deux racines complexes conjuguées.
- si $\Delta = 0$: l'équation $X^3 + pX + q = 0$ possède trois racines réelles dont l'une est multiple.
- si $\Delta < 0$: l'équation $X^3 + pX + q = 0$ possède trois racines réelles distinctes.

Nous avons donc calculé le tableau de signe associé en fonction des valeurs prises par γ .

γ	$-\infty$	$-\frac{1}{12}$	0	$\frac{1}{3}$	$+\infty$
4γ		-	0	+	
$\left(\gamma + \frac{1}{12}\right)$		-	0	+	
Δ	+	0	-	0	+

Dans la suite, nous étudions les différentes valeurs prises par γ pour calculer, dans chacun de ces cas, les racines réelles de l'équation (A.13) et donc a posteriori celles de l'équation (A.12).

Cas où $\gamma = 0$

Dans ce contexte il y a égalité des variances : $\sigma^2 = \sigma^{2*}$. D'après le tableau de signe de Δ nous devons, dans ce cas, trouver deux racines réelles (dont une double) de l'équation (A.13). Nous déterminerons ensuite celles de l'équation (A.12).

Lorsque $\gamma = 0$, les racines U^3 et V^3 du polynôme $X^2 + qX + \left(\frac{-p}{3}\right)^3$ sont égales et valent

$$U^3 = V^3 = \frac{-q}{2} = 1,$$

car $q = -2$ quand $\gamma = 0$. Nous avons précédemment posé $X = U + V$, la racine X_1 du polynôme $X^3 + pX + q$ de l'équation (A.13) vaut donc,

$$\begin{aligned}X_1 &= \sqrt[3]{1} + \sqrt[3]{1} \\ &= 2.\end{aligned}$$

Les autres racines du polynôme $X^3 + pX + q = 0$, doivent vérifier les conditions suivantes,

$$\begin{cases} X_2 + X_3 = -X_1, \\ X_2X_3 = \frac{-q}{X_1}, \end{cases}$$

c'est-à-dire que X_2 et X_3 sont solutions de l'équation

$$X^2 + X_1X - \frac{q}{X_1} = 0,$$

qui se réécrit

$$X^2 + 2X + 1 = 0,$$

car $q = -2$ et $X_1 = 2$. Le discriminant de ce polynôme est nul, ses racines X_2 et X_3 sont donc identiques et valent

$$X_2 = X_3 = \frac{-X_1}{2} = -1.$$

Finalement, quand $\gamma = 0$, les racines du polynôme de l'équation A.12 sont les suivantes,

$$\begin{aligned} \alpha_1 &= X_1 + \frac{(1+3\gamma)}{(1-3\gamma)} \\ &= X_1 + 1 \\ &= 3. \end{aligned}$$

$$\begin{aligned} \alpha_2 = \alpha_3 &= X_2 + 1 \\ &= 0. \end{aligned}$$

Cas où $\gamma = \frac{-1}{12}$

D'après le tableau de signe de Δ , dans le cas où $\gamma = \frac{-1}{12}$, nous devons trouver deux racines réelles (dont une double) de l'équation (A.13). Nous déterminerons ensuite celles de l'équation (A.12).

Lorsque $\gamma = \frac{-1}{12}$, les racines U^3 et V^3 du polynôme $X^2 + qX + \left(\frac{-p}{3}\right)^3$ sont égales et valent

$$U^3 = V^3 = \frac{-q}{2} = \frac{-8}{125}.$$

La racine X_1 du polynôme $X^3 + pX + q$ de l'équation (A.13) vaut

$$\begin{aligned} X_1 &= \sqrt[3]{\frac{-8}{125}} + \sqrt[3]{\frac{-8}{125}} \\ &= \frac{-4}{5}. \end{aligned}$$

Les autres racines du polynôme $X^3 + pX + q = 0$ vérifient les conditions suivantes,

$$\begin{cases} X_2 + X_3 = -X_1, \\ X_2X_3 = \frac{-q}{X_1}, \end{cases}$$

c'est-à-dire que X_2 et X_3 sont solutions de l'équation

$$X^2 + X_1X - \frac{q}{X_1} = 0,$$

qui se réécrit

$$X^2 - \frac{4}{5}X + \frac{4}{25} = 0.$$

car $X_1 = \frac{-4}{5}$ et $q = \frac{16}{125}$. Le discriminant de ce polynôme est nul, ses racines X_2 et X_3 sont identiques et valent

$$X_2 = X_3 = \frac{-X_1}{2} = \frac{2}{5}.$$

Finalement, quand $\gamma = \frac{-1}{12}$, les racines du polynôme de l'équation A.12 sont les suivantes,

$$\begin{aligned} \alpha_1 &= X_1 + \frac{(1+3\gamma)}{(1-3\gamma)} \\ &= X_1 + \frac{3}{5} \\ &= \frac{-1}{5}. \end{aligned}$$

$$\begin{aligned} \alpha_2 = \alpha_3 &= X_2 + \frac{3}{5} \\ &= 1. \end{aligned}$$

Cas où $-\frac{1}{12} < \gamma < 0$

D'après le tableau de signe de Δ , dans le cas où $-\frac{1}{12} < \gamma < 0$, nous devons trouver trois racines réelles distinctes de l'équation (A.13). Ces racines nous permettront de déterminer celles de l'équation (A.12).

Lorsque le discriminant est négatif, la recherche des racines réelles de l'équation (A.13) consiste à exploiter la proposition 19 et les identités trigonométriques suivantes,

$$\cos(3\theta) = 4 \cos^3(\theta) - 3 \cos(\theta),$$

$$\sin(3\theta) = 3 \sin(\theta) - 4 \sin^3(\theta),$$

qui peuvent se réécrire de la façon suivante,

$$\cos^3(\theta) = \frac{1}{4} \cos(3\theta) + \frac{3}{4} \cos(\theta),$$

$$\sin^3(\theta) = -\frac{1}{4} \sin(3\theta) + \frac{3}{4} \sin(\theta).$$

Proposition 19. *Si le discriminant Δ est strictement négatif alors,*

$$\begin{cases} p < 0, \\ \frac{3q}{2p} \sqrt{-\frac{3}{p}} \in]-1, 1[, \end{cases}$$

Nous posons la notation $X = A \cos(\theta)$, où A est un paramètre non nul et nous souhaitons désormais résoudre l'équation suivante,

$$A^3 \cos^3(\theta) + pA \cos(\theta) = -q$$

L'astuce est de faire en sorte que le membre de gauche de cette équation soit proportionnel à $\cos(3\theta)$, ce qui revient à poser $A = 2\sqrt{-\frac{p}{3}}$. En faisant cela X se réécrit,

$$X = 2\sqrt{-\frac{p}{3}} \cos(\theta),$$

et l'équation à résoudre devient alors,

$$\cos(3\theta) = \frac{3q}{2p} \sqrt{-\frac{3}{p}}.$$

Nous en déduisons,

$$X_k = 2\sqrt{-\frac{p}{3}} \cos(\theta_k),$$

avec

$$\theta_k = \frac{\beta}{3} + \frac{2k\pi}{3},$$

où $\beta = \arccos\left(\frac{3q}{2p}\sqrt{-\frac{3}{p}}\right)$. Les trois racines de l'équation $X^3 + pX + q = 0$ s'écrivent donc de la façon suivante,

$$X = 2\sqrt{-\frac{p}{3}} \cos(\theta_s), \quad \forall s \in \{0, 1, 2\}$$

$$- X_1 = 2\sqrt{-\frac{p}{3}} \cos\left(\frac{\beta}{3}\right),$$

$$- X_2 = 2\sqrt{-\frac{p}{3}} \cos\left(\frac{\beta+2\pi}{3}\right),$$

$$- X_3 = 2\sqrt{-\frac{p}{3}} \cos\left(\frac{\beta+4\pi}{3}\right).$$

Les trois racines réelles de l'équation (A.12) sont donc,

$$- \alpha_1 = X_1 - \frac{b}{3a},$$

$$- \alpha_2 = X_2 - \frac{b}{3a},$$

$$- \alpha_3 = X_3 - \frac{b}{3a}.$$

Cas où $\gamma < \frac{-1}{12}$ ou $\gamma > 0$

D'après le tableau de signe de Δ , dans le cas où $\gamma < \frac{-1}{12}$ ou $\gamma > 0$, nous devons trouver une racine réelle de l'équation (A.13). Cette racine nous permettra de déterminer la racine réelle de l'équation (A.12).

Les racines U^3 et V^3 de l'équation $X^2 + qX + \left(\frac{-p}{3}\right)^3 = 0$ valent :

$$U^3 = \frac{-q + \sqrt{\Delta}}{2} = \frac{-(18\gamma^2 + 15\gamma + 1)}{27\left(\gamma - \frac{1}{3}\right)^3} + \frac{\sqrt{\gamma\left(\gamma + \frac{1}{12}\right)}}{3\left(\gamma - \frac{1}{3}\right)^2}.$$

$$V^3 = \frac{-q - \sqrt{\Delta}}{2} = \frac{-(18\gamma^2 + 15\gamma + 1)}{27\left(\gamma - \frac{1}{3}\right)^3} - \frac{\sqrt{\gamma\left(\gamma + \frac{1}{12}\right)}}{3\left(\gamma - \frac{1}{3}\right)^2}.$$

Nous avons posé $X = U + V$, la racine X_1 de l'équation $X^3 + pX + q = 0$ vaut donc,

$$X_1 = \frac{1}{3\left(\gamma - \frac{1}{3}\right)} \left(\sqrt[3]{-(18\gamma^2 + 15\gamma + 1) + 9\left(\gamma - \frac{1}{3}\right)\sqrt{\gamma\left(\gamma + \frac{1}{12}\right)}} + \sqrt[3]{-(18\gamma^2 + 15\gamma + 1) - 9\left(\gamma - \frac{1}{3}\right)\sqrt{\gamma\left(\gamma + \frac{1}{12}\right)}} \right).$$

La racine α_1 de l'équation initiale $aX^3 + bX^2 + cX + d = 0$ vaut donc,

$$\begin{aligned} \alpha_1 &= \frac{1}{3\left(\gamma - \frac{1}{3}\right)} \left(\sqrt[3]{-(18\gamma^2 + 15\gamma + 1) + 9\left(\gamma - \frac{1}{3}\right)\sqrt{\gamma\left(\gamma + \frac{1}{12}\right)}} \right. \\ &\quad \left. + \sqrt[3]{-(18\gamma^2 + 15\gamma + 1) - 9\left(\gamma - \frac{1}{3}\right)\sqrt{\gamma\left(\gamma + \frac{1}{12}\right)}} - (3\gamma + 1) \right). \end{aligned}$$

Les autres racines de l'équation $X^3 + pX + q = 0$ doivent vérifier les conditions suivantes,

$$\begin{cases} X_2 + X_3 = -X_1, \\ X_2 X_3 = \frac{-q}{X_1}. \end{cases}$$

X_2 et X_3 sont donc les racines de l'équation suivante,

$$X^2 + X_1 X - \frac{q}{X_1} = 0.$$

Le discriminant de ce polynôme vaut $\tilde{\Delta} = X_1^2 + \frac{4q}{X_1}$. Si il est positif ($\tilde{\Delta} > 0$), les racines X_2 et X_3 de l'équation $X^3 + pX + q = 0$ sont les suivantes,

$$X_2 = \frac{-X_1 + \sqrt{\tilde{\Delta}}}{2} \quad \text{et} \quad X_3 = \frac{-X_1 - \sqrt{\tilde{\Delta}}}{2}.$$

Si ce discriminant est nul ($\tilde{\Delta} = 0$), les racines X_2 et X_3 de l'équation $X^3 + pX + q = 0$ sont

$$X_2 = X_3 = \frac{-X_1}{2}.$$

Si ce discriminant est négatif ($\tilde{\Delta} < 0$), alors les racines X_2 et X_3 du polynôme $X^2 + pX + q = 0$ sont des racines complexes.

Étude de la racine réelle α_1 en fonction des valeurs de γ

Afin d'étudier les encadrements limites de la racine réelle α_1 , nous introduisons les variables ω et β .

— Étude de $0 < \gamma < \frac{\omega + 1}{3}$ lorsque $\omega \rightarrow 0^-$

$$\text{On a } \frac{-3(3\omega + 4)}{\omega^2} < p < -3 \text{ et } \frac{2(2\omega^2 + 9\omega + 8)}{\omega^3} < q < -2.$$

De ces résultats on obtient,

$$1 < U^3 < \frac{\sqrt{4\omega^2 + 9\omega + 5}}{2\omega^2} - \frac{(2\omega^2 + 9\omega + 8)}{\omega^3}.$$

$$1 < V^3 < \frac{-\sqrt{4\omega^2 + 9\omega + 5}}{2\omega^2} - \frac{(2\omega^2 + 9\omega + 8)}{\omega^3}.$$

Comme $X_1 = \sqrt[3]{U^3} + \sqrt[3]{V^3}$, alors,

$$X_1 > 2.$$

et donc comme $\alpha_1 = X_1 - \frac{b}{3a}$, nous avons,

$$\alpha_1 > 3.$$

Lorsque $0 < \gamma < \frac{\omega + 1}{3}$ avec $\omega \rightarrow 0^-$, la seule racine réelle est positive et est supérieure à trois.

— Étude de $\frac{\omega + 1}{3} < \gamma < \beta$ lorsque $\omega \rightarrow 0^+$ et $\beta \rightarrow +\infty$

On a

$$\frac{-3(3\omega + 4)}{\omega^2} < p < \frac{-(9\beta + 1)}{3(\beta - \frac{1}{3})^2},$$

et

$$\frac{2(2\omega^2 + 9\omega + 8)}{\omega^3} < q < \frac{2(18\beta^2 + 15\beta + 1)}{27(\beta - \frac{1}{3})^3}.$$

De ces résultats, on obtient,

$$\frac{\sqrt{(\omega+1)(4\omega+5)}}{2\omega^2} - \frac{2\omega^2+9\omega+8}{\omega^3} < U^3 < \frac{\sqrt{\beta(\beta+\frac{1}{12})}}{3(\beta-\frac{1}{3})^2} - \frac{(18\beta^2+15\beta+1)}{27(\beta-\frac{1}{3})^3}.$$

$$\frac{-\sqrt{(\omega+1)(4\omega+5)}}{2\omega^2} - \frac{2\omega^2+9\omega+8}{\omega^3} < V^3 < \frac{-\sqrt{\beta(\beta+\frac{1}{12})}}{3(\beta-\frac{1}{3})^2} - \frac{(18\beta^2+15\beta+1)}{27(\beta-\frac{1}{3})^3}.$$

L'étude des limites quand $\omega \rightarrow 0^+$ et $\beta \rightarrow +\infty$ donne les résultats suivants,

$$\lim_{\omega \rightarrow 0^+} \frac{\sqrt{(\omega+1)(4\omega+5)}}{2\omega^2} - \frac{2\omega^2+9\omega+8}{\omega^3} \rightarrow -\infty.$$

$$\lim_{\omega \rightarrow 0^+} \frac{-\sqrt{(\omega+1)(4\omega+5)}}{2\omega^2} - \frac{2\omega^2+9\omega+8}{\omega^3} \rightarrow -\infty.$$

$$\lim_{\beta \rightarrow +\infty} \frac{\sqrt{\beta(\beta+\frac{1}{12})}}{3(\beta-\frac{1}{3})^2} - \frac{(18\beta^2+15\beta+1)}{27(\beta-\frac{1}{3})^3} \rightarrow 0^-.$$

$$\lim_{\beta \rightarrow +\infty} \frac{-\sqrt{\beta(\beta+\frac{1}{12})}}{3(\beta-\frac{1}{3})^2} - \frac{(18\beta^2+15\beta+1)}{27(\beta-\frac{1}{3})^3} \rightarrow 0^-.$$

Comme $X_1 = \sqrt[3]{U^3} + \sqrt[3]{V^3}$, nous avons,

$$-\infty < X_1 < 0.$$

La solution finale $\alpha_1 = X_1 - \frac{b}{3a}$ est encadrée par

$$-\infty < \alpha_1 < -1,$$

car $-\frac{b}{3a} = \frac{-(3\gamma+1)}{(3\gamma-1)}$ avec,

$$\frac{-(\omega+2)}{\omega} < \frac{-(3\gamma+1)}{(3\gamma-1)} < \frac{-(3\beta+1)}{(3\beta-1)},$$

et les limites quand $\beta \rightarrow +\infty$ et $\omega \rightarrow 0^+$ sont les suivantes,

$$\lim_{\beta \rightarrow +\infty} \frac{-(3\beta+1)}{(3\beta-1)} \rightarrow -1.$$

$$\lim_{\omega \rightarrow 0^+} \frac{-(\omega+2)}{\omega} \rightarrow -\infty.$$

Lorsque $\frac{\omega+1}{3} < \gamma < \beta$ avec $\omega \rightarrow 0^+$ et $\beta \rightarrow +\infty$, la seule racine réelle qui annule la dérivée est négative.

— Étude de $\beta < \gamma < \frac{\omega+1}{12}$ lorsque $\omega \rightarrow 0^-$ et $\beta \rightarrow -\infty$

On a

$$\frac{-12(3\omega+1)}{(\omega-5)^2} < p < \frac{-(9\beta+1)}{3(\beta-\frac{1}{3})^2},$$

et

$$\frac{2(18\beta^2+15\beta+1)}{27(\beta-\frac{1}{3})^3} < q < \frac{16(\omega^2+8\omega-1)}{(\omega-5)^3}.$$

De ces résultats on obtient,

$$\frac{4\sqrt{\omega(\omega-1)}}{(\omega-5)^2} - \frac{8(\omega^2+8\omega-1)}{(\omega-5)^3} < U^3 < \frac{\sqrt{\beta(\beta+\frac{1}{12})}}{3(\beta-\frac{1}{3})^2} - \frac{(18\beta^2+15\beta+1)}{27(\beta-\frac{1}{3})^3}.$$

$$\frac{4\sqrt{\omega(\omega-1)}}{(\omega-5)^2} - \frac{8(\omega^2+8\omega-1)}{(\omega-5)^3} < V^3 < \frac{\sqrt{\beta(\beta+\frac{1}{12})}}{3(\beta-\frac{1}{3})^2} - \frac{(18\beta^2+15\beta+1)}{27(\beta-\frac{1}{3})^3}.$$

Les limites en $\omega \rightarrow 0^-$ et $\beta \rightarrow -\infty$ s'écrivent de la façon suivante,

$$\lim_{\omega \rightarrow 0^-} \frac{4\sqrt{\omega(\omega-1)}}{(\omega-5)^2} - \frac{8(\omega^2+8\omega-1)}{(\omega-5)^3} \rightarrow -\frac{8}{125}.$$

$$\lim_{\omega \rightarrow 0^-} \frac{-4\sqrt{\omega(\omega-1)}}{(\omega-5)^2} - \frac{8(\omega^2+8\omega-1)}{(\omega-5)^3} \rightarrow -\frac{8}{125}.$$

$$\lim_{\beta \rightarrow -\infty} \frac{\sqrt{\beta(\beta+\frac{1}{12})}}{3(\beta-\frac{1}{3})^2} - \frac{(18\beta^2+15\beta+1)}{27(\beta-\frac{1}{3})^3} \rightarrow 0^+.$$

$$\lim_{\beta \rightarrow -\infty} \frac{-\sqrt{\beta(\beta+\frac{1}{12})}}{3(\beta-\frac{1}{3})^2} - \frac{(18\beta^2+15\beta+1)}{27(\beta-\frac{1}{3})^3} \rightarrow 0^+.$$

Comme $X_1 = \sqrt[3]{U^3} + \sqrt[3]{V^3}$, nous avons,

$$\frac{-4}{5} < X_1 < 0.$$

La solution finale $\alpha_1 = X_1 - \frac{b}{3a}$ est encadrée par

$$-1 < \alpha_1 < -\frac{1}{5},$$

car $-\frac{b}{3a} = \frac{-(3\gamma+1)}{(3\gamma-1)}$ avec,

$$\frac{-(3\beta+1)}{(3\beta-1)} < \frac{-(3\gamma+1)}{(3\gamma-1)} < \frac{-(\omega+3)}{(\omega-5)},$$

et les limites quand $\beta \rightarrow -\infty$ et $\omega \rightarrow 0^-$ sont les suivantes

$$\lim_{\beta \rightarrow -\infty} \frac{-(3\beta+1)}{(3\beta-1)} \rightarrow -1.$$

$$\lim_{\omega \rightarrow 0^-} \frac{-(\omega+3)}{(\omega-5)} \rightarrow \frac{3}{5}.$$

Lorsque $\beta < \gamma < \frac{\omega+1}{12}$ ave $\omega \rightarrow 0^-$ et $\beta \rightarrow -\infty$, la seule racine réelle est également négative.

A.8 Documentation package HSPOR

Dans cette annexe, nous reproduisons la documentation du package HSPOR.

Package ‘HSPOR’

August 30, 2019

Title Hidden Smooth Polynomial Regression for Rupture Detection

Version 1.1.9

Description Several functions that allow by different methods to infer a piecewise polynomial regression model under regularity constraints, namely continuity or differentiability of the link function. The implemented functions are either specific to data with two regimes, or generic for any number of regimes, which can be given by the user or learned by the algorithm. A paper describing all these methods will be submitted soon. The reference will be added to this file as soon as available.

License LGPL-3

Encoding UTF-8

LazyData true

RoxygenNote 6.1.1

Imports stats,
corpcor,
npregfast,
graphics

R topics documented:

H2SPOR	1
H2SPOR_DynProg	3
HKSPOR	5
HKSPOR_DynProg	7
Index	9

H2SPOR	<i>Inference method for two regimes</i>
--------	---

Description

H2SPOR is an inference method that estimates, under regularity constraint, the parameters of a polynomial regression model with 2 regimes.

Usage

```
H2SPOR(X, Y, deg, constraint = 1, EM = TRUE, TimeTrans_Prop = c(),
        plotG = TRUE)
```

Arguments

X	A numerical vector corresponding to the explanatory variable. X must be sorted in ascending order if this is not the case, X will be sorted in the function and the corresponding permutation will be applied to Y. The user will be notified by a warning message. In addition, if X contains NAs, they will be deleted from the data and the user will be notified by a warning message. Finally, if X contains duplicate data, the excess data will be deleted and the user will be notified by a warning message.
Y	A numerical vector corresponding to the variable to be explain. It should contain two regimes that could be modelled by polynomials. In addition, if Y contains NAs they will be deleted from the data and the user will be notified by a warning message. Finally, if X contains duplicate data, the excess data will be deleted and the value of the remaining Y will become the average of the Ys, calculated for this value of X.
deg	The degree of polynomials. The size of X and Y must be greater than $2(\text{deg}+2) + 1$.
constraint	Number that determines the regularity assumption that is applied for the parameters estimation. By default, the variable is set to 1, i. e. the parameters estimation is done under continuity constraint. If the variable is 0 or 2, the estimation of the parameters will be done without assumption of regularity (constraint = 0) or under assumption of differentiability (constraint = 2). Warning, if the differentiability assumption is not verified by the model, it is preferable not to use it to estimate the model parameters. In addition, if the degree of the polynomials is equal to 1, you cannot use the differentiability assumption.
EM	A Boolean. If EM is TRUE (default), then the function will estimate the parameters of a latent variable polynomial regression model using an EM algorithm. If EM is FALSE then the function will estimate the parameters of the initial polynomial regression model by a fixed point algorithm.
TimeTrans_Prop	A numerical vector. This vector is empty by default. If you want to estimate the model parameters for a fixed jump time value, you can propose this value here.
plotG	A Boolean. If TRUE (default) the estimation results obtained by the H2SPOR function are plotted.

Value

A dataframe that contains the estimated parameters of the polynomial regression model at two regimes: the jump time, the coefficients of the polynomials and the variances of the two regimes. If plotG = TRUE, the data (X,Y) and the estimated model will be plotted.

Examples

```
#generated data with two regimes
set.seed(1)
```

```

xgrid1 = seq(0,10,length.out=6)
xgrid2 = seq(10.2,20,length.out=6)
ygrid1 = xgrid1^2-xgrid1+1+ rnorm(length(xgrid1),0,3)
ygrid2 = rep(91,length(xgrid2))+ rnorm(length(xgrid2),0,3)
xgrid = c(xgrid1,xgrid2)
ygrid = c(ygrid1,ygrid2)

#Inference of a polynomial regression model with two regimes on these data.
#The degree of the polynomials is fixed to 2 and the parameters are estimated
#under continuity constraint.
H2SPOR(xgrid,ygrid,2,1,EM=TRUE,c())

set.seed(1)
xgrid1 = seq(0,10,by=0.2)
xgrid2 = seq(10.2,20,by=0.2)
ygrid1 = xgrid1^2-xgrid1+1+ rnorm(length(xgrid1),0,3)
ygrid2 = rep(91,length(xgrid2))+ rnorm(length(xgrid2),0,3)
xgrid = c(xgrid1,xgrid2)
ygrid = c(ygrid1,ygrid2)

#Inference of a polynomial regression model with two regimes on these data.
#The degree of the polynomials is fixed to 2 and the parameters are estimated
#under continuity constraint.
H2SPOR(xgrid,ygrid,2,1,EM=FALSE,c())
#Executed time : 9.69897 secs (intel core i7 processor)

```

H2SPOR_DynProg

Inference method that does not require a priori knowledge of the number of regimes and uses dynamic programming

Description

H2SPOR_DynProg is an inference method implemented as a binary segmentation algorithm. This method makes it possible to estimate, using dynamic programming and under regularity assumption, the parameters of a piecewise polynomial regression model when we have no a priori knowledge of the number of regimes.

Usage

```
H2SPOR_DynProg(X, Y, deg, constraint = 1, EM = TRUE, plotG = TRUE)
```

Arguments

X A numerical vector corresponding to the explanatory variable. *X* must be sorted in ascending order if this is not the case, *X* will be sorted in the function and the corresponding permutation will be applied to *Y*. The user will be notified by a warning message. In addition, if *X* contains NAs, they will be deleted from the data and the user will be notified by a warning message. Finally, if *X* contains duplicate data, the excess data will be deleted and the user will be notified by a warning message.

Y	A numerical vector corresponding to the variable to be explain. It should contain at least two regimes that could be modelled by polynomials. In addition, if Y contains NAs they will be deleted from the data and the user will be notified by a warning message. Finally, if X contains duplicate data, the excess data will be deleted and the value of the remaining Y will become the average of the Ys, calculated for this value of X.
deg	Degree of the polynomials. The size of X and Y must be greater than $2(\text{deg}+2) + 1$.
constraint	Number that determines the regularity assumption that is applied for the parameters estimation. By default, the variable is set to 1, i. e. the parameters estimation is done under continuity constraint. If the variable is 0 or 2, the estimation of the parameters will be done without assumption of regularity (constraint = 0) or under assumption of differentiability (constraint = 2). Warning, if the differentiability assumption is not verified by the model, it is preferable not to use it to estimate the model parameters. In addition, if the degree of the polynomials is equal to 1, you cannot use the differentiability assumption.
EM	A Boolean. If EM is TRUE (default), then the function will estimate the parameters of a latent variable polynomial regression model using an EM algorithm. If EM is FALSE then the function will estimate the parameters of the initial polynomial regression model by a fixed point algorithm.
plotG	A Boolean. If TRUE (default) the estimation results obtained by the H2SPOR_DynProg function are plotted.

Value

A dataframe which contains the estimated parameters of the polynomial regression model at an estimated number of regimes: the times of jump, the polynomials coefficients and the variances of an estimated number of regimes. If plotG = TRUE, the data(X,Y) and the estimated model will be plotted.

Examples

```
set.seed(1)
#generated data with two regimes
xgrid1 = seq(0,10,length.out = 6)
xgrid2 = seq(10.2,20,length.out=6)
ygrid1 = xgrid1^2-xgrid1+1+ rnorm(length(xgrid1),0,3)
ygrid2 = rep(91,length(xgrid2))+ rnorm(length(xgrid2),0,3)
xgrid = c(xgrid1,xgrid2)
ygrid = c(ygrid1,ygrid2)
# Inference of a piecewise polynomial regression model on these data.
#The degree of the polynomials is fixed to 2 and the parameters are estimated
#under continuity constraint.
H2SPOR_DynProg(xgrid,ygrid,2,1,EM=FALSE)
```

```
set.seed(1)
xgrid1 = seq(0,10,by=0.2)
xgrid2 = seq(10.2,20,by=0.2)
xgrid3 = seq(20.2,30,by=0.2)
ygrid1 = xgrid1^2-xgrid1+1+ rnorm(length(xgrid1),0,3)
ygrid2 = rep(91,length(xgrid2))+ rnorm(length(xgrid2),0,3)
ygrid3 = -10*xgrid3+300+rnorm(length(xgrid3),0,3)
datX = c(xgrid1,xgrid2,xgrid3)
```

```

datY = c(ygrid1,ygrid2,ygrid3)
#Inference of a piecewise polynomial regression model on these data.
#The degree of the polynomials is fixed to 2 and the parameters are estimated
#under continuity constraint.
H2SPOR_DynProg(datX,datY,2,1)
#Executed time : 2.349685 mins (intel core i7 processor)

```

HKSPOR

Inference method for any number K of regimes

Description

HKSPOR is an inference method that estimates, under regularity constraint, the parameters of a polynomial regression model for a number K of regimes given by the user.

Usage

```

HKSPOR(X, Y, deg, K, constraint = 1, EM = TRUE, TimeTrans_Prop = c(),
       plotG = TRUE)

```

Arguments

X	A numerical vector corresponding to the explanatory variable. X must be sorted in ascending order if this is not the case, X will be sorted in the function and the corresponding permutation will be applied to Y. The user will be notified by a warning message. In addition, if X contains NAs, they will be deleted from the data and the user will be notified by a warning message. Finally, if X contains duplicate data, the excess data will be deleted and the user will be notified by a warning message.
Y	A numerical vector corresponding to the variable to be explain. It should contain at least two regimes that could be modelled by polynomials. In addition, if Y contains NAs they will be deleted from the data and the user will be notified by a warning message. Finally, if X contains duplicate data, the excess data will be deleted and the value of the remaining Y will become the average of the Ys, calculated for this value of X.
deg	Degree of the polynomials. The size of X and Y must be greater than $K(deg+2) + K$.
K	The number of regimes. The size of X and Y must be greater than $K(deg+2) + K$.
constraint	Number that determines the regularity assumption that is applied for the parameters estimation. By default, the variable is set to 1, i. e. the parameters estimation is done under continuity constraint. If the variable is 0 or 2, the estimation of the parameters will be done without assumption of regularity (constraint = 0) or under assumption of differentiability (constraint = 2). Warning, if the differentiability assumption is not verified by the model, it is preferable not to use it to estimate the model parameters. In addition, if the degree of the polynomials is equal to 1, you cannot use the differentiability assumption.

EM	A Boolean. If EM is TRUE (default), then the function will estimate the parameters of a latent variable polynomial regression model using an EM algorithm. If EM is FALSE then the function will estimate the parameters of the initial polynomial regression model by a fixed point algorithm.
TimeTrans_Prop	A numerical vector. This vector is empty by default. If you want to estimate the model parameters for fixed jump time values, you can propose these values here. Warning, the size of this vector must be equal to K-1.
plotG	A Boolean. If TRUE (default) the estimation results obtained by the HKSPOR function are plotted.

Value

A dataframe which contains the estimated parameters of the polynomial regression model at K regimes: the times of transition, the polynomials coefficients and the variances of the K regimes. If plotG = TRUE, the data (X,Y) and the estimated model will be plotted.

Examples

```
set.seed(3)
xgrid1 = seq(0,10,by=0.2)
xgrid2 = seq(10.2,20,by=0.2)
xgrid3 = seq(20.2,30,by=0.2)
ygrid1 = xgrid1^2-xgrid1+1+ rnorm(length(xgrid1),0,3)
ygrid2 = rep(91,length(xgrid2))+ rnorm(length(xgrid2),0,3)
ygrid3 = -10*xgrid3+300+rnorm(length(xgrid3),0,3)
xgrid = c(xgrid1,xgrid2,xgrid3)
ygrid = c(ygrid1,ygrid2,ygrid3)

#Inference of a polynomial regression model with three regimes on these data.
#The degree of the polynomials is fixed to 2 and the parameters are estimated
# under continuity constraint when the times of jump are fixed to 10 and 20.

HKSPOR(xgrid,ygrid,2,3,1,EM = FALSE,c(10,20))
```

```
set.seed(3)
xgrid1 = seq(0,10,by=0.2)
xgrid2 = seq(10.2,20,by=0.2)
xgrid3 = seq(20.2,30,by=0.2)
ygrid1 = xgrid1^2-xgrid1+1+ rnorm(length(xgrid1),0,3)
ygrid2 = rep(91,length(xgrid2))+ rnorm(length(xgrid2),0,3)
ygrid3 = -10*xgrid3+300+rnorm(length(xgrid3),0,3)
xgrid = c(xgrid1,xgrid2,xgrid3)
ygrid = c(ygrid1,ygrid2,ygrid3)

#Inference of a polynomial regression model with three regimes (K=3) on these data.
#The degree of the polynomials is fixed to 2 and the parameters are estimated
#under continuity constraint.
HKSPOR(xgrid,ygrid,2,3,1)
#Executed time : 49.70051 mins (intel core i7 processor)
```

HKSPOR_DynProg	<i>Inference method for any number K of regimes using dynamic programming</i>
----------------	---

Description

HKSPOR_DynProg is an inference method implemented in the form of a Bellman algorithm that estimates, under the assumption of regularity, the parameters of a polynomial regression model for a number K of regimes given by the user..

Usage

```
HKSPOR_DynProg(X, Y, deg, K, constraint = 1, smoothing = TRUE,
               verbose = FALSE, plotG = TRUE)
```

Arguments

X	A numerical vector corresponding to the explanatory variable. X must be sorted in ascending order if this is not the case, X will be sorted in the function and the corresponding permutation will be applied to Y. The user will be notified by a warning message. In addition, if X contains NAs, they will be deleted from the data and the user will be notified by a warning message. Finally, if X contains duplicate data, the excess data will be deleted and the user will be notified by a warning message.
Y	A numerical vector corresponding to the variable to be explain. It should contain at least two regimes that could be modelled by polynomials. In addition, if Y contains NAs they will be deleted from the data and the user will be notified by a warning message. Finally, if X contains duplicate data, the excess data will be deleted and the value of the remaining Y will become the average of the Ys, calculated for this value of X.
deg	The degree of the polynomials. The size of X and Y must be greater than $K(deg+2) + K$.
K	The number of regimes. The size of X and Y must be greater than $K(deg+2) + K$.
constraint	Number that determines the regularity assumption that is applied for the parameters estimation. By default, the variable is set to 1, i. e. the parameters estimation is done under continuity constraint. If the variable is 0 or 2, the estimation of the parameters will be done without assumption of regularity (constraint = 0) or under assumption of differentiability (constraint = 2). Warning, if the differentiability assumption is not verified by the model, it is preferable not to use it to estimate the model parameters. In addition, in this dynamic programming method, to ensure that the number of constraints is not greater that the number of parameters to be estimated, the degree of the polynomials must be at least equal to 3 to be able to use the differentiability assumption.
smoothing	A Boolean. If TRUE (default), the method will estimate the parameters of a piecewise polynomial regression model with latent variable by maximizing the log-likelihood weighted by the probability of being in the latent variable regime. If FALSE, the method will estimate the parameters of the piecewise polynomial regression model.

verbose	A Boolean. If FALSE (default) the HKSPOR_Dynprog function will return only one dataframe containing the parameter estimates obtained for a model at K regimes. If TRUE, the function will return all the results obtained for a model with 1 regime up to K regimes.
plotG	A Boolean. If TRUE (default) the estimation results obtained by the HKSPOR_DynProg function are plotted.

Value

One or more dataframes depend on the verbose value. If verbose = False, the output table will contain the estimated parameters of the polynomial regression model at K regimes: jump times, polynomial coefficients and variances of K regimes. If verbose = True then there will be K dataframes in output. Each table will contain the results of the estimated parameters obtained for each value of k (k=1,...,k=K). If plotG = TRUE, the data (X,Y) and the estimated model(s) will be plotted.

Examples

```
#generated data with three regimes
set.seed(1)
xgrid1 = seq(0,10,length.out=6)
xgrid2 = seq(10.2,20,length.out=6)
ygrid1 = xgrid1^2-xgrid1+1+ rnorm(length(xgrid1),0,4)
ygrid2 = rep(91,length(xgrid2))+ rnorm(length(xgrid2),0,4)
datX = c(xgrid1,xgrid2)
datY = c(ygrid1,ygrid2)

#Inference of a polynomial regression model with two regimes (K=2) on these data.
#The degree of the polynomials is fixed to 2 and the parameters are estimated
#under continuity constraint.
HKSPOR_DynProg(datX,datY,2,2)

set.seed(2)
xgrid1 = seq(0,10,by=0.2)
xgrid2 = seq(10.2,20,by=0.2)
xgrid3 = seq(20.2,30,by=0.2)
ygrid1 = xgrid1^2-xgrid1+1+ rnorm(length(xgrid1),0,3)
ygrid2 = rep(91,length(xgrid2))+ rnorm(length(xgrid2),0,3)
ygrid3 = -10*xgrid3+300+rnorm(length(xgrid3),0,3)
datX = c(xgrid1,xgrid2,xgrid3)
datY = c(ygrid1,ygrid2,ygrid3)

#Inference of a polynomial regression model with three (K=3) regimes on these data.
#The degree of the polynomials is fixed to 2 and the parameters are estimated
#under continuity constraint.
HKSPOR_DynProg(datX,datY,2,3)
#Executed time : 3.658121 mins (intel core i7 processor)
```

