

# **La modélisation des transports: le cas des comptages routiers**

F.-X. de Rossi, OFS et KPT  
A.F. Gualtierotti, IDHEAP et HEC-UNIL  
F. Mouton, IDHEAP et UNIGE

2002

# Table des matières

<b>1</b>	<b>Motifs, contenu et résultats</b>	<b>4</b>
1.1	Motifs et contenu . . . . .	4
1.2	Résultats . . . . .	6
<b>2</b>	<b>La méthodologie de modélisation du domaine des transports</b>	<b>8</b>
2.1	Introduction . . . . .	8
2.2	La théorie du consommateur et des choix discrets . . . . .	10
2.2.1	Description du comportement individuel . . . . .	10
2.2.2	Description du comportement collectif à partir des comportements individuels . . . . .	12
2.3	Les modèles de génération . . . . .	13
2.3.1	Modèles agrégés . . . . .	13
2.3.2	Modèles désagrégés . . . . .	13
2.4	Les modèles de distribution . . . . .	14
2.4.1	Justification des modèles gravitaires . . . . .	15
2.5	Les modèles de partage modal . . . . .	16
2.6	Le choix d'itinéraire . . . . .	17
2.6.1	La formule d'Abraham . . . . .	17
2.6.2	Le recours à des principes d'équilibre . . . . .	18
<b>3</b>	<b>Les séries chronologiques des comptages routiers : descriptions par la méthode de Zeger</b>	<b>19</b>
3.1	La méthode des équations d'estimation généralisées pour données longitudinales . . . . .	19
3.2	Un modèle "Poissonien" de description des comptages . . . . .	22
3.2.1	Le modèle . . . . .	22
3.2.2	Les moments issus du modèle . . . . .	24
3.2.3	Les équations d'estimation . . . . .	24
3.2.4	Problèmes de calcul . . . . .	25
3.2.5	Estimation des paramètres de nuisance . . . . .	26
3.2.6	Biais des estimateurs de Zeger . . . . .	26

<b>4</b>	<b>Les séries chronologiques des comptages routiers : descriptions qui utilisent le maximum de vraisemblance</b>	<b>28</b>
4.1	Calcul en supposant qu'il n'y a pas de dépendance temporelle . . .	28
4.2	Calcul qui utilise une approximation de la vraisemblance . . . . .	28
4.2.1	Mise en oeuvre . . . . .	30
4.2.2	Problèmes de calcul . . . . .	38
4.3	Le modèle de Ledolter . . . . .	38
4.3.1	Les algorithmes EM et MCEM . . . . .	39
4.3.2	Le modèle de Ledolter proprement dit . . . . .	39
4.3.3	La fonction $Q_m$ . . . . .	40
4.3.4	Optimisation de la fonction $Q_m$ . . . . .	42
4.3.5	Simulation des données "complètes" sur la base des observations . . . . .	42
<b>5</b>	<b>Les méthodes qui utilisent le filtre de Kalman : le modèle de JLCSS</b>	<b>45</b>
5.1	Le modèle . . . . .	45
5.2	La structure des moments du modèle . . . . .	47
5.2.1	Moments du processus latent . . . . .	47
5.2.2	Les innovations du processus latent . . . . .	48
5.2.3	Moments du processus des comptages . . . . .	48
5.2.4	Les innovations du processus des comptages . . . . .	49
5.3	Le filtre de Kalman pour le modèle de JLCSS . . . . .	50
5.4	Le lissage de Kalman pour le modèle de JLCSS . . . . .	51
5.5	Estimation des paramètres . . . . .	52
5.5.1	La fonction de vraisemblance . . . . .	52
5.5.2	Les équations d'estimation . . . . .	53
5.5.3	L'algorithme EM (dit) de Kalman . . . . .	55
5.5.4	Estimation des paramètres de dispersion . . . . .	56
5.5.5	Un algorithme de type Newton . . . . .	57
5.6	Diagnostics . . . . .	62
5.6.1	Définition et propriétés des résidus . . . . .	62
5.6.2	Utilisation des résidus . . . . .	65
<b>6</b>	<b>Résultat des analyses par la méthode de Zeger</b>	<b>68</b>
6.1	Le modèle : description globale basée sur l'ensemble des données	68
6.2	Mise en oeuvre de la méthode de Zeger . . . . .	69
6.2.1	Valeurs initiales . . . . .	69
6.2.2	Approximation du processus stationnaire sous-jacent par un processus autorégressif . . . . .	70
6.2.3	La récursion . . . . .	73
6.2.4	Validité de la méthode . . . . .	75
6.2.5	Estimation de la matrice de covariance des paramètres . . . . .	76
6.3	Résultats . . . . .	78

<b>7</b>	<b>Annexe I : la méthode de Zeger</b>	<b>81</b>
7.1	Calculs . . . . .	81
7.1.1	Les moments issus du modèle . . . . .	81
7.1.2	Les equations d'estimation . . . . .	82
7.1.3	Problèmes de calcul : justifications de l'approximation . . . . .	83
7.2	Code MATLAB et graphiques . . . . .	88
7.2.1	Code . . . . .	88
7.2.2	Graphiques . . . . .	101
<b>8</b>	<b>Annexe II : la méthode du maximum de vraisemblance</b>	<b>155</b>
8.1	Modèle de Davis <i>et al.</i> . . . . .	155
8.1.1	L'algorithme des innovations appliqué aux processus $AR(p)$ . . . . .	155
8.1.2	Formules pour l'inverse de la matrice d'autocovariance d'un processus autorégressif . . . . .	157
8.1.3	L'approximation "standard" de la matrice des autocova- riances d'un processus stationnaire . . . . .	160
8.1.4	Illustration de la méthode du balayage . . . . .	161
8.1.5	Code MATLAB . . . . .	162
8.2	Le modèle de Ledolter . . . . .	168
8.2.1	Loi jointe des observations et du processus latent . . . . .	168
8.2.2	Les lois normales de l'échantillonneur de Gibbs . . . . .	169
<b>9</b>	<b>Annexe III : la méthode du filtre de Kalman</b>	<b>172</b>
9.1	Calcul des moments du modèle . . . . .	172
9.1.1	Espérance du processus latent . . . . .	173
9.1.2	Variance du processus latent . . . . .	173
9.1.3	Covariance et corrélation du processus latent . . . . .	174
9.1.4	Espérance et variance des comptages . . . . .	175
9.1.5	Corrélation du processus de comptage . . . . .	176
9.1.6	Innovations . . . . .	178
9.2	Les mathématiques du filtre de Kalman . . . . .	179
9.2.1	Les équations d'estimation . . . . .	182
9.2.2	Une formule de dérivation . . . . .	186

# Chapitre 1

## Motifs, contenu et résultats

### 1.1 Motifs et contenu

Il y a deux raisons fondamentales qui peuvent justifier, aux yeux d'un office de statistique, le recours à une modélisation des transports. La première est de fournir des informations plus détaillées relatives aux déplacements que celles que permettent le recensement, le microrecensement sur les transports et les comptages routiers, aussi extensifs que soient ceux-ci.<sup>1</sup> La seconde est de permettre l'évaluation de politiques publiques relatives aux transports, évaluation qui requiert le plus souvent des séries chronologiques tout aussi détaillées.

Dans ce qui suit on s'intéresse à la modélisation des comptages routiers. Il y a à cela trois raisons. Tout d'abord, la modélisation des comptages fait partie des modèles qu'il faut produire. En effet, la manière la plus directe de valider le résultat d'une modélisation est celle qui consiste à comparer les nombres produits par le modèle à ceux que l'on peut obtenir par observation directe, soit en particulier les comptages routiers. Les deux ensembles de nombres ainsi constitués ne seront que rarement très proches en valeur absolue, et même dans ce cas, cela ne saurait être que la manifestation de coïncidences numériques sans signification. On ne peut en fait assurer qu'une égalité statistique, soit celle qui tient compte de la variation. Pour cela il faut aussi avoir une description probabiliste des comptages routiers. Il y a ensuite le fait qu'une modélisation des transports qui satisfasse les exigences d'un office public de statistique est une entreprise de longue haleine au cours de laquelle de nombreuses opérations mathématiques, statistiques et informatiques doivent être menées à terme avec succès. Le domaine des comptages permet un apprentissage indispensable en la matière, domaine qui a de plus le mérite de limiter la dimension des problèmes rencontrés à un cadre compatible avec les moyens, modestes, dont on dispose. Finalement, les séries de comptages routiers présentent fréquemment des "trous"

---

1. On peut voir une telle opération comme un problème général de valeurs manquantes, soit une tâche courante dans une office de la statistique officielle!

(pannes des appareils ayant diverses causes : intempéries, dégats dus à des travaux, fermetures temporaires de routes . . . ) que les praticiens aimeraient voir comblés, et les modèles permettent de produire des estimations des valeurs qui manquent.

Ce document est structuré comme suit. Un premier chapitre contient une description rapide de ce qu'est la modélisation des transports dans sa pratique actuelle. Il a pour but de situer les problèmes et d'en souligner l'ampleur. Dans les trois chapitres qui suivent on présente trois manières, une par chapitre, de modéliser les comptages routiers. Les raisons de ce recours à de multiples modèles sont les suivantes. Les comptages routiers forment des séries chronologiques de dénombrements dont les causes qui en expliquent les valeurs sont, et cela revient au même en pratique, soit inconnues, soit trop complexes pour pouvoir être déduites analytiquement de "principes premiers." On ne peut donc recourir, pour la description, qu'à des modèles purement statistiques, et, lors d'un ajustement, il sera toujours difficile de distinguer ce qui est attribuable au modèle de ce qui est accident numérique. La comparaison des résultats issus de modèles différents sert alors de garde-fou. Comme il y a 300 stations de comptage, il n'y a pas de raison non plus de croire que le même modèle soit pertinent pour les 300 stations, ou s'il l'est, il faut pouvoir le démontrer! Par ailleurs, les nombres dont on dispose sont très abondants (il y a environ 300 stations de comptage qui produisent des comptages par heure, sur plusieurs années - une dizaine - ce qui conduit à 300 séries d'environ  $10 \times 365 \times 24$  dénombrements, ce qui fait approximativement 26'000'000 d'observations). Les méthodes disponibles n'ayant pas été testées sur des *corpus* de cette taille, il faut prévoir des solutions de rechange pour le cas où les problèmes numériques deviendraient insurmontables. Dans une troisième partie on présente le résultat d'un certain nombre d'analyses. Toute une série de formules utiles pour le calcul ou la compréhension de la matière, mais qui risquent de détourner l'attention des points essentiels, sont reportées aux annexes.

Les méthodes retenues l'ont été pour les raisons suivantes. Une première méthode, dite de Zeger, a été retenue parce que c'est l'une des premières à avoir vu le jour et parce que la démarche qui la sous-tend est conceptuellement transparente et illustre bien la démarche suivie dans d'autres approches du même problème. Une deuxième méthode, dite du maximum de vraisemblance, a été choisie parce qu'elle permet d'évaluer et corriger les biais inhérents à la méthode de Zeger tout en offrant une alternative au calcul de la covariance des paramètres estimés, calcul qui dans toutes les approches semblables pose des problèmes relatifs à la taille des jeux de données. Il est donc utile d'avoir une idée des méthodes "résistantes à la dimension."<sup>2</sup> Les méthodes de Zeger et du maximum de vraisemblance permettent la description de séries univariées.<sup>3</sup> Comme il y a 300

---

2. *Scalable*, en anglais.

3. En principe il est facile, à l'aide d'indicatrices, de ramener à ce cas l'étude des séries multivariées. Cela a cependant pour conséquence de multiplier la longueur des séries univariées par le nombre de séries individuelles des séries multivariées : les problèmes de calcul,

stations de comptage dont les comptages soit vraisemblablement corrélés, ne serait-ce que “localement,” il peut s’avérer utile, en termes de gains de précision et de pouvoir explicatif, de pouvoir traiter de front tout ou partie des 300 séries. Il semble *a priori* que seules des méthodes de type “filtre de Kalman” permettent ce genre d’exercice et c’est pourquoi la troisième méthode retenue en est une qui applique le filtre de Kalman à des séries de comptages. Parmi les méthodes de Kalman possibles on en a retenu une qui, grâce à des hypothèses très spécifiques concernant les répartitions de certaines grandeurs aléatoires, s’accompagne d’une palette assez riche de procédures de diagnostic, ce qui n’est ni le cas de la méthode de Zeger, ni de celle du maximum de vraisemblance. La méthode retenue présente en sus l’avantage qu’elle n’exige pas que soient données explicitement, ou encore soient estimées, les corrélations entre états latents du système, ce qui représente tout de même, pour 300 stations de comptage, 45’150 nombres !

Ces méthodes sont purement paramétriques dans le sens qu’elles recourent à une modélisation des espérances. Dans la “littérature” des séries chronologiques, on utilise les termes de “modèles marginaux.”<sup>4</sup> On pourrait recourir à des modèles dits “de transition,”<sup>5</sup> mais ils sont plus complexes que les “modèles marginaux,” et les modèles de type “filtre de Kalman” relèvent déjà de cette logique.

Il faut finalement remarquer que les analyses dont on présente les résultats ne tiennent pas compte du contexte complexe dans lequel les comptages ont été réalisés, et cela pour deux raisons. La première est que l’on s’intéresse tout d’abord aux conditions et à la “faisabilité technique” des ajustements,<sup>6</sup> et la seconde que ce contexte n’est pas disponible dans une forme directement utilisable.

## 1.2 Résultats

Les résultats de cette étude sont temporaires et partiels : on n’a de loin pas terminé le travail qu’exige la maîtrise opérationnelle des modèles invoqués. De plus, pour en évaluer l’utilité, il faut avoir recours à des professionnels des transports, et ce recours n’a de sens que lorsque les problèmes de technique statistique sont dominés.

On ne discute ici que des aspects les plus généraux de la procédure suivie, et des expériences faites, les détails concrets étant évoqués dans les parties du document où sont présentés les résultats des analyses spécifiques (4.2.2 et 6.3). Par ailleurs l’inventaire qui est fait est pour l’essentiel celui de l’expérience

---

importants déjà lors du traitement d’une série scalaire individuelle, en sont d’autant aggravés.

4. P.J. Diggle, K.-Y. Liang and S.L. Zeger, *Analysis of Longitudinal Data*, Clarendon Press, Oxford UK (1994), Chapitre 8.

5. *Op. Cit.*, Chapitre 10.

6. Mais on pourrait argumenter que le contexte facilite l’ajustement, plutôt qu’il ne l’entrave !

acquise lors de l'étude des modèles retenus et des analyses qui ont été menées à terme : il ne s'agit donc pas de ce que l'on pense faire, mais de ce qui a été fait.

### **Que peut-on attendre de l'ajustement des modèles retenus ?**

En principe, et dans la mesure où l'on sait obtenir les formules adéquates, l'utilité pratique et concrète des modèles est double. Ils permettent tout d'abord d'estimer les valeurs manquantes, et en particulier de procéder à des extrapolations. Le recours à l'estimation de paramètres permet ensuite de résumer et de rendre "lisibles" les caractéristiques principales des comptages, caractéristiques qui sont masquées par l'abondance et la variabilité de l'appareil numérique. Il permet également de comparer les comptages opérés par les diverses stations. On peut enfin, par ce moyen, déceler les évolutions.

### **Les problèmes de calcul**

Sauf si l'on peut ou veut raccourcir significativement les séries chronologiques avec lesquelles on travaille, l'exploitation systématique des séries de comptages demande que soit pensé et réalisé un environnement de calcul approprié, et que soient comparés systématiquement techniques statistiques et algorithmes de calcul. Ces aspects ne sont en effet que rarement et alors seulement partiellement abordés par les divers auteurs.

Les méthodes fondées sur les fonctions de vraisemblance posent des problèmes de durée de calcul insurmontables dans des environnements qui ne sont pas spécialisés et peu puissants. La raison semble en être le nombre élevé de paramètres.

L'évaluation des matrices de covariance de grande taille n'a pas, pour l'instant, trouvé de solution satisfaisante.

### **Les méthodes**

On a partiellement examiné deux des trois méthodes proposées : seule celle de Zeger a permis d'obtenir des résultats utilisables. On a aussi mis en évidence (6.2.1) que le "common wisdom" en la matière n'est pas toujours la sagesse!

### **Les problèmes liés à la production systématique des descriptions**

L'utilisation, dans une perspective de production statistique officielle, des méthodes décrites dans ce document exigera un environnement informatique bien pensé et configuré. Il faudra apporter un soin extrême à la calibration des méthodes car leur usage doit être "stable."

## Chapitre 2

# La méthodologie de modélisation du domaine des transports

### 2.1 Introduction

La modélisation des transports a une longue tradition chez les ingénieurs de cette profession et c'est cette tradition que l'on va commencer par brièvement décrire puisqu'elle sert de référence à toute modélisation de ce type. Ce qui suit est tiré de :

M. Ben-Akiva and S.R. Lerman, *Discrete Choice Analysis*, MIT Press, Cambridge, MASS (1985)

E. Quinet, *Principes d'Economie des Transports*, ECONOMICA, Paris (1998)

On trouve un survol de tout le domaine dans :

E. Cascetta, *Transportation Systems Engineering : Theory and Methods*, Kluwer, Dordrecht (2001)

J. de D. Ortúzar and L.G. Willumsen, *Modelling Transport* (2nd ed.), Wiley, Chichester (1994)

L'objectif premier<sup>1</sup> est celui d'attribuer aux voies d'un réseau de transports les flux de trafic qu'elles "portent." Cette attribution se fait à partir de modèles et

---

1. Le deuxième objectif est celui de comprendre les déterminants du trafic : seuls des modèles permettent la poursuite d'un tel objectif.

le modèle “standard” est un modèle en quatre étapes :

### **Etape 1 : génération**

La génération sert à représenter le nombre de déplacements effectués à partir d'un centre d'émission. Ces centres sont appelés *origines* et notés  $O_i$ ,  $1 \leq i \leq I$ .

### **Etape 2 : distribution**

La distribution sert à représenter la manière dont le trafic généré se répartit entre les destinations possibles. Les destinations sont notées  $D_j$ ,  $1 \leq j \leq J$ .

### **Etape 3 : choix modal**

Le nombre de trajets d'origine  $O_i$  et des destination  $D_j$  est noté  $T_{i,j}$ . Le choix modal décrit comment ces  $T_{i,j}$  trajets se répartissent entre les divers modes de transport disponibles. On écrira  $T_{i,j}^{(k)}$  pour le nombre de trajets d'origine  $O_i$  et de destination  $D_j$  qui se font à l'aide du mode de transport  $k$ ,  $1 \leq k \leq K$ .

### **Etape 4 : choix d'itinéraires**

Cette étape concerne avant tout le mode routier. Elle décrit la manière dont les usagers se répartissent entre les divers itinéraires possibles permettant de joindre une origine à une destination.

*Remarques :*

1. D'autres étapes peuvent compléter le schéma que l'on vient de décrire : on peut vouloir introduire le choix de l'heure de départ, et, en zone urbaine, la recherche d'une place de stationnement. Par ailleurs, les phénomènes que décrivent ces étapes ne sont pas indépendants et on peut être amené à représenter aussi ces liens.
2. Bien que le schéma que l'on vient de décrire soit “universel,” la variété des modèles qui existent ainsi que les possibilités d'assemblage des ces mêmes modèles d'une part et les données dont on dispose d'autre part font qu'il n'y a pas de *best practice* et que chaque situation requiert sa propre modélisation qui doit être améliorée et évoluer en permanence.
3. Certains modèles regroupent plusieurs de ces étapes.
4. Ce qui suit a pour but de donner une idée des modèles qui sont utilisés et des techniques qu'ils requièrent.

La justification du choix de certains de ces modèles se fait souvent sur la base d'une théorie du consommateur et des choix dits discrets. Cette théorie permet en particulier d'unifier les étapes de choix modal et de distribution, ce qui est,

d'un point de vue logique, plus cohérent.<sup>2</sup> C'est pourquoi on commencera par une brève description de cette dernière.

## 2.2 La théorie du consommateur et des choix discrets

### 2.2.1 Description du comportement individuel

Le consommateur est supposé pouvoir exprimer des préférences cohérentes et transitives, ce qui le dote d'une fonction d'utilité qui lui sert à opérer des choix. Les choix possibles sont notés  $\mathcal{A}$  et les choix possibles du consommateur numéro  $i$ ,  $\mathcal{A}_i$ . Les divers choix s'établissent à partir d'attributs des objets sur lesquels porte le choix ainsi que des particularités du consommateur. Comme de plus il a été maintes fois constaté que les préférences réelles des consommateurs ne sont ni cohérentes ni transitives, les utilités sont définies comme des variables aléatoires qui ont la forme générique suivante :<sup>3</sup>

$$U_j^{(i)} = \tilde{U} \left[ \underline{A}_j^{(i)}, \underline{C}_i \right] + N \left[ \underline{A}_j^{(i)}, \underline{C}_i \right] = V_j^{(i)} + E_j^{(i)},$$

où

$\underline{A}_j^{(i)}$	:=	attributs du choix $j$ pour le consommateur $i$ ,
$\underline{C}_i$	:=	caractéristiques du consommateur $i$ ,
$\tilde{U}$	:=	composante observable (systématique, représentative) de l'utilité,
$N$	:=	composante inobservable (bruit) de l'utilité (variable aléatoire).

Alors, la probabilité que l'individu  $i$  choisisse l'alternative  $j$  est donnée par l'expression suivante :

$$P(j | i) = P \left( V_j^{(i)} + E_j^{(i)} \geq V_k^{(i)} + E_k^{(i)}, k \in \mathcal{A}_i \right).$$

#### Exemple : les modèles de choix binaire

On a  $\mathcal{A}_i = \{j, k\}$ ,  $V_j^{(i)} = V \left( \underline{x}_j^{(i)} \right)$  où  $\underline{x}_j^{(i)}$  incorpore les attributs des alternatives aussi bien que les caractéristiques du consommateur, et, typiquement,

$$V \left( \underline{x}_j^{(i)} \right) = \langle \underline{x}_j^{(i)}, \underline{\theta} \rangle_{R^m}.$$

2. D. McFadden, Disaggregate Behavioural Travel Demand's RUM Side, A 30-Year Retrospective, Department of Economics, University of California, Berkeley (2000).

3.  $U_j^{(i)}$  est l'utilité qu'a, pour le consommateur numéro  $i$ , le choix numéro  $j$ .

La spécification du modèle va donc dépendre des lois de probabilité régissant le comportement de la composante inobservable de l'utilité.

**Loi uniforme**

On suppose que  $E_j^{(i)} - E_k^{(i)}$  suit une loi uniforme  $U[-S_i, +S_i]$ . Alors

$$P(j | i) = P\left(V_j^{(i)} - V_k^{(i)} \geq E_j^{(i)} - E_k^{(i)}\right)$$

$$= \begin{cases} 0 & \text{si } V_j^{(i)} - V_k^{(i)} < -S_i \\ \frac{V_j^{(i)} - V_k^{(i)} + S_i}{2S_i} & \text{si } -S_i \leq V_j^{(i)} - V_k^{(i)} \leq S_i \\ 1 & \text{si } V_j^{(i)} - V_k^{(i)} > S_i \end{cases}$$

La probabilité est donc linéaire en  $V_j^{(i)} - V_k^{(i)}$  entre  $-S_i$  et  $+S_i$ .

**Loi normale**

On suppose que  $E_j^{(i)} - E_k^{(i)}$  suit une loi normale  $\mathcal{N}(0, \sigma_i^2)$ . Alors, si  $\Phi$  désigne la fonction de répartition d'une loi  $\mathcal{N}(0, 1)$ ,

$$P(j | i) = \Phi\left(\frac{V_j^{(i)} - V_k^{(i)}}{\sigma_i}\right).$$

**Loi logistique**

On suppose que  $E_j^{(i)} - E_k^{(i)}$  suit une loi logistique  $\mathcal{L}(\mu_i)$  de densité

$$f_{\mathcal{L}(\mu_i)}(x) = \frac{\mu_i e^{-\mu_i x}}{(1 + e^{-\mu_i x})^2},$$

et de fonction de répartition

$$F_{\mathcal{L}(\mu_i)}(\alpha) = \frac{1}{1 + e^{-\mu_i \alpha}}.$$

Alors

$$P(j | i) = \frac{e^{\mu_i V_j^{(i)}}}{e^{\mu_i V_j^{(i)}} + e^{\mu_i V_k^{(i)}}}.$$

## 2.2.2 Description du comportement collectif à partir des comportements individuels

Soit  $I$  le nombre d'individus de la population dont on veut agréger les comportements individuels. Alors on dénote

$$P(j | i) = P(j | \underline{x}_i)$$

la probabilité que l'individu  $i$  choisisse l'alternative  $j$ . Le vecteur  $\underline{x}_i$  représente les valeurs des attributs influençant le choix de l'individu  $i$ . On ne spécifie pas la fonction d'utilité déterminante. Soit  $\nu_j$  le nombre escompté d'individus choisissant l'alternative  $j$ . On a :

$$\nu_j = \sum_{i=1}^I P(j | \underline{x}_i).$$

On suppose que la taille de la population,  $|I|$ , est suffisamment grande pour que  $\nu_j$  soit une bonne évaluation de la vraie valeur dans la population. En fait, comme  $\nu_j$  est l'espérance d'une somme de variables de Bernoulli indépendantes, sa variance est inférieure à  $\frac{1}{4|I|}$ . Par ailleurs, si  $\pi_j$  dénote la fraction de la population qui choisit l'alternative  $j$ , on a :

$$\pi_j = \frac{\nu_j}{|I|}.$$

Comme en pratique on ne connaît pas les valeurs du vecteur  $\underline{x}_i$ , on fait l'hypothèse que ces valeurs suivent une loi de probabilité, ce qui donne, si  $\underline{x}$  est une valeur du vecteur aléatoire  $\underline{X}$ , de densité  $f_{\underline{X}}$ ,

$$\pi_j = \int_{\mathcal{X}} P(j | \underline{x}) f_{\underline{X}}(\underline{x}) d\underline{x}$$

Mais  $f_{\underline{X}}$  étant habituellement aussi inconnue, les méthodes d'agrégation utilisées s'efforcent de fournir de bonnes approximations de  $\nu_j$  et  $\pi_j$ . La plupart des études empiriques utilisent l'une de deux méthodes dites de segmentation<sup>4</sup> et d'échantillonnage.

### La segmentation

On suppose que la population puisse être segmentée en  $p$  parties  $I_1, \dots, I_p$ , et que l'on soit capable de définir, pour chaque partie, un vecteur  $\hat{\underline{x}}_k$  de valeurs des attributs qui soit "représentatif" des valeurs des attributs des individus de la partie. On pose alors

$$\pi_j \approx \sum_{k=1}^p \frac{|I_k|}{|I|} P(j | \hat{\underline{x}}_k).$$

---

4. Il s'agit en fait de stratification : on utilise le terme "segmentation" parce que dans le langage habituel de la statistique, la stratification est une forme d'échantillonnage.

Cette approche est avant tout utilisée quand on doit obtenir des chiffres qui sont spatialement désagrégés.

### L'échantillonnage

L'échantillonnage remplace les valeurs de la population par celles d'un échantillon (qui peut être stratifié). Cependant il ne remplace pas la segmentation, car, quand le nombre de zones est grand (100 par exemple), la taille nécessaire des échantillons devient prohibitive.

## 2.3 Les modèles de génération

### 2.3.1 Modèles agrégés

On y représente le *nombre moyen* de déplacements d'une population prise dans son ensemble. Ce nombre s'exprime en fonction de variables qui caractérisent la population concernée : niveau d'activité ou de développement économique, démographie ...

*Exemple :*

$$O_i = \gamma \pi_i^\alpha R_i^\beta,$$

où

$O_i$	:=	nombre moyen de déplacements d'origine $O_i$ par unité de temps,
$\pi_i$	:=	taux de possession d'une automobile dans la zone $i$ ,
$R_i$	:=	revenu moyen par habitant dans la zone $i$ ,
$\alpha, \beta, \gamma$	:=	paramètres (à estimer).

### 2.3.2 Modèles désagrégés

La population et ses déplacements sont subdivisés en catégories homogènes à l'intérieur desquelles une certaine stabilité de la génération des déplacements peut être envisagée. Ainsi

- on peut classer les voyageurs selon plusieurs critères :
  - revenu,
  - structure et taille du ménage,
  - âge
  - possession ou non d'une voiture,
  - catégorie socioprofessionnelle,
  - caractéristiques de localisation ;
- on peut subdiviser les déplacements selon leurs *motifs* :
  - domicile  $\longleftrightarrow$  travail,
  - professionnels,

- privés pour affaires,
- loisirs de *week-end*,
- loisirs de grandes vacances.

On donne une représentation de la génération pour chaque catégorie homogène et on agrège sur toutes les catégories.

*Exemple :*

Supposons qu'il y ait  $J$  catégories de revenus,  $K$  catégories de ménages, et  $L$  classes d'âges.  $\pi_{i,j,k,l}(t)$  est la proportion d'utilisateurs de la zone  $i$  présentant les caractéristiques  $j$  de revenu,  $k$  de ménage et  $l$  d'âge, et, *mutatis mutandis*,  $\Gamma_{j,k,l}$ , le taux de génération de déplacements de cette catégorie d'utilisateurs. Alors

$$O_i(t) = \gamma \sum_{1 \leq j \leq J, 1 \leq k \leq K, 1 \leq l \leq L} \pi_{i,j,k,l}(t) \Gamma_{j,k,l}.$$

## 2.4 Les modèles de distribution

La distribution est représentée pour l'essentiel à partir de modèles dits gravitaires car on peut leur donner de nombreuses formes et justifications. La forme la plus primitive, et la plus proche de la loi de la gravitation de la physique, en est la suivante :

$$T_{i,j} = \gamma \frac{(P_i \times P_j)^\alpha}{[\delta(i,j)]^\beta},$$

où

$$\begin{aligned} P_i &:= \text{population de la zone } i, \\ \delta(i,j) &:= \text{distance entre les zones } i \text{ et } j, \\ \alpha, \beta, \gamma &:= \text{paramètres.} \end{aligned}$$

Voici une forme plus générale du modèle gravitaire :

$$T_{i,j} = \gamma O_i D_j f(c_{i,j}),$$

où

$$\begin{aligned} f &:= \text{fonction dite d'impédance (typiquement } f(x) = x^\alpha e^{-\beta x}), \\ c_{i,j} &:= \text{"coût généralisé" de circulation entre les zones } i \text{ et } j. \end{aligned}$$

Par rapport à la formule précédente,  $O_i$  correspond à  $P_i^\alpha$ ,  $D_j$  à  $P_j^\alpha$  et  $f(c_{i,j})$  à  $[\delta(i,j)]^{-\beta}$ . Ces modèles sont *constraints* ou non selon que l'on impose ou non des contraintes sur les nombres  $T_{i,j}$ ,  $O_i$  et  $D_j$ , typiquement <sup>5</sup>

$$T_{i,\bullet} = O_i, \quad T_{\bullet,j} = D_j.$$

---

5.  $T_{\bullet,j}$  := somme sur la colonne numéro  $j$  du tableau des valeurs  $T_{i,j}$

## 2.4.1 Justification des modèles gravitaires

### ... sur la base du principe d'entropie

Etant donnée une matrice fixe  $[T_{i,j}]$  et  $T_{\bullet,\bullet}$  individus, on cherche à répartir ces individus dans un tableau de taille  $I \times J$  de telle sorte que la case  $(i, j)$  comporte exactement  $T_{i,j}$  individus. Le nombre des répartitions possibles est

$$W([T_{i,j}]) = \frac{T_{\bullet,\bullet}!}{\prod_{1 \leq i \leq I, 1 \leq j \leq J} (T_{i,j}!)}.$$

Une application de la formule de Stirling donne l' "approximation" :

$$\ln \{W([T_{i,j}])\} \approx \ln \{T_{\bullet,\bullet}!\} - \sum_{1 \leq i \leq I, 1 \leq j \leq J} (T_{i,j} \ln \{T_{i,j}\} - T_{i,j}).$$

On cherche alors le maximum de cette fonction (d'entropie) sous toutes ou partie des contraintes suivantes :<sup>6</sup>

$$\begin{aligned} T_{i,\bullet} &= O_i, \\ T_{\bullet,j} &= D_j, \\ \sum_{1 \leq i \leq I, 1 \leq j \leq J} c_{i,j} T_{i,j} &= C. \end{aligned}$$

Le résultat est un modèle gravitaire.

### ... sur la base des choix discrets

L'exemple suivant illustre la démarche dans le cas des déplacements domicile-travail. Il fait intervenir la loi de Gumbel dont les principales propriétés sont les suivantes.<sup>7</sup>  $G \sim \mathcal{G}(\eta, \mu)$  quand

$$P(G \leq \alpha) = e^{-e^{-\mu[\alpha-\eta]}}, \mu > 0.$$

1. Le mode de la répartition est en  $\eta$ .
2.  $E[G] = \eta + \frac{\gamma}{\mu}$ ,  $V[G] = \frac{\pi^2}{6\mu^2}$ .
3. Si  $\alpha > 0$ , alors  $\alpha G + \beta \sim \mathcal{G}(\alpha\eta + \beta, \frac{\mu}{\alpha})$ .
4. Si  $G_1 \sim \mathcal{G}(\eta_1, \mu)$  et  $G_2 \sim \mathcal{G}(\eta_2, \mu)$  sont indépendantes, on a alors
  - (a) pour  $L = G_2 - G_1$ ,  $P(L \leq \alpha) = \frac{1}{1 + e^{\mu(\eta_2 - \eta_1 - \alpha)}}$
  - (b) pour  $M = G_1 \vee G_2$ ,  $M \sim \mathcal{G}(\frac{1}{\mu} \ln \{e^{\mu\eta_1} + e^{\mu\eta_2}\}, \mu)$ .

*Mutatis mutandis*, cette expression reste vraie pour un nombre quelconque, mais fini, de variables indépendantes.

6.  $C$  est le niveau fixé du "coût global" des transports.

7.  $\mathcal{G}$  abrègé "Gumbel."  $\gamma$  est la constante d'Euler :  $\gamma = 0.577\dots$

L'exemple est alors comme suit. Un individu issu de la zone  $i$  veut maximiser l'utilité qu'il retire d'un emploi dans la zone  $j$  : son utilité est faite du salaire dans l'emploi  $k$ , soit  $s_k$ , et du coût de déplacement, soit  $c_{i,j}$ . Les salaires ont une répartition indépendante de la localisation de l'emploi et seul leur niveau contribue à l'utilité :<sup>8</sup>

$$S = s + G, \quad G \sim \mathcal{G}(0, \mu).$$

On suppose qu'il y a  $D_j$  emplois dans la zone  $j$ . L'utilité de chacun de ces emplois est  $s_k - c_{i,j} = s - c_{i,j} + g_k$  qui a une loi  $\mathcal{G}(s - c_{i,j}, \mu)$ . L'utilité maximale associée à la région  $j$  a donc une loi (maximum de  $D_j$  variables aléatoires de Gumbel indépendantes et identiquement réparties)

$$\mathcal{G}\left(\frac{1}{\mu} \ln \left\{ D_j e^{\mu[s - c_{i,j}]} \right\}, \mu\right) := \mathcal{G}(\eta_j, \mu).$$

Par suite  $\eta_j = s - c_{i,j} + \frac{\ln\{D_j\}}{\mu}$  et la probabilité que la destination  $j$  soit choisie est

$$p_j = \frac{e^{\mu\eta_j}}{\sum_{k=1}^J e^{\mu\eta_k}} = \frac{D_j e^{\mu[s - c_{i,j}]}}{\sum_{k=1}^J D_k e^{\mu[s - c_{i,k}]}} = \frac{D_j e^{-\mu c_{i,j}}}{\sum_{k=1}^J D_k e^{-\mu c_{i,k}}}.$$

Pour le trafic total on obtient bien

$$T_{i,j} = O_i p_j = O_i \frac{D_j e^{-\mu c_{i,j}}}{\sum_{k=1}^J D_k e^{-\mu c_{i,k}}}.$$

## 2.5 Les modèles de partage modal

La description du partage modal part de l'idée que ce partage dépend soit de la différence, soit encore du rapport, des "coûts des transports." Typiquement, si le choix est entre deux modes, on postule que la probabilité  $p_1$  du choix du premier mode est donnée par une formule du type

$$p_1 = \frac{1}{1 + e^{\mu[c_1 - c_2]}}, \quad \mu > 0,$$

où  $c_1$  et  $c_2$  sont les coûts respectifs des deux modes. Selon la valeur de  $\mu$ , la probabilité  $p_1$  est plus ou moins fortement influencée par l'importance de  $|c_1 - c_2|$ . Une telle formule peut être obtenue à partir d'un modèle de choix discret.

---

8.  $s_k$  est une "observation" de  $S$ .

## 2.6 Le choix d'itinéraire

Les différentes approches distinguent d'une part entre approches pour débits faibles (coûts du trafic peu influencés par son niveau) et approches pour débits élevés, et, d'autre part, entre approches déterministes et approches stochastiques.

### 2.6.1 La formule d'Abraham

Elle concerne le trafic interurbain à débit faible. Elle postule que le coût généralisé sur l'itinéraire numéro  $l$ ,  $\Gamma_l$ , s'exprime à l'aide de la formule suivante :

$$\Gamma_l = M_l + F_D D_l + F_C C_l,$$

où

$M_l$	:=	coût monétaire du trajet,
$D_l$	:=	durée du trajet,
$C_l$	:=	évaluation quantitative du confort du trajet,
$F_D, F_C$	:=	valorisation monétaire de la durée et du confort.

Il est postulé de plus que

$$\begin{aligned} N &\sim \mathcal{N}(0, \sigma_N^2), \\ M_l &= (1 + N) M_l^0, \\ D_l &= (1 + N) D_l^0, \\ C_l &= (1 + N) C_l^0, \\ \ln[L] &\sim \mathcal{N}(\mu_L, \sigma_L^2), \\ F_D &= L^\alpha F_D^0, \\ F_C &= L^\alpha F_C^0, \\ \alpha &= \text{paramètre.} \end{aligned}$$

Alors, la proportion d'utilisateurs choisissant l'itinéraire 1 par rapport à l'itinéraire 2 est égale à la probabilité que le coût généralisé sur l'itinéraire 1, soit  $C_1$ , soit inférieur à celui sur l'itinéraire 2, soit  $C_2$ . On calcule que, quand  $L$  est constante, et  $T_l$  désigne le trafic sur l'itinéraire  $l$ ,

$$\frac{T_1}{T_2} \approx \left( \frac{\Gamma_1}{\Gamma_2} \right)^{-f(\sigma_N)}.$$

Cette formule peut également s'obtenir à partir d'un modèle de choix discret dans lequel

$$\Gamma_1 = \Gamma_1^0 + U_1, \quad \Gamma_2 = \Gamma_2^0 + U_2, \quad U_1, U_2 \text{ indépendantes et uniformes.}$$

## 2.6.2 Le recours à des principes d'équilibre

Ces principes ont été énoncés par Wardrop.<sup>9</sup> Le premier principe s'énonce comme suit :

*Under equilibrium conditions traffic arranges itself in congested networks in such a way that no individual trip maker can reduce his paths costs by switching routes.*

Supposons qu'il n'y ait qu'un couple  $(O, D)$  et deux arcs (itinéraires)  $a_1$  et  $a_2$ ;  $n_1$  et  $n_2$  sont les flux respectifs sur les arcs  $a_1$  et  $a_2$  et  $n$  est le flux total :  $n_1 + n_2 = n$ . Le coût du transport sur l'arc  $a_i$  est donné par la fonction  $\Gamma(n_i)$ . Soit

$$\gamma(n_1, n_2) = \Gamma(n_1) \wedge \Gamma(n_2).$$

Soient  $n_1^e$  et  $n_2^e$  les valeurs d'équilibre. On doit, selon le principe de Wardrop, avoir :

$$\Gamma(n_i^e) - \gamma(n_1^e, n_2^e) = \begin{cases} c_i = 0 & \text{si } n_i^e > 0 \\ c_i \geq 0 & \text{si } n_i^e = 0 \end{cases},$$

ce qui conduit à la formule

$$\Gamma(n_1^e)[n_1 - n_1^e] + \Gamma(n_2^e)[n_2 - n_2^e] \geq \gamma(n_1^e, n_2^e)([n_1 - n_1^e] + [n_2 - n_2^e]) = 0,$$

qui est la condition (optimum) d'équilibre pour l'utilisateur.

Le deuxième principe d'équilibre veut que les usagers soient conduits à choisir les itinéraires qui minimisent globalement les coûts, soit à réaliser

$$\min_{n_1, n_2} \{n_1 \Gamma(n_1) + n_2 \Gamma(n_2)\}.$$

Les deux principes ne conduisent en général pas aux mêmes solutions !

Il est possible de recourir à des méthodes stochastiques pures pour lesquelles les coûts sont aléatoires, méthodes qui conduisent à des modèles de type *logit* ou *probit*, ou encore à des méthodes mixtes (*SUE* := *stochastic user equilibrium*).

---

9. J.G. Wardrop, Some theoretical aspects of road traffic research, Proceedings of the Institution of Civil Engineers, Part II, 1(1952), 325-362.

## Chapitre 3

# Les séries chronologiques des comptages routiers : descriptions par la méthode de Zeger

### 3.1 La méthode des équations d'estimation généralisées pour données longitudinales

Les références sont :

1. S.L. Zeger and K.-Y. Liang, Longitudinal Data Analysis for Discrete and Continuous Outcomes, *Biometrics* 42 (1986), 121-130.
2. K.-Y. Liang and S.L. Zeger, Longitudinal Data Analysis Using Generalized Linear Models, *Biometrika* 73 (1986), 13-22 .

On suppose que l'on observe  $m$  "sujets" (l'indice est  $i$ ) à des instants qui sont, pour le sujet numéro  $i$ ,  $t_{i,j}$ ,  $1 \leq j \leq n_i$ . Les variables aléatoires qui sont à l'origine des observations relatives au sujet numéro  $i$  sont  $Y_{i,j}$ ,  $1 \leq j \leq n_i$ . A chaque instant on dispose de  $p$  variables "explicatives" ou "exogènes"  $X_{i,j,k}$ ,  $1 \leq k \leq p$ . On "résume" ces données comme suit :

$\underline{t}_i$  := instants d'observation relatifs au sujet numéro  $i$ ,  
 $\underline{Y}_i$  := observations pour le sujet numéro  $i$ ,  
 $X_i$  := matrice  $n_i \times p$  des variables "explicatives" pour le sujet numéro  $i$ .

Soient  $h$  une “fonction de lien” et  $\underline{X}_{i,j}$  le vecteur à  $p$  composantes

$$\underline{X}_{i,j} = \begin{bmatrix} X_{i,j,1} \\ X_{i,j,2} \\ X_{i,j,3} \\ \vdots \\ X_{i,j,p} \end{bmatrix}.$$

Si l'on dénote  $\langle \underline{u}, \underline{v} \rangle_{\mathbb{R}^q} = \sum_{l=1}^q u_l v_l$  le produit scalaire dans  $\mathbb{R}^q$ , et  $\underline{\beta}$  un vecteur de  $p$  paramètres, on donne à  $E[Y_{i,j}] = \mu_{i,j}$  la représentation

$$\mu_{i,j} = h\left(\langle \underline{X}_{i,j}, \underline{\beta} \rangle_{\mathbb{R}^p}\right).$$

Soit  $g$  une fonction connue, à valeurs positives ou nulles : on pose encore  $g_{i,j} = g(\mu_{i,j})$ , et

$$G_i = \begin{bmatrix} g_{i,1} & 0 & \cdots & \cdots & 0 \\ 0 & g_{i,2} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & g_{i,n_i} \end{bmatrix}.$$

$\Gamma_i(\underline{\alpha})$  désigne une matrice qui, sans être nécessairement la matrice de corrélation de  $\underline{Y}_i$ , joue ce rôle,  $\underline{\alpha}$  étant un vecteur de paramètres à estimer.  $\phi$  est une constante. On pose alors que la matrice qui tient lieu de matrice de covariance de  $\underline{Y}_i$  est la matrice

$$\Sigma_i = \frac{1}{\phi} G_i^{\frac{1}{2}} \Gamma_i(\underline{\alpha}) G_i^{\frac{1}{2}}.$$

Le vecteur  $\underline{\mu}_i$  a les valeurs  $\mu_{i,j}$  comme composantes, l'expression  $D_{\underline{\beta}} \left[ \underline{\mu}_i \right]$  représente la matrice  $n_i \times p$

$$\begin{bmatrix} \frac{\partial \mu_{i,1}}{\partial \beta_1} & \cdots & \cdots & \frac{\partial \mu_{i,1}}{\partial \beta_p} \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial \mu_{i,n_i}}{\partial \beta_1} & \cdots & \cdots & \frac{\partial \mu_{i,n_i}}{\partial \beta_p} \end{bmatrix},$$

et  $\underline{Z}_i$  l'écart  $\underline{Y}_i - \underline{\mu}_i$ . L'équation d'estimation s'écrit alors ( $M^t$  désignant la transposée de  $M$ )

$$\sum_{i=1}^m D_{\underline{\beta}} \left[ \underline{\mu}_i \right]^t \Sigma_i^{-1} \underline{Z}_i = \underline{0}.$$

Cette famille d'équations est conçue pour assurer, avec un minimum de contraintes, la consistance des estimateurs des coefficients de régression chaque fois que la

“fonction de lien” est correctement spécifiée. Comme les termes matriciels de l’expression sont indépendants des observations, les racines de l’équation sont des estimateurs consistants dès que  $E[\underline{Z}_i] = \underline{0}$ ,  $1 \leq i \leq m$ .

Les équations d’estimation dépendent en fait de  $\underline{\alpha}$ , de  $\underline{\beta}$  et de  $\phi$ . Mais, si l’on remplace successivement  $\underline{\alpha}$  par un estimateur  $\sqrt{m}$ -consistant, puis  $\phi$  par un estimateur qui est aussi  $\sqrt{m}$ -consistant, alors les équations d’estimation permettent d’estimer  $\underline{\beta}$ . On obtient alors que

1. la solution  $\hat{\underline{\beta}}_\Gamma$  des équations d’estimation est un estimateur consistant de  $\underline{\beta}$ ,<sup>1</sup>
2.  $\sqrt{m}(\hat{\underline{\beta}}_\Gamma - \underline{\beta})$  est une grandeur asymptotiquement normale de covariance  $\Sigma_\Gamma$  qui vaut, avec les notations<sup>2</sup>

$$\Sigma_0 = \sum_{i=1}^m D_{\underline{\beta}}[\underline{\mu}_i]^t \Sigma_i^{-1} \Sigma_{\underline{Y}_i} \Sigma_i^{-1} D_{\underline{\beta}}[\underline{\mu}_i],$$

$$\Sigma_1 = \sum_{i=1}^m D_{\underline{\beta}}[\underline{\mu}_i]^t \Sigma_i^{-1} D_{\underline{\beta}}[\underline{\mu}_i],$$

$$\Sigma_\Gamma = \lim_{m \rightarrow \infty} m (\Sigma_1^{-1} \Sigma_0 \Sigma_1^{-1}).$$

En pratique, pour le calcul de la variance asymptotique, on remplace  $\Sigma_{\underline{Y}_i}$  par  $\underline{Z}_i \underline{Z}_i^t$  et les valeurs des paramètres par leurs estimations.

La procédure d’estimation se fait itérativement et en alternance : partant d’estimations de  $\underline{\alpha}$  et de  $\phi$ , on obtient une estimation de  $\underline{\beta}$  par moindres carrés (re-)pondérés, ce qui permet une nouvelle estimation de  $\underline{\alpha}$  et de  $\phi$ , et ainsi de suite. L’estimation de  $\underline{\alpha}$  et  $\phi$  s’appuie sur les résidus

$$r_{i,j} = \frac{Y_{i,j} - \hat{\mu}_{i,j}}{\sqrt{\hat{\Sigma}_i^{-1}(j,j)}}.$$

Les estimateurs obtenus sont robustes relativement au choix des matrices  $\Gamma_i$ , si bien que les variances sont asymptotiquement “correctes.” Un choix qui est proche des “vraies” valeurs assure cependant une meilleure efficacité.

#### Application aux comptages routiers :

1.  $m$  représente le nombre de stations de comptage ;
2.  $n_{i,j} = n$ , le nombre d’heures pour lesquelles un comptage existe (comme ces comptages sont automatiques, on a des comptages par heure, pendant plusieurs années) ;

---

1. L’indice  $\Gamma$  rappelle que l’on travaille avec une “pseudo-matrice” de corrélation.  
2.  $\Sigma_{\underline{Y}_i}$  dénote la vraie matrice de covariance de  $\underline{Y}_i$ .

3. les variables “explicatives” ou “exogènes” sont les variables temporelles (type d’heure, type de jour, année), les variables qui désignent le type de route (autoroute, nationale . . . ) et le type d’environnement (rural, urbain).

## 3.2 Un modèle “Poissonien” de description des comptages

Les flux de trafic sont fréquemment représentés à l’aide d’une loi de Poisson dont la limitation première tient au fait que l’espérance et la variance sont dans ce cas égaux.<sup>3</sup> Le premier mérite de la méthode présentée ci-dessous tient à ce que l’on travaille avec un modèle “surdispersé.” Les références sont :

1. S.L. Zeger, A Regression Model for Time Series of Counts, *Biometrika* 75 (1988), 621-629.
2. R.A. Davis, Y. Wang and W.T.M. Dunsmuir, Modelling Time Series of Count Data, in *ASYMPTOTICS, NONPARAMETRICS, AND TIME-SERIES*, S. Ghosh, Editor, Marcel Dekker, New York-Basel (1999).
3. R.A. Davis, W.T.M. Dunsmuir and Y. Wang, On Autocorrelation in a Poisson Regression Model, *Biometrika* 87 (2000), 491-505.

L’annexe consacré à la méthode de Zeger contient un certain nombre de calculs qui permettent de comprendre un peu mieux les formules qui apparaissent ci-dessous.

### 3.2.1 Le modèle

Soit  $\underline{B} = \{B_n, n \in \mathbb{N}\}$ , un processus stationnaire à valeurs positives ou nulles tel que

$$E[B_n] = 1, \text{ cov}(B_n, B_{n+p}) = \sigma_B^2 \Gamma_B(p).$$

A chaque instant  $n \in \mathbb{N}$ , on peut observer un vecteur de  $\mathbb{R}^p$ ,

$$\underline{X}_n = \begin{bmatrix} X_1^{(n)} \\ \vdots \\ X_i^{(n)} \\ \vdots \\ X_p^{(n)} \end{bmatrix},$$

de variables “explicatives” ou “exogènes.” On pose

$$\lambda_n = e^{\langle \underline{\theta}, \underline{X}_n \rangle}$$

---

3. F.L. Mannering and W.P. Kilareski, *Principles of Highway Engineering and Traffic Analysis*, 2nd. ed., Wiley, New York (1998), section 5.4 et en particulier 5.4.2.

où  $\underline{\theta}$  est un vecteur de  $p$  paramètres  $\theta_1, \dots, \theta_p$ , qu'il faut estimer. On considère alors une série chronologique de comptages

$$Y_1, \dots, Y_n,$$

tels que, si

$$p_{n,k|\underline{B}} = P(Y_n = k | \underline{B}),$$

1.  $p_{n,k|\underline{B}} = \frac{(\lambda_n B_n)^k}{k!} e^{-\lambda_n B_n},$
2.  $P(Y_1 = k_1, \dots, Y_n = k_n | \underline{B}) = \prod_{l=1}^n p_{l,k_l|\underline{B}}.$

On a donc que

$$P(Y_i = j) = \int_0^\infty e^{-\lambda_i x} \frac{(\lambda_i x)^j}{j!} P_{B_i}(dx),$$

soit que la loi de  $Y_i$  est un mélange, "sur le paramètre," de lois de Poisson.

*Exemple :*

Soit  $0 < \rho < 1$ , et  $\{W_n, -\infty < n < \infty\}$  une suite de variables aléatoires indépendantes et identiquement réparties. On pose alors

$$B_n = \sum_{i=0}^{\infty} \rho^i W_{n-i},$$

ce qui définit un processus stationnaire. Si  $W_n$  a une loi exponentielle de paramètre  $\alpha$ ,<sup>4</sup>  $B_n$  est une variable aléatoire positive. Alors

$$E[B_n] = \frac{E[W_n]}{1-\rho} = \frac{1}{\alpha(1-\rho)}.$$

Donc si  $\alpha = \frac{1}{1-\rho}$ ,  $E[B_n] = 1$ . En particulier, quand on pose  $p = \frac{\lambda_i}{\alpha + \lambda_i}$ , on a que<sup>5</sup>

$$P(Y_i \leq [\lambda_i]) = 1 - p^{[\lambda_i]+1},$$

si bien que l'on observera un plus ou moins grand nombre de "comptages" en deçà ou au delà de  $\lambda_i$  selon les valeurs que prennent  $\rho$  et  $\lambda_i$ .<sup>6</sup>

---

4.  $f_{W_n}(x) = \begin{cases} \alpha e^{-\alpha x} & \text{si } x > 0, \\ 0 & \text{si } x \leq 0, \end{cases}$

5.  $[x]$  représente la valeur entière de  $x$ .

6. Ce genre d'observation peut aider à comprendre la forme certains des graphiques de validation que l'on rencontre plus loin, et par suite de servir de moyen de diagnostic.

### 3.2.2 Les moments issus du modèle

On a :

$$\begin{aligned} E[Y_n] &= \lambda_n \\ V[Y_n] &= \lambda_n + \lambda_n^2 \sigma_B^2 \\ \text{corr}(Y_n, Y_{n+p}) &= \frac{\Gamma_B(p)}{\sqrt{\left(1 + \frac{1}{\sigma_B^2 \lambda_n}\right) \left(1 + \frac{1}{\sigma_B^2 \lambda_{n+p}}\right)}} \end{aligned}$$

### 3.2.3 Les équations d'estimation

Soient  $\underline{\lambda}_n$  le vecteur (colonne) de composantes  $\lambda_1, \dots, \lambda_n$ ,  $\Lambda_n$  la matrice diagonale dont les éléments diagonaux sont les composantes de  $\underline{\lambda}_n$ , et  $X_n$  la matrice  $(p, n)$  définie par l'égalité suivante :  $X_n = [\underline{X}_1 \mid \dots \mid \underline{X}_n]$ . On a alors que :

1. Les  $p$  équations pour l'estimation du paramètre  $\underline{\theta}$  sont données par l'égalité suivante :

$$\left\{ \frac{\partial}{\partial \underline{\theta}} [\underline{\lambda}_n] \right\}^t \Sigma_{\underline{Y}_n}^{-1} (\underline{Y}_n - \underline{\lambda}_n) = \underline{0},$$

qui s'écrit :

$$X_n \Lambda_n \Sigma_{\underline{Y}_n}^{-1} (\underline{Y}_n - \underline{\lambda}_n) = \underline{0}.$$

2. Conditionnellement aux paramètres de nuisance, le paramètre  $\underline{\theta}$  se calcule récursivement à l'aide des formules suivantes :

$$\underline{Z}_n = \left\{ \frac{\partial}{\partial \underline{\theta}} [\underline{\lambda}_n] \right\} \underline{\theta} + (\underline{Y}_n - \underline{\lambda}_n),$$

$$\hat{\underline{\theta}}_{m+1} = \left\{ \left\{ \frac{\partial}{\partial \underline{\theta}} [\underline{\lambda}_n] \right\}^t \Sigma_{\underline{Y}_n}^{-1} \left\{ \frac{\partial}{\partial \underline{\theta}} [\underline{\lambda}_n] \right\} \right\}^{-1} \left\{ \left\{ \frac{\partial}{\partial \underline{\theta}} [\underline{\lambda}_n] \right\}^t \Sigma_{\underline{Y}_n}^{-1} \underline{Z}_n \right\} \left[ \hat{\underline{\theta}}_m \right],$$

où le "terme"  $\left[ \hat{\underline{\theta}}_m \right]$  de cette dernière égalité indique que ce qui précède est évalué en  $\hat{\underline{\theta}}_m$ .

3. La matrice de covariance asymptotique de  $\hat{\underline{\theta}}$  est donnée par l'expression

$$\Sigma_{\hat{\underline{\theta}}} = \lim_n \left( n \left\{ \left\{ \frac{\partial}{\partial \underline{\theta}} [\underline{\lambda}_n] \right\}^t \Sigma_{\underline{Y}_n}^{-1} \left\{ \frac{\partial}{\partial \underline{\theta}} [\underline{\lambda}_n] \right\} \right\}^{-1} \right).$$

### 3.2.4 Problèmes de calcul

Des comptages horaires sur 10 ans produisent quelque  $10 \times 365 \times 24 = 87600$  observations et il est évident que des matrices carrées de telles dimensions peuvent causer de difficiles problèmes de calcul. Zeger a prévu, pour les cas où l'on rencontre ces difficultés, les modifications suivantes de sa méthode.

Soit  $D_n$  la matrice diagonale dont les éléments diagonaux sont  $\lambda_i + \sigma_B^2 \lambda_i^2$ ,  $1 \leq i \leq n$ . On pose alors

$$\Sigma_{\underline{Y}_n} \approx \tilde{\Sigma}_{\underline{Y}_n} \equiv D_n^{\frac{1}{2}} \Sigma_p^{(n)} D_n^{\frac{1}{2}},$$

où  $\Sigma_p^{(n)}$  est la matrice d'autocorrélation d'un processus autorégressif d'ordre  $p$ . On a que

$$D_n^{\frac{1}{2}} = \Lambda_n^{\frac{1}{2}} (I_n + \sigma_B^2 \Lambda_n)^{\frac{1}{2}},$$

si bien que l'approximation choisie remplace

$$I_n + \sigma_B^2 \Lambda_n^{\frac{1}{2}} \Sigma_p^{(n)} \Lambda_n^{\frac{1}{2}} \quad \text{par} \quad (I_n + \sigma_B^2 \Lambda_n)^{\frac{1}{2}} \Sigma_p^{(n)} (I_n + \sigma_B^2 \Lambda_n)^{\frac{1}{2}}.$$

Si  $L_{n,p}$  désigne la matrice suivante, de dimensions  $(n-p, n)$  :

$$L_{n,p} = \begin{bmatrix} 1 & -\alpha_1 & \cdots & -\alpha_p & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 1 & -\alpha_1 & \cdots & -\alpha_p \end{bmatrix},$$

alors

$$\left\{ \Sigma_p^{(n)} \right\}^{-1} \approx \kappa_p L_{n,p}^t L_{n,p}.$$

où  $\kappa_p$  est une constante qui dépend de l'ordre du processus autorégressif et de ses paramètres. Quand  $p = 1$ ,  $\kappa_1 = (1 - \alpha^2)^{-1}$ . Donc

$$\Sigma_{\underline{Y}_n}^{-1} \approx \tilde{\Sigma}_{\underline{Y}_n}^{-1} \approx \kappa_p D_n^{-\frac{1}{2}} L_{n,p}^t L_{n,p} D_n^{-\frac{1}{2}}.$$

La formule de récursion pour l'estimation de  $\underline{\theta}$  devient alors, les  $\kappa_p$  se simplifiant :

$$\begin{aligned} \tilde{\underline{\theta}}_{m+1} &= \left\{ \left( L_{n,p} D_n^{-\frac{1}{2}} \frac{\partial}{\partial \underline{\theta}} [\underline{\lambda}_n] \right)^t \left( L_{n,p} D_n^{-\frac{1}{2}} \frac{\partial}{\partial \underline{\theta}} [\underline{\lambda}_n] \right) \right\}^{-1} \\ &\times \left( L_{n,p} D_n^{-\frac{1}{2}} \frac{\partial}{\partial \underline{\theta}} [\underline{\lambda}_n] \right)^t \left( L_{n,p} D_n^{-\frac{1}{2}} \underline{Z}_n \right) [\tilde{\underline{\theta}}_m]. \end{aligned}$$

La matrice de covariance asymptotique de  $\hat{\underline{\theta}}$  est alors donnée par l'expression

$$\Sigma_{\hat{\underline{\theta}}} \approx \Sigma_{\tilde{\underline{\theta}}} = \Sigma_0^{-1} \Sigma_1 \Sigma_0^{-1}$$

où

$$\Sigma_0 = \lim_n \left( n \left\{ \left\{ \frac{\partial}{\partial \underline{\theta}} [\underline{\lambda}_n] \right\}^t \tilde{\Sigma}_{\underline{Y}_n}^{-1} \left\{ \frac{\partial}{\partial \underline{\theta}} [\underline{\lambda}_n] \right\} \right\}^{-1} \right),$$

et

$$\Sigma_1 = \lim_n \left( n \left\{ \left\{ \frac{\partial}{\partial \underline{\theta}} [\underline{\lambda}_n] \right\}^t \tilde{\Sigma}_{\underline{Y}_n}^{-1} \Sigma_{\underline{Y}_n} \tilde{\Sigma}_{\underline{Y}_n}^{-1} \left\{ \frac{\partial}{\partial \underline{\theta}} [\underline{\lambda}_n] \right\} \right\}^{-1} \right).$$

### 3.2.5 Estimation des paramètres de nuisance

Les paramètres de nuisance sont estimés comme suit :

$$\hat{\sigma}_B^2 = \frac{\sum_{i=1}^n (Y_i - \hat{\lambda}_i)^2 - \hat{\lambda}_i}{\sum_{i=1}^n \hat{\lambda}_i^2},$$

$$\hat{\Gamma}_B(m) = \frac{1}{\hat{\sigma}_B^2} \frac{\sum_{i=m+1}^n (Y_i - \hat{\lambda}_i) (Y_{i-m} - \hat{\lambda}_{i-m})}{\sum_{i=m+1}^n \hat{\lambda}_i \hat{\lambda}_{i-m}}.$$

Il faut savoir que ces formules peuvent produire une variance négative et des corrélations qui, en valeur absolue, dépassent un ! Finalement, si l'on utilise une approximation autorégressive, on peut utiliser les relations de Yule-Walker.

### 3.2.6 Biais des estimateurs de Zeger

Les espérances  $\lambda_i$  sont estimées avec un biais,<sup>7</sup> puisque

$$E[\hat{\lambda}_i] \approx \lambda_i \exp \left\{ \frac{1}{2} (\langle \underline{X}_i, M_n \underline{X}_i \rangle) \right\},$$

où  $M_n = \Omega_{1,n}^{-1} + \Omega_{1,n}^{-1} \Omega_{2,n} \Omega_{1,n}^{-1}$  et

$$\Omega_{1,n} = \sum_{i=1}^n \hat{\lambda}_i \underline{X}_i \underline{X}_i^t$$

$$\Omega_{2,n} = \sum_{i=1}^n \sum_{j=1}^n \hat{\lambda}_i \hat{\lambda}_j \hat{\Gamma}_B(|i-j|) \underline{X}_i \underline{X}_j^t.$$

$\hat{\Gamma}_B$  est l'estimation de l'autocovariance. Pour alléger le calcul, on utilise l'approximation suivante (pour  $m$  approprié<sup>8</sup>) :

$$\Omega_{2,n} \approx \hat{\Gamma}_B(0) \sum_{i=1}^n \hat{\lambda}_i^2 \underline{X}_i \underline{X}_i^t + 2 + 2 \sum_{i=1}^m \hat{\Gamma}_B(i) \sum_{j=1}^{n-i} \hat{\lambda}_j \hat{\lambda}_{j+i} \underline{X}_j \underline{X}_{j+i}^t.$$

7. Davis et al. (2000). Mais il faudrait établir que les formules sont valides pour le modèle saturé que l'on utilise ici (non fonctionnel, alors que les résultats de l'article le sont pour des fonctions continues).

8. Etant donnée la nature particulière de la matrice  $X_n$ , il n'est pas évident que  $m$  puisse être choisi "petit".

Cela produit alors les estimations ( $1 \leq i \leq n$ )

$$\hat{\lambda}_i = \exp \left\{ -\frac{1}{2} \langle \underline{X}_i, M_n \underline{X}_i \rangle \right\} \lambda_i.$$

*Remarque* : En prenant en compte ces biais, on peut espérer améliorer les estimations des autocorrélations relatives au processus stationnaire  $B$  comme suit. Soient

$$\begin{aligned} g_i^{(n)} &= e^{\langle \underline{X}_i, M_n \underline{X}_i \rangle}, \\ g_{i,j}^{(n)} &= e^{\langle [\underline{X}_i + \underline{X}_j], M_n [\underline{X}_i + \underline{X}_j] \rangle}. \end{aligned}$$

Alors

$$\hat{\sigma}_B^2 = \frac{\sum_{i=1}^n \left\{ (Y_i - \hat{\lambda}_i)^2 + \left( 1 - \frac{2}{\{g_i^{(n)}\}^{\frac{3}{2}}} + \frac{1}{\{g_i^{(n)}\}^2} \right) \hat{\lambda}_i^2 - \hat{\lambda}_i \right\}}{\sum_{i=1}^n \frac{\hat{\lambda}_i^2}{\{g_i^{(n)}\}^2}}$$

et

$$\hat{\Gamma}_B(k) = \frac{\sum_{i=1}^{n-k} \left\{ (Y_i - \hat{\lambda}_i) (Y_{i+k} - \hat{\lambda}_{i+k}) + g_{i,i+k}^{(n)} \left( 1 - \sqrt{g_i^{(n)}} - \sqrt{g_{i+k}^{(n)}} + \frac{1}{g_{i,i+k}} \right) \right\}}{\sum_{i=1}^{n-k} g_{i,i+k} \hat{\lambda}_i \hat{\lambda}_{i+k}}.$$

## Chapitre 4

# Les séries chronologiques des comptages routiers : descriptions qui utilisent le maximum de vraisemblance

### 4.1 Calcul en supposant qu’il n’y a pas de dépendance temporelle

Les améliorations à la procédure de Zeger proposées par Davis *et al.* (2000) sont produites en recourant à des résultats généraux relatifs à la consistance et à la normalité asymptotique des estimateurs du maximum de vraisemblance obtenus à partir de la fonction de vraisemblance suivante (à minimiser) qui ignore la présence du processus stationnaire latent :

$$l(\underline{\theta}_p) = - \sum_{i=1}^n e^{\langle \underline{\theta}_p, \underline{X}_i \rangle} + \sum_{i=1}^n Y_i \langle \underline{\theta}_p, \underline{X}_i \rangle.$$

C’est cette approche qui permet la correction des biais de la méthode de Zeger. Mais Davis *et al.* proposent une alternative qui fait l’objet de la section suivante.

### 4.2 Calcul qui utilise une approximation de la vraisemblance

Le modèle est toujours le suivant ( $\mathcal{IP}$  signifiant “indépendantes et de loi de Poisson”) :

$$Y_i | B_i, \underline{X}_i \sim \mathcal{IP} \left( B_i e^{\langle \underline{\theta}_p, \underline{X}_i \rangle} \right), \quad 1 \leq i \leq n,$$

où  $B_i = e^{W_i}$  dénote l'exponentielle d'un processus stationnaire autorégressif tel que (*IIN* signifiant "indépendantes de même loi normale")

$$W_i = \tau_1 W_{i-1} + \dots + \tau_m W_{i-m} + N_i, \quad N_i \sim IIN(0, \sigma_N^2).$$

Une approximation de la vraisemblance, qui consiste à remplacer les exponentielles par un développement à l'ordre 2, conduit alors à l'algorithme suivant.

1. On suppose données les valeurs initiales suivantes :<sup>1</sup>

$$\underline{\tau}_m [0] = \begin{bmatrix} \tau_1 [0] \\ \vdots \\ \tau_m [0] \end{bmatrix}, \quad \sigma_N^2 [0], \quad \underline{w}_n [0] = \begin{bmatrix} w_1 [0] \\ w_2 [0] \\ w_3 [0] \\ \vdots \\ w_n [0] \end{bmatrix}.$$

On définit

$$\underline{b}_n [0] = \begin{bmatrix} e^{w_1 [0]} \\ e^{w_2 [0]} \\ e^{w_3 [0]} \\ \vdots \\ e^{w_n [0]} \end{bmatrix}, \quad X_n = [\underline{X}_1 | \dots | \underline{X}_n], \quad \underline{\lambda}_n (\underline{\theta}_p) = \begin{bmatrix} e^{\langle \underline{\theta}_p, \underline{X}_1 \rangle_{\mathbb{R}^p}} \\ \vdots \\ e^{\langle \underline{\theta}_p, \underline{X}_n \rangle_{\mathbb{R}^p}} \end{bmatrix}.$$

2. Soit

$$\mathcal{L}_0 (\underline{\theta}_p) = \langle \underline{\theta}_p, X_n \underline{Y}_n \rangle_{\mathbb{R}^p} - \langle \underline{\lambda}_n (\underline{\theta}_p), \underline{b}_n [0] \rangle_{\mathbb{R}^n}.$$

On résout

$$\underline{\theta}_p [1] = \arg [\max \{ \mathcal{L}_0 (\underline{\theta}_p) \}],$$

et calcule

$$\lambda_i [1] = e^{\langle \underline{\theta}_p [1], \underline{X}_i \rangle_{\mathbb{R}^p}}, \quad 1 \leq i \leq n.$$

3. Soit  $\Sigma_{\underline{W}_n}$  la matrice de covariance de  $\underline{W}_n$  : c'est une fonction de  $\underline{\tau}_m$  et de  $\sigma_N^2$ . C'est pourquoi on écrit  $\Sigma^{\underline{W}_n} (\underline{\tau}_m, \sigma_N^2)$  pour son inverse. On pose alors

$$\begin{aligned} \underline{\lambda}_n [1] &= \text{vecteur de composantes } \lambda_i [1], \\ \underline{\mathcal{B}}_n [0] &= \text{diag} \{ \underline{b}_n [0] \}, \\ \underline{\Lambda}_n [1] &= \text{diag} \{ \underline{\lambda}_n [1] \}, \\ \tilde{\underline{Y}}_n [0] &= \underline{Y}_n - \underline{\Lambda}_n [1] \underline{b}_n [0] + \underline{\mathcal{B}}_n [0] \underline{\Lambda}_n [1] \underline{w}_n [0], \end{aligned}$$

---

1. Aucune indication n'est fournie sur la manière de choisir ces valeurs initiales !

$$\begin{aligned}
& \mathcal{T}_n [1] (\mathcal{I}_m, \sigma_N^2) \\
&= \ln \left\{ \frac{|\Sigma^{\mathcal{W}_n} (\mathcal{I}_m, \sigma_N^2)|}{|\mathcal{B}_n [0] \Lambda_n [1] + \Sigma^{\mathcal{W}_n} (\mathcal{I}_m, \sigma_N^2)|} \right\} \\
&+ \langle \tilde{\mathcal{Y}}_n [0], (\mathcal{B}_n [0] \Lambda_n [1] + \Sigma^{\mathcal{W}_n} (\mathcal{I}_m, \sigma_N^2))^{-1} \tilde{\mathcal{Y}}_n [0] \rangle_{\mathbb{R}^n}.
\end{aligned}$$

On résout

$$(\mathcal{I}_m [1], \sigma_N^2 [1]) = \arg [\max \{ \mathcal{T}_n [1] (\mathcal{I}_m, \sigma_N^2) \}].$$

4. Soient

$$\underline{b}_n [\underline{w}_n] = \begin{bmatrix} e^{w_1} \\ e^{w_2} \\ e^{w_3} \\ \vdots \\ e^{w_n} \end{bmatrix}, \quad \mathcal{B}_n [\underline{w}_n] = \text{diag} \{ \underline{b}_n [\underline{w}_n] \},$$

$$\tilde{\mathcal{Y}}_n [\underline{w}_n] = \underline{\mathcal{Y}}_n - \Lambda_n [1] \underline{b}_n [\underline{w}_n] + \mathcal{B}_n [\underline{w}_n] \Lambda_n [1] \underline{w}_n.$$

On résout alors itérativement, pour  $\underline{w}_n$ , l'équation

$$\underline{w}_n = (\mathcal{B}_n [\underline{w}_n] \Lambda_n [1] + \Sigma^{\mathcal{W}_n} (\mathcal{I}_m [1], \sigma_N^2 [1]))^{-1} \tilde{\mathcal{Y}}_n [\underline{w}_n],$$

dont le résultat est  $\underline{w}_n [1]$ .

5. On retourne en 1. avec

$$\begin{aligned}
\underline{w}_n [0] &\leftarrow \underline{w}_n [1], \\
\mathcal{I}_m [0] &\leftarrow \mathcal{I}_m [1], \\
\sigma_N^2 [0] &\leftarrow \sigma_N^2 [1],
\end{aligned}$$

et on continue ainsi jusqu'à convergence.

### 4.2.1 Mise en oeuvre

#### Valeurs initiales

1. *Choix de l'ordre du processus autorégressif latent*

On a choisi un processus d'ordre 1 pour deux raisons : d'une part les difficultés de calcul rencontrées avec cette méthode poussent à travailler avec les modèles les plus simples, et, d'autre part, les calculs faits avec les autres méthodes, qui choisissent empiriquement l'ordre, ont toujours donné l'ordre 1 comme satisfaisant.

## 2. Choix des valeurs initiales

### (a) Paramètres du processus autorégressif latent

Suite à divers essais “empiriques,” on a constaté que les valeurs issues de la méthode de Zeger donnent les meilleurs résultats.

### (b) Valeurs du processus autorégressif latent

On a simulé le processus autorégressif et utilisé les valeurs simulées.

## Calcul des déterminants et des inverses des matrices de l’algorithme

Le problème majeur est celui de la concernées, problème déjà rencontré avec la méthode de Zeger, mais de manière plus limitée puisqu’ il y est confiné au calcul de la matrice de covariance des estimateurs, alors qu’ici il est au coeur du processus d’estimation de ces derniers. On va donc passer en revue un certain nombre de méthodes disponibles même si leur pertinence, pour les calculs qu’il faut faire, doit encore être démontrée.

### A) L’algorithme des innovations

C’est une méthode qui revient à effectuer une orthogonalisation de type Gram-Schmidt. On en trouve une description dans

1. P.J. Brockwell and R.A. Davis, Time Series : Theory and Methods, Springer, New York (1987), page 165,
2. P.J. Brockwell and R.A. Davis, Introduction to Time Series and Forecasting, Springer, New York (1996), page 70.

Soit  $X$  un processus du second ordre, d’espérance nulle.  $\hat{X}_k$  désigne la meilleure estimation linéaire de  $X_k$ , à partir de  $X_1, \dots, X_{k-1}$ , soit la projection orthogonale dans  $L_2[P]$  de  $X_k$  sur le sous-espace engendré par de  $X_1, \dots, X_{k-1}$ . On a, “génériquement,”

$$\hat{X}_k = \theta_1^{(k-1)} X_{k-1} + \dots + \theta_{k-1}^{(k-1)} X_1.$$

Par suite, si

$$-\theta_l^{(k-1)} := a_l^{(k-1)}, \quad 1 \leq l \leq k-1, \quad \underline{a}_k = \begin{bmatrix} a_{k-1}^{(k-1)} \\ \vdots \\ a_1^{(k-1)} \\ 1 \end{bmatrix}, \quad \underline{X}_k = \begin{bmatrix} X_1 \\ \vdots \\ X_k \end{bmatrix},$$

alors

$$X_k - \hat{X}_k = X_k - \left\{ \theta_1^{(k-1)} X_{k-1} + \dots + \theta_{k-1}^{(k-1)} X_1 \right\} = \langle \underline{a}_k, \underline{X}_k \rangle_{\mathbb{R}^k}.$$

Soit donc  $A_n$  la matrice définie par

$$A_n = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ a_1^{(1)} & 1 & 0 & 0 & \cdots & 0 & 0 \\ a_2^{(2)} & a_1^{(2)} & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ a_{n-1}^{(n-1)} & a_{n-2}^{(n-1)} & a_{n-3}^{(n-1)} & a_{n-4}^{(n-1)} & \cdots & a_1^{(n-1)} & 1 \end{bmatrix}.$$

et  $I_k$  l'innovation à l'instant  $k$  :  $I_k = X_k - \hat{X}_k$ . On a, quand

$$\underline{I}_n = \begin{bmatrix} I_1 \\ \vdots \\ I_n \end{bmatrix}, \quad \hat{\underline{X}}_n = \begin{bmatrix} \hat{X}_1 \\ \vdots \\ \hat{X}_n \end{bmatrix},$$

que<sup>2</sup>

$$\begin{aligned} \underline{I}_n &= \underline{X}_n - \hat{\underline{X}}_n = A_n \underline{X}_n, \\ \hat{\underline{X}}_n &= \underline{X}_n - \underline{I}_n = A_n^{-1} \underline{I}_n - \underline{I}_n = (A_n^{-1} - I_n) \underline{I}_n := B_n \underline{I}_n. \end{aligned}$$

Comme  $A_n^{-1}$  a la forme

$$A_n^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ \alpha_1^{(1)} & 1 & 0 & 0 & \cdots & 0 & 0 \\ \alpha_2^{(2)} & \alpha_1^{(2)} & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ \alpha_{n-1}^{(n-1)} & \alpha_{n-2}^{(n-1)} & \alpha_{n-3}^{(n-1)} & \alpha_{n-4}^{(n-1)} & \cdots & \alpha_1^{(n-1)} & 1 \end{bmatrix},$$

$$\hat{\underline{X}}_n = B_n \underline{I}_n = B_n (\underline{X}_n - \hat{\underline{X}}_n), \text{ où}$$

$$B_n = \begin{bmatrix} 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ \alpha_1^{(1)} & 0 & 0 & 0 & \cdots & 0 & 0 \\ \alpha_2^{(2)} & \alpha_1^{(2)} & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ \alpha_{n-1}^{(n-1)} & \alpha_{n-2}^{(n-1)} & \alpha_{n-3}^{(n-1)} & \alpha_{n-4}^{(n-1)} & \cdots & \alpha_1^{(n-1)} & 0 \end{bmatrix}.$$

Il s'en suit que (c'est l'algorithme des innovations)

$$\hat{X}_{k+1} = \begin{cases} 0 & \text{si } k = 0, \\ \sum_{l=1}^k \alpha_l^{(k)} (X_{k+1-l} - \hat{X}_{k+1-l}) & \text{si } k \geq 1. \end{cases}$$

---

2.  $I_n$  est la matrice identité de dimension  $n$ .

Si l'on pose

$$\hat{s}_k^2 = E \left[ \left( X_{k+1} - \hat{X}_{k+1} \right)^2 \right], \text{ et si } \sigma_{i,j}^X = E[X_i X_j],$$

les paramètres  $\alpha_i^{(j)}$  se calculent récursivement comme suit :

$$\begin{aligned} \hat{s}_0^2 &= \sigma_{1,1}^X \\ \alpha_{n-k}^{(n)} &= \frac{\sigma_{n+1,k+1}^X - \sum_{l=0}^{k-1} \alpha_{k-l}^{(k)} \alpha_{n-l}^{(n)} \hat{s}_l^2}{\hat{s}_k^2}, \quad 0 \leq k \leq n-1, \\ \hat{s}_n^2 &= \sigma_{n+1,n+1}^X - \sum_{l=0}^{n-1} \left( \alpha_{n-l}^{(n)} \right)^2 \hat{s}_l^2. \end{aligned}$$

L'ordre des calculs est le suivant :

$$\begin{aligned} &\hat{s}_0^2, \\ &\alpha_1^{(1)}, \hat{s}_1^2, \\ &\alpha_2^{(2)}, \alpha_1^{(2)}, \hat{s}_2^2, \\ &\alpha_3^{(3)}, \alpha_2^{(3)}, \alpha_2^{(3)}, \hat{s}_3^2, \\ &\dots \end{aligned}$$

Soit  $\Sigma_{\underline{X}_n}$  la matrice de covariance de  $\underline{X}_n$  et soit  $\underline{\sigma}_n^X$  le vecteur de composantes

$$E[X_n X_{n+1}], \dots, E[X_1 X_{n+1}].$$

Soit  $\underline{\gamma}_n$  le vecteur de coefficients

$$\gamma_1^{(n)}, \dots, \gamma_n^{(n)}$$

qui est solution de l'équation

$$[\Sigma_{\underline{X}_n}] \underline{\gamma}_n = \underline{\sigma}_n^X.$$

On a que<sup>3</sup>

$$\hat{X}_{n+1} = \sum_{i=1}^n \gamma_i^{(n)} X_{n+1-i}$$

est la meilleure approximation linéaire de  $X_{n+1}$  par moindres carrés qui utilise  $X_1, \dots, X_n$ , et  $\gamma_n^{(n)}$  est la corrélation partielle d'écart  $n$ . Soit  $C_n$  la matrice

---

3. P.J. Brockwell et R.A. Davis, Time Series : Theory and Methods, Springer, New York (1987), page 102.

$$C_n = \begin{bmatrix} 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ \gamma_1^{(1)} & 0 & 0 & 0 & \cdots & 0 & 0 \\ \gamma_2^{(2)} & \gamma_1^{(2)} & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ \gamma_{n-1}^{(n-1)} & \gamma_{n-2}^{(n-1)} & \gamma_{n-3}^{(n-1)} & \gamma_{n-4}^{(n-1)} & \cdots & \gamma_1^{(n-1)} & 0 \end{bmatrix}.$$

Alors

$$\hat{\underline{X}}_n = C_n \underline{X}_n.$$

Comme on a aussi que  $\hat{\underline{X}}_n = B_n (\underline{X}_n - \hat{\underline{X}}_n)$ ,

$$\hat{\underline{X}}_n = (I_n + B_n)^{-1} B_n \underline{X}_n = A_n^{-1} B_n \underline{X}_n,$$

d'où  $C_n = A_n^{-1} B_n$ , et la solution de  $[\Sigma_{\underline{X}_n}] \underline{\gamma}_n = \underline{\sigma}_n^X$  peut s'obtenir à partir de l'algorithme des innovations : on la trouve dans la dernière ligne de  $A_{n+1}^{-1} B_{n+1}$ .

Par ailleurs, les innovations sont des variables aléatoires non corrélées, d'espérance nulle et de variances respectives

$$\hat{s}_0^2, \hat{s}_1^2, \dots, \hat{s}_{n-1}^2.$$

De l'égalité  $\underline{L}_n = A_n \underline{X}_n$  on tire la suivante :

$$\text{diag} [\hat{s}_0^2, \hat{s}_1^2, \dots, \hat{s}_{n-1}^2] = A_n \Sigma_{\underline{X}_n} A_n^t,$$

et par suite les égalités

$$|\Sigma_{\underline{X}_n}| = \prod_{i=1}^{n-1} \hat{s}_i^2,$$

et

$$\Sigma_{\underline{X}_n}^{-1} = A_n^{-1} \{\text{diag} [\hat{s}_0^2, \hat{s}_1^2, \dots, \hat{s}_{n-1}^2]\}^{-1} (A_n^t)^{-1}.$$

De plus,

$$\langle \underline{X}_n, \Sigma_{\underline{X}_n}^{-1} \underline{X}_n \rangle_{\mathbb{R}^n} = \langle A_n^{-1} \underline{L}_n, \Sigma_{\underline{X}_n}^{-1} A_n^{-1} \underline{L}_n \rangle_{\mathbb{R}^n} = \langle \underline{L}_n, \Sigma_{\underline{L}_n}^{-1} \underline{L}_n \rangle_{\mathbb{R}^n} = \sum_{i=1}^n \frac{(X_i - \hat{X}_i)^2}{\hat{s}_{i-1}^2}.$$

*Remarque* : Cet algorithme demande que soient conservées en mémoire toutes les valeurs  $\alpha_i^{(j)}$  : donc, sauf dans certains cas particuliers, il sera d'usage difficile pour les séries de grande taille.

*Le cas particulier des processus autorégressifs*

Quand  $X$  est autorégressif d'ordre  $m$ , l'algorithme des innovations se simplifie. On recourt à cette fin aux définitions et à la notation suivantes :

$$\sigma_{i,j}^X = E[X_i X_j] = \sigma_X (i - j),$$

$$Y_n = \begin{cases} \frac{X_n}{\sigma_N} & \text{si } 1 \leq n \leq m, \\ \frac{X_n - \tau_1 X_{n-1} - \dots - \tau_m X_{n-m}}{\sigma_N} & \text{si } n > m, \end{cases}$$

$$\sigma_{i,j}^Y = E[Y_i Y_j] = \sigma_Y (i - j).$$

On a alors que, à cause des relations de Yule-Walker,

$$\sigma_{i,j}^Y = \begin{cases} \frac{\sigma_X(i-j)}{\sigma_N^2} & \text{si } 1 \leq i, j \leq m, \\ 1 & \text{si } i = j, i > m, \\ 0 & \text{dans les autres cas.} \end{cases}$$

L'algorithme d'innovation est calculé avec la matrice  $\Sigma_{\underline{Y}_n}$  de coefficients  $\sigma_{i,j}^Y$ . Soient  $\tilde{\alpha}_i^{(j)}$  et  $\tilde{s}_i^2$  les grandeurs issues de ce calcul. On a alors que

$$\hat{X}_{k+1} = \begin{cases} \sum_{l=1}^k \tilde{\alpha}_l^{(k)} (X_{k+1-l} - \hat{X}_{k+1-l}) & \text{si } 1 \leq k < m, \\ \tau_1 X_k + \dots + \tau_m X_{k+1-m} & \text{si } k \geq m, \end{cases}$$

$$\hat{s}_k^2 = \sigma_N^2 \tilde{s}_k^2.$$

On trouve dans l'Annexe II les formules explicites pour les cas  $m = 1, 2, 3$ . On trouve également dans cet annexe les formules qui donnent explicitement l'inverse de la matrice de covariance d'un processus autorégressif ainsi que l'approximation standard de la matrice des autocovariances d'un processus du même type.

*Remarque :* Ces simplifications ne s'appliquent que partiellement à la méthode de Davis *et al.* puisque les matrices en cause, avec une exception, ne sont pas celles d'un processus autorégressif (la diagonale des matrices n'est pas constante).

*Applications à l'algorithme de Davis et al. (2000)*

1. *Etape 3 :* Il s'agit de déterminer les valeurs de  $\underline{\mathcal{I}}_m$  et de  $\sigma_N^2$  qui maximisent la fonction

$$f(\underline{\mathcal{I}}_m, \sigma_N^2) = \ln \left\{ \frac{|\Sigma^{\underline{W}_n}(\underline{\mathcal{I}}_m, \sigma_N^2)|}{|\mathcal{B}_n[0] \Lambda_n[1] + \Sigma^{\underline{W}_n}(\underline{\mathcal{I}}_m, \sigma_N^2)|} \right\}$$

$$+ \langle \tilde{\underline{Y}}_n[0], (\mathcal{B}_n[0] \Lambda_n[1] + \Sigma^{\underline{W}_n}(\underline{\mathcal{I}}_m, \sigma_N^2))^{-1} \tilde{\underline{Y}}_n[0] \rangle_{\mathbb{R}^n}.$$

A cette fin, il faut exprimer

$$|\mathcal{B}_n [0] \Lambda_n [1] + \Sigma^{W_n} (\underline{\tau}_m, \sigma_N^2)|$$

et

$$\langle \tilde{\underline{Y}}_n [0], (\mathcal{B}_n [0] \Lambda_n [1] + \Sigma^{W_n} (\underline{\tau}_m, \sigma_N^2))^{-1} \tilde{\underline{Y}}_n [0] \rangle_{\mathbb{R}^n}$$

explicitement comme des fonctions des paramètres  $\underline{\tau}_m$  et  $\sigma_N^2$ . Cela signifie l'une de deux choses :

- (a) soit l'on évalue numériquement  $\Sigma_{W_n} (\underline{\tau}_m, \sigma_N^2)$ , matrice que l'on inverse numériquement ensuite, mais c'est là ce que l'on veut éviter étant donnée la taille des matrices,
- (b) soit on dispose de formules qui permettent d'exprimer explicitement  $\Sigma^{W_n} (\underline{\tau}_m, \sigma_N^2)$  en fonction des paramètres  $\underline{\tau}_m$  et  $\sigma_N^2$ .

Une autre solution consiste à utiliser une (bonne) approximation de la matrice  $\Sigma_{W_n} (\underline{\tau}_m, \sigma_N^2)$ , que l'on peut inverser facilement. Ces approches sont toutes possibles en principe. Mais le recours à l'algorithme des innovations n'est possible pratiquement que si la structure particulière des matrices concernées peut être exploitée.

2. *Etape 4* : Il s'agit de calculer

$$(\mathcal{B}_n [\underline{w}_n] \Lambda_n [1] + \Sigma^{W_n} (\underline{\tau}_m [1], \sigma_N^2 [1]))^{-1} \tilde{\underline{Y}}_n [\underline{w}_n].$$

Mais on a vu que c'est la solution de l'équation

$$(\mathcal{B}_n [\underline{w}_n] \Lambda_n [1] + \Sigma^{W_n} (\underline{\tau}_m [1], \sigma_N^2 [1])) \underline{x}_n = \tilde{\underline{Y}}_n [\underline{w}_n],$$

qui s'obtient également à partir de l'algorithme des innovations. Les remarques faites ci-dessus s'appliquent également ici.

#### B) L'algorithme du balayage<sup>4</sup>

Cet algorithme paraît particulièrement intéressant pour le problème de Davis *et al.* puisque chaque ligne de la matrice du problème ne contient qu'un petit nombre de valeurs non nulles, et que ces valeurs sont toujours les mêmes, sauf en ce qui concerne la diagonale principale. Voici la description de l'algorithme et de l'application appropriée telle qu'elles sont données par Lange (voir la note de bas de page).

Soit  $M$  une matrice symétrique de dimension  $p$ . Soit  $m_{i,i} > 0$  l'élément diagonal numéro  $i$  de  $M$ . "Balayer sur  $m_{i,i}$ " c'est produire la matrice  $\widetilde{M}$  dont les éléments

---

4. K. Lange, Numerical Analysis for Statisticians, Springer, New York (1999) : "Although there are faster and numerically more stable algorithms for inverting a matrix or solving a least-squares problem, no algorithm matches the conceptual simplicity and utility of sweeping." (page 79)

sont donnés par les formules suivantes :

$$\tilde{m}_{i,i} = -\frac{1}{m_{i,i}}, \quad \tilde{m}_{j,i} = \frac{m_{j,i}}{m_{i,i}}, \quad \tilde{m}_{i,j} = \frac{m_{i,j}}{m_{i,i}}, \quad \tilde{m}_{j,k} = m_{j,k} - \frac{m_{j,i}m_{i,k}}{m_{i,i}}, \quad j \neq i, \quad k \neq i.$$

Les **faits** importants sont les suivants :<sup>5</sup>

$M$  est positive définie si et seulement si il est possible de “balayer sur” chaque élément successif de la diagonale principale et que cet élément est positif tant que le “balayage sur” cet élément n’a pas eu lieu. Quand le “balayage” a eu lieu, l’élément diagonal ainsi créé est négatif et le demeure au cours des “balayages” suivants. Le produit des éléments diagonaux tels qu’ils sont juste avant le “balayage” fournit le déterminant de  $M$ . L’annexe contient une illustration de cet algorithme. De plus, si l’on a que

$$M = \begin{pmatrix} M_{1,1} & M_{1,2} \\ M_{2,1} & M_{2,2} \end{pmatrix},$$

où  $M_{1,1}$  et  $M_{2,2}$  sont carrées mais de dimensions non nécessairement égales, et que l’on “balayer sur” la diagonale de  $M_{1,1}$ ,<sup>6</sup> on obtient

$$\tilde{M} = \begin{pmatrix} -M_{1,1}^{-1} & M_{1,1}^{-1}M_{1,2} \\ M_{2,1}M_{1,1}^{-1} & M_{2,2} - M_{2,1}M_{1,1}^{-1}M_{1,2} \end{pmatrix},$$

*Application*

Soit la matrice

$$\begin{pmatrix} \Sigma & \underline{x} - \underline{\mu} \\ \underline{x}^t - \underline{\mu}^t & 0 \end{pmatrix}.$$

Le “balayage sur”  $\Sigma$  produit, en position  $\tilde{M}_{2,2}$ , la valeur du produit scalaire

$$-\langle \Sigma^{-1} (\underline{x}^t - \underline{\mu}^t), \underline{x}^t - \underline{\mu}^t \rangle.$$

Au cours du processus il est possible d’obtenir, par cumul des logarithmes, le logarithme du déterminant.

Soit encore le vecteur  $\begin{bmatrix} \underline{U} \\ \underline{V} \end{bmatrix}$ , de covariance

$$\begin{bmatrix} \Sigma_{\underline{U},\underline{U}} & \Sigma_{\underline{U},\underline{V}} & \underline{\mu}_{\underline{U}} - \underline{u} \\ \Sigma_{\underline{V},\underline{U}} & \Sigma_{\underline{V},\underline{V}} & \underline{\mu}_{\underline{V}} \\ \underline{\mu}_{\underline{U}}^t - \underline{u}^t & \underline{\mu}_{\underline{V}}^t & 0 \end{bmatrix}.$$

5. Lange, *Op. Cit.*, Proposition 7.5.3, page 83.

6. On suppose que cela est possible!

Le “balayage sur”  $\Sigma_{\underline{U}, \underline{U}}$  produit

$$\begin{aligned} E[\underline{V} | \underline{U} = \underline{u}] &= \underline{\mu}_V + \Sigma_{\underline{V}, \underline{U}} \Sigma_{\underline{U}, \underline{U}}^{-1} (\underline{u} - \underline{\mu}_U), \\ V[\underline{V} | \underline{U} = \underline{u}] &= \Sigma_{\underline{V}, \underline{V}} - \Sigma_{\underline{V}, \underline{U}} \Sigma_{\underline{U}, \underline{U}}^{-1} \Sigma_{\underline{U}, \underline{V}}. \end{aligned}$$

### C) L’exploitation de la forme particulière de la matrice

Comme l’on a une “matrice bande,” on peut recourir aux algorithmes spécifiques exposés dans

G.H. Golub and C.F. Van Loan, *Matrix Computations*, Johns Hopkins U Press, Baltimore (1983), 96-98.

## 4.2.2 Problèmes de calcul

Il y a deux fonctions de vraisemblance dont on cherche l’optimum,  $\mathcal{L}_0$  et  $\mathcal{T}_n$ . Les difficultés rencontrées sont relatives à l’optimisation individuelle de chacune des fonctions, mais aussi à leur enchaînement, car les optimisations sont alternées.

Le problème avec  $\mathcal{L}_0$  tient au nombre de paramètres (plusieurs dizaines), ce qui entraîne une convergence très lente de l’algorithme d’optimisation. Il est donc nécessaire de procéder à un très grand nombre d’itérations qui, pour une optimisation, nécessitent à leur tour plusieurs heures de calcul dans un environnement “standard.”

Les difficultés de l’optimisation de  $\mathcal{T}_n$  sont d’un autre ordre : il y a deux paramètres, mais il faut calculer le déterminant d’une très grande matrice ainsi que résoudre un système linéaire de très grande taille. Comme on l’a déjà remarqué l’algorithme des innovations ne peut pas être invoqué dans un tel cas. En revanche, les matrices étant des matrices bandes, les procédures de calcul de MATLAB s’avèrent efficaces, quitte à normaliser ces matrices pour avoir des déterminants dont la valeur numérique n’excède pas les capacités de calcul.

On a essayé de réduire les problèmes de calcul relatifs à  $\mathcal{L}_0$  en limitant le nombre d’itérations. Mais on a alors constaté que le valeur du déterminant qui apparaît dans  $\mathcal{T}_n$ , même après normalisation, est très instable et acquiert rapidement une valeur qui dépasse les capacités de la machine utilisée. Cela semble être la conséquence d’une première étape de calcul de précision insuffisante.

## 4.3 Le modèle de Ledolter

La méthode de Ledolter est une méthode de vraisemblance qui fait appel à des calculs de type Monte-Carlo, et qui en cela évite le calcul d’inverses et de déterminants. La méthode de Davis *et al.* a été développée pour éviter les calculs

de type Monte-Carlo, mais elle conduit au calcul d'inverses et de déterminants. L'avantage de l'une des méthodes par rapport à l'autre dans le contexte des comptages ne peut être établie qu'expérimentalement.

### 4.3.1 Les algorithmes EM et MCEM

$\underline{Y}$  représente les données que l'on peut observer et  $\underline{X}$  représente les données "complètes :". On suppose que  $\underline{Y}$  est une fonction mesurable de  $\underline{X}$ . La densité de  $\underline{X}$  et le logarithme de cette densité sont notés respectivement

$$f_{\underline{X}}(\underline{x} | \underline{\theta}) \text{ et } L_{\underline{X}}(\underline{\theta} | \underline{x}).$$

$\underline{\theta}$  est un paramètre que l'on veut estimer. La densité de  $\underline{X}$  quand  $\underline{Y} = \underline{y}$  est notée

$$f_{\underline{X}|\underline{Y}=\underline{y}}(\underline{x} | \underline{\theta}),$$

et, quand  $F$  dénote une fonction de  $\underline{x}$  appropriée, on écrit

$$E_{\underline{\theta}} [F(\underline{X}) | \underline{Y} = \underline{y}] \text{ pour } \int F(\underline{x}) f_{\underline{X}|\underline{Y}=\underline{y}}(\underline{x} | \underline{\theta}) d\underline{x}.$$

L'algorithme EM consiste à itérer les deux étapes suivantes ( $\underline{\theta}$  est variable,  $\underline{\theta}^*$  est constant, le temps d'une itération) :

1. Etape **E** : calculer  $Q(\underline{\theta} | \underline{\theta}^*) := E_{\underline{\theta}^*} [L_{\underline{X}}(\underline{\theta} | \underline{X}) | \underline{Y} = \underline{y}]$ .
2. Etape **M** : trouver  $\underline{\theta}_{max}$  qui maximise  $\underline{\theta} \mapsto Q(\underline{\theta} | \underline{\theta}^*)$ .

On pose alors  $\underline{\theta}^* \leftarrow \underline{\theta}_{max}$ , et on reconduit les étapes **E** et **M**.

En pratique, le calcul de **E** est difficile et on le remplace par le calcul suivant. On observe les valeurs  $\{\underline{x}_i\}$  d'une chaîne de Markov dont la loi invariante est celle dont la densité est

$$f_{\underline{X}|\underline{Y}=\underline{y}}(\underline{x} | \underline{\theta}^*).$$

Pour obtenir l'algorithme MCEM on itère alors

1. Etape **MCE** : calculer  $Q_m(\underline{\theta} | \underline{\theta}^*) := \frac{1}{m} \sum_{i=1}^m L_{\underline{X}}(\underline{\theta} | \underline{x}_i)$ .
2. Etape **MCM** : trouver  $\underline{\theta}_{max}$  qui maximise  $\underline{\theta} \mapsto Q_m(\underline{\theta} | \underline{\theta}^*)$ .

### 4.3.2 Le modèle de Ledolter proprement dit

Soit  $\underline{B} = \{B_n, n \in \mathbb{N}\}$ , un bruit blanc Gaussien de variance  $\sigma_B^2$ . Soit  $\underline{W} = \{W_n, n \in \mathbb{Z}\}$  le processus stationnaire latent obtenu en posant

$$W_n = \rho W_{n-1} + B_n.$$

Alors

$$P(Y_1 = k_1, \dots, Y_n = k_n | \underline{W}) = \prod_{i=1}^n P(Y_i = k_i | \underline{W})$$

et

$$P(Y_i = k_i | \underline{W}) = e^{-\lambda_i} \frac{\lambda_i^{k_i}}{k_i!}, \quad \ln[\lambda_i] = \langle \underline{\alpha}, \underline{X}_i \rangle_{\mathbb{R}^p} + W_i.$$

Soient

$$\underline{\theta} = \begin{bmatrix} \underline{\alpha} \\ \rho \\ \sigma_B^2 \end{bmatrix} \in \mathbb{R}^{p+2},$$

$$\underline{Y}_n^t = (Y_1 \cdots Y_n), \quad \underline{W}_n^t = (W_1 \cdots W_n),$$

$$\underline{V}_n = \begin{bmatrix} \underline{Y}_n \\ \underline{W}_n \end{bmatrix}.$$

$\underline{V}_n$  est la donnée ‘‘complète’’ et  $\underline{Y}_n$  représente ce qui est observé. On a que (voir annexe)

$$\begin{aligned} f_{W_2, \dots, W_n | W_1}(w_2, \dots, w_n) P(Y_1 = k_1, \dots, Y_n = k_n | W_1 = w_1, \dots, W_n = w_n) \\ = \frac{e^{-\frac{1}{2\sigma_B^2} \sum_{i=1}^{n-1} (w_{i+1} - \rho w_i)^2}}{(2\pi\sigma_B^2)^{\frac{n-1}{2}}} \prod_{i=1}^n e^{-\lambda_i} \frac{\lambda_i^{k_i}}{k_i!}. \end{aligned}$$

### 4.3.3 La fonction $Q_m$

Soit  $L_{\underline{V}_n | W_1 = w_1}(\underline{\theta} | \underline{k}_n, \underline{w}_n)$  le logarithme de la densité de  $\underline{V}_n$  connaissant  $W_1$ .

On a :

$$\begin{aligned} L_{\underline{V}_n | W_1 = w_1}(\underline{\theta} | \underline{k}_n, \underline{w}_n) &= \ln \left( f_{\underline{V}_n | W_1 = w_1}(\underline{k}_n, \underline{w}_n | \underline{\theta}) \right), \\ L_{\underline{V}_n | W_1 = w_1}(\underline{\theta} | \underline{k}_n, \underline{w}_n) &= \gamma - \frac{n-1}{2} \ln[\sigma_B^2] \\ &\quad - \frac{1}{2\sigma_B^2} \sum_{i=1}^{n-1} (w_{i+1} - \rho w_i)^2 + \sum_{i=1}^n \{k_i \ln[\lambda_i] - \lambda_i\}. \end{aligned}$$

Comme  $\lambda_i = e^{\langle \underline{\alpha}, \underline{X}_i \rangle_{\mathbb{R}^p} + W_i}$ ,

$$\begin{aligned} L_{\underline{V}_n | W_1 = w_1}(\underline{\theta} | \underline{k}_n, \underline{w}_n) &= \gamma - \frac{n-1}{2} \ln[\sigma_B^2] - \frac{1}{2\sigma_B^2} \sum_{i=1}^{n-1} (w_{i+1} - \rho w_i)^2 \\ &\quad + \sum_{i=1}^n \left\{ k_i (\langle \underline{\alpha}, \underline{X}_i \rangle_{\mathbb{R}^p} + w_i) - e^{\langle \underline{\alpha}, \underline{X}_i \rangle_{\mathbb{R}^p} + w_i} \right\}. \end{aligned}$$

Si l'on introduit les symboles suivants, regroupant les "données" de l'échantillon, il vient :

$$\begin{aligned}
A_i &= e^{w_i} \\
B &= \sum_{i=1}^{n-1} w_{i+1}^2 \\
C &= \sum_{i=1}^{n-1} w_i^2 \\
D &= \sum_{i=1}^{n-1} w_i w_{i+1} \\
E &= \sum_{i=1}^n k_i w_i
\end{aligned}$$

et

$$\begin{aligned}
L_{\underline{V}_n | W_1}(\underline{\theta} | \underline{k}_n, \underline{w}_n) &= \gamma - \frac{n-1}{2} \ln [\sigma_B^2] - \frac{1}{2\sigma_B^2} (B - 2\rho D + \rho^2 C)^2 \\
&+ E + \sum_{i=1}^n \left\{ k_i (\langle \underline{\alpha}, \underline{X}_i \rangle_{\mathbb{R}^p}) - A_i e^{\langle \underline{\alpha}, \underline{X}_i \rangle_{\mathbb{R}^p}} \right\}.
\end{aligned}$$

Comme

$$Q_m(\underline{\theta} | \underline{\theta}^*) = \frac{1}{m} \sum_{i=1}^m L_{\underline{V}_n | W_1 = w_1}(\underline{\theta} | \underline{k}_n^{(i)}, \underline{w}_n^{(i)}),$$

où l'on écrit, par exemple  $\underline{v}_n^{(i)}$  pour l'échantillon numéro  $i$ , de composantes  $\underline{k}_n^{(i)}$  et  $\underline{w}_n^{(i)}$ ,  $\underline{\theta}^*$  étant la valeur du paramètre utilisée pour obtenir les valeurs de la chaîne de Markov, si l'on définit

$$\overline{A}_i^{(m)}, \overline{B}^{(m)}, \overline{C}^{(m)}, \overline{D}^{(m)}, \overline{E}^{(m)},$$

en posant par exemple

$$\overline{A}_i^{(m)} = \frac{1}{m} \sum_{j=1}^m A_i^{(j)}, \quad A_i^{(j)} = e^{w_i^{(j)}},$$

alors

$$Q_m(\underline{\theta} | \underline{\theta}^*) = \gamma - \frac{n-1}{2} \ln [\sigma_B^2]$$

$$\begin{aligned}
& - \frac{1}{2\sigma_B^2} \left( \overline{B}^{(m)} - 2\rho\overline{D}^{(m)} + \rho^2\overline{C}^{(m)} \right)^2 + \overline{E}^{(m)} \\
& + \sum_{i=1}^n \left\{ k_i (\langle \underline{\alpha}, \underline{X}_i \rangle_{\mathbb{R}^p}) - \overline{A}_i^{(m)} e^{\langle \underline{\alpha}, \underline{X}_i \rangle_{\mathbb{R}^p}} \right\}.
\end{aligned}$$

#### 4.3.4 Optimisation de la fonction $Q_m$

On a :

$$\begin{aligned}
\rho_{max}(\underline{\theta}^*) &= \frac{\overline{D}^{(m)}}{\overline{C}^{(m)}}, \\
\{\sigma_B^2\}_{max}(\underline{\theta}^*) &= \frac{\overline{B}^{(m)} - [\overline{D}^{(m)}]^2}{(n-1)\overline{C}^{(m)}}, \\
\underline{\alpha}_{max}(\underline{\theta}^*) &= \underline{\alpha}^*,
\end{aligned}$$

où  $\underline{\alpha}^*$  s'obtient numériquement à l'aide de l'option DFPMIN de l'algorithme de Davidon-Fletcher-Powell.<sup>7</sup>

#### 4.3.5 Simulation des données “complètes” sur la base des observations

Il faut calculer la densité de  $P(\underline{Y}_n \in A \mid \underline{Y}_n = \underline{m}_n^{(0)})$ . Quand on choisit un ensemble  $A$  de la forme  $A = A_1 \times A_2$ , il vient

$$\begin{aligned}
P(\underline{Y}_n \in A \mid \underline{Y}_n = \underline{m}_n^{(0)}) &= \frac{P(\underline{Y}_n \in A, \underline{Y}_n = \underline{m}_n^{(0)})}{P(\underline{Y}_n = \underline{m}_n^{(0)})} \\
&= I_{A_1}(\underline{m}_n^{(0)}) \frac{P(\underline{W}_n \in A_2, \underline{Y}_n = \underline{m}_n^{(0)})}{P(\underline{Y}_n = \underline{m}_n^{(0)})} \\
&= I_{A_1}(\underline{m}_n^{(0)}) \int_{A_2} d\underline{w}_n f_{\underline{W}_n}(\underline{w}_n) \prod_{i=1}^n e^{-\lambda_i(w_i)} \frac{[\lambda_i(w_i)]^{m_i}}{m_i!} \\
&= \int_{A_1} \delta_{\underline{m}_n^{(0)}}(d\underline{m}_n) \int_{A_2} d\underline{w}_n f_{\underline{W}_n}(\underline{w}_n) \prod_{i=1}^n e^{-\lambda_i(w_i)} \frac{[\lambda_i(w_i)]^{m_i}}{m_i!}.
\end{aligned}$$

---

7. W.H. Press, B.P. Flannery, S.A. Teukolsky, et W.T. Vetterling, Numerical Recipes, Cambridge University Press, Cambridge, UK (1986).

Ici

$$\delta_{\underline{m}_n^{(0)}}(A_1) = \begin{cases} 0 & \text{quand } \underline{m}_n^{(0)} \notin A_1, \\ 1 & \text{quand } \underline{m}_n^{(0)} \in A_1, \end{cases} \quad \text{et } \lambda_i(w_i) = e^{\langle \underline{\alpha}, \underline{X}_i \rangle_{\mathbb{R}^p} + w_i}.$$

La densité recherchée est donc

$$f_{\underline{W}_n}(\underline{w}_n) \prod_{i=1}^n e^{-\lambda_i(w_i)} \frac{[\lambda_i(w_i)]^{m_i^{(0)}}}{m_i^{(0)!}.$$

La simulation de  $\underline{W}_n$  conditionnellement aux valeurs de  $\underline{Y}_n$  peut donc se faire à partir de l'échantillonneur de Gibbs qui fonctionne comme suit.<sup>8</sup>

Soit  $\underline{z}_n^{(k)}$  un vecteur de valeurs simulées de  $\underline{Z}_n$ .

1. Obtenir une valeur simulée  $z_1^{(k+1)}$  de  $Z_1$  à partir de la loi conditionnelle

$$f_{Z_1|Z_2=z_2^{(k)}, Z_3=z_3^{(k)}, \dots, Z_n=z_n^{(k)}}.$$

2. Obtenir une valeur simulée  $z_2^{(k+1)}$  de  $Z_2$  à partir de la loi conditionnelle

$$f_{Z_2|Z_1=z_1^{(k+1)}, Z_3=z_3^{(k)}, \dots, Z_n=z_n^{(k)}}.$$

3. Obtenir une valeur simulée  $z_3^{(k+1)}$  de  $Z_3$  à partir de la loi conditionnelle

$$f_{Z_3|Z_1=z_1^{(k+1)}, Z_2=z_2^{(k+1)}, \dots, Z_n=z_n^{(k)}}.$$

4. ....

5. Obtenir une valeur simulée  $z_n^{(k+1)}$  de  $Z_n$  à partir de la loi conditionnelle

$$f_{Z_n|Z_1=z_1^{(k+1)}, Z_2=z_2^{(k+1)}, \dots, Z_{n-1}=z_{n-1}^{(k+1)}}.$$

6. On pose que  $\underline{z}_n^{(k+1)}$  est le vecteur de composantes  $z_1^{(k+1)}, \dots, z_n^{(k+1)}$ , et on recommence l'opération.

Pour le cas qui est présentement traité, les valeurs de  $\underline{Y}_n$  sont constantes et égales à  $\underline{m}_n^{(0)}$ . Si l'on suppose alors que le vecteur simulé courant ait les composantes

$$m_1^{(0)}, \dots, m_n^{(0)}, w_1^{(k)}, \dots, w_n^{(k)},$$

on obtiendra

---

8. C.P. Robert et G. Casella, Monte Carlo Statistical Methods, Springer, New York (1999), page 285).

1. la valeur  $w_1^{(k+1)}$  en simulant une variable de densité

$$\frac{f_{W_1, \dots, W_n} \left( x, w_2^{(k)}, \dots, w_n^{(k)} \right)}{f_{W_2, \dots, W_n} \left( w_2^{(k)}, \dots, w_n^{(k)} \right)} e^{-\lambda_1(x)} \frac{[\lambda_1(x)]^{m_1^{(0)}}}{m_1^{(0)}!}$$

$$= f_{W_1|W_2=w_2^{(k)}, \dots, W_n=w_n^{(k)}}(x) e^{-\lambda_1(x)} \frac{[\lambda_1(x)]^{m_1^{(0)}}}{m_1^{(0)}!}$$

2. la valeur  $w_2^{(k+1)}$  en simulant une variable de densité

$$\frac{f_{W_1, \dots, W_n} \left( w_1^{(k+1)}, x, w_3^{(k)}, \dots, w_n^{(k)} \right)}{f_{W_1, W_3, \dots, W_n} \left( w_1^{(k+1)}, w_3^{(k)}, \dots, w_n^{(k)} \right)} e^{-\lambda_2(x)} \frac{[\lambda_2(x)]^{m_2^{(0)}}}{m_2^{(0)}!}$$

$$= f_{W_2|W_1=w_1^{(k+1)}, W_3=w_3^{(k)}, \dots, W_n=w_n^{(k)}}(x) e^{-\lambda_2(x)} \frac{[\lambda_2(x)]^{m_2^{(0)}}}{m_2^{(0)}!}$$

3. et ainsi de suite ...

Les densités conditionnelles qui apparaissent ci-dessus sont des lois normales et, en posant,

$$\underline{W}_n^{(1)} = \begin{bmatrix} W_2 \\ \vdots \\ W_n \end{bmatrix}, \quad \underline{W}_n^{(i)} = \begin{bmatrix} W_1 \\ \vdots \\ W_{i-1} \\ W_{i+1} \\ \vdots \\ W_n \end{bmatrix}, \quad \text{si } 1 < i < n, \quad \underline{W}_n^{(n)} = \begin{bmatrix} W_1 \\ \vdots \\ W_{n-1} \end{bmatrix}.$$

on a (le calcul est en annexe) :

$$W_1 | \underline{W}_n^{(1)} \sim \mathcal{N}(\rho w_2, \sigma_N^2),$$

$$W_i | \underline{W}_n^{(i)} \sim \mathcal{N}\left(\frac{\rho}{1+\rho^2} (w_{i-1} + w_{i+1}), \frac{\sigma_N^2}{1+\rho^2}\right),$$

$$W_n | \underline{W}_n^{(n)} \sim \mathcal{N}(\rho w_{n-1}, \sigma_N^2).$$

*Remarques :*

1. L'approche proposée se généralise à des processus qui contiennent les processus  $ARMA(p, q)$ .
2. L'approche proposée permet la prévision.
3. La méthode permet aussi d'évaluer des valeurs manquantes et peut être adaptée à des observations irrégulièrement espacées dans le temps.

## Chapitre 5

# Les méthodes qui utilisent le filtre de Kalman : le modèle de JLCSS

JLCSS abrège les noms des auteurs de la référence suivante : B. Jørgensen, S. Lundbye-Christensen, X.-K. Song et L. Sun, A state space model for multivariate longitudinal count data, *Biometrika* 86 (1999), 169-181.

### 5.1 Le modèle

On suppose que l'on a  $p$  séries chronologiques de vecteurs de dimension  $m$  (représentant  $m$  "catégories") de comptages qui sont réalisés en  $p$  "situations" et en  $n_j$ ,  $1 \leq j \leq p$ , instants successifs :

$$\underline{N}_1^{(j)}, \dots, \underline{N}_{n_j}^{(j)}, \underline{N}_i^{(j)} \in \mathbb{N}_0^m, 1 \leq i \leq n_j, 1 \leq j \leq p.$$

On pose

$$\underline{N}^{(j)} = \begin{bmatrix} \underline{N}_1^{(j)} \\ \vdots \\ \underline{N}_{n_j}^{(j)} \end{bmatrix}.$$

On suppose que les  $m$  comptages, en chaque instant, reflètent, par "situation," les mêmes "déterminants" qui entrent dans le modèle par le biais d'un processus de Markov scalaire latent ( $\theta_0^{(j)}$  représente une "valeur initiale")

$$\theta_0^{(j)}, \theta_1^{(j)}, \dots, \theta_{n_j}^{(j)}, \theta_i^{(j)} \in \mathbb{R}, 0 \leq i \leq n_j, 1 \leq j \leq p.$$

Par ailleurs, il y a 3 catégories de "co-variables"

$$\left\{ \underline{\mathbf{u}}_i^{(j)}, \underline{\mathbf{v}}_i^{(j)}, \underline{\mathbf{w}}^{(j)}, 1 \leq i \leq n_j, 1 \leq j \leq p \right\}.$$

Ces “co-variables” représentent, respectivement, des effets de court-terme, de long terme, et de structure (constants).

On pose successivement ( $\mathcal{L}(X | Y)$  dénote la loi de  $X$  conditionnellement à  $Y$ ) :

1.  $N_i^{(j,k)}$  : comptage dans la situation  $j$ , à l’instant  $i$ , pour la catégorie  $k$ .
2.  $\mathcal{P}(\lambda)$  : loi de Poisson de paramètre  $\lambda$ .
3.  $G(\mu, \kappa^2)$  : loi Gamma de densité

$$g(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, \quad x > 0, \quad \alpha > 0, \quad \lambda > 0,$$

avec,  $\mu$  étant l’espérance,  $\sigma^2$  la variance et  $\kappa$  le coefficient de variation  $\frac{\sigma}{\mu}$ ,

$$\mu = \frac{\alpha}{\lambda}, \quad \sigma^2 = \frac{\alpha}{\lambda^2}, \quad \kappa^2 = \frac{\sigma^2}{\mu^2} = \frac{1}{\alpha}.$$

$$4. \quad \mathcal{L}(\theta_0^{(j)}) = G(g_j, \tau^2), \quad \text{où } g_j = e^{\langle \underline{w}^{(j)}, \underline{\gamma} \rangle_{\mathbb{R}^c}}.$$

$\tau^2$  représente la variation aléatoire entre “situations” et  $\underline{w}^{(j)}$ , la variation systématique.  $\underline{\gamma}$  est un paramètre à estimer.

$$5. \quad \mathcal{L}(\theta_i^{(j)} | \theta_{i-1}^{(j)}) = G\left(b_i^{(j)} \theta_{i-1}^{(j)}, \frac{\eta^2}{\theta_{i-1}^{(j)}}\right), \quad \text{où}$$

$$\begin{aligned} b_i^{(j)} &= e^{\langle \Delta \underline{v}_i^{(j)}, \underline{\beta}^{(j)} \rangle_{\mathbb{R}^{b_j}}}, \\ \Delta \underline{v}_i^{(j)} &= \underline{v}_i^{(j)} - \underline{v}_{i-1}^{(j)}, \quad \underline{v}_0^{(j)} = \underline{0}. \end{aligned}$$

$$6. \quad \mathcal{L}(N_i^{(j,k)} | \theta_i^{(j)}) = \mathcal{P}(a_i^{(j,k)} \theta_i^{(j)}), \quad \text{où } a_i^{(j,k)} = e^{\langle \underline{w}_i^{(j)}, \underline{\alpha}_k^{(j)} \rangle_{\mathbb{R}^{a_j}}}.$$

$$7. \quad \underline{\theta}^{(j)} = \begin{bmatrix} \theta_0^{(j)} \\ \theta_1^{(j)} \\ \vdots \\ \theta_{n_j}^{(j)} \end{bmatrix}, \quad \underline{\theta}_{[-0]}^{(j)} = \begin{bmatrix} \theta_1^{(j)} \\ \vdots \\ \theta_{n_j}^{(j)} \end{bmatrix}.$$

On suppose alors que la loi des observations est la suivante. Soit

$$\{B_{i,j,k} \subseteq \mathbb{R}, \quad 1 \leq i \leq n_j, \quad 1 \leq j \leq p, \quad 1 \leq k \leq m\}$$

une famille quelconque de Boréliens. Alors

$$\begin{aligned} & P \left( N_i^{(j,k)} \in B_{i,j,k}, 1 \leq j \leq p, 1 \leq k \leq m, 1 \leq i \leq n_j \mid \underline{\theta}^{(j)}, 1 \leq j \leq p \right) \\ &= \prod_{1 \leq j \leq p, 1 \leq k \leq m, 1 \leq i \leq n_j} P \left( N_i^{(j,k)} \in B_{i,j,k} \mid \underline{\theta}^{(j)} \right) \\ &= \prod_{1 \leq j \leq p, 1 \leq k \leq m, 1 \leq i \leq n_j} P \left( N_i^{(j,k)} \in B_{i,j,k} \mid \theta_i^{(j)} \right), \end{aligned}$$

et

$$P \left( N_i^{(j,k)} \in B_{i,j,k} \mid \theta_i^{(j)} \right) = \sum_{l \in B_{i,j,k}} e^{-[a_i^{(j,k)} \theta_i^{(j)}]} \frac{[a_i^{(j,k)} \theta_i^{(j)}]^l}{l!}.$$

*Interprétation :* Les catégories pourraient être celles des divers types de véhicules, les “situations,” les diverses stations de comptage, les effets de long terme les types de jours et l’année, les effets de court terme, l’heure, et les effets de structure des variables relatives aux caractéristiques des stations de comptage (type de route). Une autre interprétation, les catégories pouvant être corrélées positivement, consisterait à définir les stations de comptage comme étant les catégories.

## 5.2 La structure des moments du modèle

Les calculs sont en annexe.

### 5.2.1 Moments du processus latent

#### Espérance du processus latent

$$\ln \left( E \left[ \theta_i^{(j)} \right] \right) = \ln (g_j) + \sum_{l=1}^i \langle \Delta \underline{v}_l^{(j)}, \underline{\beta}^{(j)} \rangle_{\mathbb{R}^{b_j}} = \langle \underline{w}^{(j)}, \underline{\gamma} \rangle_{\mathbb{R}^c} + \langle \underline{v}_i^{(j)}, \underline{\beta}^{(j)} \rangle_{\mathbb{R}^{b_j}}.$$

On écrira  $\mu_j^\theta(i) = E \left[ \theta_i^{(j)} \right]$ .

#### Variance du processus latent

Si  $\zeta_i^{(j)}$  et  $\xi_i^{(j)}$  sont définis respectivement par

$$\begin{aligned} \zeta_i^{(j)} &= b_i^{(j)} + b_i^{(j)} b_{i-1}^{(j)} + b_i^{(j)} b_{i-1}^{(j)} b_{i-2}^{(j)} + \dots + b_i^{(j)} b_{i-1}^{(j)} b_{i-2}^{(j)} \dots b_3^{(j)} b_2^{(j)} b_1^{(j)}, \\ \xi_i^{(j)} &= \frac{V \left[ \theta_i^{(j)} \right]}{E \left[ \theta_i^{(j)} \right]} = \zeta_i^{(j)} \eta^2 + \mu_j^\theta(i) \tau^2, \end{aligned}$$

alors

$$V \left[ \theta_i^{(j)} \right] = \xi_i^{(j)} \mu_j^\theta(i) = \xi_i^{(j)} E \left[ \theta_i^{(j)} \right].$$

### Corrélation du processus latent

On a que

$$\rho \left( \theta_i^{(j)}, \theta_{i+l}^{(j)} \right) = \frac{\xi_i^{(j)} \mu_j^\theta(i+l)}{\sqrt{\xi_i^{(j)} \mu_j^\theta(i)} \sqrt{\xi_{i+l}^{(j)} \mu_j^\theta(i+l)}} = \sqrt{\frac{\xi_i^{(j)} \mu_j^\theta(i+l)}{\xi_{i+l}^{(j)} \mu_j^\theta(i)}}.$$

### 5.2.2 Les innovations du processus latent

Soit  $\hat{\theta}_i^{(j)} := E \left[ \theta_i^{(j)} \mid \theta_0^{(j)}, \theta_1^{(j)}, \dots, \theta_{i-1}^{(j)} \right] = b_i^{(j)} \theta_{i-1}^{(j)}$ . L'innovation  $I_i^{(j)}$  est définie par l'expression

$$I_i^{(j)} = \theta_i^{(j)} - \hat{\theta}_i^{(j)} = \theta_i^{(j)} - b_i^{(j)} \theta_{i-1}^{(j)}.$$

Si  $X$  est adaptée à la tribu  $\sigma \left( \theta_0^{(j)}, \theta_1^{(j)}, \dots, \theta_{i-1}^{(j)} \right)$ , alors

$$\begin{aligned} \text{cov} \left( I_i^{(j)}, X \right) &= \text{cov} \left( \theta_i^{(j)} - \hat{\theta}_i^{(j)}, X \right) \\ &= \text{cov} \left( \theta_i^{(j)}, X \right) - \text{cov} \left( \hat{\theta}_i^{(j)}, X \right) \\ &= \text{cov} \left( \hat{\theta}_i^{(j)}, X \right) - \text{cov} \left( \hat{\theta}_i^{(j)}, X \right) \\ &= 0. \end{aligned}$$

Donc

$$\text{cov} \left( I_{i+l}^{(j)}, I_i^{(j)} \right) = 0, \quad l \geq 1,$$

c'est-à-dire que les innovations ont une structure de corrélation qui est celle d'un processus  $AR(1)$ , que leurs covariances "strictes" sont nulles et que leurs variances sont égales à  $b_i^{(j)} \mu_j^\theta(i) \eta^2$ .

### 5.2.3 Moments du processus des comptages

#### Espérance des comptages

$$E \left[ \underline{N}_i^{(j)} \right] = \mu_j^\theta(i) \underline{a}_i^{(j)}.$$

Sous forme "standard" le modèle a donc l'expression suivante :

$$\ln \left( E \left[ N_i^{(j,k)} \right] \right) = \langle \underline{u}_i^{(j)}, \underline{\alpha}_k^{(j)} \rangle_{\mathbb{R}^{a_j}} + \langle \underline{v}_i^{(j)}, \underline{\beta}^{(j)} \rangle_{\mathbb{R}^{b_j}} + \langle \underline{w}^{(j)}, \underline{\gamma} \rangle_{\mathbb{R}^c}.$$

### Variance des comptages

$$V \left[ \underline{N}_i^{(j)} \right] = \mu_j^\theta(i) \left\{ A_i^{(j)} + \xi_i^{(j)} \underline{a}_i^{(j)} \left( \underline{a}_i^{(j)} \right)^t \right\}.$$

Il s'en suit en particulier que, pour  $k \neq l$ ,

$$\rho \left( N_i^{(j,k)}, N_i^{(j,l)} \right) = \sqrt{\frac{\xi_i^{(j)} a_i^{(j,k)} a_i^{(j,l)}}{\left(1 + \xi_i^{(j)} a_i^{(j,k)}\right) \left(1 + \xi_i^{(j)} a_i^{(j,l)}\right)}}.$$

### Corrélation du processus des comptages

$$\rho \left( N_i^{(j,k)}, N_{i+m}^{(j,l)} \right) = \sqrt{\frac{\left(\xi_i^{(j)}\right)^2 \mu_j^\theta(i+m) a_i^{(j,k)} a_{i+m}^{(j,l)}}{\mu_j^\theta(i) \left(1 + \xi_i^{(j)} a_i^{(j,k)}\right) \left(1 + \xi_{i+m}^{(j)} a_{i+m}^{(j,l)}\right)}}.$$

On a de plus

$$\rho \left( N_{i+m}^{(j,k)}, \theta_i^{(j)} \right) = \sqrt{\frac{\xi_i^{(j)} \mu_j^\theta(i+m) a_{i+m}^{(j,k)}}{\mu_j^\theta(i) \left(1 + \xi_{i+m}^{(j)} a_{i+m}^{(j,k)}\right)}},$$

et encore

$$\text{cov} \left( \underline{N}_i^{(j)}, \theta_{i+m}^{(j)} \right) = \xi_i^{(j)} \mu_j^\theta(i+m) \underline{a}_i^{(j)}.$$

### 5.2.4 Les innovations du processus des comptages

Le processus d'innovation est défini comme suit :

$$\underline{J}_i^{(j)} := \underline{N}_i^{(j)} - E \left[ \underline{N}_i^{(j)} \mid \underline{\theta}^{(j)} \right] = \underline{N}_i^{(j)} - \theta_i^{(j)} \underline{a}_i^{(j)}.$$

On a que

$$V \begin{bmatrix} \underline{J}_1^{(j)} \\ \vdots \\ \underline{J}_{n_j}^{(j)} \end{bmatrix} = \begin{bmatrix} \theta_1^{(j)} A_1^{(j)} & \cdots & O \\ \vdots & \vdots & \vdots \\ O & \cdots & \theta_{n_j}^{(j)} A_{n_j}^{(j)} \end{bmatrix}$$

et que

$$\text{cov} \left( \underline{J}^{(j)}, \underline{\theta}^{(j)} \right) = O.$$

### 5.3 Le filtre de Kalman pour le modèle de JLCSS

De ce qui précède, on tire

$$\begin{aligned} \text{Equation d'état} & : \theta_i^{(j)} = b_i^{(j)} \theta_{i-1}^{(j)} + I_i^{(j)}, \\ \text{Observations} & : \underline{N}_i^{(j)} = \theta_i^{(j)} \underline{a}_i^{(j)} + \underline{J}_i^{(j)}, \end{aligned}$$

ce qui est déjà sous forme de filtre de Kalman. Dans la suite, on utilise la notation suivante :

$$\underline{X} | \underline{Y} \sim [\underline{m}_{X|Y}; C_{X|Y}],$$

$$\begin{aligned} \underline{m}_{X|Y} & = E[\underline{X}] + \Sigma_{X,Y} \Sigma_Y^{-1} \{\underline{Y} - E[\underline{Y}]\} \\ C_{X|Y} & = \Sigma_X - \Sigma_{X,Y} \Sigma_Y^{-1} \Sigma_{Y,X}. \end{aligned}$$

Alors, quand

$$\underline{N}_{[i]}^{(j)} = \begin{bmatrix} \underline{N}_1^{(j)} \\ \vdots \\ \underline{N}_i^{(j)} \end{bmatrix}, \quad [\mu_0^{(j)}; \gamma_0^{(j)}] = [g_j; g_j^2 \tau^2], \quad \text{et} \quad \theta_i^{(j)} | \underline{N}_{[i]}^{(j)} \sim [\mu_i^{(j)}; \gamma_i^{(j)}],$$

$$\begin{aligned} \underline{N}_i^{(j)} | \underline{N}_{[i-1]}^{(j)} & \sim [\underline{\nu}_i^{(j)}; \Gamma_i^{(j)}], \\ \underline{\nu}_i^{(j)} & = b_i^{(j)} \mu_{i-1}^{(j)} \underline{a}_i^{(j)}, \\ d_i^{(j)} & = b_i^{(j)} \left\{ \mu_j^\theta (i-1) \eta^2 + \gamma_{i-1}^{(j)} \right\}, \\ \Gamma_i^{(j)} & = b_i^{(j)} d_i^{(j)} \underline{a}_i^{(j)} \left( \underline{a}_i^{(j)} \right)^t + \mu_j^\theta (i) A_i^{(j)}, \end{aligned}$$

et

$$\begin{aligned} \theta_i^{(j)} | \underline{N}_{[i-1]}^{(j)} & \sim [b_i^{(j)} \mu_{i-1}^{(j)}; b_i^{(j)} d_i^{(j)}], \\ \mu_i^{(j)} & = b_i^{(j)} \mu_{i-1}^{(j)} + b_i^{(j)} d_i^{(j)} \langle \underline{a}_i^{(j)}, \left( \Gamma_i^{(j)} \right)^{-1} \left\{ \underline{N}_i^{(j)} - b_i^{(j)} \mu_{i-1}^{(j)} \underline{a}_i^{(j)} \right\} \rangle_{\mathbb{R}^{a_j}} \\ & = b_i^{(j)} \mu_{i-1}^{(j)} + \frac{d_i \sum_{k=1}^{a_j} \left( N_i^{(j,k)} - \nu_i^{(j,k)} \right)}{\mu_j^\theta (i-1) + d_i^{(j)} \sum_{k=1}^{a_j} a_i^{(j,k)}}, \\ \gamma_i^{(j)} & = b_i^{(j)} d_i^{(j)} - \left( b_i^{(j)} d_i^{(j)} \right)^2 \langle \underline{a}_i^{(j)}, \left( \Gamma_i^{(j)} \right)^{-1} \underline{a}_i^{(j)} \rangle_{\mathbb{R}^{a_j}} \\ & = \frac{b_i^{(j)} d_i^{(j)} \mu_j^\theta (i-1)}{\mu_j^\theta (i-1) + d_i^{(j)} \sum_{k=1}^{a_j} a_i^{(j,k)}}. \end{aligned}$$

Ces résultats, comme les suivants, s'obtiennent par application directe des faits, relatifs au calcul des moments conditionnels, énoncés dans l'annexe.

## 5.4 Le lissage de Kalman pour le modèle de JLCSS

On calcule "rétroactivement" (soit avec, successivement, les valeurs suivantes :  $i = n_j - 1, \dots, i = 0$ )

$$\theta_i^{(j)} \mid \underline{N}^{(j)} \sim [m_i^{(j)}; c_i^{(j)}]$$

où

$$p_i^{(j)} = \frac{\gamma_i^{(j)}}{d_{i+1}^{(j)}},$$

$$m_{n_j}^{(j)} = \mu_{n_j}^{(j)},$$

$$m_i^{(j)} = \mu_i^{(j)} + p_i^{(j)} \{m_{i+1}^{(j)} - b_{i+1}^{(j)} \mu_i^{(j)}\},$$

$$c_{n_j}^{(j)} = \gamma_{n_j}^{(j)},$$

$$c_i^{(j)} = \gamma_i^{(j)} + (p_i^{(j)})^2 \{c_{i+1}^{(j)} - b_{i+1}^{(j)} d_{i+1}^{(j)}\}.$$

Soit

$$\underline{m}^{(j)} = \begin{bmatrix} m_0^{(j)} \\ m_1^{(j)} \\ \vdots \\ m_{n_j}^{(j)} \end{bmatrix}.$$

La matrice de covariance des erreurs de lissage

$$E_{liss}^{(j)} = E \left[ \left( \underline{\theta}^{(j)} - \underline{m}^{(j)} \right) \left( \underline{\theta}^{(j)} - \underline{m}^{(j)} \right)^t \right]$$

a les éléments diagonaux

$$e_{i,i}^{liss,j} = c_i^{(j)}$$

et les éléments hors diagonale

$$e_{i,i+k}^{liss,j} = E \left[ \left( \theta_i^{(j)} - m_i^{(j)} \right) \left( \theta_{i+k}^{(j)} - m_{i+k}^{(j)} \right) \right] = c_{i+k}^{(j)} \prod_{l=1}^k \frac{\gamma_{i+l-1}^{(j)}}{d_{i+l}^{(j)}}.$$

## 5.5 Estimation des paramètres

Soient les vecteurs respectivement des comptages, du processus latent et des paramètres :

$$\underline{N} = \begin{bmatrix} \underline{N}^{(1)} \\ \vdots \\ \underline{N}^{(p)} \end{bmatrix} = \begin{bmatrix} \underline{N}_1^{(1)} \\ \vdots \\ \underline{N}_{n_1}^{(1)} \\ \hline \underline{N}_1^{(2)} \\ \vdots \\ \underline{N}_{n_2}^{(2)} \\ \hline \vdots \\ \hline \underline{N}_1^{(p)} \\ \vdots \\ \underline{N}_{n_p}^{(p)} \end{bmatrix}, \quad \underline{\theta}_0 = \begin{bmatrix} \theta_0^{(1)} \\ \vdots \\ \theta_0^{(p)} \end{bmatrix}, \quad \underline{\theta} = \begin{bmatrix} \underline{\theta}^{(1)} \\ \vdots \\ \underline{\theta}^{(p)} \end{bmatrix} = \begin{bmatrix} \theta_0^{(1)} \\ \vdots \\ \theta_{n_1}^{(1)} \\ \hline \theta_0^{(2)} \\ \vdots \\ \theta_{n_2}^{(2)} \\ \hline \vdots \\ \hline \theta_0^{(p)} \\ \vdots \\ \theta_{n_p}^{(p)} \end{bmatrix},$$

$$\underline{\alpha} = \begin{bmatrix} \underline{\alpha}^{(1)} \\ \vdots \\ \underline{\alpha}^{(p)} \end{bmatrix} = \begin{bmatrix} \alpha_1^{(1)} \\ \vdots \\ \alpha_n^{(1)} \\ \hline \vdots \\ \hline \alpha_1^{(p)} \\ \vdots \\ \alpha_n^{(p)} \end{bmatrix}, \quad \underline{\beta} = \begin{bmatrix} \beta^{(1)} \\ \vdots \\ \beta^{(p)} \end{bmatrix}.$$

### 5.5.1 La fonction de vraisemblance

La vraisemblance s'écrit

$$\mathcal{L}_{\underline{N}|\underline{\theta}}(\underline{\alpha}, \underline{\beta}, \underline{\gamma}, \tau^2, \eta^2) = \mathcal{L}_{\underline{N}|\underline{\theta}}(\underline{\alpha}) + \mathcal{L}_{\underline{\theta}|\underline{\theta}_0}(\underline{\beta}, \eta^2) + \mathcal{L}_{\underline{\theta}_0}(\underline{\gamma}, \tau^2),$$

avec, en posant  $\phi = \frac{1}{\eta^2}$  et  $\psi = \frac{1}{\tau^2}$ ,

$$\mathcal{L}_{\underline{N}|\underline{\theta}}(\underline{\alpha}) = \sum_{k=1}^m \sum_{j=1}^p \sum_{i=1}^{n_j} \left\{ N_i^{j,k} \langle \underline{u}_i^{(j)}, \underline{\alpha}_k^{(j)} \rangle_{\mathbb{R}^{a_j}} - \theta_i^{(j)} e^{\langle \underline{u}_i^{(j)}, \underline{\alpha}_k^{(j)} \rangle_{\mathbb{R}^{a_j}}} \right\},$$

$$\begin{aligned}
\mathcal{L}_{\underline{\theta}|\underline{\theta}_0}(\underline{\beta}, \eta^2) &= \sum_{j=1}^p \sum_{i=1}^{n_j} \\
&\{ \\
&\quad -\phi \theta_i^{(j)} e^{-\langle \Delta \underline{v}_i^{(j)}, \underline{\beta}^{(j)} \rangle_{\mathbb{R}^{b_j}}} - \phi \theta_{i-1}^{(j)} \langle \Delta \underline{v}_i^{(j)}, \underline{\beta}^{(j)} \rangle_{\mathbb{R}^{b_j}} \\
&\quad + \phi \theta_{i-1}^{(j)} \ln(\theta_i^{(j)}) + \phi \theta_{i-1}^{(j)} \ln(\phi) - \ln(\Gamma(\phi \theta_{i-1}^{(j)})) \\
&\}, \\
\mathcal{L}_{\underline{\theta}_0}(\underline{\gamma}, \tau^2) &= \sum_{j=1}^p \\
&\{ \\
&\quad -\psi \theta_0^{(j)} e^{-\langle \underline{w}^{(j)}, \underline{\gamma} \rangle_{\mathbb{R}^c}} - \psi \langle \underline{w}^{(j)}, \underline{\gamma} \rangle_{\mathbb{R}^c} \\
&\quad + \psi \ln(\theta_0^{(j)}) + \psi \ln(\psi) - \ln(\Gamma(\psi)) \\
&\}.
\end{aligned}$$

## 5.5.2 Les équations d'estimation

La recherche des paramètres produisant le maximum de la vraisemblance se fait en partie à partir des équations qui suivent. Tout d'abord,

$$\frac{\partial}{\partial \underline{\alpha}} \mathcal{L}_{N, \underline{\theta}}(\underline{\alpha}, \underline{\beta}, \underline{\gamma}, \tau^2, \eta^2) = \frac{\partial}{\partial \underline{\alpha}} \mathcal{L}_{N|\underline{\theta}}(\underline{\alpha}).$$

Si l'on pose

$$\underline{u}_1^{(1)} = \begin{bmatrix} u_1^{(1,1)} \\ \vdots \\ u_1^{(1,a)} \end{bmatrix}, \dots, \underline{u}_{n_1}^{(1)} = \begin{bmatrix} u_{n_1}^{(1,1)} \\ \vdots \\ u_{n_1}^{(1,a)} \end{bmatrix},$$

puis  $U_1 =$

$$\left[ \begin{array}{cccccc|cccccc|ccc|cccccc} \underline{u}_1^{(1)} & \underline{0} & \underline{0} & \cdots & \underline{0} & \underline{0} & \underline{u}_2^{(1)} & \underline{0} & \underline{0} & \cdots & \underline{0} & \underline{0} & \cdots & \underline{u}_{n_1}^{(1)} & \underline{0} & \underline{0} & \cdots & \underline{0} & \underline{0} \\ \underline{0} & \underline{u}_1^{(1)} & \underline{0} & \cdots & \underline{0} & \underline{0} & \underline{0} & \underline{u}_2^{(1)} & \underline{0} & \cdots & \underline{0} & \underline{0} & \cdots & \underline{0} & \underline{u}_{n_1}^{(1)} & \underline{0} & \cdots & \underline{0} & \underline{0} \\ \vdots & \vdots \\ \underline{0} & \underline{0} & \underline{0} & \cdots & \underline{0} & \underline{u}_1^{(1)} & \underline{0} & \underline{0} & \underline{0} & \cdots & \underline{0} & \underline{u}_2^{(1)} & \cdots & \underline{0} & \underline{0} & \underline{0} & \cdots & \underline{0} & \underline{u}_{n_1}^{(1)} \end{array} \right],$$

et que l'on définit de même  $U_2, \dots, U_p$ , puis

$$U = \begin{bmatrix} U_1 & O_{am, mn_2} & O_{am, mn_3} & \cdots & O_{am, mn_{p-1}} & O_{am, mn_p} \\ O_{am, mn_1} & U_2 & O_{am, mn_3} & \cdots & O_{am, mn_{p-1}} & O_{am, mn_p} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ O_{am, mn_1} & O_{am, mn_2} & O_{am, mn_3} & \cdots & O_{am, mn_{p-1}} & U_p \end{bmatrix},$$

si l'on forme encore la matrice

$$A_1 = \begin{bmatrix} \underline{a}_1^{(1)} & \underline{0} & \underline{0} & \cdots & \underline{0} & \underline{0} \\ \underline{0} & \underline{a}_2^{(1)} & \underline{0} & \cdots & \underline{0} & \underline{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \underline{0} & \underline{0} & \underline{0} & \cdots & \underline{0} & \underline{a}_{n_1}^{(1)} \end{bmatrix},$$

puis de manière similaire les matrices  $A_2, \dots, A_p$ , et enfin la matrice

$$A = \begin{bmatrix} A_1 & O_{mn_1, n_2} & O_{mn_1, n_2} & \cdots & O_{mn_1, n_{p-1}} & O_{mn_p, n_p} \\ O_{mn_2, n_1} & A_2 & O_{mn_2, n_3} & \cdots & O_{mn_2, n_{p-1}} & O_{mn_2, n_p} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ O_{mn_p, n_1} & O_{mn_p, n_2} & O_{mn_p, n_3} & \cdots & O_{mn_p, n_{p-1}} & A_p \end{bmatrix},$$

on a

$$\frac{\partial}{\partial \underline{\alpha}} \mathcal{L}_{\underline{N}, \underline{\theta}}(\underline{\alpha}, \underline{\beta}, \underline{\gamma}, \tau^2, \eta^2) = U \left[ \underline{N} - A \underline{\theta}_{[-0]} \right].$$

De manière semblable,

$$\frac{\partial}{\partial \underline{\beta}} \mathcal{L}_{\underline{N}, \underline{\theta}}(\underline{\alpha}, \underline{\beta}, \underline{\gamma}, \tau^2, \eta^2) = \frac{\partial}{\partial \underline{\beta}} \mathcal{L}_{\underline{\theta} | \underline{\theta}_0}(\underline{\beta}, \eta^2).$$

Alors on introduit les matrices suivantes, , pour  $1 \leq j \leq p$ , :  $B_j$  est la matrice diagonale dont les éléments diagonaux sont respectivement  $b_1^{(j)}, \dots, b_{n_j}^{(j)}$ ,

$$\tilde{B}_j = \begin{bmatrix} -b_1^{(j)} & 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & -b_2^{(j)} & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & b_3^{(j)} & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & -b_{n_j}^{(j)} & 1 \end{bmatrix},$$

$$\Delta V_j = \left[ \Delta \underline{v}_1^{(j)} \mid \cdots \mid \Delta \underline{v}_{n_j}^{(j)} \right].$$

Soit  $\underline{D}_j$  le vecteur de composantes respectives  $D_{j,1}, \dots, D_{j,b}$ . On a que :

$$\underline{D}_j = \Delta V_j B_j^{-1} \tilde{B}_j \underline{\theta}^{(j)}.$$

Soient maintenant les matrices suivantes :

$$B = \begin{bmatrix} B_1 & O_{n_1, n_1} & O_{n_1, n_2} & \cdots & O_{n_1, n_{p-1}} & O_{n_1, n_p} \\ O_{n_2, n_1} & B_2 & O_{n_2, n_3} & \cdots & O_{n_1, n_{p-1}} & O_{n_1, n_p} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ O_{n_p, n_1} & O_{n_p, n_2} & O_{n_p, n_3} & \cdots & O_{n_p, n_{p-1}} & B_p \end{bmatrix},$$

$\Delta V$  et  $\tilde{B}$  construites de manière analogue. **On a alors**

$$\frac{\partial}{\partial \underline{\beta}} \mathcal{L}_{\underline{N}, \underline{\theta}}(\underline{\alpha}, \underline{\beta}, \underline{\gamma}, \tau^2, \eta^2) = \Delta V B^{-1} \tilde{B} \underline{\theta}.$$

Finalemment

$$\frac{\partial}{\partial \underline{\gamma}} \mathcal{L}_{\underline{N}, \underline{\theta}}(\underline{\alpha}, \underline{\beta}, \underline{\gamma}, \tau^2, \eta^2) = \frac{\partial}{\partial \underline{\gamma}} \mathcal{L}_{\underline{\theta}_0}(\underline{\gamma}, \tau^2).$$

Si alors  $\underline{g}$  est le vecteur de composantes  $g_1, \dots, g_p$ ,  $G$  la matrice diagonale dont les éléments diagonaux sont les composantes de  $\underline{g}$ , et qu'enfin

$$W = \left[ \underline{w}^{(1)} \mid \dots \mid \underline{w}^{(p)} \right],$$

on obtient

$$\frac{\partial}{\partial \underline{\gamma}} \mathcal{L}_{\underline{N}, \underline{\theta}}(\underline{\alpha}, \underline{\beta}, \underline{\gamma}, \tau^2, \eta^2) = W G^{-1} [\underline{\theta}_0 - \underline{g}].$$

Si maintenant

$$\begin{aligned} \underline{\eta} &= \begin{bmatrix} \underline{\alpha} \\ \underline{\beta} \\ \underline{\gamma} \end{bmatrix}, \\ \underline{F}_1(\underline{\eta}) &= U \left[ \underline{N} - A \underline{\theta}_{[-0]} \right], \\ \underline{F}_2(\underline{\eta}) &= \Delta V B^{-1} \tilde{B} \underline{\theta}, \\ \underline{F}_3(\underline{\eta}) &= W G^{-1} [\underline{\theta}_0 - \underline{g}], \\ \underline{F}(\underline{\eta}) &= \begin{bmatrix} \underline{F}_1(\underline{\eta}) \\ \underline{F}_2(\underline{\eta}) \\ \underline{F}_3(\underline{\eta}) \end{bmatrix}, \end{aligned}$$

les **équations d'estimation** sont données par l'expression  $\underline{F}(\underline{\eta}) = \underline{0}$ .

### 5.5.3 L'algorithme EM (dit) de Kalman

Le terme  $\ln \left( \Gamma \left( \phi \theta_{i-1}^{(j)} \right) \right)$  de la vraisemblance fait qu'elle n'est celle d'une famille exponentielle que si  $\eta^2$  est connu. L'algorithme EM habituel comprend deux étapes :

#### le E-step

qui implique le calcul des espérances, conditionnellement aux comptages, des variables<sup>1</sup>

$$\theta_i^{(j)}, \ln \left( \theta_i^{(j)} \right), \theta_{i-1}^{(j)} \ln \left( \theta_i^{(j)} \right), \theta_{i-1}^{(j)} \Psi \left( \phi \theta_{i-1}^{(j)} \right);$$

---

1.  $\Psi$  est la fonction *digamma* définie par l'expression  $\Psi(z) = \frac{d}{dz} \ln(\Gamma(z))$ .

### le M-step

qui implique la solution des trois équations linéaires résumées en  $\underline{F}(\underline{\eta}) = \underline{0}$ .

Dans l'algorithme EM de Kalman (K-EM dans la suite), le E-step est remplacé par le lissage de Kalman.

Quand le K-EM converge, il converge vers la solution de l'équation  $\tilde{\underline{F}}(\underline{\eta}) = \underline{0}$  définie comme suit :

$$\begin{aligned}\tilde{\underline{F}}_1(\underline{\eta}) &= U \left[ \underline{N} - A\underline{m}_{[-0]} \right], \\ \tilde{\underline{F}}_2(\underline{\eta}) &= \Delta V B^{-1} \tilde{\underline{B}} \underline{m}, \\ \tilde{\underline{F}}_3(\underline{\eta}) &= W G^{-1} [\underline{m}_0 - \underline{g}], \\ \tilde{\underline{F}}(\underline{\eta}) &= \begin{bmatrix} \tilde{\underline{F}}_1(\underline{\eta}) \\ \tilde{\underline{F}}_2(\underline{\eta}) \\ \tilde{\underline{F}}_3(\underline{\eta}) \end{bmatrix},\end{aligned}$$

le processus  $\underline{m}$  étant issu de l'opération de lissage. Cette équation produit un estimateur sans biais.

### 5.5.4 Estimation des paramètres de dispersion

Parce que les paramètres de dispersion ne sont pas linéaires en  $\underline{\theta}$ , ils doivent être estimés séparément, et la méthode retenue est celle (dite) des "estimateurs de Pearson ajustés." Cette méthode corrige le biais introduit dans l'estimateur de Pearson des modèles linéaires généralisés, quand on remplace  $\underline{\theta}$  par  $\underline{m}_0$ , en utilisant l'estimateur de la variance issu du lissage de Kalman. Or, comme

$$E \left[ V \left[ \theta_i^{(j)} \mid \theta_{i-1}^{(j)} \right] \right] = E \left[ \left( \theta_i^{(j)} - b_i^{(j)} \theta_{i-1}^{(j)} \right)^2 \right] = b_i^{(j)} \mu_j^\theta(i) \eta^2,$$

on peut écrire

$$\begin{aligned}E \left[ \left( \theta_i^{(j)} - b_i^{(j)} \theta_{i-1}^{(j)} \right)^2 \right] &= E \left[ \left( m_i^{(j)} - b_i^{(j)} m_{i-1}^{(j)} \right)^2 \right] \\ &+ E \left[ \left\{ \theta_i^{(j)} - m_i^{(j)} - b_i^{(j)} \left( \theta_{i-1}^{(j)} - m_{i-1}^{(j)} \right) \right\}^2 \right] \\ &= E \left[ \left( m_i^{(j)} - b_i^{(j)} m_{i-1}^{(j)} \right)^2 \right] \\ &+ e_{i,i}^{liss,j} + \left( b_i^{(j)} \right)^2 e_{i-1,i-1}^{liss,j} - 2b_i^{(j)} e_{i,i-1}^{liss,j}.\end{aligned}$$

Par suite, si  $n := n_1 + \dots + n_p$ ,

$$\begin{aligned}\hat{\eta}^2 &:= \frac{1}{n} \sum_{j=1}^p \sum_{i=1}^{n_j} \frac{\left(m_i^{(j)} - b_i^{(j)} m_{i-1}^{(j)}\right)^2}{b_i^{(j)} \mu_j^\theta(i)} \\ &+ \frac{1}{n} \sum_{j=1}^p \sum_{i=1}^{n_j} \frac{e_{i,i}^{liss,j} + \left(b_i^{(j)}\right)^2 e_{i-1,i-1}^{liss,j} - 2b_i^{(j)} e_{i,i-1}^{liss,j}}{b_i^{(j)} \mu_j^\theta(i)}\end{aligned}$$

De manière semblable,

$$E \left[ \left( \theta_0^{(j)} - g_j \right)^2 \right] = g_j^2 \tau^2,$$

et

$$E \left[ \left( \theta_0^{(j)} - g_j \right)^2 \right] = E \left[ \left( \theta_0^{(j)} - m_0^{(j)} + m_0^{(j)} - g_j \right)^2 \right] = e_{0,0}^{liss,j} + E \left[ \left( m_0^{(j)} - g_j \right)^2 \right],$$

si bien que

$$\hat{\tau}^2 = \frac{1}{p} \sum_{j=1}^p \frac{\left(m_0^{(j)} - g_j\right)^2}{g_j^2} + \frac{1}{p} \sum_{j=1}^p \frac{e_{0,0}^{liss,j}}{g_j^2}.$$

Pour de petits échantillons,  $\hat{\eta}^2$  et  $\hat{\tau}^2$  présentent encore un biais qui peut être corrigé en remplaçant  $n$  par  $n - pb$  et  $p$  par  $p - c$ , où  $b_1 = \dots = b_p = b$ .

### 5.5.5 Un algorithme de type Newton

Quand le nombre de données est grand, l'algorithme EM est lent. C'est pourquoi la solution de  $\underline{\tilde{F}}(\underline{\eta}) = \underline{0}$  se cherche à l'aide d'un algorithme de type Newton.

Si l'on suppose connus les paramètres de dispersion, la matrice de stabilité de  $\underline{\eta}$ , soit  $S$ , est définie comme suit :

$$S(\underline{\eta}) = E_{\underline{\eta}} \left[ \frac{\partial \underline{\tilde{F}}}{\partial \underline{\eta}} \right].$$

L'algorithme s'exprime par la récursion suivante :

$$\underline{\eta}_{new} = \underline{\eta}_{old} - S^{-1}(\underline{\eta}_{old}) \underline{\tilde{F}}(\underline{\eta}_{old}).$$

La matrice de covariance des estimateurs est l'inverse de la matrice d'information de Godambe,  $\Gamma(\underline{\eta})$ , définie par les relations suivantes :

$$\begin{aligned}\Phi(\underline{\eta}) &= E_{\underline{\eta}} \left[ \underline{\tilde{F}}(\underline{\eta}) \underline{\tilde{F}}(\underline{\eta})^t \right], \\ \Gamma(\underline{\eta}) &= S^t(\underline{\eta}) \Phi^{-1}(\underline{\eta}) S(\underline{\eta}).\end{aligned}$$

**La matrice  $\Phi$**

On a :

$$\Phi_{i,j} = E_{\underline{\eta}} \left[ \tilde{\underline{F}}_i(\underline{\eta}) \tilde{\underline{F}}_j^t(\underline{\eta}) \right], \quad \Phi = \begin{bmatrix} \Phi_{1,1} & \Phi_{1,2} & \Phi_{1,3} \\ \Phi_{2,1} & \Phi_{2,2} & \Phi_{2,3} \\ \Phi_{3,1} & \Phi_{3,2} & \Phi_{3,3} \end{bmatrix},$$

$$\begin{aligned} \Phi_{1,1} &= U \Sigma_{\{\underline{N}-A\underline{m}_{[-0]}\}} U^t, \\ \Phi_{1,2} &= U \Sigma_{\{\underline{N}-A\underline{m}_{[-0]}, \tilde{\underline{B}}\underline{m}\}} B^{-1}(\Delta V)^t, \\ \Phi_{1,3} &= U \Sigma_{\{\underline{N}-A\underline{m}_{[-0]}, \underline{m}_0\}} G^{-1} W^t, \\ \Phi_{2,2} &= \Delta V B^{-1} \tilde{\underline{B}} \Sigma_{\{\underline{m}\}} \tilde{\underline{B}}^t B^{-1}(\Delta V)^t, \\ \Phi_{2,3} &= \Delta V B^{-1} \tilde{\underline{B}} \Sigma_{\{\underline{m}, \underline{m}_0\}} G^{-1} W^t, \\ \Phi_{3,3} &= W G^{-1} \Sigma_{\{\underline{m}_0\}} G^{-1} W^t. \end{aligned}$$

**La matrice  $S$**

On définit tout d'abord :

$$\begin{aligned} D_{\underline{\alpha}^{(j)}}^{\mu} \left[ i; \underline{\alpha}^{(j)}, \underline{\beta}^{(j)}, \underline{\gamma}, \tau^2, \eta^2 \right] &= D_{\underline{\alpha}^{(j)}}^{\mu} [i] = E_{\underline{\eta}} \left[ \frac{\partial \mu_i^{(j)}}{\partial \underline{\alpha}^{(j)}} \right], \\ D_{\underline{\alpha}^{(j)}}^m \left[ i; \underline{\alpha}^{(j)}, \underline{\beta}^{(j)}, \underline{\gamma}, \tau^2, \eta^2 \right] &= D_{\underline{\alpha}^{(j)}}^m [i] = E_{\underline{\eta}} \left[ \frac{\partial m_i^{(j)}}{\partial \underline{\alpha}^{(j)}} \right]. \end{aligned}$$

Si  $\frac{\partial \underline{a}_i^{(j)}}{\partial \underline{\alpha}^{(j)}}$

$$= \begin{bmatrix} \frac{\partial a_i^{(j,1)}}{\partial \alpha_1^{(j,1)}} & \cdots & \frac{\partial a_i^{(j,1)}}{\partial \alpha_1^{(j,a)}} & \left| \cdots \right| & \frac{\partial a_i^{(j,1)}}{\partial \alpha_k^{(j,1)}} & \cdots & \frac{\partial a_i^{(j,1)}}{\partial \alpha_k^{(j,a)}} & \left| \cdots \right| & \frac{\partial a_i^{(j,1)}}{\partial \alpha_m^{(j,1)}} & \cdots & \frac{\partial a_i^{(j,1)}}{\partial \alpha_m^{(j,a)}} \\ \vdots & \vdots & \vdots & \left| \cdots \right| & \vdots & \vdots & \vdots & \left| \cdots \right| & \vdots & \vdots & \vdots \\ \frac{\partial a_i^{(j,m)}}{\partial \alpha_1^{(j,1)}} & \cdots & \frac{\partial a_i^{(j,m)}}{\partial \alpha_1^{(j,a)}} & \left| \cdots \right| & \frac{\partial a_i^{(j,m)}}{\partial \alpha_k^{(j,1)}} & \cdots & \frac{\partial a_i^{(j,m)}}{\partial \alpha_k^{(j,a)}} & \left| \cdots \right| & \frac{\partial a_i^{(j,m)}}{\partial \alpha_m^{(j,1)}} & \cdots & \frac{\partial a_i^{(j,m)}}{\partial \alpha_m^{(j,a)}} \end{bmatrix}$$

$$= \begin{bmatrix} u_i^{(j,1)} a_i^{(j,1)} & \cdots & u_i^{(j,a)} a_i^{(j,1)} & \left| \cdots \right| & 0 & \cdots & 0 & \left| \cdots \right| & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \left| \cdots \right| & \vdots & \vdots & \vdots & \left| \cdots \right| & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & \left| \cdots \right| & 0 & \cdots & 0 & \left| \cdots \right| & u_i^{(j,1)} a_i^{(j,m)} & \cdots & u_i^{(j,a)} a_i^{(j,m)} \end{bmatrix}$$

$$= \left[ \begin{array}{c|ccc|c} a_i^{(j,1)} \{ \underline{u}_i^{(j)} \}^t & \cdots & \underline{0}^t & \cdots & \underline{0}^t \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ \underline{0}^t & \cdots & \underline{0}^t & \cdots & a_i^{(j,m)} \{ \underline{u}_i^{(j)} \}^t \end{array} \right],$$

on obtient alors la récursion :

$$\begin{aligned} D_{\underline{\alpha}^{(j)}}^\mu [0] &= \underline{0}^t, \\ D_{\underline{\alpha}^{(j)}}^\mu [i] &= \left\{ \frac{\gamma_i^{(j)}}{d_i^{(j)}} \right\} D_{\underline{\alpha}^{(j)}}^\mu [i-1] - b_i^{(j)} d_i^{(j)} \mu_j^\theta(i) \{ \underline{a}_i^{(j)} \}^t \{ \Gamma_i^{(j)} \}^{-1} \frac{\partial \underline{a}_i^{(j)}}{\partial \underline{\alpha}^{(j)}}. \end{aligned}$$

$\{ \underline{a}_i^{(j)} \}^t$  a les dimensions  $(1, m)$ ,  $\{ \Gamma_i^{(j)} \}^{-1}$  les dimensions  $(m, m)$ , et  $\frac{\partial \underline{a}_i^{(j)}}{\partial \underline{\alpha}^{(j)}}$  les dimensions  $(m, a \times m)$ .  $D_{\underline{\alpha}^{(j)}}^\mu$  a donc les dimensions  $(1, a \times m)$ . On a de manière semblable :

$$\begin{aligned} D_{\underline{\beta}^{(j)}}^\mu [0] &= \underline{0}^t, \\ D_{\underline{\beta}^{(j)}}^\mu [i] &= \left\{ \frac{\gamma_i^{(j)}}{b_i^{(j)} d_i^{(j)}} \right\} \left\{ \mu_j^\theta(i) \{ \underline{v}_i^{(j)} \}^t + b_i^{(j)} D_{\underline{\beta}^{(j)}}^\mu [i-1] \right\}, \end{aligned}$$

qui est une matrice de dimensions  $(1, b)$ , et

$$\begin{aligned} D_{\underline{\gamma}}^\mu [0] &= g_j \{ \underline{w}^{(j)} \}^t, \\ D_{\underline{\gamma}}^\mu [i] &= \left\{ \frac{\gamma_i^{(j)}}{d_i^{(j)}} \right\} D_{\underline{\gamma}}^\mu [i-1]. \end{aligned}$$

On a également les récursions suivantes :

$$\begin{aligned} D_{\underline{\alpha}^{(j)}}^m [n_j] &= D_{\underline{\alpha}^{(j)}}^\mu [n_j], \\ D_{\underline{\alpha}^{(j)}}^m [i] &= D_{\underline{\alpha}^{(j)}}^\mu [i] + \left\{ \frac{\gamma_i^{(j)}}{d_{i+1}^{(j)}} \right\} \left\{ D_{\underline{\alpha}^{(j)}}^m [i+1] - b_{i+1}^{(j)} D_{\underline{\alpha}^{(j)}}^\mu [i] \right\}, \end{aligned}$$

$$\begin{aligned} D_{\underline{\beta}^{(j)}}^m [n_j] &= D_{\underline{\beta}^{(j)}}^\mu [n_j], \\ D_{\underline{\beta}^{(j)}}^m [i] &= D_{\underline{\beta}^{(j)}}^\mu [i] + \left\{ \frac{\gamma_i^{(j)}}{d_{i+1}^{(j)}} \right\} \left\{ D_{\underline{\beta}^{(j)}}^m [i+1] - b_{i+1}^{(j)} D_{\underline{\beta}^{(j)}}^\mu [i] - \mu_j^\theta(i+1) \{ \underline{v}_{i+1}^{(j)} \}^t \right\}, \end{aligned}$$

$$D_{\underline{\gamma}}^m [n_j] = D_{\underline{\gamma}}^\mu [n_j],$$

$$D_{\underline{\gamma}}^m [i] = D_{\underline{\gamma}}^\mu [i] + \left\{ \frac{\gamma_i^{(j)}}{d_{i+1}^{(j)}} \right\} \left\{ D_{\underline{\gamma}}^m [i+1] - b_{i+1}^{(j)} D_{\underline{\gamma}}^\mu [i] \right\}.$$

La matrice  $S$  s'obtient alors à l'aide des expressions suivantes :

$$D_{\underline{\alpha}}^m = \begin{bmatrix} D_{\underline{\alpha}^{(1)}}^m \\ \vdots \\ D_{\underline{\alpha}^{(p)}}^m \end{bmatrix}, \quad D_{\underline{\beta}}^m = \begin{bmatrix} D_{\underline{\beta}^{(1)}}^m \\ \vdots \\ D_{\underline{\beta}^{(p)}}^m \end{bmatrix},$$

$$\underline{\eta}_1 := \underline{\alpha}, \quad \underline{\eta}_2 := \underline{\beta}, \quad \underline{\eta}_3 := \underline{\gamma}, \quad S_{i,j} = E_{\underline{\eta}} \left[ \frac{\partial \tilde{F}_i(\underline{\eta})}{\partial \underline{\eta}_j} \right].$$

Mais alors, en utilisant la formule de dérivation donnée en annexe,

$$\begin{aligned} \frac{\partial \tilde{F}_1}{\partial \underline{\alpha}} &= \frac{\partial}{\partial \underline{\alpha}} \left\{ U \left[ \underline{N} - A \underline{m}_{[-0]} \right] \right\} \\ &= -U \left\{ \frac{\partial}{\partial \underline{\alpha}} \left[ A \underline{m}_{[-0]} \right] \right\} \\ &= -U \left\{ \dot{A} + A \times \frac{\partial \underline{m}_{[-0]}}{\partial \underline{\alpha}} \right\}. \end{aligned}$$

et

$$\begin{aligned} S_{1,1} &= E_{\underline{\eta}} \left[ \frac{\partial \tilde{F}_1}{\partial \underline{\alpha}} \right] \\ &= -U \left\{ \dot{A}|_{E_{\underline{\eta}}[\underline{m}_{[-0]}}] + A \times D_{\underline{\alpha}}^m \right\} \\ &= -U \left\{ \dot{A}|_{\underline{\mu}_{[-0]}^\theta} + A \times D_{\underline{\alpha}}^m \right\}, \end{aligned}$$

puis, successivement,

$$\begin{aligned} \frac{\partial \tilde{F}_1}{\partial \underline{\beta}} &= \frac{\partial}{\partial \underline{\beta}} \left\{ U \left[ \underline{N} - A \underline{m}_{[-0]} \right] \right\} \\ &= -U \left\{ \frac{\partial}{\partial \underline{\beta}} \left[ A \underline{m}_{[-0]} \right] \right\} \\ &= -U \left\{ O + A \times \frac{\partial \underline{m}_{[-0]}}{\partial \underline{\beta}} \right\}, \end{aligned}$$

et

$$\begin{aligned} S_{1,2} &= E_{\underline{\eta}} \left[ \frac{\partial \tilde{F}_1}{\partial \underline{\beta}} \right] \\ &= -U \times A \times D_{\underline{\beta}}^m, \end{aligned}$$

$$\begin{aligned} S_{1,3} &= E_{\underline{\eta}} \left[ \frac{\partial \tilde{F}_1}{\partial \underline{\gamma}} \right] \\ &= -U \times A \times D_{\underline{\gamma}}^m. \end{aligned}$$

Ensuite,

$$\begin{aligned} \frac{\partial \tilde{F}_2}{\partial \underline{\beta}} &= \frac{\partial}{\partial \underline{\beta}} \left\{ \Delta V B^{-1} \tilde{B} \underline{m} \right\} \\ &= \Delta V B^{-1} \left\{ \frac{\partial}{\partial \underline{\beta}} [\tilde{B} \underline{m}] \right\} \\ &= \Delta V B^{-1} \left\{ \dot{\tilde{B}} + \tilde{B} \times \frac{\partial \underline{m}}{\partial \underline{\beta}} \right\}, \end{aligned}$$

et

$$\begin{aligned} S_{2,2} &= E_{\underline{\eta}} \left[ \frac{\partial \tilde{F}_2}{\partial \underline{\beta}} \right] \\ &= \Delta V B^{-1} \left\{ \dot{\tilde{B}}|_{E_{\underline{\eta}}[\underline{m}]} + \tilde{B} \times D_{\underline{\beta}}^m \right\} \\ &= \Delta V B^{-1} \left\{ \dot{\tilde{B}}|_{\underline{\mu}^\theta} + \tilde{B} \times D_{\underline{\beta}}^m \right\}. \end{aligned}$$

De même,

$$\begin{aligned} S_{2,1} &= E_{\underline{\eta}} \left[ \frac{\partial \tilde{F}_2}{\partial \underline{\alpha}} \right] \\ &= \Delta V B^{-1} \tilde{B} \times D_{\underline{\alpha}}^m, \end{aligned}$$

et

$$S_{2,3} = E_{\underline{\eta}} \left[ \frac{\partial \tilde{F}_2}{\partial \underline{\gamma}} \right]$$

$$= \Delta V B^{-1} \tilde{B} \times D_{\underline{\gamma}}^m.$$

Finalement,

$$\begin{aligned} \frac{\partial \tilde{F}_3}{\partial \underline{\gamma}} &= \frac{\partial}{\partial \underline{\gamma}} \{ W G^{-1} [m_0 - \underline{g}] \} \\ &= W G^{-1} \left\{ \frac{\partial m_0}{\partial \underline{\gamma}} - \frac{\partial \underline{g}}{\partial \underline{\gamma}} \right\}, \end{aligned}$$

et

$$\begin{aligned} S_{3,3} &= E_{\underline{\eta}} \left[ \frac{\partial \tilde{F}_3}{\partial \underline{\gamma}} \right] \\ &= W G^{-1} \left\{ D_{\underline{\gamma}}^{m_0} - \frac{\partial \underline{g}}{\partial \underline{\gamma}} \right\}, \end{aligned}$$

$$\begin{aligned} S_{3,1} &= E_{\underline{\eta}} \left[ \frac{\partial \tilde{F}_3}{\partial \underline{\alpha}} \right] \\ &= W G^{-1} \times D_{\underline{\alpha}}^{m_0}, \end{aligned}$$

$$\begin{aligned} S_{3,2} &= E_{\underline{\eta}} \left[ \frac{\partial \tilde{F}_3}{\partial \underline{\beta}} \right] \\ &= W G^{-1} \times D_{\underline{\beta}}^{m_0}. \end{aligned}$$

## 5.6 Diagnostics

L'instrument de diagnostic est le résidu. Dans ce qui suit un résidu est toujours standardisé.

### 5.6.1 Définition et propriétés des résidus

La méthode reconnaît divers types de résidus d'usages divers. Ce sont les résidus

1. conditionnels := valeurs prédites des innovations

origine : filtrage	$r_i^{(j,k)} := \frac{N_i^{(j,k)} - \nu_i^{(j,k)}}{\sqrt{\Gamma_i^{(j)}[k,k]}}$ $r_i^{(j)} := \frac{\mu_i^{(j)} - b_i^{(j)} \mu_{i-1}^{(j)}}{\sqrt{b_i^{(j)} d_i^{(j)} - \gamma_i^{(j)}}}$
origine : lissage	$\rho_i^{(j,k)} := \frac{N_i^{(j,k)} - m_i^{(j)} a_i^{(j,k)}}{\sqrt{a_i^{(j,k)} \mu_j^\theta(i) - \{a_i^{(j,k)}\}^2 c_i^{(j)}}}$ $\rho_i^{(j)} := \frac{m_i^{(j)} - b_i^{(j)} m_{i-1}^{(j)}}{\sqrt{b_i^{(j)} \mu_j^\theta(i) \eta^2 - c_i^{(j)} - \{b_i^{(j)}\}^2 c_{i-1}^{(j)} + 2b_i^{(j)} e_{i,i-1}^{liss,j}}}$

Propriétés :

- (a)  $V[\underline{r}_i^{(j)}] = \Gamma_i^{(j)}$ .
- (b)  $\text{cov}(\underline{r}_i^{(j)}, \underline{r}_{i+k}^{(j)}) = O$ .
- (c)  $V[r_i^{(j)}] = b_i^{(j)} d_i^{(j)} - \gamma_i^{(j)}$ .
- (d)  $\text{cov}(r_i^{(j)}, r_{i+k}^{(j)}) = 0$ .
- (e)  $V[\underline{\rho}_i^{(j)}] = \mu_j^\theta(i) A_i^{(j)} - c_i^{(j)} \underline{a}_i^{(j)} (\underline{a}_i^{(j)})^t$ .
- (f)  $\text{cov}(\underline{\rho}_i^{(j)}, \underline{\rho}_{i+k}^{(j)}) = -e_{i,i+k}^{liss,j} \underline{a}_i^{(j)} (\underline{a}_{i+k}^{(j)})^t$ .
- (g)  $\text{cov}(\rho_i^{(j)}, \rho_{i+k}^{(j)}) =$ 

$$\begin{cases} b_i^{(j)} \mu_j^\theta(i) \eta^2 - c_i^{(j)} - \{b_i^{(j)}\}^2 c_{i-1}^{(j)} + 2b_i^{(j)} e_{i,i-1}^{liss,j} & \text{quand } k = 0, \\ -e_{i,i+k}^{liss,j} + b_i^{(j)} e_{i-1,i+k}^{liss,j} + b_{i+k}^{(j)} e_{i,i+k-1}^{liss,j} - b_i^{(j)} b_{i+k}^{(j)} e_{i-1,i+k-1}^{liss,j} & \text{quand } k > 0. \end{cases}$$

2. marginaux := différence entre valeurs observées et valeurs escomptées

origine : observations	$\delta_i^{(j)} := \frac{N_i^{(j)} - \mu_j^\theta(i) a_i^{(j,k)}}{\sqrt{a_i^{(j,k)} \mu_j^\theta(i) [1 + a_i^{(j,k)} \zeta_i^{(j)} \eta^2]}}$
origine : filtrage	$\widehat{\delta}_i^{(j)} := \frac{\mu_i^{(j)} - \mu_j^\theta(i)}{\sqrt{\mu_j^\theta(i) \zeta_i^{(j)} \eta^2 - \gamma_i^{(j)}}}$
origine : lissage	$\widetilde{\delta}_i^{(j)} := \frac{m_i^{(j)} - \mu_j^\theta(i)}{\sqrt{\mu_j^\theta(i) \zeta_i^{(j)} \eta^2 - c_i^{(j)}}}$

*Propriétés :*

1.  $V \left[ \underline{\delta}_i^{(j)} \right] = \mu_j^\theta(i) \left[ A_i^{(j)} + \eta^2 \zeta_i^{(j)} \underline{a}_i^{(j)} \left( \underline{a}_i^{(j)} \right)^t \right].$
2.  $\text{cov} \left( \underline{\delta}_i^{(j)}, \underline{\delta}_{i+k}^{(j)} \right) = \eta^2 \zeta_i^{(j)} \mu_j^\theta(i+k) \underline{a}_i^{(j)} \left( \underline{a}_{i+k}^{(j)} \right)^t.$
3.  $V \left[ \widehat{\delta}_i^{(j)} \right] = \mu_j^\theta(i) \zeta_i^{(j)} \eta^2 - \gamma_i^{(j)}.$
4.  $\text{cov} \left( \widehat{\delta}_i^{(j)}, \widehat{\delta}_{i+k}^{(j)} \right) = \eta^2 \zeta_i^{(j)} \mu_j^\theta(i+k) - b_{i+1}^{(j)} \cdots b_{i+k}^{(j)} \gamma_i^{(j)}.$
5.  $V \left[ \widetilde{\delta}_i^{(j)} \right] = \mu_j^\theta(i) \zeta_i^{(j)} \eta^2 - c_i^{(j)}.$
6.  $\text{cov} \left( \widetilde{\delta}_i^{(j)}, \widetilde{\delta}_{i+k}^{(j)} \right) = \eta^2 \zeta_i^{(j)} \mu_j^\theta(i+k) - e_{i,i+k}^{liss,j}.$

*Remarques :*

1. Les résidus conditionnels doivent servir à contrôler le détail de la structure du modèle, alors que les résidus marginaux doivent servir à contrôler les hypothèses concernant les répartitions.
2. Les propriétés recensées des résidus ont été obtenues en considérant que les valeurs des paramètres sont connues : leur comportement quand ces paramètres sont remplacés par leur estimation n'est pas connu.
3. Les moments des résidus marginaux  $\delta_i^{(j,k)}$  sont conditionnels aux valeurs  $\theta_0^{(j)}$ .
4. L'indice  $j$  étant fixé, la suite des  $m_i^{(j)}$  et celle des  $\tilde{\delta}_i^{(j)}$  forment des processus de Markov d'ordre 1.
5. Il n'y a pas de corrélation entre les éléments  $\underline{r}_i^{(j)}$  et  $b_i^{(j)} \mu_{i-1}^{(j)} \underline{a}_i^{(j)}$  d'une part, et les éléments  $r_i^{(j)}$  et  $b_i^{(j)} \mu_{i-1}^{(j)}$  d'autre part.
6. Les variances des résidus du filtrage peuvent être grandes, en particulier pour le début des séries. Ce n'est pas le cas des résidus du lissage qui ont des variances plus petites que celles des résidus du filtrage.
7. La présence de comptages faibles a une influence significative sur l'aspect des graphiques relatifs aux résidus, ce qui rend leur interprétation difficile.

## 5.6.2 Utilisation des résidus

### Evaluation de la structure de corrélation du modèle

Les résidus utiles sont les résidus du filtrage car ils ne sont pas corrélés.

**Faire :**

1. Contrôle de la structure markovienne du processus latent : graphique

$$(x, y) \leftarrow \left( r_{i-1}^{(j)}, r_i^{(j)} \right).$$

2. Examen de la fonction d'autocorrélation de la suite en  $i, r_i^{(j)}$ ,  $j$  étant fixé.
3. Examen de la fonction d'autocorrélation partielle de la suite en  $i, \tilde{\delta}_i^{(j)}$ ,  $j$  étant fixé.
4. Contrôle de l'indépendance conditionnelle dans le temps des comptages : graphiques (pour chaque  $k$ )

$$(x, y) \leftarrow \left( r_{i-1}^{(j,k)}, r_i^{(j,k)} \right).$$

5. Examen de la fonction d'autocorrélation de la suite en  $i, r_i^{(j,k)}$ ,  $j$  étant fixé (pour chaque  $k$ ).

6. Contrôle de l'indépendance conditionnelle des catégories relativement au processus latent : examen, pour  $j$  fixé, de la matrice

$$\frac{1}{n_j} \sum_{i=1}^{n_j} \left( \Gamma_i^{(j)} \right)^{-\frac{1}{2}} \underline{r}_i^{(j)} \left( \underline{r}_i^{(j)} \right)^t \left\{ \left( \Gamma_i^{(j)} \right)^{-\frac{1}{2}} \right\}^t$$

pour laquelle on a

$$E \left[ \frac{1}{n_j} \sum_{i=1}^{n_j} \left( \Gamma_i^{(j)} \right)^{-\frac{1}{2}} \underline{r}_i^{(j)} \left( \underline{r}_i^{(j)} \right)^t \left\{ \left( \Gamma_i^{(j)} \right)^{-\frac{1}{2}} \right\}^t \right] = I_m.$$

Les éléments hors diagonale de cette matrice empirique ont des écart-type valant  $\frac{1}{\sqrt{n_j}}$ .

7. Contrôle de la corrélation entre les éléments  $r_i^{(j)}$  et  $\underline{r}_i^{(j)}$ .

### Evaluation des hypothèses relatives aux répartitions

On utilise aussi bien les résidus du filtrage que ceux du lissage.

Faire:

1. Contrôle de la validité de l'hypothèse relative à la loi Gamma du processus latent (fonction de variance) : graphique

$$(x, y) \longleftarrow \left( \ln \left( b_i^{(j)} \mu_{i-1}^{(j)} \right), r_i^{(j)} \right).$$

2. Contrôle de la validité de l'hypothèse relative à la loi de Poisson (fonction de variance) : graphique

$$(x, y) \longleftarrow \left( \ln \left( a_i^{(j,k)} b_i^{(j)} \mu_{i-1}^{(j)} \right), r_i^{(j,k)} \right).$$

3. Contrôle de la validité de l'hypothèse relative à la loi Gamma des différences entre "situations" (fonction de variance :  $g_j = \mu_j^\theta(0)$ ) : graphique

$$(x, y) \longleftarrow \left( g_j, \tilde{\delta}_0^{(j)} \right).$$

### Evaluation de la pertinence de la structure donnée aux variables explicatives

Faire:

1. Contrôle de linéarité : graphique

$$(x, y) \longleftarrow \left( \text{variable explicative}, \rho_i^{j,k} \right).$$

*N.B.* On utilise donc les résidus issus du lissage.

2. Contrôle du lien logarithmique : graphique

$$(x, y) \leftarrow \left( \ln \left\{ m_i^{(j)} a_i^{(j,k)} \right\}, \ln \left\{ N_i^{(j,k)} \right\} \right).$$

Le graphique devrait être proche d'une horizontale : toute autre forme indique un problème dans la définition du lien.

3. Contrôle du choix "long terme/court terme" pour les variables explicatives :

Une variable explicative de court terme, définie de manière inadéquate comme variable de long terme, devrait "montrer" une association avec les résidus "Poissoniens," alors qu'une variable de long terme, définie de manière inadéquate comme variable de court terme, devrait "montrer" une association avec les résidus "gamma."

Si le graphique

$$(x, y) \leftarrow \left( u_i^{(j,k)} - u_{i-1}^{(j,k)}, r_i^j \right)$$

montre une association, il y a lieu de penser que la variable  $u_i^{(j,k)}$  est une variable de long terme. De même. si le graphique

$$(x, y) \leftarrow \left( v_i^{(j,k)} - u_{i-1}^{(j,k)}, r_i^{j,k} \right)$$

montre une association, il y a lieu de penser que la variable  $v_i^{(j,k)}$  est une variable de court terme.

## Chapitre 6

# Résultat des analyses par la méthode de Zeger

Un examen des comptages faits par quelques stations (Rolle (001), Monteceneri (137), Gotthardtunnel (150)) met clairement en évidence

- l’existence des tendances (par exemple : une tendance faible à la décroissance, suivie d’une légère remontée),
- des effets journaliers, hebdomadaires et annuels,
- mais aussi des problèmes de “régularité :” les années n’ont pas le même nombre de jours, les types de jours ne se suivent pas (on peut avoir un dimanche suivi d’un mardi<sup>1</sup>),
- et de longues plages de “valeurs manquantes :” on a trouvé ainsi une série (Seelisbergtunnel (156)) où il y a tout d’abord 122 jours consécutifs pour lesquels il y a, pour chaque heure, un unique véhicule recensé, et ensuite 30 jours consécutifs, quelques années plus tard, pour lesquels le même phénomène se répète.<sup>2</sup>

### 6.1 Le modèle : description globale basée sur l’ensemble des données

On tient compte de l’absence de “régularité” des données en ajustant un “modèle saturé :”<sup>3</sup> on a un paramètre pour chaque heure de comptage, chaque type de jour et chaque année. Cela donne :<sup>4</sup>

---

1. Cela signifie souvent que le lundi précédant le mardi en question a été un jour férié qui a été, en tant que tel, assimilé à un dimanche.

2. Cela pourrait correspondre à des fermetures du tunnel.

3. C’est aussi un modèle qui tient compte de la nature discrète des données : comment justifier le choix d’une courbe continue pour décrire les totaux horaires sur 24 heures ?

4.  $C_0$  représente le “niveau général” des comptages. On a “omis” les paramètres  $\alpha_1$ ,  $\beta_1$  et  $A_{90}$  pour pouvoir obtenir une matrice de *design* qui possède une inverse. Le “niveau général” est donc celui qui correspond à la première heure du premier lundi de 1990.

$$\begin{aligned}
C(i, j, k) &= C_0 \\
&+ \alpha_2 H_2(k) + \cdots + \alpha_{24} H_{24}(k) \\
&+ \beta_2 J_2(j) + \cdots + \beta_7 J_7(j) \\
&+ \gamma_{91} A_1(i) + \cdots + \gamma_{98} A_{98}(i),
\end{aligned}$$

où  $C(i, j, k)$  est le nombre de véhicules dénombré au cours de l'heure numéro  $i$ , du jour de type  $j$  de l'année numéro  $k$ , et

$$H_l(k) = \begin{cases} 0 & \text{si } k \neq l \\ 1 & \text{si } k = l \end{cases}, \quad J_l(j) = \begin{cases} 0 & \text{si } j \neq l \\ 1 & \text{si } j = l \end{cases}, \quad A_l(k) = \begin{cases} 0 & \text{si } i \neq l \\ 1 & \text{si } i = l \end{cases}.$$

Ce modèle devrait donner le meilleur ajustement possible, puisqu'on paramétrise tout ce qui est susceptible de l'être.<sup>5</sup> Par ailleurs, il permet d'éviter la difficulté de modélisation de la "trajectoire journalière des comptages" qui est plutôt complexe. Finalement, un des aspects utiles de cette approche est que la matrice  $X_n$  est non singulière, et, en principe, bien conditionnée.

La matrice de covariance de  $\underline{Y}_n$  est une matrice de dimensions  $(n, n)$ , ce qui signifie qu'il faut inverser une matrice (carrée) de dimension 78'888. La méthode simplifiée semble plus accessible, bien qu'il faille aussi procéder à une multiplication par une matrice (carrée) de dimension 78'888. Une simplification supplémentaire (aux conséquences inconnues) consiste à utiliser la matrice de covariance  $\Sigma_{\tilde{g}} = \Sigma_0^{-1}$ , ce qui revient à poser  $\Sigma_{\underline{Y}_n} = \tilde{\Sigma}_{\underline{Y}_n}$ .

Le problème avec cette approche est celui de la taille des matrices à manipuler. Seule l'expérience peut donner des indications sur sa faisabilité. On a commencé par utiliser MATLAB.

## 6.2 Mise en oeuvre de la méthode de Zeger

### 6.2.1 Valeurs initiales

Les valeurs initiales servent d'"entrée" à l'estimation, d'une part, de  $\tilde{\theta}_0$  et, d'autre part, de  $\sigma_B^2$ , et de  $\Gamma_B$ . Ce sont les valeurs  $\hat{\lambda}_1^{(0)}, \dots, \hat{\lambda}_n^{(0)}$ .

La méthode habituellement recommandée pour obtenir les valeurs initiales consiste à ajuster un modèle "log-linéaire" qui "ignore" les corrélations des comptages successifs.<sup>6</sup> L'alternative la plus simple est de calculer la moyenne, par année, sur les "couples" (type de jour, type d'heure), ce qui donne des moyennes prises sur environ 50 observations (on suppose donc que les semaines se répètent identiques à elles-mêmes, dans l'année). Cela donne des valeurs initiales assez proches de celles que produit l'ajustement "log-linéaire."<sup>7</sup> Une autre alternative consiste à obtenir les valeurs initiales à l'aide d'une procédure de lissage

5. Cela peut changer si l'on sait introduire le contexte ainsi qu'il l'a été mentionné dans l'introduction.

6. Mais on verra plus loin que cette option peut ne pas exister!

7. Et donc, c'est aussi un option qui peut faire défaut.

robuste,<sup>8</sup> ce qui donne des valeurs initiales assez différentes des précédentes. On peut finalement utiliser le maximum de vraisemblance selon la logique qui prévaut pour l'ajustement "log-linéaire." Cette dernière approche semble exiger (on a fait des essais) un temps de calcul très long (plusieurs heures) parce qu'il y faut un très grand nombre d'itérations.

Pour les valeurs initiales de  $\underline{\theta}$ , soit  $\tilde{\underline{\theta}}_0$ , on peut procéder comme suit. On interprète la relation

$$\ln [\hat{\lambda}_i^{(0)}] = \langle \underline{\theta}, \underline{X}_i \rangle$$

comme une régression usuelle qu'on ajuste par moindres carrés. La matrice  $X_n$  doit alors être non seulement non singulière, mais encore bien conditionnée. Le recours aux pseudo-inverses ne semble avoir aucun effet bénéfique. On n'a pas essayé de techniques de régression robuste.

Les valeurs  $\hat{\lambda}_1^{(0)}, \dots, \hat{\lambda}_n^{(0)}$  permettent aussi de calculer, en tenant compte des comptages réalisés  $Y_1, \dots, Y_n$ , les valeurs

$$\hat{\sigma}_B^2 [0] \text{ et } \hat{\gamma}_B^{(p)}(i) [0].$$

### Effet des valeurs initiales sur la convergence de l'algorithme utilisé

On a constaté les faits suivants relatifs à l'effet des valeurs initiales sur la convergence-divergence de l'algorithme utilisé :

Lieu	Valeurs initiales :		
	Lissage	Moyenne	Log-Linéaire
Rolle	converge	converge	converge
Monteceneri	converge	diverge	diverge
Gotthardtunnel	converge	diverge	diverge

Quand il y a convergence, celle-ci a lieu essentiellement vers les mêmes valeurs. Une explication pour les convergences et divergences constatées pourrait être que le lissage donne des valeurs initiales plus proches des "vraies" valeurs que les deux autres méthodes qui "conduiraient" l'algorithme vers des plages de divergence. Le fait que les deux méthodes "moyenne" et "log-linéaire" convergent ou divergent simultanément est "rassurant" puisqu'elles produisent des valeurs initiales "proches."

### 6.2.2 Approximation du processus stationnaire sous-jacent par un processus autorégressif

Pour essayer de minimiser les problèmes de calcul, on peut commencer par utiliser, comme processus stationnaire, une autorégression de la forme

$$B_n = \alpha_1 B_{n-1} + \alpha_2 B_{n-2} + \dots + \alpha_p B_{n-p} + W_n.$$

---

8. MINITAB, Resistant smooth 4253H

### La méthode d'Anderson pour le calcul des coefficients $\alpha_i$ , $1 \leq i \leq p$

Si le processus considéré est d'ordre  $p$ , et les corrélations estimées à l'étape précédente notées  $\hat{\gamma}_B(k)$ , les équations d'estimation sont :

1.  $p = 2$  :

$$\begin{aligned}\hat{\alpha}_1^{(2)} &= \frac{1 - \hat{\gamma}_B(2)}{1 - \hat{\gamma}_B^2(1)} \hat{\gamma}_B(1) \\ \hat{\alpha}_2^{(2)} &= \frac{\hat{\gamma}_B(2) - \hat{\gamma}_B^2(1)}{1 - \hat{\gamma}_B^2(1)}\end{aligned}$$

2.  $p = 3$  :

$$\begin{aligned}\hat{\alpha}_3^{(3)} &= \frac{\hat{\gamma}_B(3) - \hat{\gamma}_B(2)\hat{\alpha}_1^{(2)} - \hat{\gamma}_B(1)\hat{\alpha}_2^{(2)}}{1 - \hat{\gamma}_B(2)\hat{\alpha}_2^{(2)} - \hat{\gamma}_B(1)\hat{\alpha}_1^{(2)}} \\ \hat{\alpha}_1^{(3)} &= \hat{\alpha}_1^{(2)} - \hat{\alpha}_2^{(2)}\hat{\alpha}_3^{(3)} \\ \hat{\alpha}_2^{(3)} &= \hat{\alpha}_2^{(2)} - \hat{\alpha}_1^{(2)}\hat{\alpha}_3^{(3)}\end{aligned}$$

3.  $p = 4$  :

$$\begin{aligned}\hat{\alpha}_4^{(4)} &= \frac{\hat{\gamma}_B(4) - \hat{\gamma}_B(3)\hat{\alpha}_1^{(3)} - \hat{\gamma}_B(2)\hat{\alpha}_2^{(3)} - \hat{\gamma}_B(1)\hat{\alpha}_3^{(3)}}{1 - \hat{\gamma}_B(3)\hat{\alpha}_3^{(3)} - \hat{\gamma}_B(2)\hat{\alpha}_2^{(3)} - \hat{\gamma}_B(1)\hat{\alpha}_1^{(3)}} \\ \hat{\alpha}_1^{(4)} &= \hat{\alpha}_1^{(3)} - \hat{\alpha}_3^{(3)}\hat{\alpha}_4^{(4)} \\ \hat{\alpha}_2^{(4)} &= \hat{\alpha}_2^{(3)} - \hat{\alpha}_2^{(3)}\hat{\alpha}_4^{(4)} \\ \hat{\alpha}_3^{(4)} &= \hat{\alpha}_3^{(3)} - \hat{\alpha}_1^{(3)}\hat{\alpha}_4^{(4)}\end{aligned}$$

4. ...

### L'ajustement par itération

Il faut décider de la valeur de  $p$ , l'ordre de l'autorégression. A cette fin, on peut utiliser le fait, qu'asymptotiquement,  $\hat{\underline{\alpha}}^{(p)}$ , le vecteur de composantes

$$\hat{\alpha}_1^{(p)}, \dots, \hat{\alpha}_p^{(p)}$$

est tel que

$$\underline{\alpha}^{(p)} \sim N \left( \underline{\alpha}_p, \frac{1}{n} \times \frac{\sigma_N^2}{\sigma_B^2} \times \left\{ \Sigma_B^{(p)} \right\}^{-1} \right),$$

où  $\underline{\alpha}_p$  est le vecteur des “vraies” valeurs des paramètres du processus autorégressif ajusté, et  $\sigma_N^2$  est la variance du bruit blanc qui définit  $B$ . Un coefficient  $\alpha_i$  est “nul” si

$$\frac{\hat{\alpha}_i^{(p)}}{\sqrt{V \left[ \hat{\alpha}_i^{(p)} \right]}}$$

est “nul” pour une loi de Student à  $(n - p)$  degrés de liberté.  $\sigma_N^2$  est estimé d’habitude à l’aide de l’expression

$$\hat{\sigma}_N^2 = \hat{\sigma}_B^2 \left\{ 1 - \langle \hat{\underline{\alpha}}_B^{(p)}, \hat{\underline{\alpha}}^{(p)} \rangle \right\},$$

où

$$\begin{aligned} \hat{\underline{\alpha}}_B^{(p)} &:= \text{vecteur de dimension } p \text{ dont les composantes sont } \hat{\gamma}_B(i), 1 \leq i \leq p, \\ \langle \underline{u}_p, \underline{v}_p \rangle &:= \text{produit scalaire dans } \mathbb{R}^p. \end{aligned}$$

De ce qui précède, on tire aussi la valeur initiale  $L_{n,p}^{(0)}$  de  $L_{n,p}$ .

### Contrôle de la stationnarité

On le fait à l’aide du test dit de Schur qui lui-même repose sur le théorème suivant :

Soient  $\alpha_1, \dots, \alpha_p$ , des constantes,  $U$  et  $V$  les matrices “bas-triangulaires” de Toeplitz, de dimensions  $(p, p)$ , que l’on obtient à partir, respectivement, des vecteurs

$$\begin{bmatrix} 1 \\ \alpha_1 \\ \vdots \\ \alpha_{p-1} \end{bmatrix} \text{ et } \begin{bmatrix} \alpha_p \\ \vdots \\ \alpha_1 \end{bmatrix}.$$

Alors,

1. Les zéros de  $g(z) = \sum_{i=0}^p \alpha_i z^i$  ont tous un module plus grand que 1 si et seulement si la matrice  $S = UU^t - VV^t$  est positive définie.
2. Quand  $S$  est positive définie, la matrice  $\Gamma_p = \sigma^2 S^{-1}$  est la matrice de covariance d’un processus AR  $(p, \underline{\alpha}, \sigma^2)$ .
3.  $\Gamma_p$  est positive définie si et seulement si les autocorrélations partielles aux retards allant de 1 à  $p$  ont toutes une valeur absolue inférieure à 1.

Si le coefficient d’autocorrélation partielle à l’ordre  $k$  est noté  $\lambda_k(k)$ , on a la récursion suivante :

1.  $\lambda_p(k) = -\alpha_k, 1 \leq k \leq p.$
2. Pour  $i = p - 1, \dots, 1,$

$$\lambda_i(k) = \frac{\lambda_{i+1}(k) + \lambda_{i+1}(i+1)\lambda_{i+1}(i+1-k)}{1 - \lambda_{i+1}^2(i+1)}, k = 1, \dots, i.$$

### 6.2.3 La récursion

#### Première étape

La formule de récursion

$$\begin{aligned} \tilde{\underline{\theta}}_1 &= \left\{ \left( L_{n,p} D_n^{-\frac{1}{2}} \frac{\partial}{\partial \underline{\theta}} [\underline{\Delta}_n] \right)^t \left( L_{n,p} D_n^{-\frac{1}{2}} \frac{\partial}{\partial \underline{\theta}} [\underline{\Delta}_n] \right) \right\}^{-1} \\ &\times \left( L_{n,p} D_n^{-\frac{1}{2}} \frac{\partial}{\partial \underline{\theta}} [\underline{\Delta}_n] \right)^t \left( L_{n,p} D_n^{-\frac{1}{2}} \underline{Z}_n \right) [\tilde{\underline{\theta}}_0] \end{aligned}$$

peut être écrite sous la forme suivante. Comme

$$\underline{Z}_n = \left\{ \frac{\partial}{\partial \underline{\theta}} [\underline{\Delta}_n] \right\} \underline{\theta} + (\underline{Y}_n - \underline{\Delta}_n),$$

$$\begin{aligned} \tilde{\underline{\theta}}_1 &= \tilde{\underline{\theta}}_0 \\ &= \left\{ \left( L_{n,p}^{(0)} \{ D_n^{(0)} \}^{-\frac{1}{2}} \Lambda_n^{(0)} X_n^t \right)^t \left( L_{n,p}^{(0)} \{ D_n^{(0)} \}^{-\frac{1}{2}} \Lambda_n^{(0)} X_n^t \right) \right\}^{-1} \\ &\times \left( L_{n,p}^{(0)} \{ D_n^{(0)} \}^{-\frac{1}{2}} \Lambda_n^{(0)} X_n^t \right)^t \left[ L_{n,p}^{(0)} \{ D_n^{(0)} \}^{-\frac{1}{2}} (\underline{Y}_n - \underline{\Delta}_n^{(0)}) \right]. \end{aligned}$$

Finalement, le produit  $\left\{ D_n^{(0)} \right\}^{-\frac{1}{2}} \Lambda_n^{(0)}$  est une matrice diagonale d'éléments diagonaux

$$\sqrt{\frac{\lambda_i^{(0)}}{1 + \sigma_B^2 [0] \lambda_i^{(0)}}}, 1 \leq i \leq n.$$

Soit  $\tilde{D}_n^{(0)}$  cette matrice. Alors

$$\begin{aligned} \tilde{\underline{\theta}}_1 &= \tilde{\underline{\theta}}_0 \\ &= \left\{ \left( L_{n,p}^{(0)} \tilde{D}_n^{(0)} X_n^t \right)^t \left( L_{n,p}^{(0)} \tilde{D}_n^{(0)} X_n^t \right) \right\}^{-1} \\ &\times \left( L_{n,p}^{(0)} \tilde{D}_n^{(0)} X_n^t \right)^t \left[ L_{n,p}^{(0)} \{ D_n^{(0)} \}^{-\frac{1}{2}} (\underline{Y}_n - \underline{\Delta}_n^{(0)}) \right]. \end{aligned}$$

On calcule successivement, en écrivant  $\hat{\lambda}_n^{(0)}$  pour le vecteur de composantes

$$\hat{\lambda}_1^{(0)}, \dots, \hat{\lambda}_n^{(0)} :$$

1. Calcul des valeurs  $\hat{\lambda}_n^{(0)}$  : recours à une routine de lissage.<sup>9</sup>
2. Calcul de  $X_n$  :

Les 24 première colonnes de  $X_n^t$  correspondent à l'heure de comptage, les 6 colonnes suivantes au type de jour (lundi, mardi, ...) durant lequel le comptage a lieu (le lundi correspond à des indicatrices du type de jour qui ont toutes la valeur 0), et les 8 dernières à l'année durant laquelle le comptage a eu lieu (l'année 1990 correspond à des indicatrices qui toutes valent 0). Le rang de  $X_n$  est 38, et la matrice semble bien conditionnée.

3. Calcul de  $\tilde{\theta}_0$  :

La relation

$$\ln \left[ \hat{\lambda}_n^{(0)} \right] = \langle \tilde{\theta}_0, \underline{X}_n \rangle$$

s'écrit sous la forme

$$\ln \left[ \hat{\lambda}_n^{(0)} \right] = X_n^t \tilde{\theta}_0,$$

où  $\ln \left[ \hat{\lambda}_n^{(0)} \right]$  est le vecteur de composantes  $\ln \left[ \hat{\lambda}_n^{(0)} \right]$ . On a donc

$$\tilde{\theta}_0 = \{X_n X_n^t\}^{-1} X_n \ln \left[ \hat{\lambda}_n^{(0)} \right].$$

4. Calcul de  $\sigma_B^2 [0]$  et de  $\hat{\alpha}_i^{(p)} [0]$  :

On utilise donc une approximation autorégressive d'ordre  $p$ . Il faut commencer par calculer les autocorrélations empiriques selon les formules de la section 5.2.2. Puis on utilise les récursions d'Anderson.

La méthode d'Anderson commence avec un processus d'ordre 2. Si on veut commencer avec un processus d'ordre 1, on utilise les formules suivantes :

$$\hat{\alpha}_1^{(1)} [0] = \hat{\gamma}_B (1) [0],$$

$$\hat{\sigma}_W^2 [0] = \frac{\hat{\sigma}_B^2}{1 - \hat{\gamma}_B^2 (1)} [0].$$

5. Construction de la matrice  $L_{n,p}^{(0)}$  :

---

9. Comme indiqué, dans un premier temps, on a utilisé MINITAB, Resistant smooth 4253H.

6. Calcul de la matrice  $L_{n,p}^{(0)} \{D_n^{(0)}\}^{-\frac{1}{2}}$  :

La matrice  $\{D_n^{(0)}\}^{-\frac{1}{2}}$  est diagonale et ses éléments diagonaux ont la forme suivante :

$$\frac{1}{\sqrt{\hat{\lambda}_i^{(0)} + \hat{\sigma}_B^2 [0] (\hat{\lambda}_i^{(0)})^2}}$$

7. Calcul du vecteur  $L_{n,p}^{(0)} \{D_n^{(0)}\}^{-\frac{1}{2}} (\underline{Y}_n - \hat{\underline{\lambda}}_n^{(0)})$  :

8. Calcul de la matrice  $L_{n,p}^{(0)} \tilde{D}_n^{(0)} X_n^t$  :

La matrice  $\tilde{D}_n^{(0)}$  est une matrice diagonale dont les éléments diagonaux sont donnés par l'expression suivante :

$$\sqrt{\frac{\hat{\lambda}_i^{(0)}}{1 + \hat{\sigma}_B^2 \hat{\lambda}_i^{(0)}}}$$

9. Calcul de  $\tilde{\underline{\theta}}_1$  :

### Deuxième étape

De la première étape, on retire  $\tilde{\underline{\theta}}_1$ . Cela permet de calculer

$$\lambda_i^{(1)} = e^{\langle \tilde{\underline{\theta}}_1, \underline{X}_n \rangle}.$$

On répète alors les calculs de la première étape. Le seul élément invariant du calcul est la matrice  $X_n$ .

### 6.2.4 Validité de la méthode

Davis et al.<sup>10</sup> font remarquer que les espérances  $\lambda_i$  sont estimées avec un biais,<sup>11</sup> puisque

$$E [\hat{\lambda}_i] \approx \lambda_i \exp \left\{ \frac{1}{2} (\langle \underline{X}_i, M_n \underline{X}_i \rangle) \right\},$$

où  $M_n = \Omega_{1,n}^{-1} + \Omega_{1,n}^{-1} \Omega_{2,n} \Omega_{1,n}^{-1}$  et

$$\begin{aligned} \Omega_{1,n} &= \sum_{i=1}^n \hat{\lambda}_i \underline{X}_i \underline{X}_i^t \\ \Omega_{2,n} &= \sum_{i=1}^n \sum_{j=1}^n \hat{\lambda}_i \hat{\lambda}_j \hat{\Gamma}_B(|i-j|) \underline{X}_i \underline{X}_j^t. \end{aligned}$$

10. R.D. Davis, W.T.M. Dunsmuir and Y. Wang, On autocorrelation in a Poisson regression model, *Biometrika* 87 (2000), 491-505

11. Mais il faudrait établir que les formules sont valides pour le modèle saturé que l'on utilise ici (non fonctionnel, alors que les résultats de l'article le sont pour des fonctions continues).

$\hat{\Gamma}_B$  est l'estimation de l'autocovariance. Pour alléger le calcul, on utilise l'approximation suivante (pour  $m$  approprié<sup>12</sup>) :

$$\Omega_{2,n} \approx \hat{\Gamma}_B(0) \sum_{i=1}^n \hat{\lambda}_i^2 \underline{X}_i \underline{X}_i^t + 2 \sum_{i=1}^m \hat{\Gamma}_B(i) \sum_{j=1}^{n-i} \hat{\lambda}_j \hat{\lambda}_{j+i} \underline{X}_j \underline{X}_{j+i}^t.$$

Cela produit alors les estimations ( $1 \leq i \leq n$ )

$$\hat{\lambda}_i = \exp \left\{ -\frac{1}{2} \langle (\underline{X}_i, M_n \underline{X}_i) \rangle \right\} \hat{\lambda}_i.$$

*Remarque* : En prenant en compte ces biais, on peut espérer améliorer les estimations des autocorrélations relatives au processus stationnaire  $B$  comme suit.<sup>13</sup>

Soient

$$\begin{aligned} g_i^{(n)} &= e^{\langle \underline{X}_i, M_n \underline{X}_i \rangle}, \\ g_{i,j}^{(n)} &= e^{\langle [\underline{X}_i + \underline{X}_j], M_n [\underline{X}_i + \underline{X}_j] \rangle}. \end{aligned}$$

Alors

$$\hat{\sigma}_B^2 = \frac{\sum_{i=1}^n \left\{ (Y_i - \hat{\lambda}_i)^2 + \left( 1 - \frac{2}{\{g_i^{(n)}\}^{\frac{3}{2}}} + \frac{1}{\{g_i^{(n)}\}^2} \right) \hat{\lambda}_i^2 - \hat{\lambda}_i \right\}}{\sum_{i=1}^n \frac{\hat{\lambda}_i^2}{\{g_i^{(n)}\}^2}}$$

et

$$\hat{\Gamma}_B(k) = \frac{\sum_{i=1}^{n-k} \left\{ (Y_i - \hat{\lambda}_i) (Y_{i+k} - \hat{\lambda}_{i+k}) + g_{i,i+k}^{(n)} \left( 1 - \sqrt{g_i^{(n)}} - \sqrt{g_{i+k}^{(n)}} + \frac{1}{g_{i,i+k}} \right) \right\}}{\sum_{i=1}^{n-k} g_{i,i+k} \hat{\lambda}_i \hat{\lambda}_{i+k}}.$$

## 6.2.5 Estimation de la matrice de covariance des paramètres

### 1. Méthode "naïve :"

On a<sup>14</sup> que  $\hat{\Sigma}_{\hat{\theta}} = \Sigma_0^{-1} \Sigma_1 \Sigma_0^{-1}$ , où  $\Sigma_0$  et  $\Sigma_1$  s'obtiennent comme des limites quand le nombre d'observations devient infini. On commence donc par

12. Etant donnée la nature particulière de la matrice  $X_n$ , il n'est pas évident que  $m$  puisse être choisi "petit". Par ailleurs le calcul est long :  $m = 1'000$  a requis une vingtaine d'heures de calcul.

13. Un essai rapide a montré que l'amélioration obtenue est peu significative, mais ici encore une étude systématique reste à faire.

14. (2.1)

“ignorer” le passage à la limite, pour écrire :

$$\hat{\Sigma}_0 = n \left\{ \left\{ \frac{\partial}{\partial \underline{\theta}} [\underline{\lambda}_n] \right\}^t \tilde{\Sigma}_{\underline{Y}_n}^{-1} \left\{ \frac{\partial}{\partial \underline{\theta}} [\underline{\lambda}_n] \right\} \right\}^{-1},$$

et

$$\hat{\Sigma}_1 = n \left\{ \left\{ \frac{\partial}{\partial \underline{\theta}} [\underline{\lambda}_n] \right\}^t \tilde{\Sigma}_{\underline{Y}_n}^{-1} \Sigma_{\underline{Y}_n} \tilde{\Sigma}_{\underline{Y}_n}^{-1} \left\{ \frac{\partial}{\partial \underline{\theta}} [\underline{\lambda}_n] \right\} \right\}^{-1}.$$

Mais on a encore les “égalités” suivantes :

$$\frac{\partial}{\partial \underline{\theta}} [\underline{\lambda}_n] = \Lambda_n X_n^t,$$

$$\tilde{\Sigma}_{\underline{Y}_n}^{-1} \approx \kappa_p D_n^{-\frac{1}{2}} L_{n,p}^t L_{n,p} D_n^{-\frac{1}{2}},$$

si bien que

$$\left\{ \frac{\partial}{\partial \underline{\theta}} [\underline{\lambda}_n] \right\}^t \tilde{\Sigma}_{\underline{Y}_n}^{-1} \left\{ \frac{\partial}{\partial \underline{\theta}} [\underline{\lambda}_n] \right\} = \kappa_p X_n \Lambda_n D_n^{-\frac{1}{2}} L_{n,p}^t L_{n,p} D_n^{-\frac{1}{2}} \Lambda_n X_n^t,$$

et

$$\tilde{\Sigma}_{\underline{Y}_n}^{-1} \left\{ \frac{\partial}{\partial \underline{\theta}} [\underline{\lambda}_n] \right\} = \kappa_p D_n^{-\frac{1}{2}} L_{n,p}^t L_{n,p} D_n^{-\frac{1}{2}} \Lambda_n X_n^t.$$

Le calcul le plus simple s’obtient en posant  $\hat{\Sigma}_{\underline{Y}_n} \approx \left( \tilde{\Sigma}_{\underline{Y}_n}^{-1} \right)^{-1}$ , ce qui donne  $\hat{\Sigma}_1 = \hat{\Sigma}_0$ , et

$$\hat{\Sigma}_{\underline{\theta}} = \hat{\Sigma}_0^{-1} = \frac{\kappa_p}{n} \left\{ X_n \Lambda_n D_n^{-\frac{1}{2}} L_{n,p}^t L_{n,p} D_n^{-\frac{1}{2}} \Lambda_n X_n^t \right\}.$$

## 2. Méthode utilisant l’approximation autorégressive quand celle-ci est d’ordre 1

L’autocovariance du processus  $\underline{B}$  a été écrite sous la forme  $\sigma_B^2 \Gamma_B$ . Si l’on a donc  $B_n + \alpha B_{n-1} = W_n$ , la covariance de  $\underline{B}$  à l’écart  $k$  vaut

$$\frac{\sigma_W^2}{1 - \alpha^2} (-\alpha)^k,$$

si bien que  $\Gamma_B(k) = (-\alpha)^k$ . On peut alors utiliser la remarque de 2.1 pour obtenir une approximation calculable de  $\Sigma_{\underline{Y}}$ . En effet, il faut choisir  $k$  tel que  $\beta^k \approx 0$  et remplacer la matrice  $L$  par la matrice  $\tilde{L}$  obtenue en

remplaçant par 0 les termes  $\beta^l$ ,  $l \geq k$ . Si la décomposition de Cholesky approchée (Remarque de 2.1) est écrite sous la forme  $\tilde{L}D\tilde{L}^t$ , alors

$$\begin{aligned}\hat{\Sigma}_{\underline{\theta}} &= \frac{1}{n} \left[ \kappa_p \left\{ X_n \Lambda_n D_n^{-\frac{1}{2}} L_{n,p}^t L_{n,p} D_n^{-\frac{1}{2}} \Lambda_n X_n^t \right\} \right. \\ &\quad \times \kappa_p^{-2} \left\{ X_n \Lambda_n D_n^{-\frac{1}{2}} L_{n,p}^t L_{n,p} D_n^{-\frac{1}{2}} D_n^{\frac{1}{2}} \tilde{L} D \tilde{L}^t D_n^{\frac{1}{2}} D_n^{-\frac{1}{2}} L_{n,p}^t L_{n,p} D_n^{-\frac{1}{2}} \Lambda_n X_n^t \right\}^{-1} \\ &\quad \left. \times \kappa_p \left\{ X_n \Lambda_n D_n^{-\frac{1}{2}} L_{n,p}^t L_{n,p} D_n^{-\frac{1}{2}} \Lambda_n X_n^t \right\} \right].\end{aligned}$$

Si  $M_n := L_{n,p} D_n^{-\frac{1}{2}} \Lambda_n X_n^t$ , on obtient

$$\hat{\Sigma}_{\underline{\theta}} = \frac{1}{n} \{M_n^t M_n\} \left\{ M_n^t L_{n,p} \tilde{L} D \tilde{L}^t L_{n,p}^t M_n \right\}^{-1} \{M_n^t M_n\}.$$

On voit que ce calcul est de même nature que celui qu'il faut faire pour estimer  $\underline{\theta}$ .

### 6.3 Résultats

Les graphiques et tableaux sont dans l'Annexe I (Chapitre 7). On abrège Gotthard en G, Monteceneri en M et Rolle en R.

#### Examen des paramètres

La contribution la plus importante aux "moyennes" de trafic est due aux paramètres représentant les heures. Ceux qui représentent les types de jours et les années n'interviennent que "marginalelement," les paramètres relatifs au jours intervenant surtout négativement. Pour ce qui est des paramètres relatifs aux années, ils sont négatifs surtout pour R, et cela pour les années 93-98.

1. *Heures* : valeur du paramètre en fonction de l'heure

L'ordre décroissant d'importance du trafic est M, G, et R. La configuration est qualitativement la même pour les trois stations de comptage, mais, selon la station, cette configuration est plus ou moins marquée. Les configurations pour G et M sont plus proches l'une de l'autre qu'elles ne le sont de celle de R. On distingue aisément trois périodes : celle des heures du petit matin, puis celle de la matinée, et enfin celles de l'après-midi et du soir. Géométriquement, la courbe est, au cours de ces trois périodes, approximativement et successivement, convexe, concave et concave. La variation pour R est bien plus marquée que pour G et M. Il y a des différences de détail : ainsi le minimum pour les courbes G et M est atteint à quatre heures du matin, alors que celui de la courbe R survient à cinq heures. A partir de sept heures les courbes sont à peu près parallèles.

2. *Type de jour* : valeur du paramètre en fonction du type de jour

Les trois configurations sont assez dissemblables :

- (a) celle de G est stable mardi et mercredi, puis croît linéairement jusqu'à samedi, et retombe dimanche ;
- (b) celle de M décroît de mardi à mercredi, puis reste stable jusqu'à vendredi, et croît ensuite linéairement jusqu'à dimanche ;
- (c) celle de R décroît du mardi au vendredi, pour remonter "exponentiellement" durant les jours fériés.

Les niveaux des paramètres s'entrecroisent.

### 3. *Années* : valeur du paramètre en fonction de l'année

Les configurations pour G et M sont semblables et croissantes, avec un palier de 94 à 96, alors que la configuration pour R décroît jusqu'en 94, croît les deux années suivantes, puis retombe.

### Examen des écarts-type des paramètres

Ces écarts-type sont difficiles à calculer car leur calcul fait intervenir des matrices pleines de très grandes dimensions ( $78'888 \times 78'888$ ) qui exigent la recherche d'algorithmes appropriés, ce qui n'a pas été fait, le travail étant dans une phase de "défrichage" plutôt que de "consolidation." On a donc procédé à deux calculs, un calcul "naïf" qui en fait consiste à n'utiliser qu'une de trois matrices qui entrent dans un produit, et un calcul plus élaboré, mais toujours approximatif, pour déceler le biais introduit par le recours à la méthode "naïve." Le constat est que probablement on sous-estime ainsi la variation des estimations des paramètres : c'est vrai pour les heures et les années, mais pas pour les jours. Les différences sont cependant faibles.

Comme on travaille dans un univers "Poissonien,"<sup>15</sup> il devrait y avoir un lien entre les estimations des paramètres et leur écart-type.<sup>16</sup> Cela apparaît surtout dans les écarts-types relatifs aux heures pour R, qui présentent une grande variation, et dans les écarts entre ces écarts-type, qui devraient dépendre "directement" du niveau des comptages. On trouve bien ce genre de situation. Pour les

1. *heures* : Les écarts-type pour R sont environ le double de ceux de G et M, sauf aux petites heures, et ceux de M sont plus élevés que ceux de G d'environ 25%.
2. *jours* et les *années* : Ces écarts-type sont parallèles, avec, en ordre décroissant, R, M et G. Comme le niveau des comptages est, toujours par ordre décroissant, M, G et R, cela signifie que l'incertitude sur les estimations relatives à R est plus importante que celle relative à G et M.

Si on examine les coefficients de variation, on constate les faits suivants :

---

15. Les comptages, en un instant donné, sont des mélanges, sur l'espérance, de lois de Poisson.

16. Même si les paramètres ne sont qu'indirectement liés à ces mélanges.

1. Les coefficients de variation sont faibles en ce qui concerne les heures (maximum  $< 1$ ), mais très importants en ce qui concerne les jours et les années. Statistiquement donc les paramètres sont significatifs pour les heures, mais probablement pas pour les jours et les années.
2. En ce qui concerne les heures, les coefficients de variation sont presque identiques pour G et M, et bien plus élevés pour R.
3. En ce qui concerne les jours, les coefficients de variation sont assez proches (moins de 10), mais il y a trois exceptions : le Mercredi, R est “excentrique,” le jeudi c’est G, et le samedi, M.
4. En ce qui concerne les années, il y a une tendance à la décroissance des coefficients. L’année 91 est très “atypique,” avec des coefficients très différents pour G, M et R.

### Comparaison des comptages obtenus et de l’estimation de leur espérance

Ces comparaisons sont assez difficiles puisque, comme on l’a déjà relevé dans l’exemple qui suit la définition du modèle, ce sont les valeurs des paramètres qui déterminent la position des comptages par rapport à leur espérance. Par ailleurs, sur les graphiques, on superpose des tirages issus de lois différentes. Néanmoins on tire de ces graphiques l’impression que les estimations sont “raisonnables.”

Par ailleurs le graphique qui donne les comptages en fonction de leur espérance semble aussi compatible avec le modèle.

### Influence de la durée d’observation sur les comptages

On a estimé les paramètres du modèle sur trois périodes :<sup>17</sup>  $[90 : 98]$ ,  $[93 : 98]$ , et  $[96 : 98]$ . On constate que, pour ce qui concerne les

1. *heures*, que la “courbe” des paramètres reste remarquablement stable tout en indiquant une tendance à l’augmentation du trafic, sauf en fin de journée,
2. et, pour les *jours*, que les courbes ont tendance à “pivoter,” le trafic augmentant les premiers jours de la semaine, et diminuant ensuite.

Pour ce qui concerne la variation des estimations, pour ce qui est des heures, on constate une plus grande variation quand les séries sont plus courtes. L’image est moins “nette” quand il s’agit des jours. On constate une inversion importante pour le paramètre du jeudi.

### Utilisation des estimations des espérances

Le recours aux espérances permet d’étudier directement les caractéristiques du trafic. On a fait quelques graphiques pour illustrer les possibilités, mais c’est l’utilisateur qui doit se prononcer là-dessus!

---

17.  $[m : n] := \{m, m + 1, m + 2, \dots, n - 2, n - 1, n\}$ .

# Chapitre 7

## Annexe I : la méthode de Zeger

### 7.1 Calculs

#### 7.1.1 Les moments issus du modèle

On a :

1. **Fait 1:**  $E[Y_n] = \lambda_n$
2. **Fait 2:**  $E[Y_n^2] = \lambda_n + (1 + \sigma_B^2) \lambda_n^2$

En effet

$$\begin{aligned} E[Y_n^2] &= E[E[Y_n^2 | \underline{B}]] \\ &= E[V[Y_n | \underline{B}] + E^2[Y_n | \underline{B}]] \\ &= E[E[Y_n | \underline{B}] + E^2[Y_n | \underline{B}]] \\ &= E[\lambda_n B_n + (\lambda_n B_n)^2] \\ &= \lambda_n + \lambda_n^2 E[B_n^2] \\ &= \lambda_n + \lambda_n^2 \{V[B_n] + E^2[B_n]\} \\ &= \lambda_n + \lambda_n^2 \{\sigma_B^2 \Gamma_B(0) + 1\} \\ &= \lambda_n + \lambda_n^2 (1 + \sigma_B^2) \end{aligned}$$

3. **Fait 3:**  $V[Y_n] = \lambda_n + \lambda_n^2 \sigma_B^2$
4. **Fait 4:**  $E[Y_n Y_{n+p}] = \lambda_n \lambda_{n+p} \sigma_B^2 \Gamma_B(p) + \lambda_n \lambda_{n+p}$

En effet

$$E[Y_n Y_{n+p}] = E[E[Y_n Y_{n+p} | \underline{B}]]$$

$$\begin{aligned}
&= E[E[Y_n | \underline{B}] E[Y_{n+p} | \underline{B}]] \\
&= E[(\lambda_n B_n)(\lambda_{n+p} B_{n+p})] \\
&= \lambda_n \lambda_{n+p} E[B_n B_{n+p}] \\
&= \lambda_n \lambda_{n+p} \{\text{cov}(B_n, B_{n+p}) + E[B_n] E[B_{n+p}]\} \\
&= \lambda_n \lambda_{n+p} \{\sigma_B^2 \Gamma_B(p) + 1\}
\end{aligned}$$

5. **Fait 5:**  $\text{cov}(Y_n, Y_{n+p}) = \lambda_n \lambda_{n+p} \sigma_B^2 \Gamma_B(p)$

6. **Fait 6:**  $\text{corr}(Y_n, Y_{n+p}) = \frac{\Gamma_B(p)}{\sqrt{\left(1 + \frac{1}{\sigma_B^2 \lambda_n}\right) \left(1 + \frac{1}{\sigma_B^2 \lambda_{n+p}}\right)}}$

7. Soit  $\underline{Y}_n$  le vecteur de composantes  $Y_1, \dots, Y_n$ . Sa matrice de covariance est notée  $\Sigma_{\underline{Y}_n}$ . La matrice  $\Sigma_B^{(n)}$  est éfinie par l'égalité suivante :

$$\Sigma_B^{(n)} = \begin{bmatrix} 1 & \Gamma_B(1) & \cdots & \Gamma_B(n-1) \\ \Gamma_B(1) & 1 & \cdots & \Gamma_B(n-2) \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma_B(n-1) & \Gamma_B(n-2) & \cdots & 1 \end{bmatrix}.$$

8. **Fait 7:** Si  $\Lambda_n$  désigne la matrice d'éléments diagonaux  $\lambda_1, \dots, \lambda_n$ , on a alors :

$$\Sigma_{\underline{Y}_n} = \Lambda_n + \sigma_B^2 \Lambda_n \Sigma_B^{(n)} \Lambda_n.$$

*Remarque :* On retrouve le cas général en posant  $h(x) = e^x, g(x) = x$ , et

$$\Gamma = \phi \left[ I_n + \sigma_B^2 \Lambda_n^{\frac{1}{2}} \Sigma_B^{(n)} \Lambda_n^{\frac{1}{2}} \right].$$

Il faut aussi remarquer que l'on ne distingue pas  $\Sigma_{\underline{Y}_n}$  de  $\Sigma$ !

### 7.1.2 Les equations d'estimation

Soient  $\underline{\lambda}_n$  le vecteur (colonne) de composantes  $\lambda_1, \dots, \lambda_n$ , et  $X_n$  la matrice  $(p, n)$  définie par l'égalité suivante :  $X_n = [\underline{X}_1 | \cdots | \underline{X}_n]$ .  $\underline{e}_k^p$  désigne le k-ème vecteur de la base canonique de  $\mathbb{R}^p$ . On a alors que

1. **Fait 1:**  $\frac{\partial}{\partial \theta_i} \lambda_j = \lambda_j X_i^{(j)}$

2. **Fait 2:**  $\frac{\partial}{\partial \theta_i} [\underline{\lambda}_n] = \begin{bmatrix} \frac{\partial}{\partial \theta_i} \lambda_1 \\ \vdots \\ \frac{\partial}{\partial \theta_i} \lambda_j \\ \vdots \\ \frac{\partial}{\partial \theta_i} \lambda_n \end{bmatrix} = \begin{bmatrix} \lambda_1 X_i^{(1)} \\ \vdots \\ \lambda_j X_i^{(j)} \\ \vdots \\ \lambda_n X_i^{(n)} \end{bmatrix}$

3. **Fait 3:**  $\{\underline{e}_i^p\}^t X_n = [X_i^{(1)} \dots X_i^{(j)} \dots X_i^{(n)}]$

4. **Fait 4:**  $\frac{\partial}{\partial \theta_i} [\underline{\lambda}_n] = \Lambda_n X_n^t \underline{e}_i^p$

5. **Fait 5:** Si l'on définit la matrice  $(n, p)$  suivante :

$$\frac{\partial}{\partial \underline{\theta}} [\underline{\lambda}_n] = \left[ \frac{\partial}{\partial \theta_1} [\underline{\lambda}_n] \mid \dots \mid \frac{\partial}{\partial \theta_p} [\underline{\lambda}_n] \right],$$

alors

$$\frac{\partial}{\partial \underline{\theta}} [\underline{\lambda}_n] = \Lambda_n X_n^t.$$

### 7.1.3 Problèmes de calcul : justifications de l'approximation

#### Cas d'un processus autorégressif d'ordre 1

Si l'ordre du processus est  $p = 1$ , la matrice d'autocorrélation a la forme suivante, avec  $\beta = -\alpha$  :<sup>1</sup>

$$\begin{bmatrix} 1 & \beta & \beta^2 & \beta^3 & \dots \\ \beta & 1 & \beta & \beta^2 & \dots \\ \beta^2 & \beta & 1 & \beta & \dots \\ \beta^3 & \beta^2 & \beta & 1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}.$$

Cette matrice a une décomposition de Cholesky de la forme  $LDL^t =$

$$\begin{bmatrix} 1 & 0 & 0 & 0 & \dots \\ \beta & 1 & 0 & 0 & \dots \\ \beta^2 & \beta & 1 & 0 & \dots \\ \beta^3 & \beta^2 & \beta & 1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 & \dots \\ 0 & 1 - \beta^2 & 0 & 0 & \dots \\ 0 & 0 & 1 - \beta^2 & 0 & \dots \\ 0 & 0 & 0 & 1 - \beta^2 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} 1 & \beta & \beta^2 & \beta^3 & \dots \\ 0 & 1 & \beta & \beta^2 & \dots \\ 0 & 0 & 1 & \beta & \dots \\ 0 & 0 & 0 & 1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}.$$

De plus

$$\begin{bmatrix} 1 & 0 & 0 & 0 & \dots \\ \beta & 1 & 0 & 0 & \dots \\ \beta^2 & \beta & 1 & 0 & \dots \\ \beta^3 & \beta^2 & \beta & 1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 & \dots \\ -\beta & 1 & 0 & 0 & \dots \\ 0 & -\beta & 1 & 0 & \dots \\ 0 & 0 & -\beta & 1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}.$$

Donc

$$\left[ \Sigma_1^{(n)} \right]^{-1} = [L^{-1}]^t D^{-1} L^{-1} =$$

---

1. Il faut faire attention au signe! En effet, Zeger utilise la représentation  $B_n = \alpha B_{n-1} + W_n$ , alors que l'on utilise implicitement la représentation  $B_n + \alpha B_{n-1} = W_n$ .

$$\begin{bmatrix} 1 & -\beta & 0 & 0 & \cdots \\ 0 & 1 & -\beta & 0 & \cdots \\ 0 & 0 & 1 & -\beta & \cdots \\ 0 & 0 & 0 & 1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots \\ 0 & \frac{1}{1-\beta^2} & 0 & 0 & \cdots \\ 0 & 0 & \frac{1}{1-\beta^2} & 0 & \cdots \\ 0 & 0 & 0 & \frac{1}{1-\beta^2} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots \\ -\beta & 1 & 0 & 0 & \cdots \\ 0 & -\beta & 1 & 0 & \cdots \\ 0 & 0 & -\beta & 1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

Selon l'article de Zeger, il faut utiliser  $(1 - \beta^2)^{-1} L_{n,1}^t L_{n,1}$ , ce qui correspond à  $(1 - \beta^2)^{-1}$  fois le produit

$$\begin{bmatrix} 1 & 0 & 0 & 0 & \cdots \\ -\beta & 1 & 0 & 0 & \cdots \\ 0 & -\beta & 1 & 0 & \cdots \\ 0 & 0 & -\beta & 1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} 1 & -\beta & 0 & 0 & \cdots \\ 0 & 1 & -\beta & 0 & \cdots \\ 0 & 0 & 1 & -\beta & \cdots \\ 0 & 0 & 0 & 1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} = \begin{bmatrix} 1 & -\beta & 0 & 0 & \cdots \\ -\beta & 1 + \beta^2 & -\beta & 0 & \cdots \\ 0 & -\beta & 1 + \beta^2 & -\beta & \cdots \\ 0 & 0 & -\beta & 1 + \beta^2 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

qui donne le même résultat. Bien que les deux démarches soient très proches et que la première semble plus "naturelle," les raisons du choix de la seconde semblent tenir au fait qu'il y a une matrice de moins à multiplier.

### Cas général

Plus généralement, soit  $\mathcal{X} := \{X_n\}_{n=-\infty}^{n=+\infty}$  un processus stationnaire au sens large, d'espérance nulle et de covariance

$$\gamma(|m - n|) = E[X_m - X_n].$$

La ligne numéro  $i$  de la matrice de covariance  $\Sigma_{\underline{X}_n}$  des variables aléatoires  $X_1, \dots, X_n$ , notée  $\Sigma_{\underline{X}_n}[i, \bullet]$ , a la forme

$$\gamma(i-1), \gamma(i-2), \dots, \gamma(1), \gamma(0), \gamma(1), \dots, \gamma(n-i).$$

Si  $\underline{\gamma}[m : n]$  désigne le vecteur de composantes  $\gamma(m), \dots, \gamma(n)$  et que  $p$  soit un entier valant au moins un, cette ligne peut s'écrire comme suit :<sup>2</sup>

$$\underline{\mathbb{L}}_i^* := \{\underline{\gamma}[i-1, 1]\}^*, \{\underline{\gamma}[0, p]\}^*, \{\underline{\gamma}[p+1, n]\}^*.$$

Si l'on suppose que  $\mathcal{X}$  est autorégressif d'ordre  $p$ , soit que

$$X_n + \sum_{i=1}^p \alpha_i X_{n-i} = B_n,$$

et que l'on définisse les vecteurs suivants :

$$\underline{\alpha} = \begin{bmatrix} 1 \\ \alpha_1 \\ \vdots \\ \alpha_p \end{bmatrix} \in \mathbb{R}^{p+1}, \underline{\gamma}_n = \begin{bmatrix} \gamma(n) \\ \gamma(n-1) \\ \vdots \\ \gamma(n-p) \end{bmatrix} \in \mathbb{R}^{p+1}, n \geq 0,$$

2.  $\star$  dénote le transposé. On fait l'hypothèse que  $p+1 < n$ .

on a alors<sup>3</sup> la propriété d'autorégression suivante :

$$\langle \underline{\alpha}, \underline{\gamma}_n \rangle_{\mathbb{R}^{p+1}} = \begin{cases} \sigma_B^2 & \text{si } n = 0 \\ 0 & \text{si } n > 0 \end{cases} .$$

Soit maintenant  $T$  la matrice dont la colonne numéro  $j$ , notée  $\underline{C}_j$ , a la forme<sup>4</sup>

$$\begin{bmatrix} \underline{0}_{j-1} \\ \underline{\alpha} \\ \underline{0}_{n-p-j} \end{bmatrix} .$$

Alors

$$\langle \underline{L}_i, \underline{C}_j \rangle_{\mathbb{R}^n} = \begin{cases} 0 & \text{si } i > j \\ \sigma_B^2 & \text{si } i = j \\ \langle \underline{\gamma}[j-i, j-i+p], \underline{\alpha} \rangle_{\mathbb{R}^{p+1}} & \text{si } i < j \end{cases} .$$

La matrice  $\Sigma_{\underline{X}_n} T$  est donc “triangulaire” avec des zéros sous la “diagonale” et la valeur  $\sigma_B^2$  dans la “diagonale” elle-même. Pour voir ce qui se passe dans la partie hors “diagonale” et non nulle de  $\Sigma_{\underline{X}_n} T$ , on examine le cas particulier pour lequel  $n = 6$  et  $p = 2$ . On écrit, quand  $|i - j| = k$ ,

$$\gamma(|i - j|) := \gamma_k .$$

Alors

$$\Sigma_{\underline{X}_6} = \begin{bmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \gamma_3 & \gamma_4 & \gamma_5 \\ \gamma_1 & \gamma_0 & \gamma_1 & \gamma_2 & \gamma_3 & \gamma_4 \\ \gamma_2 & \gamma_1 & \gamma_0 & \gamma_1 & \gamma_2 & \gamma_3 \\ \gamma_3 & \gamma_2 & \gamma_1 & \gamma_0 & \gamma_1 & \gamma_2 \\ \gamma_4 & \gamma_3 & \gamma_2 & \gamma_1 & \gamma_0 & \gamma_1 \\ \gamma_5 & \gamma_4 & \gamma_3 & \gamma_2 & \gamma_1 & \gamma_0 \end{bmatrix}, \quad T = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \alpha_1 & 1 & 0 & 0 \\ \alpha_2 & \alpha_1 & 1 & 0 \\ 0 & \alpha_2 & \alpha_1 & 1 \\ 0 & 0 & \alpha_2 & \alpha_1 \\ 0 & 0 & 0 & \alpha_2 \end{bmatrix},$$

et

$$\Sigma_{\underline{X}_6} T = \begin{bmatrix} \sigma_B^2 & \langle \underline{\gamma}[1:3], \underline{\alpha} \rangle_{\mathbb{R}^3} & \langle \underline{\gamma}[2:4], \underline{\alpha} \rangle_{\mathbb{R}^3} & \langle \underline{\gamma}[3:5], \underline{\alpha} \rangle_{\mathbb{R}^3} \\ 0 & \sigma_B^2 & \langle \underline{\gamma}[1:3], \underline{\alpha} \rangle_{\mathbb{R}^3} & \langle \underline{\gamma}[2:4], \underline{\alpha} \rangle_{\mathbb{R}^3} \\ 0 & 0 & \sigma_B^2 & \langle \underline{\gamma}[1:3], \underline{\alpha} \rangle_{\mathbb{R}^3} \\ 0 & 0 & 0 & \sigma_B^2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} .$$

3. W.A. Fuller, Introduction to Statistical Time Series, 2nd ed., Wiley, New York (1996), Corollary 2.6.1.1., page 61.

4. Quand  $j-1 = 0$ , ou quand  $n-p-j = 0$ , il n'y a pas de zéros dans la partie correspondante du vecteur.

Le calcul de  $\Sigma_{\underline{X}_6} TT^*$  donne

Terme	Formule
(1, 1)	$\sigma_B^2$
(1, 2)	$\alpha_1 \sigma_B^2 + \langle \underline{\gamma}[1 : 3], \underline{\alpha} \rangle_{\mathbb{R}^3}$
(1, 3)	$\alpha_2 \sigma_B^2 + \alpha_1 \langle \underline{\gamma}[1 : 3], \underline{\alpha} \rangle_{\mathbb{R}^3} + \langle \underline{\gamma}[2 : 4], \underline{\alpha} \rangle_{\mathbb{R}^3}$
(1, 4)	$\alpha_2 \langle \underline{\gamma}[1 : 3], \underline{\alpha} \rangle_{\mathbb{R}^3} + \alpha_1 \langle \underline{\gamma}[2 : 4], \underline{\alpha} \rangle_{\mathbb{R}^3} + \langle \underline{\gamma}[3 : 5], \underline{\alpha} \rangle_{\mathbb{R}^3}$
(1, 5)	$\alpha_2 \langle \underline{\gamma}[2 : 4], \underline{\alpha} \rangle_{\mathbb{R}^3} + \alpha_1 \langle \underline{\gamma}[3 : 5], \underline{\alpha} \rangle_{\mathbb{R}^3}$
(1, 6)	$\alpha_2 \langle \underline{\gamma}[3 : 5], \underline{\alpha} \rangle_{\mathbb{R}^3}$
(2, 1)	0
(2, 2)	$\sigma_B^2$
(2, 3)	$\alpha_1 \sigma_B^2 + \langle \underline{\gamma}[1 : 3], \underline{\alpha} \rangle_{\mathbb{R}^3}$
(2, 4)	$\alpha_2 \sigma_B^2 + \alpha_1 \langle \underline{\gamma}[1 : 3], \underline{\alpha} \rangle_{\mathbb{R}^3} + \langle \underline{\gamma}[2 : 4], \underline{\alpha} \rangle_{\mathbb{R}^3}$
(2, 5)	$\alpha_2 \langle \underline{\gamma}[1 : 3], \underline{\alpha} \rangle_{\mathbb{R}^3} + \alpha_1 \langle \underline{\gamma}[2 : 4], \underline{\alpha} \rangle_{\mathbb{R}^3}$
(2, 6)	$\alpha_2 \langle \underline{\gamma}[2 : 4], \underline{\alpha} \rangle_{\mathbb{R}^3}$
(3, 1)	0
(3, 2)	0
(3, 3)	$\sigma_B^2$
(3, 4)	$\alpha_1 \sigma_B^2 + \langle \underline{\gamma}[1 : 3], \underline{\alpha} \rangle_{\mathbb{R}^3}$
(3, 5)	$\alpha_2 \sigma_B^2 + \alpha_1 \langle \underline{\gamma}[1 : 3], \underline{\alpha} \rangle_{\mathbb{R}^3}$
(3, 6)	$\alpha_2 \langle \underline{\gamma}[1 : 3], \underline{\alpha} \rangle_{\mathbb{R}^3}$
(4, 1)	0
(4, 2)	0
(4, 3)	0
(4, 4)	$\sigma_B^2$
(4, 5)	$\alpha_1 \sigma_B^2$
(4, 6)	$\alpha_2 \sigma_B^2$
(5, 1)	0
⋮	
(6, 6)	0

Il faut alors remarquer que les termes de la forme

$$\alpha_2 \langle \underline{\gamma}[1 : 3], \underline{\alpha} \rangle_{\mathbb{R}^3} + \alpha_1 \langle \underline{\gamma}[2 : 4], \underline{\alpha} \rangle_{\mathbb{R}^3} + \langle \underline{\gamma}[3 : 5], \underline{\alpha} \rangle_{\mathbb{R}^3}$$

sont nuls. En effet, ils s'écrivent

$$\begin{aligned} & \alpha_2 [\gamma_1 + \alpha_1 \gamma_2 + \alpha_2 \gamma_3] \\ & + \alpha_1 [\gamma_2 + \alpha_1 \gamma_3 + \alpha_2 \gamma_4] \\ & + 1 [\gamma_3 + \alpha_1 \gamma_4 + \alpha_2 \gamma_5] \\ & = 0, \end{aligned}$$

en invoquant la propriété d'autorégression déjà utilisée. Comme en général  $n \gg p$ , les termes de ce type sont "majoritaires," si bien que  $\Sigma_{\underline{X}_n} TT^*$  est "pour l'essentiel" une matrice diagonale, et donc, puisque le calcul de  $TT^*\Sigma_{\underline{X}_n}$  est "identique," on a bien, avec  $TT^*$ , une approximation de l'inverse de  $\Sigma_{\underline{X}_n}$ .

## 7.2 Code MATLAB et graphiques

### 7.2.1 Code

#### MATLAB script

```
% -----  
%                               SCRIPT FOR ZEGER'S METHOD  
% -----  
  
global Tsdata n X_n Counts Lambda0  
  
% -----  
%                               Table of global variables  
% -----  
% Tsdata      matrix with the following columns  
%             column 1 : counts  
%             column 2 : year of counts  
%             column 3 : day type of counts  
%             column 4 : hour of counts  
%             column 5 : "first Lambda0"  
%             either (s) computed with robust smooth (MINITAB)  
%                   may contain negative values  
%             or (p) computed by Poisson approx  
%             or (m) computed by means  
%             this matrix is deleted after definition of the following matrices :  
% n           number of observations (number of counts)  
% X_n        design matrix  
%            (see design(...) function)  
% Counts     response matrix  
% Lambda0    "first Lambda0" - min("first Lambda0") +1  
%            [because will have to compute log(Lambda0)]  
% -----
```

```

%-----
%                               Table of variables of this script
%-----
%  str                          string for the name of data file
%  thold, max_iter              threshold and maximum iteration number for the function estimates()
%  Thetahat, Lambdahat         estimates that are computed iteratively by the function estimates()
%  Alpha, sigmaB2              parameters of the auto-regressive approximation to the hidden
%                               stationary process of the model and estimated noise variance
%                               (see function estimates())
%  Lnp                          Intermediate matrix
%                               (see function makeLnp())
%  C, approx                    C is an approximation of the auto-regressive process' autocovariance
%                               and approx is an approximation threshold for this computation
%                               (see function ARCoV())
%  NaiveCov, ZegerCov          estimations for the covariance of theta estimates
%-----

```

```

% -----
%                               Table of functions
% -----
% design → X_n
%                               Uses information in tsdata about year, type of day and hour
%                               to produce a fully parametrized design matrix
% estimates(thold,max_iter) → Thetahat, Lambdahat, Alpha, sigmaB2, Lnp
%                               Calls on functions approxARP(), makeLnp(), iterTheta()
%                               Estimates iteratively the unknown design parameters using iterTheta.
%                               At each iteration hidden stationary process is estimated anew
%                               It stops when improvement becomes negligible
%                               or runs out of allowed iterations
% approxARP(Lambda, max_p) → Alpha, sigmaB2, p, stationary
%                               Calls on function stationarity()
%                               Uses counts and Lambdas so that a correlation matrix and an
%                               autoregressive approximation to the hidden stationary process
%                               of the model can be computed
% stationarity(Alphatmp) → stationary
%                               Receives the parameters of an autoregressive
%                               process and uses Schur's test to decide whether or not
%                               the process is stationary
% makeLnp(Alpha,n) → Lnp
%                               Construction of the matrix Lnp (see code for details)
% iterTheta(Theta,Lambda,sigmaB2,Lnp) → newTheta
%                               Computes an adjusted estimation of Theta given a previous value
% ARCoV(Alpha,approx) → C
%                               Computes an approximation to the autocovariance function of an
%                               autoregressive process knowing the process' parameters and using
%                               approx as threshold
% ThetaCov(sigmaB2,Lambdahat,Alpha,Lnp,C) → NaiveCov, ZegerCov
%                               Computes two different approximations of theta's covariance
% -----

```

```

%-----
%                               Step 0 : load the initial data
%-----

str      = input('Please enter the name of the (ascii) data file (within quotes) : ');
Tsdata  = load(str, '-ascii'); % here dim(Tsdata) = (78888,5)

disp('***** Step 0 complete *****');

%-----
%                               Step 1 :
%                               get the number of counts
%                               the design matrix
%                               ( here : saturated model
%                               24 for hour, 6 for day, 8 for year )
%                               the response matrix
%                               and the new Lambda0
%-----

tmp      = size(Tsdata);
n        = tmp(1); % number of counts
X_n     = design; % here X_n is a (38,78888)-matrix
Counts  = Tsdata(:,1); % here response is a (78888,1)-matrix
Lambda0 = Tsdata(:,5)-min(Tsdata(:,5))+1; % eliminates negative and null values

clear Tsdata; % saves memory

disp('***** Step 1 complete *****');

%-----
%                               Step 2 : run the iteration for parameter estimation
%-----

thold    = 0.001;
max_iter = 20;
[Thetahat,Lambdahat,Alpha,sigmaB2,Lnp] = estimates(thold,max_iter);
save(['Theta_',str], 'Thetahat', '-ascii', '-double');

disp('***** Step 2 complete *****');

%-----
%                               Step 3 : compute the covariance matrix of the estimations
%-----

approx   = 0.2;
C        = ARCoV(Alpha,approx);
[NaiveCov,ZegerCov] = ThetaCov(sigmaB2,Lambdahat,Alpha,Lnp,C);
save(['Naivecov_',str], 'NaiveCov', '-ascii', '-double');
save(['Zegercov_',str], 'ZegerCov', '-ascii', '-double');

disp('***** Step 3 complete *****');

```

### La fonction : design

```

%-----
%                               function [X_n] = design();
%-----

```

```

global Tsdata n
% Uses information in Tsdata about
% year, type of day and hour
% to produce a fully parametrized design matrix

% preallocation of memory for tmp
% WARNING needs to be changed with change of years (see also below) :
disp('Years 90 to 98');
tmp = zeros(n,38);

% disp('Years 93 to 98');
% tmp = zeros(n,35);

% disp('Years 96 to 98');
% tmp = zeros(n,32);

% disp('Only year 98');
% tmp = zeros(n,30);

% hour indicators
for i = 1 :24
    tmp(:,i) = (Tsdata(:,4)==i);
end

% type of day indicators
for i = 2 :7
    tmp(:,i+23) = (Tsdata(:,3)==i);
end

% year indicators
% WARNING needs to be changed with change of years (see also above) :
%disp('Years 90 to 98');
for i = 91 :98
    tmp(:,i-60) = (Tsdata(:,2)==i);
end

% disp('Years 93 to 98');
% for i = 94 :98
%     tmp(:,i-63) = (Tsdata(:,2)==i);
% end

% disp('Years 96 to 98');
% for i = 97 :98
%     tmp(:,i-66) = (Tsdata(:,2)==i);
% end

% disp('Only year 98');
% get the design matrix
X_n = sparse(tmp');
clear tmp;

```

## La fonction : estimates

```
% -----  
function [Thetahat,Lambdahat,Alpha,sigmaB2,Lnp] = estimates(thold,max_iter)  
% -----  
  
global n Counts X_n Lambda0  
  
    % Calls on functions approxARP(), makeLnp(), iterTheta()  
    % Estimates iteratively the unknown design parameters using iterTheta()  
    % At each iteration hidden stationary process is estimated anew  
    % It stops when improvement becomes negligible or runs out of allowed iterations  
  
    % set max order of approximating AR process :  
  
    max_p = 10;  
    % initial values :  
        Lambdahat      = Lambda0;  
        Thetahat       = X_n'  
        log(Lambda0);  
    % loop variables :  
        dist           = 1000000000;  
        iter_nb = 1;  
        while (dist >= thold) & (iter_nb <= max_iter)  
            disp(sprintf('Starting iteration number %d',iter_nb));  
            [Alpha,sigmaB2,p,stationary] = approxARP(Lambdahat,max_p);  
            Lnp = makeLnp(Alpha);  
            newThetahat = iterTheta(Thetahat,Lambdahat,sigmaB2,Lnp);  
            dist = norm(newThetahat-Thetahat,inf);  
            disp(sprintf('Distance between the two last estimations of Thetahat is %g',dist));  
            disp('-----');  
            Thetahat = newThetahat;  
            Lambdahat = exp(X_n'*Thetahat);  
            iter_nb = iter_nb+1;  
        end  
    disp('Theta estimate is :');  
    disp(Thetahat);  
  
    Function calls :  
    1. approxARP  
    2. makeLnp  
    3. iterTheta
```

## La fonction : approxARP

```
function [Alpha,sigmaB2,p,stationary] = approxARP(Lambda,max_p)
```

```
global n Counts
```

```
% Calls on function stationarity().  
% This function uses Counts and Lambda so that a correlation matrix and an  
% autoregressive approximation to the hidden stationary process of the model  
% can be computed.  
% It is also given a maximum order for the process in  
% case one cannot obtain a stationary process by increasing  
% the number of parameters in the autoregression  
%  
% This function returns :  
%   the estimated parameters  
%   the maximum order of the fitted autoregressive process  
%   the result of the last stationarity test  
%  
% It stops as soon as a stationary process is obtained  
%  
% Anderson's recursive method is used and max_p is at least 1  
% Zeger's formulae are used  
% Required number of observations for the computation of correlations  
min_nbobs = 100 ;  
% To estimate max_p parameters, one needs at least max_p correlations  
% (Yule-Walker equations)  
% to compute the correlation at lag k, one has n-k data couples available  
% thus one must have :
```

```

% n-max_p >= min_nbobs
    if (n-max_p <= min_nbobs)
        disp('Ordre du processus autogressif trop lev : arrt');
        return;
    else
        end

% Computation of noise variance
sigmaB2 = sum((Counts-Lambda).^2-Lambda)/sum(Lambda.^2);

% Computation of correlations
for i = 1 :max_p
    Firstl = Lambda(1 :n-i,1);
    Firsty = Counts(1 :n-i,1);
    Secondl = Lambda(i+1 :n,1);
    Secondy = Counts(i+1 :n,1);
    Corrvect(i) = sum((Firsty-Firstl).*(Secondy-Secondl))/.../(sum(Firstl.*Secondl)*sigmaB2);
end

% Computation of alpha parameters
q = 1; % order of investigated model
Alphatmp = Corrvect(1); % approximated first-order model
stationary = stationarity(Alphatmp); % is it stationary?

while ((q < max_p) & ~stationary);
    q = q+1;

    % The estimation of parameters for model of order q
    % needs parameters for model of order q-1 and
    % correlations from lag 1 to lag q

    Alphatmp_prev = Alphatmp;

    % Calculates th q-th parameter first :

    numerator = Corrvect(q);
    denominator = 1;
    for i = 1 :(q-1)
        numerator = numerator - Alphatmp_prev(i)*Corrvect(q-1);
        denominator = denominator - Alphatmp_prev(i)*Corrvect(i);
    end
    Alphatmp(q) = numerator/denominator;

    % Then calculates the others :

    for i = 1 :(q-1)
        Alphatmp(i) = Alphatmp_prev(i)-Alphatmp(q)*Alphatmp_prev(q-i);
    end

    stationary = stationarity(Alphatmp);
end

if stationary
    p = q;
    Alpha = [1 ;-Alphatmp'];
    disp(sprintf('Fitted stationary model is : p = %d, sigma2 = %g, and alpha is'... ,p,sigmaB2));
    disp(Alphatmp);
else
    p = 0;
    Alpha = [];
    disp('Not able to produce an AR approximation of order less than max_p');
end

```

**Function calls :** stationarity

## La fonction : stationarity

```
% -----  
function stationary = stationarity(Alphatmp)  
% -----  
  
% This function receives the parameters of an autoregressive process.  
% It uses Schur's test to decide whether or not the process is stationary.  
% The test is based on the size of the partial autocorrelations,  
% which are obtained recursively.  
  
% The partial at highest lag is simply the  
% parameter at highest lag  
  
p = length(Alphatmp);  
Partials(p) = Alphatmp(p);  
  
% The recursion goes through one less step  
% than parameters  
Partial1 = Alphatmp';  
for i = 1:(p-1)  
    % Get the working dimensions of vectors and matrices :  
    dim = p-i;  
  
    % Computation of the recursion vector  
    constant = Partial1(dim+1);  
    denominator = 1-constant^2;  
  
    %Use formula of page 90 (6.2.2)  
    Partial2 = (1/denominator)*(Partial1(1:dim)+constant*flipup(Partial1(1:dim)));  
    Partial1 = Partial2;  
  
    Partials(dim) = Partial1(dim);  
  
stationary = (max(Partials)<=1);
```

## La fonction : makeLnp

```
% -----  
function Lnp = makeLnp(Alpha,n)  
% -----  
  
global n  
  
% it is assumed Alpha is in the form :  
% [1 -a1 -a2 ... -ap]' (transposition is crucial here!)  
% the L matrix is constructed by adding one  
% "diagonal" at a time :  
% the first "diagonal" is made of ones  
% the next "diagonal" is made of -a1 ...
```

```

% padding with zeros must be added
    Tmp          = size(Alpha);
    p            = Tmp(1)-1;
    B            = zeros(n-p,p+1);      % preallocation of memory
    for i = 1 :(p+1)
        B(:,i)   = Alpha(i)*ones(n-p,1);
    end
    Lnp          = spdiags(B,0 :p,n-p,n);

```

## La fonction : iterTheta

```
% -----  
function newTheta = iterTheta(Theta,Lambda,sigmaB2,Lnp)  
% -----  
  
global n Counts X_n  
  
% computes an adjusted estimation of Theta given a previous value  
% compute the term Lnp*inv(sqrt(D)) : get M1  
D2 = spdiags(sqrt(1./(Lambda+sigmaB2*Lambda.*Lambda)),0,n,n);  
M1 = Lnp*D2;  
clear D2;  
  
% compute the term inv(sqrt(D))*Lambda  
  
D1 = spdiags(sqrt(Lambda./(1+sigmaB2.*Lambda)),0,n,n);  
% compute the term Lnp*inv(sqrt(D))*Lambda*X^t  
M2 = Lnp*D1*X_n';  
clear D1;  
  
% prepare the new estimation of Theta  
M3 = inv(M2'*M2);  
M4 = M2'*M1;  
clear M2;  
clear M1;  
  
% compute new value of Theta  
newTheta = Theta+M3*M4*(Counts-Lambda);  
clear M3;  
clear M4;
```

## La fonction : ARcov

```
% -----  
function C = ARcov(Alpha,approx)  
% -----  
  
global n  
  
% Only in the case of order one AR-processes!!!  
  
% Computes an approximation to the autocovariance function of an autoregressive  
% process knowing the process' parameters.  
% The approximation is obtained by keeping from the full matrix  
% only a number of diagonals (whose power is "significantly different from zero").  
  
% n := number of values in time series  
% alphasmp := MINUS second component of Alpha vector  
% approx := value that "defines" zero  
  
alphatmp = -Alpha(2);  
  
% get the number of non zero diagonals  
nz = ceil(log(abs(approx))/log(abs(alphatmp)));  
disp(sprintf('The number of non zero diagonals will be %d',nz));  
if (nz > 30) % too many iterations for available machine  
    nz = 30;  
end;  
Tmp = ones(n,nz); % preallocation of memory  
for i = 2:nz  
    Tmp(:,nz-i+1) = alphasmp*Tmp(:,nz-i+2);  
end  
L = spdiags(Tmp,(1-nz):0,n,n);  
clear Tmp;  
disp('Computation of L done!');  
D = spdiags([1;(1-alphasmp^2)*ones(n-1,1)],0,n,n);  
disp('Computation of D done!');  
  
C = L*D*L';  
disp('Computation of C done!');
```

## La fonction : ThetaCov

```
% -----  
function [NaiveCov,ZegerCov] = ThetaCov(sigmaB2,Lambdahat,Alpha,Lnp,C)  
% -----  
  
global n X_n  
  
% Computes two different approximations of theta's covariance.  
% Notations here are those of the paper.  
% M equals Lnp*(D^(-1/2)*Lambda)*X' :  
M = Lnp*spdiags(sqrt(Lambdahat./(1+sigmaB2*Lambdahat)),0,n,n)*X_n';  
disp('Computation of M done!');  
  
% Sigma0 inverse up to a multiplicative constant :  
S0inv = M'*M;  
disp('Computation of inv(S0) done!');  
  
% Sigma1 up to a multiplicative constant :  
M2 = Lnp'*M;  
clear M;  
S1 = (M2'*C*M2)^(-1);  
clear M2;  
disp('Computation of S1 done!');  
  
% results ZegerCov = full((1/n)*S0inv*S1*S0inv);  
k= 1/(1-Alpha(2)^2);  
NaiveCov = full((k/n)*S0inv);  
  
disp(sprintf('Max distance between temrs of the two matrices is %g'... ,max(max(abs(NaiveCov-  
ZegerCov)))));
```

## 7.2.2 Graphiques

Voici une description rapide du contenu des graphiques qui suivent.

Pages	Catégorie	Description
124-126	Station	Valeurs des paramètres selon les heures, jours, années
127-129	Station	Valeurs des écarts-types des paramètres selon les heures, jours, années
130-132	Mode de calcul	Valeurs des écarts-types des paramètres selon les heures, jours, années
133-135	Station	Valeurs des coefficients de variation des paramètres selon les heures, jours, années
136-142	Jour	Répartition des comptages par heure et valeur de l'espérance (Rolle :1990)
143-149	Jour	Répartition des comptages par heure et valeur de l'espérance (Rolle :1998)
150	Rolle	Lien entre comptages et estimations
151-157	Jour	Répartition des comptages par heure et valeur de l'espérance (Gotthard :1990)
158-164	Jour	Répartition des comptages par heure et valeur de l'espérance (Gotthard :1998)
165-166	Année	Estimations horaires selon le jour (Gotthard)
167-173	Jour	Estimations horaires selon l'année (Gotthard)
174-175	Heures, jours	Influence du nombre d'années sur la valeur des paramètres

Graphique 1

Graphique 2

Graphique 3

Graphique 4

Graphique 5

Graphique 6

Graphique 7

Graphique 8

Graphique 9

Graphique 10

Graphique 11

Graphique 12

Graphique 13

Graphique 14

Graphique 15

Graphique 16

Graphique 17

Graphique 18

Graphique 19

Graphique 20

Graphique 21

Graphique 22

Graphique 23

Graphique 24

Graphique 25

Graphique 26

Graphique 27

Graphique 28

Graphique 29

Graphique 30

Graphique 31

Graphique 32

Graphique 33

Graphique 34

Graphique 35

Graphique 36

Graphique 37

Graphique 38

Graphique 39

Graphique 40

Graphique 41

Graphique 42

Graphique 43

Graphique 44

Graphique 45

Graphique 46

Graphique 47

Graphique 48

Graphique 49

Graphique 50

Graphique 51

Graphique 52

Graphique 53

## Chapitre 8

# Annexe II : la méthode du maximum de vraisemblance

### 8.1 Modèle de Davis *et al.*

#### 8.1.1 L'algorithme des innovations appliqué aux processus $AR(p)$

1. Le cas  $m = 1$  :  $W_i = \tau_1 W_{i-1} + N_i$

On a que  $\sigma_W(0) = \frac{\sigma_N^2}{1-\tau_1^2}$  et par suite,  $\underline{Q}_n$  désignant un vecteur dont les  $n$  composantes valent 0,

$$\Sigma_{\underline{Y}_n} = \begin{bmatrix} \frac{\sigma_W(0)}{\sigma_N^2} & \underline{0}_{n-1}^t \\ \underline{Q}_{n-1} & I_{n-1} \end{bmatrix} = \begin{bmatrix} \frac{1}{1-\tau_1^2} & \underline{0}_{n-1}^t \\ \underline{Q}_{n-1} & I_{n-1} \end{bmatrix}.$$

On voit alors que les coefficients  $\tilde{\alpha}_i^{(j)}$  sont tous nuls et que

$$\tilde{s}_0^2 = \frac{1}{1-\tau_1^2}, \quad \tilde{s}_i = 1, \quad i > 1.$$

Par suite

$$\hat{s}_0^2 = \frac{\sigma_N^2}{1-\tau_1^2}, \quad \hat{s}_i^2 = \sigma_N^2, \quad i > 1,$$

et

$$|\Sigma_{\underline{W}_n}(\tau_1, \sigma_N^2)| = \frac{\sigma_N^{2n}}{1-\tau_1^2}.$$

2. Le cas  $m = 2$  :  $W_i = \tau_1 W_{i-1} + \tau_2 W_{i-2} + N_i$

On a que

$$\sigma_W(0) = \frac{1-\tau_2}{1-\tau_1} \times \frac{\sigma_N^2}{(1-\tau_2)^2 - \tau_1^2},$$

$$\begin{aligned}
\sigma_W(1) &= \frac{\tau_1}{1-\tau_2} \times \sigma_W(0) \\
&= \frac{\tau_1}{1-\tau_1} \times \frac{\sigma_N^2}{(1-\tau_2)^2 - \tau_1^2}.
\end{aligned}$$

Soit  $M_2$  la matrice

$$M_2 = \begin{bmatrix} \frac{1-\tau_2}{1-\tau_1} \times \frac{1}{(1-\tau_2)^2 - \tau_1^2} & \frac{\tau_1}{1-\tau_1} \times \frac{1}{(1-\tau_2)^2 - \tau_1^2} \\ \frac{\tau_1}{1-\tau_1} \times \frac{1}{(1-\tau_2)^2 - \tau_1^2} & \frac{1-\tau_2}{1-\tau_1} \times \frac{1}{(1-\tau_2)^2 - \tau_1^2} \end{bmatrix} := \begin{bmatrix} m_{1,1}^{(2)} & m_{1,2}^{(2)} \\ m_{1,2}^{(2)} & m_{1,1}^{(2)} \end{bmatrix}.$$

Alors

$$\Sigma_{\underline{Y}_n} = \begin{bmatrix} M_2 & \underline{0}_{n-2}^t \\ \underline{0}_{n-2} & I_{n-2} \end{bmatrix}.$$

Par suite,

$$\begin{aligned}
\tilde{s}_0^2 &= m_{1,1}^{(2)}, \\
\tilde{\alpha}_1^{(1)} &= \frac{m_{1,2}^{(2)}}{m_{1,1}^{(2)}} = \frac{\tau_1}{1-\tau_2}, \\
\tilde{s}_1^2 &= m_{1,1}^{(2)}, \\
\tilde{\alpha}_2^{(2)} &= 0, \\
\tilde{\alpha}_1^{(2)} &= 0, \\
\tilde{s}_2^2 &= 1.
\end{aligned}$$

On a finalement que

$$|\Sigma_{\underline{W}_n}(\tau_1, \tau_2, \sigma_N^2)| = \sigma_N^{2n} \times \frac{1-\tau_2}{1-\tau_1} \times \frac{1}{(1-\tau_2)^2 - \tau_1^2}.$$

3. Le cas  $m > 2$  :

Il faut utiliser les formules de Yule-Walker pour obtenir l'équivalent de la matrice  $M_2$ , puis appliquer l'algorithme des innovations. Quand  $\Sigma_{\underline{W}_m}$  est la matrice des covariances de  $W_1, \dots, W_m$ , dont les éléments ont la forme  $\sigma_W(i-j)$ , et  $\underline{\alpha}_m^W$  le vecteur d'éléments

$$\sigma_W(1), \dots, \sigma_W(m),$$

les équations de Yule-Walker s'expriment sous la forme

$$\begin{aligned}\Sigma_{\underline{W}_m} \times \underline{\tau}_m &= \underline{\sigma}_m^W, \\ \sigma_N^2 &= \sigma_W(0) - \langle \underline{\tau}_m, \underline{\sigma}_m^W \rangle_{\mathbb{R}^m}.\end{aligned}$$

Ayant exprimé la variance  $\sigma_W(0)$  et le vecteur  $\underline{\sigma}_m^W$  comme fonctions des paramètres  $\underline{\tau}_m$  et  $\sigma_N^2$ , on calcule la matrice  $\Sigma_{\underline{W}_m}$  à l'aide de l'expression

$$\Sigma_{\underline{W}_m} = \sigma_W(0) I_m + \sum_{i=1}^m \sigma_W(i) \left[ B^i + (B^t)^i \right],$$

où  $B$  est la matrice dont tous les coefficients sont nuls, sauf ceux qui se trouvent dans les positions  $(i, i+1)$ ,  $1 \leq i \leq m-1$ .

### 8.1.2 Formules pour l'inverse de la matrice d'autocovariance d'un processus autorégressif

Ces formules sont tirées de :

R.F. Galbraith and J.I. Galbraith, On the inverses of some patterned matrices arising in the theory of stationary time series, J. Appl. Prob. 11 (1974), 63-71.

$W_1, \dots, W_n$  représentent les valeurs 1 à  $n$  d'un processus  $AR(p)$  d'espérance nulle, pour lequel  $n \gg m$  :

$$W_n = \tau_1 W_{n-1} + \dots + \tau_m W_{n-m} + N_n.$$

$\Sigma_{\underline{W}_n}$  est la matrice de covariance de  $\underline{W}_n$ , le vecteur dont les composantes sont  $W_1, \dots, W_n$ . On pose :

$$\Sigma_{\underline{W}_n}^{-1} = \frac{1}{\sigma_N^2} \Sigma_n$$

et on cherche une expression explicite de  $\Sigma_n$  en fonction de  $\tau_1, \dots, \tau_m$ . Comme  $\Sigma_n$  est symétrique, on peut se contenter de donner la valeur des composantes  $\gamma_{i,j}$ ,  $j \geq i$ . Or, si  $j - i > m$ ,  $\gamma_{i,j} = 0$ . On peut donc se contenter de donner les valeurs

$$\gamma_{i,i}, \gamma_{i,i+1}, \dots, \gamma_{i,i+m}.$$

De plus, tant que  $n+1 > 2m$ , ce qui est le cas ici,

$$\gamma_{i,j} = \sum_{k=0}^{n_{i,j}} \tau_k \tau_{k+j-i}, \text{ pour } n_{i,j} = \min \{i-1, m+i-j, n-j\}.$$

Comme il faut tenir compte de la valeur de  $m$ , on va donner explicitement la matrice pour  $m = 1, 2, 3, \dots$ . On pose :  $\tau_0 = -1$ ,  $\tau_l = 0$  pour  $l > m$ .

1. Cas  $m = 1$  :

$i$	$j$	$n_{i,j}$	$\gamma_{i,j}$	
1	1	0	$\tau_0^2$	= 1
	2	0	$\tau_0 \tau_1$	= $-\tau_1$
	3	-1	0	
	4	-2	0	
	$\vdots$	$\vdots$	$\vdots$	
2	2	1	$\tau_0^2 + \tau_1^2$	= $1 + \tau_1^2$
	3	0	$\tau_0 \tau_1$	= $-\tau_1$
	4	-1	0	
	$\vdots$	$\vdots$	$\vdots$	
	$\vdots$	$\vdots$	$\vdots$	
$n-1$	$n-1$	1	$\tau_0^2 + \tau_1^2$	= $1 + \tau_1^2$
	$n$	0	$\tau_0 \tau_1$	= $-\tau_1$
$n$	$n$	0	$\tau_0^2$	= 1

2. Cas  $m = 2$  :

$i$	$j$	$n_{i,j}$	$\gamma_{i,j}$	
1	1	0	$\tau_0^2$	= 1
	2	0	$\tau_0 \tau_1$	= $-\tau_1$
	3	0	$\tau_0 \tau_2$	= $-\tau_2$
	4	-1	0	
	$\vdots$	$\vdots$	$\vdots$	
2	2	1	$\tau_0^2 + \tau_1^2$	= $1 + \tau_1^2$
	3	1	$\tau_0 \tau_1 + \tau_1 \tau_2$	= $\tau_1 (\tau_2 - 1)$
	4	0	$\tau_0 \tau_2$	= $-\tau_2$
	5	-1	0	
	$\vdots$	$\vdots$	$\vdots$	
3	3	2	$\tau_0^2 + \tau_1^2 + \tau_2^2$	= $1 + \tau_1^2 + \tau_2^2$
	4	1	$\tau_0 \tau_1 + \tau_1 \tau_2$	= $\tau_1 (\tau_2 - 1)$
	5	0	$\tau_0 \tau_2$	= $-\tau_2$
	6	-1	0	
	$\vdots$	$\vdots$	$\vdots$	
4	4	2	$\tau_0^2 + \tau_1^2 + \tau_2^2$	= $1 + \tau_1^2 + \tau_2^2$
	5	1	$\tau_0 \tau_1 + \tau_1 \tau_2$	= $\tau_1 (\tau_2 - 1)$
	6	0	$\tau_0 \tau_2$	= $-\tau_2$
	7	-1	0	
	$\vdots$	$\vdots$	$\vdots$	
$n-2$	$n-2$	2	$\tau_0^2 + \tau_1^2 + \tau_2^2$	= $1 + \tau_1^2 + \tau_2^2$
	$n-1$	1	$\tau_0 \tau_1 + \tau_1 \tau_2$	= $\tau_1 (\tau_2 - 1)$
	$n$	0	$\tau_0 \tau_2$	= $-\tau_2$
$n-1$	$n-1$	1	$\tau_0^2 + \tau_1^2$	= $1 + \tau_1^2$
	$n$	0	$\tau_0 \tau_1$	= $-\tau_1$
$n$	$n$	0	$\tau_0^2$	= 1

3. Cas  $m = 3$  :

$i$	$j$	$n_{i,j}$	$\gamma_{i,j}$
1	1	0	$\tau_0^2$ = 1
	2	0	$\tau_0 \tau_1$ = $-\tau_1$
	3	0	$\tau_0 \tau_2$ = $-\tau_2$
	4	0	$\tau_0 \tau_3$ = $-\tau_3$
	5	-1	0
	⋮	⋮	⋮
2	2	1	$\tau_0^2 + \tau_1^2$ = $1 + \tau_1^2$
	3	1	$\tau_0 \tau_1 + \tau_1 \tau_2$ = $\tau_1 \tau_2 - \tau_1$
	4	1	$\tau_0 \tau_2 + \tau_1 \tau_3$ = $\tau_1 \tau_3 - \tau_2$
	5	0	$\tau_0 \tau_3$ = $-\tau_3$
	6	-1	0
	⋮	⋮	⋮
3	3	2	$\tau_0^2 + \tau_1^2 + \tau_2^2$ = $1 + \tau_1^2 + \tau_2^2$
	4	2	$\tau_0 \tau_1 + \tau_1 \tau_2 + \tau_2 \tau_3$ = $\tau_2 \tau_3 + \tau_1 \tau_2 - \tau_1$
	5	1	$\tau_0 \tau_2 + \tau_1 \tau_3$ = $\tau_1 \tau_3 - \tau_2$
	6	0	$\tau_0 \tau_3$ = $-\tau_3$
	7	-1	0
	⋮	⋮	⋮
4	4	3	$\tau_0^2 + \tau_1^2 + \tau_2^2 + \tau_3^2$ = $1 + \tau_1^2 + \tau_2^2 + \tau_3^2$
	5	2	$\tau_0 \tau_1 + \tau_1 \tau_2 + \tau_2 \tau_3$ = $\tau_2 \tau_3 + \tau_1 \tau_2 - \tau_1$
	6	1	$\tau_0 \tau_2 + \tau_1 \tau_3$ = $\tau_1 \tau_3 - \tau_2$
	7	0	$\tau_0 \tau_3$ = $-\tau_3$
	8	-1	0
	⋮	⋮	⋮
5	5	3	$\tau_0^2 + \tau_1^2 + \tau_2^2 + \tau_3^2$ = $1 + \tau_1^2 + \tau_2^2 + \tau_3^2$
	6	2	$\tau_0 \tau_1 + \tau_1 \tau_2 + \tau_2 \tau_3$ = $\tau_2 \tau_3 + \tau_1 \tau_2 - \tau_1$
	7	1	$\tau_0 \tau_2 + \tau_1 \tau_3$ = $\tau_1 \tau_3 - \tau_2$
	8	0	$\tau_0 \tau_3$ = $-\tau_3$
	⋮	⋮	⋮
	⋮	⋮	⋮
$n-3$	$n-3$	3	$\tau_0^2 + \tau_1^2 + \tau_2^2 + \tau_3^2$ = $1 + \tau_1^2 + \tau_2^2 + \tau_3^2$
	$n-2$	2	$\tau_0 \tau_1 + \tau_1 \tau_2 + \tau_2 \tau_3$ = $\tau_2 \tau_3 + \tau_1 \tau_2 - \tau_1$
	$n-1$	1	$\tau_0 \tau_2 + \tau_1 \tau_3$ = $\tau_1 \tau_3 - \tau_2$
	$n$	0	$\tau_0 \tau_3$ = $-\tau_3$
	⋮	⋮	⋮
	⋮	⋮	⋮
$n-2$	$n-2$	2	$\tau_0^2 + \tau_1^2 + \tau_2^2$ = $1 + \tau_1^2 + \tau_2^2$
	$n-1$	1	$\tau_0 \tau_1 + \tau_1 \tau_2$ = $\tau_1 \tau_2 - \tau_1$
	$n$	0	$\tau_0 \tau_2$ = $-\tau_2$
	⋮	⋮	⋮
	⋮	⋮	⋮
	⋮	⋮	⋮
$n-1$	$n-1$	1	$\tau_0^2 + \tau_1^2$ = $1 + \tau_1^2$
	$n$	0	$\tau_0 \tau_1$ = $-\tau_1$
	⋮	⋮	⋮
	⋮	⋮	⋮
	⋮	⋮	⋮
	⋮	⋮	⋮
$n$	$n$	0	$\tau_0^2$ = 1

Le cas de  $\Sigma_m$

C'est un cas utile, puisque  $|\Sigma_n| = |\Sigma_m|$  et que l'on doit calculer ce déterminant.

Les formules sont alors les suivantes ( $1 \leq i \leq j \leq m$ ) :

$$\gamma_{i,j} = \sum_{k=0}^{i-1} \tau_k \tau_{k+j-i} - \sum_{k=m+1-j}^{m+i-j} \tau_k \tau_{k+j-i}.$$

Cela donne :

1. Le cas  $m = 1$  :

$i$	$j$	$i-1$	$m+1-j$	$m+i-j$	$\gamma_{i,j}$
1	1	0	1	1	$\tau_0^2 - \tau_1^2 = 1 - \tau_1^2$

2. Le cas  $m = 2$  :

$i$	$j$	$i-1$	$m+1-j$	$m+i-j$	$\gamma_{i,j}$
1	1	0	2	2	$\tau_0^2 - \tau_2^2 = 1 - \tau_2^2$
1	2	0	1	1	$\tau_0 \tau_1 - \tau_1 \tau_2 = -\tau_1(1 + \tau_2)$
2	2	1	1	2	$\tau_0^2 + \tau_1^2 - \tau_1^2 - \tau_2^2 = 1 - \tau_2^2$

3. Le cas  $m = 3$  :

$i$	$j$	$i-1$	$m+1-j$	$m+i-j$	$\gamma_{i,j}$
1	1	0	3	3	$\tau_0^2 - \tau_3^2 = 1 - \tau_3^2$
1	2	0	2	2	$\tau_0 \tau_1 - \tau_2 \tau_3 = -\tau_1 - \tau_2 \tau_3$
1	3	0	1	1	$\tau_0 \tau_2 - \tau_1 \tau_3 = -\tau_2 - \tau_1 \tau_3$
2	2	1	2	3	$\tau_0^2 + \tau_1^2 - \tau_2^2 - \tau_3^2 = 1 - \tau_1^2 - \tau_2^2 - \tau_3^2$
2	3	1	1	2	$\tau_0 \tau_1 + \tau_1 \tau_2 - \tau_1 \tau_2 - \tau_2 \tau_3 = -\tau_1 - \tau_2 \tau_3$
3	3	2	1	3	$\tau_0^2 + \tau_1^2 + \tau_2^2 - \tau_1^2 - \tau_2^2 - \tau_3^2 = 1 - \tau_3^2$

### 8.1.3 L'approximation "standard" de la matrice des auto-covariances d'un processus stationnaire

La densité spectrale du processus autorégressif d'ordre  $m$ ,  $W$ , est une fonction explicite des paramètres du processus puisque, avec la définition

$$\tau(z) = 1 - \tau_1 z - \dots - \tau_m z^m,$$

$$f_W(\lambda) = \frac{\sigma_N^2}{2\pi |\tau(e^{-i\lambda})|^2}.$$

Quand on introduit les expressions suivantes :

$$\omega_k^{(n)} = \frac{2\pi k}{n},$$

$$\underline{u}_0^t = \sqrt{\frac{1}{2}} [1 \ 1 \ 1 \ \dots \ 1],$$

$$\underline{u}_k^t = \sqrt{\frac{2}{n}} [1 \ \cos(\omega_k^{(n)}) \ \cos(2\omega_k^{(n)}) \ \cos(3\omega_k^{(n)}) \ \dots \ \cos([n-1]\omega_k^{(n)})],$$

$$\underline{u}_k^t = \sqrt{\frac{2}{n}} \left[ 1 \quad \sin(\omega_k^{(n)}) \quad \sin(2\omega_k^{(n)}) \quad \sin(3\omega_k^{(n)}) \quad \cdots \quad \sin([n-1]\omega_k^{(n)}) \right],$$

$$P_n^t = \left[ \underline{u}_0 \quad \underline{u}_1 \quad \underline{u}_1 \quad \underline{u}_2 \quad \underline{u}_2 \quad \cdots \quad \underline{u}_{[\frac{n}{2}]} \quad \underline{u}_{[\frac{n}{2}]} \right],$$

$$D_n = \begin{cases} \text{si } n \text{ est pair :} \\ \text{diag} \left[ f_W(0), f_W(\omega_1^{(n)}), f_W(\omega_1^{(n)}), \dots, f_W(\omega_{[\frac{n}{2}]}^{(n)}), f_W(\omega_{[\frac{n}{2}]}^{(n)}) \right], \\ \text{si } n \text{ est impair :} \\ \text{diag} \left[ f_W(0), f_W(\omega_1^{(n)}), f_W(\omega_1^{(n)}), \dots, f_W(\omega_{[\frac{n-2}{2}]}^{(n)}), f_W(\omega_{[\frac{n-2}{2}]}^{(n)}) \right], \end{cases}$$

on obtient que  $P_n$  est une matrice orthogonale et que  $P_n \Sigma_{\underline{W}_n} P_n^t - 2\pi D_n$  converge uniformément vers une matrice infinie d'éléments nuls.

#### 8.1.4 Illustration de la méthode du balayage

Soit la matrice

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix},$$

et soit

$$A^{(i)} = \begin{bmatrix} a_{1,1}^{(i)} & a_{1,2}^{(i)} \\ a_{2,1}^{(i)} & a_{2,2}^{(i)} \end{bmatrix},$$

le résultat du “balayage sur”  $a_{i,i}$  ( $A^{(0)} = A$ ). Alors

1. *Etape 1* : On a

$$a_{1,1}^{(1)} = -\frac{1}{a_{1,1}}, \quad a_{1,2}^{(1)} = \frac{a_{1,2}}{a_{1,1}}, \quad a_{2,1}^{(1)} = \frac{a_{2,1}}{a_{1,1}}, \quad a_{2,2}^{(1)} = a_{2,2} - \frac{a_{2,1}a_{1,2}}{a_{1,1}}.$$

2. *Etape 2* : On a

$$a_{1,1}^{(2)} = a_{1,1}^{(1)} - \frac{a_{1,2}^{(1)}a_{2,1}^{(1)}}{a_{2,2}^{(1)}}, \quad a_{1,2}^{(2)} = \frac{a_{1,2}^{(1)}}{a_{2,2}^{(1)}}, \quad a_{2,1}^{(2)} = \frac{a_{2,1}^{(1)}}{a_{2,2}^{(1)}}, \quad a_{2,2}^{(2)} = -\frac{1}{a_{2,2}^{(1)}}.$$

3. *Déterminant* : Il vaut :

$$a_{1,1}^{(0)} \times a_{2,2}^{(1)} = a_{1,1}^{(0)} \left( a_{2,2}^{(0)} - \frac{a_{2,1}^{(0)}a_{1,2}^{(0)}}{a_{1,1}^{(0)}} \right) = a_{1,1}a_{2,2} - a_{1,2}a_{2,1}.$$

## 8.1.5 Code MATLAB

### MATLAB script

```
% -----  
%                               SCRIPT FOR DAVIS' METHOD  
% -----  
  
global Tsdata n  
  
% -----  
%                               Table of variables  
% -----  
% CONVENTION : capital letters for vectors and matrices, non-capitals for scalars  
% NOTATION   : exactly that of the paper  
  
%   Tsdata, n           data matrix Tsdata(n,?), sample size n.  
%                       The matrix has at least four columns  
%                       column 1 : counts (response)  
%                       column 2 : year of counts  
%                       column 3 : day type of counts  
%                       column 4 : hour of counts  
%                       and any additional column is ignored.  
%                       This matrix is deleted after definition of the  
%                       two following matrices :  
%  
%   X, p               design matrix X(p,n) and nb of parameters p  
%   Y                 response matrix Y(n,1)  
%  
%   W, m, Tau_sig2    auto-regressive process W(n,1), order m and  
%                       ARP' parameters Tau_sig2(m+1,1) (m-vector Tau followed by  
%                       noise-noise variance sigmaN2)  
%  
%   nbeval_L , nb_eval_T  maximum number of evaluations for search of max of L and T  
%  
%   nb_iter_W         maximum number of iterations for the calculation of W  
%  
%   Theta, Lambda     parameters Theta(p,1) and result Lambda(n,1) = exp(X'*Theta)  
%  
%   Blambda           intermediate calculation Blambda(n,1) = exp(W).*Lambda  
%  
%   Ytilde            intermediate calculation Ytilde(n,1) = Y+Blambda.*(W-ones(n,1))  
%  
%   Auxiliary variables : str, Tmp, tau, sig, i, w, Theta_0, Tau_sig2_0, options, fval, exitflag,output M, Inv  
% -----
```

```

%-----
%                               Table of functions
%-----
%
% design(Tsdata,n)              → X
%   uses information in Tsdata about
%   year, type of day and hour
%   to produce a fully parametrized
%   design matrix
%
% ms_L_zero(Theta,W,X,Y)        → ms_likelihood
%   minus the function L
%
% ms_T_un(Tau_sig2,m,n,Blambda,Ytilde) → ms_likelihood
%   minus the function T
%
% inv_AR_cov(Tau_sig2,m,n)      → M
%   inverse of auto-covariance for
%   order m and parameters
%   Tau_sig2 AR process
%-----

%-----
%                               Step 0 : load the initial data
%-----
%
str      = input('Please enter the name of
the (ascii) data file (within quotes) : ');
Tsdata  = load(str, 'ascii'); % here dim(Tsdata) = (78888,5)
disp('***** Step 0 complete *****');

%-----
%                               Step 1 :
%                               get the number of counts
%                               the design matrix
%                               ( here : saturated model
%                               24 for hour, 6 for day, 8 for year )
%                               and the response matrix
%-----

Tmp      = size(Tsdata);
n        = Tmp(1);      % sample size

X        = design;      % here X_n is a (38,78888)-matrix
Y        = Tsdata(:,1); % here response is a (78888,1)-matrix

Tmp      = size(X);
p        = Tmp(1);      % number of parameters

clear Tsdata;          % saves memory
disp('***** Step 1 complete *****');

```

```

%-----
% Step 2 : Initialisation of m, Tau_sig2, W and Theta
%-----
    m      = 1;          % order of AR process
    Tau_sig2 = [0.9 0.1]; % m-vector Tau followed by scalar sigmaN2

% simulation of W
% WARNING : to be rewritten in the case m<>1 :
    tau     = Tau_sig2(1);
    sig     = sqrt(Tau_sig2(m+1));

    w       = 0;
    for i   = 1:100
        w   = tau*w + sig*randn;
    end;

    W       = zeros(n,1);
    for i   = 1:n
        w   = tau*w + sig*randn;
        W(i) = w;
    end;

    Theta   = ones(p,1);

disp('***** Step 2 complete *****');

```

```

% -----
% Step 3 : Iterative estimations
% -----
nb_eval_L = 10000;
nb_eval_T = 100;
nb_iter_W = 10;

% WHILE not convergence do iteratively the following
% (Up to now only simple iteration was tested and min research is problematic)
Theta_0 = Theta;
% a priori value for the minimum
options = optimset('MaxFunEvals',nb_eval_L);
% 'Display','iter');
[Theta,fval,exitflag,output] = fminsearch('ms_L_zero',Theta_0,options,W,X,Y);

Lambda = exp(X'*Theta);
Blambda = exp(W).*Lambda;
Ytilde = Y + Blambda.*(W-ones(n,1));

Tau_sig2_0 = Tau_sig2;
% a priori value for the minimum
options = optimset('MaxFunEvals',nb_eval_T);
[Tau_sig2,fval,exitflag,output] = fminsearch('ms_T_un',Tau_sig2_0,options,m,n,Blambda,Ytilde);
Tau_sig2(m+1) = abs(Tau_sig2(m+1));
% cf the absolute value in ms_T_un

Inv = inv_AR_cov(Tau_sig2,m,n);
sig2 = Tau_sig2(m+1);

M = Inv + sig2*spdiags(Blambda,[0],n,n);
W = sig2*(M\Ytilde);

% To be completed : iteration for W to be put here

disp('***** Step 3 complete *****');

```

### La fonction : ms\_L\_zero

```
%-----  
function ms_likelihood = ms_L_zero(Theta,W,X,Y)  
%-----  
    Tmp      = Theta'*X;  
    ms_likelihood = sum(exp(Tmp'+W))-Tmp*Y;  
    clear Tmp;
```

### La fonction : ms\_T\_un

```
%-----  
function ms_likelihood = ms_T_un(Tau_sig2,m,n,Blambda,Ytilde)  
%-----  
% absolute value below for sigmaN2 > 0  
    sig2 = abs(Tau_sig2(m+1));  
  
    M = inv_AR_cov(Tau_sig2,m,n);  
    d1 = det(M);  
    M = M + sig2*spdiags(Blambda,[0],n,n);  
    d2 = det(M);  
  
% disp('det of inv_AR_cov is :');  
% disp(d1);  
% disp(d2);  
  
    Tmp      = M\Ytilde;  
    clear M;  
    s        = Ytilde'*Tmp;  
    clear Tmp;  
    s        = sig2*s;  
  
    disp(sprintf('s = %f, d2 = %g, d1 = %g, ln = %f',-s,d2,d1,log(d2/d1)));  
    % Tau = %f, sig2 = %f Tau_sig2(1),sig2  
  
    ms_likelihood = log(d2/d1)-s;
```

### La fonction :inv\_AR\_cov

```
% -----  
%           function M = inv_AR_cov(Tau_sig2,m,n)  
% -----  
% inverse of the covariance matrix of an order m AR process  
% of parameters Tau_sig2(i), 1<=i<=m, IN THE CASE sigmaN2 = 1  
  
if      m      ==      1  
    B      =      [-Tau_sig2(1)*ones(n,1) [1; (1+Tau_sig2(1)^2)*ones(n-2,1);1  
                  -Tau_sig2(1)*ones(n,1)];  
    M      =      spdiags(B,[-1 0 1],n,n);  
clear B;  
else    disp('The case of AR processes of order bigger than one is not available yet');  
end;
```

## 8.2 Le modèle de Ledolter

### 8.2.1 Loi jointe des observations et du processus latent

On la calcule conditionnellement à la première valeur du processus latent. On a que

$$\begin{aligned}
 & P(W_i \in A_i, Y_i = k_i, 1 \leq i \leq n) \\
 &= \int_{A_1} \cdots \int_{A_n} dw_1 \cdots dw_n \\
 & f_{W_1}(w_1) f_{W_2, \dots, W_n | W_1}(w_2, \dots, w_n) \\
 & P(Y_1 = k_1, \dots, Y_n = k_n | W_1 = w_1, \dots, W_n = w_n),
 \end{aligned}$$

et il faut obtenir une expression explicite pour  $f_{W_2, \dots, W_n | W_1}$ . Or, par récursion, on obtient :

$$\begin{bmatrix} W_2 \\ W_3 \\ W_4 \\ \vdots \\ W_n \end{bmatrix} = \rho W_1 \begin{bmatrix} 1 \\ \rho \\ \rho^2 \\ \vdots \\ \rho^{n-2} \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ \rho & 1 & 0 & 0 & \cdots & 0 \\ \rho^2 & \rho & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho^{n-2} & \rho^{n-3} & \rho^{n-4} & \rho^{n-5} & \cdots & 1 \end{bmatrix} \begin{bmatrix} B_2 \\ B_3 \\ B_4 \\ \vdots \\ B_n \end{bmatrix}.$$

On a donc

$$\underline{W}_n^{(1)} = \rho W_1 \underline{\rho}_{n-1} + R_{n-1} \underline{B}_n^{(1)},$$

si bien que

$$\underline{W}_n^{(1)} | W_1 = w_1 \sim N \left( \rho w_1 \underline{\rho}_{n-1}, \sigma_B^2 R_{n-1} R_{n-1}^t \right).$$

Pour obtenir la densité, il faut calculer

$$\begin{aligned}
 & \langle (\sigma_B^2 R_{n-1} R_{n-1}^t)^{-1} [\underline{w}_n^{(1)} - \rho w_1 \underline{\rho}_{n-1}], [\underline{w}_n^{(1)} - \rho w_1 \underline{\rho}_{n-1}] \rangle_{\mathbb{R}^{n-1}} \\
 &= \frac{1}{\sigma_B^2} \left\| R_{n-1}^{-1} [\underline{w}_n^{(1)} - \rho w_1 \underline{\rho}_{n-1}] \right\|_{\mathbb{R}^{n-1}}^2.
 \end{aligned}$$

Mais

$$R_{n-1}^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ -\rho & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & -\rho & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & -\rho & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & 1 & 0 \\ 0 & 0 & 0 & 0 & \cdots & -\rho & 1 \end{bmatrix},$$

si bien que

$$\left\| R_{n-1}^{-1} \left[ \underline{w}_n^{(1)} - \rho w_1 \underline{\rho}_{n-1} \right] \right\|_{\mathbb{R}^{n-1}}^2 = \sum_{i=1}^{n-1} (w_{i+1} - \rho w_i)^2.$$

Comme  $\det(R_{n-1} R_{n-1}^t) = 1$ , on a finalement :

$$f_{W_2, \dots, W_n | W_1}(w_2, \dots, w_n) P(Y_1 = k_1, \dots, Y_n = k_n | W_1 = w_1, \dots, W_n = w_n)$$

$$= \frac{e^{-\frac{1}{2\sigma_B^2} \sum_{i=1}^{n-1} (w_{i+1} - \rho w_i)^2}}{(2\pi\sigma_B^2)^{\frac{n-1}{2}}} \prod_{i=1}^n e^{-\lambda_i} \frac{\lambda_i^{k_i}}{k_i!}.$$

## 8.2.2 Les lois normales de l'échantillonneur de Gibbs

Leur calcul se fait comme suit. On a tout d'abord le fait que

$$f_{\underline{W}_n}(w_1, \dots, w_n) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma_{\underline{W}_n}|^{\frac{1}{2}}} e^{-\frac{1}{2} \langle \Sigma_{\underline{W}_n}^{-1} \underline{w}_n, \underline{w}_n \rangle_{\mathbb{R}^n}}$$

avec

$$\Sigma_{\underline{W}_n}^{-1} = \frac{1}{\sigma_N^2} \begin{bmatrix} 1 & -\rho & 0 & 0 & \dots & 0 & 0 & 0 \\ -\rho & 1 + \rho^2 & -\rho & 0 & \dots & 0 & 0 & 0 \\ 0 & -\rho & 1 + \rho^2 & \rho & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & -\rho & 1 + \rho^2 & \rho \\ 0 & 0 & 0 & 0 & \dots & 0 & -\rho & 1 \end{bmatrix},$$

ce qui donne

$$f_{\underline{W}_n}(w_1, \dots, w_n) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma_{\underline{W}_n}|^{\frac{1}{2}}} e^{-\frac{1}{2\sigma_N^2} \sum_{i=1}^{n-1} (w_i - \rho w_{i+1})^2}.$$

On utilise encore les notations suivantes :

$$\underline{W}_n^{(1)} = \begin{bmatrix} W_2 \\ \vdots \\ W_n \end{bmatrix}, \quad \underline{W}_n^{(i)} = \begin{bmatrix} W_1 \\ \vdots \\ W_{i-1} \\ W_{i+1} \\ \vdots \\ W_n \end{bmatrix}, \quad \text{si } 1 < i < n, \quad \underline{W}_n^{(n)} = \begin{bmatrix} W_1 \\ \vdots \\ W_{n-1} \end{bmatrix}.$$

Alors

$$f_{\underline{W}_n^{(1)}}(w_2, \dots, w_n) = \frac{\sqrt{2\pi}\sigma_N^2}{(2\pi)^{\frac{n}{2}} |\Sigma_{\underline{W}_n}|^{\frac{1}{2}}} e^{-\frac{1}{2\sigma_N^2} \sum_{i=2}^{n-1} (w_i - \rho w_{i+1})^2},$$

et

$$f_{W_1 | \underline{W}_n^{(1)}}(w_1 | w_2, \dots, w_n) = \frac{f_{\underline{W}_n}(w_1, w_2, \dots, w_n)}{f_{\underline{W}_n^{(1)}}(w_2, \dots, w_n)} = \frac{1}{\sqrt{2\pi}\sigma_N^2} e^{-\frac{1}{2\sigma_N^2} (w_1 - \rho w_2)^2}.$$

De même, en posant

$$\begin{aligned} g_n &:= g_n(w_1, \dots, w_{i-1}, w_{i+1}, \dots, w_n) \\ &= \frac{\sqrt{2\pi}\sigma_N^2}{(2\pi)^{\frac{n}{2}} |\Sigma_{\underline{W}_n}|^{\frac{1}{2}}} e^{-\frac{1}{2\sigma_N^2} \sum_{j=1}^{i-2} (w_j - \rho w_{j+1})^2 - \frac{1}{2\sigma_N^2} \sum_{j=i+1}^{n-1} (w_j - \rho w_{j+1})^2}, \end{aligned}$$

$$f_{\underline{W}_n^{(i)}}(w_1, \dots, w_{i-1}, w_{i+1}, \dots, w_n) = g_n \int_{\mathbb{R}} e^{-\frac{1}{2\sigma_N^2} \{(w_{i-1} - \rho w_i)^2 + (w_i - \rho w_{i+1})^2\}} dw_i.$$

Mais, avec le changement de variable  $x = \sqrt{1 + \rho^2} w_i$ , on obtient, par intégration,

$$\begin{aligned} f_{\underline{W}_n^{(i)}}(w_1, \dots, w_{i-1}, w_{i+1}, \dots, w_n) \\ = g_n \sqrt{\frac{2\pi\sigma_N^2}{1 + \rho^2}} e^{\frac{1}{2\sigma_N^2} \left\{ \frac{\rho^2 (w_{i-1} + w_{i+1})^2}{1 + \rho^2} - w_{i-1}^2 - \rho^2 w_{i+1}^2 \right\}}. \end{aligned}$$

Une simplification produit l'expression

$$f_{\underline{W}_n^{(i)}}(w_1, \dots, w_{i-1}, w_{i+1}, \dots, w_n) = g_n \sqrt{\frac{2\pi\sigma_N^2}{1 + \rho^2}} e^{-\frac{1}{2(1 + \rho^2)^2 \sigma_N^2} (w_{i-1} - \rho^2 w_{i+1})^2}.$$

Par suite

$$\begin{aligned} f_{W_i | \underline{W}_n^{(i)}}(w_i | w_1, \dots, w_{i-1}, w_{i+1}, \dots, w_n) \\ = \sqrt{\frac{1 + \rho^2}{2\pi\sigma_N^2}} e^{\frac{1}{2\sigma_N^2} \left\{ \frac{1}{1 + \rho^2} (w_{i-1} - \rho^2 w_{i+1})^2 - (w_{i-1} - \rho w_i)^2 - (w_i - \rho w_{i+1})^2 \right\}}. \end{aligned}$$

L'exposant de l'exponentielle se simplifie alors en

$$-\frac{1 + \rho^2}{2\sigma_N^2} \left\{ w_i - \frac{\rho}{1 + \rho^2} (w_{i-1} + w_{i+1}) \right\}^2.$$

On a finalement que

$$W_1 | \underline{W}_n^{(1)} \sim \mathcal{N}(\rho w_2, \sigma_N^2),$$

$$W_i | \underline{W}_n^{(i)} \sim \mathcal{N}\left(\frac{\rho}{1+\rho^2} (w_{i-1} + w_{i+1}), \frac{\sigma_N^2}{1+\rho^2}\right),$$

$$W_n | \underline{W}_n^{(n)} \sim \mathcal{N}(\rho w_{n-1}, \sigma_N^2).$$

## Chapitre 9

# Annexe III : la méthode du filtre de Kalman

### 9.1 Calcul des moments du modèle

Par hypothèse,

$$E \left[ \theta_i^{(j)} \mid \theta_0^{(j)}, \theta_1^{(j)}, \dots, \theta_{i-1}^{(j)} \right] = E \left[ \theta_i^{(j)} \mid \theta_{i-1}^{(j)} \right] = b_i^{(j)} \theta_{i-1}^{(j)},$$

et <sup>1</sup>

$$V \left[ \theta_i^{(j)} \mid \theta_0^{(j)}, \theta_1^{(j)}, \dots, \theta_{i-1}^{(j)} \right] = V \left[ \theta_i^{(j)} \mid \theta_{i-1}^{(j)} \right] = \left[ b_i^{(j)} \right]^2 \theta_{i-1}^{(j)} \eta^2.$$

Par ailleurs, si

$$\underline{a}_i^{(j)} = \begin{bmatrix} a_i^{j,1} \\ a_i^{j,2} \\ a_i^{j,3} \\ \vdots \\ a_i^{j,m} \end{bmatrix}, \quad A_i^{(j)} = \begin{bmatrix} a_i^{j,1} & 0 & 0 & \cdots & 0 \\ 0 & a_i^{j,2} & 0 & \cdots & 0 \\ 0 & 0 & a_i^{j,3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a_i^{j,m} \end{bmatrix},$$

$$E \left[ \underline{N}_i^{(j)} \mid \theta_0^{(j)}, \theta_1^{(j)}, \dots, \theta_i^{(j)} \right] = E \left[ \underline{N}_i^{(j)} \mid \theta_i^{(j)} \right] = \theta_i^{(j)} \underline{a}_i^{(j)}$$

et

$$V \left[ \underline{N}_i^{(j)} \mid \theta_0^{(j)}, \theta_1^{(j)}, \dots, \theta_i^{(j)} \right] = V \left[ \underline{N}_i^{(j)} \mid \theta_i^{(j)} \right] = \theta_i^{(j)} A_i^{(j)}.$$

---

1. Pour une loi Gamma,  $\sigma^2 = \mu^2 \kappa^2$ , si bien qu'ici  $\sigma^2 = \left[ b_i^{(j)} \theta_{i-1}^{(j)} \right]^2 \times \frac{\eta^2}{\theta_{i-1}^{(j)}}$ .

### 9.1.1 Espérance du processus latent

Par récursion,  $E \left[ \theta_i^{(j)} \right] = g_j \prod_{l=1}^i b_l^{(j)}$ , si bien que

$$\ln \left( E \left[ \theta_i^{(j)} \right] \right) = \ln(g_j) + \sum_{l=1}^i \langle \Delta \underline{v}_l^{(j)}, \underline{\beta}^{(j)} \rangle_{\mathbb{R}^{b_j}} = \langle \underline{w}^{(j)}, \underline{\gamma} \rangle_{\mathbb{R}^c} + \langle \underline{v}_i^{(j)}, \underline{\beta}^{(j)} \rangle_{\mathbb{R}^{b_j}}.$$

On écrira  $\mu_j^\theta(i) = E \left[ \theta_i^{(j)} \right]$ .

### 9.1.2 Variance du processus latent

On utilise à cette fin la relation

$$V[X] = E[V[X | Y]] + V[E[X | Y]]$$

avec

$$X = \theta_i^{(j)} \text{ et } Y = \theta_{i-1}^{(j)},$$

qui donne la récurrence

$$\begin{aligned} V \left[ \theta_i^{(j)} \right] &= E \left[ \left( b_i^{(j)} \right)^2 \theta_{i-1}^{(j)2} \eta^2 \right] + V \left[ b_i^{(j)} \theta_{i-1}^{(j)} \right] \\ &= \left( b_i^{(j)} \right)^2 \mu_j^\theta(i-1) \eta^2 + \left( b_i^{(j)} \right)^2 V \left[ \theta_{i-1}^{(j)} \right] \\ &= b_i^{(j)} \mu_j^\theta(i) \eta^2 + \left( b_i^{(j)} \right)^2 V \left[ \theta_{i-1}^{(j)} \right], \end{aligned}$$

car  $E \left[ \theta_i^{(j)} \right] = b_i^{(j)} E \left[ \theta_{i-1}^{(j)} \right]$ . Alors, puisque  $\mu_j^\theta(1) = g_j b_1^{(j)}$ ,

$$\begin{aligned} V \left[ \theta_0^{(j)} \right] &= g_j^2 \tau^2, \\ V \left[ \theta_1^{(j)} \right] &= b_1^{(j)} \mu_j^\theta(1) \eta^2 + \left( b_1^{(j)} \right)^2 V \left[ \theta_0^{(j)} \right] \\ &= b_1^{(j)} \mu_j^\theta(1) \eta^2 + \left( \mu_j^\theta(1) \right)^2 \tau^2, \\ V \left[ \theta_2^{(j)} \right] &= b_2^{(j)} \mu_j^\theta(2) \eta^2 + \left( b_2^{(j)} \right)^2 V \left[ \theta_1^{(j)} \right] \\ &= b_2^{(j)} \mu_j^\theta(2) \eta^2 + \left( b_2^{(j)} \right)^2 \left\{ b_1^{(j)} \mu_j^\theta(1) \eta^2 + \left( \mu_j^\theta(1) \right)^2 \tau^2 \right\} \\ &= \left\{ b_2^{(j)} + b_2^{(j)} b_1^{(j)} \right\} \mu_j^\theta(2) \eta^2 + \left( \mu_j^\theta(2) \right)^2 \tau^2, \\ V \left[ \theta_3^{(j)} \right] &= b_3^{(j)} \mu_j^\theta(3) \eta^2 + \left( b_3^{(j)} \right)^2 V \left[ \theta_2^{(j)} \right] \\ &= b_3^{(j)} \mu_j^\theta(3) \eta^2 + \left( b_3^{(j)} \right)^2 \left\{ \left\{ b_2^{(j)} + b_2^{(j)} b_1^{(j)} \right\} \mu_j^\theta(2) \eta^2 + \left( \mu_j^\theta(2) \right)^2 \tau^2 \right\} \\ &= \left\{ b_3^{(j)} + b_3^{(j)} b_2^{(j)} + b_3^{(j)} b_2^{(j)} b_1^{(j)} \right\} \mu_j^\theta(3) \eta^2 + \left( \mu_j^\theta(3) \right)^2 \tau^2, \end{aligned}$$

et ainsi de suite. Si  $\zeta_i^{(j)}$  et  $\xi_i^{(j)}$  sont définis respectivement par

$$\begin{aligned}\zeta_i^{(j)} &= b_i^{(j)} + b_i^{(j)} b_{i-1}^{(j)} + b_i^{(j)} b_{i-1}^{(j)} b_{i-2}^{(j)} + \dots + b_i^{(j)} b_{i-1}^{(j)} b_{i-2}^{(j)} \dots b_3^{(j)} b_2^{(j)} b_1^{(j)}, \\ \xi_i^{(j)} &= \frac{V[\theta_i^{(j)}]}{E[\theta_i^{(j)}]} = \zeta_i^{(j)} \eta^2 + \mu_j^\theta(i) \tau^2,\end{aligned}$$

alors

$$V[\theta_i^{(j)}] = \xi_i^{(j)} \mu_j^\theta(i) = \xi_i^{(j)} E[\theta_i^{(j)}].$$

### 9.1.3 Covariance et corrélation du processus latent

On a les récursions immédiates suivantes :

$$\begin{aligned}\text{cov}(\theta_i^{(j)}, \theta_{i+1}^{(j)}) &= E\left[\left(\theta_i^{(j)} - \mu_j^\theta(i)\right) E\left[\left(\theta_{i+1}^{(j)} - \mu_j^\theta(i+1) \mid \theta_i^{(j)}\right)\right]\right] \\ &= E\left[\left(\theta_i^{(j)} - \mu_j^\theta(i)\right) \left(b_{i+1} \theta_i^{(j)} - \mu_j^\theta(i+1)\right)\right] \\ &= b_{i+1}^{(j)} V[\theta_i^{(j)}] \\ &= b_{i+1}^{(j)} \left\{ \zeta_i^{(j)} \mu_j^\theta(i) \eta^2 + (\mu_j^\theta(i))^2 \tau^2 \right\} \\ &= \zeta_i^{(j)} \mu_j^\theta(i+1) \eta^2 + \mu_j^\theta(i) \mu_j^\theta(i+1) \tau^2,\end{aligned}$$

et

$$\begin{aligned}\text{cov}(\theta_i^{(j)}, \theta_{i+2}^{(j)}) &= E\left[\left(\theta_i^{(j)} - \mu_j^\theta(i)\right) E\left[\left(\theta_{i+2}^{(j)} - \mu_j^\theta(i+2) \mid \theta_{i+1}^{(j)}\right)\right]\right] \\ &= E\left[\left(\theta_i^{(j)} - \mu_j^\theta(i)\right) \left(b_{i+2} \theta_{i+1}^{(j)} - \mu_j^\theta(i+2)\right)\right] \\ &= b_{i+2}^{(j)} \text{cov}(\theta_i^{(j)}, \theta_{i+1}^{(j)}) \\ &= b_{i+2}^{(j)} \left\{ \zeta_i^{(j)} \mu_j^\theta(i+1) \eta^2 + \mu_j^\theta(i) \mu_j^\theta(i+1) \tau^2 \right\} \\ &= \zeta_i^{(j)} \mu_j^\theta(i+2) \eta^2 + \mu_j^\theta(i) \mu_j^\theta(i+2) \tau^2 \\ &= \xi_i^{(j)} E[\theta_{i+2}]\end{aligned}$$

En conséquence

$$\rho(\theta_i^{(j)}, \theta_{i+l}^{(j)}) = \frac{\xi_i^{(j)} \mu_j^\theta(i+l)}{\sqrt{\xi_i^{(j)} \mu_j^\theta(i)} \sqrt{\xi_{i+l}^{(j)} \mu_j^\theta(i+l)}} = \sqrt{\frac{\xi_i^{(j)} \mu_j^\theta(i+l)}{\xi_{i+l}^{(j)} \mu_j^\theta(i)}}.$$

*Remarque :*

Soit  $\hat{\theta}_i^{(j)} := E \left[ \theta_i^{(j)} \mid \theta_0^{(j)}, \theta_1^{(j)}, \dots, \theta_{i-1}^{(j)} \right] = b_i^{(j)} \theta_{i-1}^{(j)}$ . L'innovation  $I_i^{(j)}$  est définie par l'expression

$$I_i^{(j)} = \theta_i^{(j)} - \hat{\theta}_i^{(j)} = \theta_i^{(j)} - b_i^{(j)} \theta_{i-1}^{(j)}.$$

Si  $X$  est adaptée à la tribu  $\sigma \left( \theta_0^{(j)}, \theta_1^{(j)}, \dots, \theta_{i-1}^{(j)} \right)$ , alors

$$\begin{aligned} \text{cov} \left( I_i^{(j)}, X \right) &= \text{cov} \left( \theta_i^{(j)} - \hat{\theta}_i^{(j)}, X \right) \\ &= \text{cov} \left( \theta_i^{(j)}, X \right) - \text{cov} \left( \hat{\theta}_i^{(j)}, X \right) \\ &= \text{cov} \left( \hat{\theta}_i^{(j)}, X \right) - \text{cov} \left( \hat{\theta}_i^{(j)}, X \right) \\ &= 0. \end{aligned}$$

Donc

$$\text{cov} \left( I_{i+l}^{(j)}, I_i^{(j)} \right) = 0, \quad l \geq 1,$$

c'est-à-dire que les innovations ont une structure de corrélation qui est celle d'un processus  $AR(1)$ , que leurs covariances "strictes" sont nulles et que leurs variances sont égales à  $b_i^{(j)} \mu_j^\theta(i) \eta^2$ .

### 9.1.4 Espérance et variance des comptages

On a que

$$\begin{aligned} E \left[ \underline{N}_i^{(j)} \right] &= E \left[ E \left[ \underline{N}_i^{(j)} \mid \theta_i^{(j)} \right] \right] = E \left[ \theta_i^{(j)} \underline{a}_i^{(j)} \right] = \mu_j^\theta(i) \underline{a}_i^{(j)}, \\ V \left[ \underline{N}_i^{(j)} \right] &= E \left[ V \left[ \underline{N}_i^{(j)} \mid \theta_i^{(j)} \right] \right] + V \left[ E \left[ \underline{N}_i^{(j)} \mid \theta_i^{(j)} \right] \right] \\ &= E \left[ \theta_i^{(j)} A_i^{(j)} \right] + V \left[ \theta_i^{(j)} \underline{a}_i^{(j)} \right] \\ &= \mu_j^\theta(i) A_i^{(j)} + \underline{a}_i^{(j)} V \left[ \theta_i^{(j)} \right] \left( \underline{a}_i^{(j)} \right)^t \\ &= \mu_j^\theta(i) A_i^{(j)} + \xi_i^{(j)} \mu_j^\theta(i) \underline{a}_i^{(j)} \left( \underline{a}_i^{(j)} \right)^t \\ &= \mu_j^\theta(i) \left\{ A_i^{(j)} + \xi_i^{(j)} \underline{a}_i^{(j)} \left( \underline{a}_i^{(j)} \right)^t \right\}. \end{aligned}$$

Il s'en suit en particulier que, pour  $k \neq l$ ,

$$\rho \left( N_i^{(j,k)}, N_i^{(j,l)} \right) = \sqrt{\frac{\xi_i^{(j)} a_i^{(j,k)} a_i^{(j,l)}}{\left( 1 + \xi_i^{(j)} a_i^{(j,k)} \right) \left( 1 + \xi_i^{(j)} a_i^{(j,l)} \right)}}.$$

Sous forme “standard” le modèle a l’expression suivante :

$$\ln \left( E \left[ N_i^{(j,k)} \right] \right) = \langle \underline{\mathbf{u}}_i^{(j)}, \underline{\mathbf{a}}_k^{(j)} \rangle_{\mathbb{R}^{a_j}} + \langle \underline{\mathbf{u}}_i^{(j)}, \underline{\beta}^{(j)} \rangle_{\mathbb{R}^{b_j}} + \langle \underline{\mathbf{w}}^{(j)}, \underline{\gamma} \rangle_{\mathbb{R}^c}.$$

### 9.1.5 Corrélation du processus de comptage

On a, par indépendance, conditionnellement aux valeurs du processus latent,

$$\begin{aligned} \text{cov} \left( \underline{N}_i^{(j)}, \underline{N}_{i+1}^{(j)} \right) &= E \left[ \left( \underline{N}_i^{(j)} - E \left[ \underline{N}_i^{(j)} \right] \right) \left( \underline{N}_{i+1}^{(j)} - E \left[ \underline{N}_{i+1}^{(j)} \right] \right)^t \right] \\ &= E \left[ E \left[ \underline{N}_i^{(j)} - \mu_j^\theta(i) \underline{\mathbf{a}}_i^{(j)} \mid \underline{\theta}_i^{(j)} \right] \right. \\ &\quad \left. \times E \left[ \left( \underline{N}_{i+1}^{(j)} - \mu_j^\theta(i+1) \underline{\mathbf{a}}_{i+1}^{(j)} \right)^t \mid \underline{\theta}_i^{(j)} \right] \right]. \end{aligned}$$

Or,

$$\begin{aligned} E \left[ \underline{N}_i^{(j)} - \mu_j^\theta(i) \underline{\mathbf{a}}_i^{(j)} \mid \underline{\theta}_i^{(j)} \right] &= \underline{\theta}_i^{(j)} \underline{\mathbf{a}}_i^{(j)} - \mu_j^\theta(i) \underline{\mathbf{a}}_i^{(j)} \\ &= \left\{ \underline{\theta}_i^{(j)} - \mu_j^\theta(i) \right\} \underline{\mathbf{a}}_i^{(j)} \\ &= \left\{ \underline{\theta}_i^{(j)} - E \left[ \underline{\theta}_i^{(j)} \right] \right\} \underline{\mathbf{a}}_i^{(j)}, \end{aligned}$$

et

$$\begin{aligned} &E \left[ \left( \underline{N}_{i+1}^{(j)} - \mu_j^\theta(i+1) \underline{\mathbf{a}}_{i+1}^{(j)} \right)^t \mid \underline{\theta}_i^{(j)} \right] \\ &= E \left[ E \left[ \left( \underline{N}_{i+1}^{(j)} - \mu_j^\theta(i+1) \underline{\mathbf{a}}_{i+1}^{(j)} \right)^t \mid \underline{\theta}_{i+1}^{(j)} \right] \mid \underline{\theta}_i^{(j)} \right] \\ &= E \left[ \left( \underline{\theta}_{i+1}^{(j)} \underline{\mathbf{a}}_{i+1}^{(j)} - \mu_j^\theta(i+1) \underline{\mathbf{a}}_{i+1}^{(j)} \right)^t \mid \underline{\theta}_i^{(j)} \right] \\ &= \left\{ b_{i+1}^{(j)} \underline{\theta}_i^{(j)} - \mu_j^\theta(i+1) \right\} \left( \underline{\mathbf{a}}_{i+1}^{(j)} \right)^t \\ &= b_{i+1}^{(j)} \left\{ \underline{\theta}_i^{(j)} - E \left[ \underline{\theta}_i^{(j)} \right] \right\} \left( \underline{\mathbf{a}}_{i+1}^{(j)} \right)^t. \end{aligned}$$

Mais alors

$$\begin{aligned} \text{cov} \left( \underline{N}_i^{(j)}, \underline{N}_{i+1}^{(j)} \right) &= b_{i+1}^{(j)} V \left[ \underline{\theta}_i^{(j)} \right] \underline{\mathbf{a}}_i^{(j)} \left( \underline{\mathbf{a}}_{i+1}^{(j)} \right)^t \\ &= b_{i+1}^{(j)} \xi_i^{(j)} \mu_j^\theta(i) \underline{\mathbf{a}}_i^{(j)} \left( \underline{\mathbf{a}}_{i+1}^{(j)} \right)^t \\ &= \xi_i^{(j)} \mu_j^\theta(i+1) \underline{\mathbf{a}}_i^{(j)} \left( \underline{\mathbf{a}}_{i+1}^{(j)} \right)^t. \end{aligned}$$

Un calcul semblable donne

$$\begin{aligned} \text{cov} \left( \underline{N}_i^{(j)}, \underline{N}_{i+2}^{(j)} \right) &= E \left[ E \left[ \underline{N}_i^{(j)} - \mu_j^\theta(i) \underline{a}_i^{(j)} \mid \underline{\theta}_i^{(j)} \right] \right. \\ &\quad \left. \times E \left[ \left( \underline{N}_{i+2}^{(j)} - \mu_j^\theta(i+2) \underline{a}_{i+2}^{(j)} \right)^t \mid \underline{\theta}_i^{(j)} \right] \right], \end{aligned}$$

et

$$\begin{aligned} &E \left[ \left( \underline{N}_{i+2}^{(j)} - \mu_j^\theta(i+2) \underline{a}_{i+2}^{(j)} \right)^t \mid \underline{\theta}_i^{(j)} \right] \\ &= E \left[ E \left[ \left( \underline{N}_{i+2}^{(j)} - \mu_j^\theta(i+2) \underline{a}_{i+2}^{(j)} \right)^t \mid \underline{\theta}_{i+2}^{(j)} \right] \mid \underline{\theta}_i^{(j)} \right] \\ &= E \left[ \left( \underline{\theta}_{i+2}^{(j)} \underline{a}_{i+2}^{(j)} - \mu_j^\theta(i+2) \underline{a}_{i+2}^{(j)} \right)^t \mid \underline{\theta}_i^{(j)} \right] \\ &= E \left[ E \left[ \left( \underline{\theta}_{i+2}^{(j)} \underline{a}_{i+2}^{(j)} - \mu_j^\theta(i+2) \underline{a}_{i+2}^{(j)} \right)^t \mid \underline{\theta}_{i+1}^{(j)} \right] \mid \underline{\theta}_i^{(j)} \right] \\ &= E \left[ \left( b_{i+2}^{(j)} \underline{\theta}_{i+1}^{(j)} \underline{a}_{i+2}^{(j)} - \mu_j^\theta(i+2) \underline{a}_{i+2}^{(j)} \right)^t \mid \underline{\theta}_i^{(j)} \right] \\ &= b_{i+2}^{(j)} b_{i+1}^{(j)} \left\{ \underline{\theta}_i^{(j)} - \mu_j^\theta(i) \right\} \left( \underline{a}_{i+2}^{(j)} \right)^t. \end{aligned}$$

Par suite

$$\begin{aligned} \text{cov} \left( \underline{N}_i^{(j)}, \underline{N}_{i+2}^{(j)} \right) &= E \left[ \left( \underline{\theta}_i^{(j)} - E \left[ \underline{\theta}_i^{(j)} \right] \right) \underline{a}_i^{(j)} b_{i+2}^{(j)} b_{i+1}^{(j)} \left( \underline{\theta}_i^{(j)} - E \left[ \underline{\theta}_i^{(j)} \right] \right) \left( \underline{a}_{i+2}^{(j)} \right)^t \right] \\ &= b_{i+2}^{(j)} b_{i+1}^{(j)} V \left[ \underline{\theta}_i^{(j)} \right] \underline{a}_i^{(j)} \left( \underline{a}_{i+2}^{(j)} \right)^t \\ &= b_{i+2}^{(j)} b_{i+1}^{(j)} \xi_i^{(j)} \mu_j^\theta(i) \underline{a}_i^{(j)} \left( \underline{a}_{i+2}^{(j)} \right)^t \\ &= \xi_i^{(j)} \mu_j^\theta(i+2) \underline{a}_i^{(j)} \left( \underline{a}_{i+2}^{(j)} \right)^t. \end{aligned}$$

En particulier,

$$\rho \left( Y_i^{(j,k)}, Y_{i+m}^{(j,l)} \right) = \sqrt{\frac{\left( \xi_i^{(j)} \right)^2 \mu_j^\theta(i+m) a_i^{(j,k)} a_{i+m}^{(j,l)}}{\mu_j^\theta(i) \left( 1 + \xi_i^{(j)} a_i^{(j,k)} \right) \left( 1 + \xi_{i+m}^{(j)} a_{i+m}^{(j,l)} \right)}}.$$

De même,

$$\begin{aligned} \text{cov} \left( \underline{N}_{i+1}^{(j)}, \underline{\theta}_i^{(j)} \right) &= E \left[ \left( \underline{N}_{i+1}^{(j)} - E \left[ \underline{N}_{i+1}^{(j)} \right] \right) \left( \underline{\theta}_i^{(j)} - E \left[ \underline{\theta}_i^{(j)} \right] \right) \right] \\ &= E \left[ E \left[ \underline{N}_{i+1}^{(j)} - \mu_j^\theta(i+1) \underline{a}_{i+1}^{(j)} \mid \underline{\theta}_i^{(j)} \right] \left( \underline{\theta}_i^{(j)} - \mu_j^\theta(i) \right) \right] \end{aligned}$$

$$\begin{aligned}
&= E \left[ b_{j+1}^{(j)} \left\{ \theta_i^{(j)} - E \left[ \theta_i^{(j)} \right] \right\} \underline{a}_{i+1}^{(j)} \left\{ \theta_i^{(j)} - E \left[ \theta_i^{(j)} \right] \right\} \right] \\
&= b_{j+1}^{(j)} V \left[ \theta_i^{(j)} \right] \underline{a}_{i+1}^{(j)} \\
&= \xi_i^{(j)} \mu_j^\theta (i+1) \underline{a}_{i+1}^{(j)},
\end{aligned}$$

et

$$\begin{aligned}
\text{cov} \left( \underline{N}_{i+2}^{(j)}, \theta_i^{(j)} \right) &= E \left[ \left( \underline{N}_{i+2}^{(j)} - E \left[ \underline{N}_{i+2}^{(j)} \right] \right) \left( \theta_i^{(j)} - E \left[ \theta_i^{(j)} \right] \right) \right] \\
&= E \left[ E \left[ \underline{N}_{i+2}^{(j)} - \mu_j^\theta (i+2) \underline{a}_{i+2}^{(j)} \mid \theta_i^{(j)} \right] \left( \theta_i^{(j)} - \mu_j^\theta (i) \right) \right] \\
&= E \left[ b_{i+2}^{(j)} b_{i+1}^{(j)} \left( \theta_i^{(j)} - \mu_j^\theta (i) \right) \underline{a}_{i+2}^{(j)} \left( \theta_i^{(j)} - \mu_j^\theta (i) \right) \right] \\
&= b_{i+2}^{(j)} b_{i+1}^{(j)} V \left[ \theta_i^{(j)} \right] \underline{a}_{i+2}^{(j)} \\
&= b_{i+2}^{(j)} b_{i+1}^{(j)} \xi_i^{(j)} \mu_j^\theta (i) \underline{a}_{i+2}^{(j)} \\
&= \xi_i^{(j)} \mu_j^\theta (i+2) \underline{a}_{i+2}^{(j)}.
\end{aligned}$$

On a donc

$$\rho \left( \underline{N}_{i+m}^{(j,k)}, \theta_i^{(j)} \right) = \sqrt{\frac{\xi_i^{(j)} \mu_j^\theta (i+m) \underline{a}_{i+m}^{(j,k)}}{\mu_j^\theta (i) \left( 1 + \xi_{i+m}^{(j)} \underline{a}_{i+m}^{(j,k)} \right)}}.$$

On obtient de manière analogue

$$\text{cov} \left( \underline{N}_i^{(j)}, \theta_{i+m}^{(j)} \right) = \xi_i^{(j)} \mu_j^\theta (i+m) \underline{a}_i^{(j)}.$$

### 9.1.6 Innovations

Le processus d'innovation est défini comme suit :

$$\underline{J}_i^{(j)} := \underline{N}_i^{(j)} - E \left[ \underline{N}_i^{(j)} \mid \underline{\theta}^{(j)} \right] = \underline{N}_i^{(j)} - \theta_i^{(j)} \underline{a}_i^{(j)}.$$

On a que

$$V \begin{bmatrix} \underline{J}_1^{(j)} \\ \vdots \\ \underline{J}_{n_j}^{(j)} \end{bmatrix} = \begin{bmatrix} \theta_1^{(j)} A_i^{(j)} & \cdots & O \\ \vdots & \vdots & \vdots \\ O & \cdots & \theta_{n_j}^{(j)} A_{n_j}^{(j)} \end{bmatrix}$$

et que

$$\text{cov} \left( \underline{J}^{(j)}, \underline{\theta}^{(j)} \right) = O.$$

## 9.2 Les mathématiques du filtre de Kalman

La référence pour ce qui suit est : D.E. Catlin, Estimation, Control, and the Discrete Kalman Filter, Springer, New York (1989).

Dans ce qui suit,  $\mathcal{P} = (\Omega, \mathcal{A}, P)$  dénote un espace de probabilité, et  $L_2[\mathcal{P}]$  l'espace de Hilbert des (classes d'équivalence de) variables aléatoires de carré intégrable qui lui est associé.  $Y_1, \dots, Y_n$  sont des éléments de  $L_2[\mathcal{P}]$  représentés par le symbole  $\underline{Y}_n$ .  $\mathcal{F}_n$  est un espace vectoriel de fonctions  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  qui sont mesurables au sens de Borel et pour lesquelles on a  $f(\underline{Y}) \in L_2[\mathcal{P}]$ .  $L_2[\underline{Y}_n, \mathcal{F}_n] \subseteq L_2[\mathcal{P}]$  est l'ensemble des (classes d'équivalence de) variables aléatoires  $\{f(\underline{Y}_n), f \in \mathcal{F}_n\}$  et on suppose que c'est un sous-espace de Hilbert fermé de  $L_2[\mathcal{P}]$ .

On est amené à considérer les familles suivantes :

$$\begin{aligned} \mathcal{F}_n^{all} &:= \text{ensemble de toutes les fonctions possibles} \\ \mathcal{F}_n^{lin} &:= \text{ensemble des fonctions linéaires} \\ \mathcal{F}_n^{aff} &:= \text{ensemble des fonctions affines} \end{aligned}$$

Soit  $X \in L_2[\mathcal{P}]$  et soit  $\Pi_{L_2[\underline{Y}_n, \mathcal{F}_n]}(X)$  la projection orthogonale de  $X$  sur  $L_2[\underline{Y}_n, \mathcal{F}_n]$ . On a que :

$$\begin{aligned} \Pi_{L_2[\underline{Y}_n, \mathcal{F}_n^{all}]}(X) &= E[X | \sigma(\underline{Y}_n)] \\ \Pi_{L_2[\underline{Y}_n, \mathcal{F}_n^{lin}]}(X) &= \text{estimateur linéaire de variance minimum (BLMVE)} \\ \Pi_{L_2[\underline{Y}_n, \mathcal{F}_n^{aff}]}(X) &= \text{estimateur affine de variance minimum (BAMVE)} \end{aligned}$$

$L_2^{\oplus m}[\mathcal{P}]$  dénote la somme directe de  $m$  copies de  $L_2[\mathcal{P}]$ . C'est un espace de Hilbert dont les éléments sont notés  $\underline{X}_m$ , un vecteur à  $m$  composantes contenues dans  $L_2[\mathcal{P}]$ .  $L_2^{\oplus m}[\underline{Y}_n, \mathcal{F}_n]$  est le produit cartésien de  $m$  copies de  $L_2[\underline{Y}_n, \mathcal{F}_n]$ . C'est un sous-espace de Hilbert fermé de  $L_2^{\oplus m}[\mathcal{P}]$ . On a que  $\Pi_{L_2^{\oplus m}[\underline{Y}_n, \mathcal{F}_n]}(\underline{X}_m)$  est l'élément de  $L_2^{\oplus m}[\mathcal{P}]$  dont les composantes respectives sont les (classes d'équivalence des) variables aléatoires  $\Pi_{L_2[\underline{Y}_n, \mathcal{F}_n]}(X_i)$ ,  $1 \leq i \leq m$ .

Le théorème de Gauss-Markov dit que le BLMVE de  $\underline{X}_m$  connaissant  $\underline{Y}_n$  est

$$\hat{\underline{X}}_m = \Pi_{L_2^{\oplus m}[\underline{Y}_n, \mathcal{F}_n^{lin}]}(\underline{X}_m) = M\underline{Y}_n,$$

où  $M$  est une matrice de constantes de dimensions  $m \times n$ . De plus :

$$M = E[\underline{X}_m \underline{Y}_n^t] \{E[\underline{Y}_n \underline{Y}_n^t]\}^+,$$

où le symbole  $+$  indique qu'il s'agit de la pseudoinverse de  $E[\underline{Y}_n \underline{Y}_n^t]$ , et

$$E\left[\left(\hat{\underline{X}}_m - \underline{X}_m\right)\left(\hat{\underline{X}}_m - \underline{X}_m\right)^t\right] = E[\underline{X}_m \underline{X}_m^t] - ME[\underline{Y}_n \underline{X}_m^t].$$

Soit  $\underline{Z}$  le vecteur dont les  $m$  premières composantes sont celles de  $\underline{X}_m$  et les  $n$  suivantes celles de  $\underline{Y}_n$ . La matrice de covariance de  $\underline{Z}$ ,  $\Sigma_{\underline{Z}}$ , est donnée par la formule suivante :

$$\Sigma_{\underline{Z}} = \begin{bmatrix} \Sigma_{\underline{X}_m} & \Sigma_{\underline{X}_m, \underline{Y}_n} \\ \Sigma_{\underline{Y}_n, \underline{X}_m} & \Sigma_{\underline{Y}_n} \end{bmatrix}.$$

Alors le BAMVE de  $\underline{X}_m$  connaissant  $\underline{Y}_n$  est

$$\hat{\underline{X}}_m = M\underline{Y}_n + \underline{c}_m,$$

où

$$M = \Sigma_{\underline{X}_m, \underline{Y}_n} \Sigma_{\underline{Y}_n}^+$$

et

$$\underline{c}_m = \underline{\mu}_{\underline{X}_m} - M\underline{\mu}_{\underline{Y}_n}, \quad \underline{\mu}_{\underline{X}_m} = E[\underline{X}_m], \quad \underline{\mu}_{\underline{Y}_n} = E[\underline{Y}_n].$$

Cela donne la formule

$$\hat{\underline{X}}_m = \underline{\mu}_{\underline{X}_m} + \Sigma_{\underline{X}_m, \underline{Y}_n} \Sigma_{\underline{Y}_n}^+ \left\{ \underline{Y}_n - \underline{\mu}_{\underline{Y}_n} \right\}.$$

Finalement,

$$E \left[ \left( \hat{\underline{X}}_m - \underline{X}_m \right) \left( \hat{\underline{X}}_m - \underline{X}_m \right)^t \right] = \Sigma_{\underline{X}_m} - \Sigma_{\underline{X}_m, \underline{Y}_n} \Sigma_{\underline{Y}_n}^+ \Sigma_{\underline{Y}_n, \underline{X}_m}.$$

On utilisera les notations suivantes :

$$\begin{aligned} \underline{m}_{X|Y} &:= \hat{\underline{X}}_m, \\ C_{X|Y} &:= E \left[ \left( \hat{\underline{X}}_m - \underline{X}_m \right) \left( \hat{\underline{X}}_m - \underline{X}_m \right)^t \right], \\ \underline{X}_m | \underline{Y}_n &\sim \left[ \underline{m}_{X|Y}; C_{X|Y} \right]. \end{aligned}$$

On a les propriétés suivantes pour le BAMVE :

1. Il est sans biais.
2. Si  $A$  est une matrice et  $\underline{b}$  un vecteur de dimensions appropriées, alors

$$(A\underline{X}_m + \underline{b}) | \underline{Y}_n \sim \left[ A \left( \underline{m}_{X|Y} \right) + \underline{b}; AC_{X|Y}A^t \right].$$

3.  $\Sigma_{(\underline{X}_m - \underline{m}_{X|Y}), \underline{Y}_n} = O_{m,n}$ , où  $O_{m,n}$  est une matrice de zéros à  $m$  lignes et  $n$  colonnes.

4. Si  $\Sigma_{\underline{Y}_n, \underline{Z}_p} = O_{n,p}$ , et si  $\underline{U}_{n+p} = \begin{bmatrix} \underline{Y}_n \\ \underline{Z}_p \end{bmatrix}$ , alors

$$\begin{aligned} \underline{m}_{X|U} &= \underline{\mu}_{\underline{X}_m} - \Sigma_{\underline{X}_m, \underline{Y}_n} \Sigma_{\underline{Y}_n}^+ (\underline{Y}_n - \underline{\mu}_{\underline{Y}_n}) - \Sigma_{\underline{X}_m, \underline{Z}_p} \Sigma_{\underline{Z}_p}^+ (\underline{Z}_p - \underline{\mu}_{\underline{Z}_p}), \\ C_{X|U} &= \Sigma_{\underline{X}_m} - \Sigma_{\underline{X}_m, \underline{Y}_n} \Sigma_{\underline{Y}_n}^+ \Sigma_{\underline{Y}_n, \underline{X}_m} - \Sigma_{\underline{X}_m, \underline{Z}_p} \Sigma_{\underline{Z}_p}^+ \Sigma_{\underline{Z}_p, \underline{X}_m} \end{aligned}$$

5. Dans ce qui suit on a recours à la notation suivante :

$$\underline{U}_{m+n} := \begin{bmatrix} \underline{X}_m \\ \underline{Y}_n \end{bmatrix}, \quad \underline{V}_{n+p} := \begin{bmatrix} \underline{Y}_n \\ \underline{Z}_p \end{bmatrix}.$$

Alors, posant  $C_{U|Z} = \Sigma_{\underline{X}_m, \underline{Y}_n} - \Sigma_{\underline{X}_m, \underline{Z}_p} \Sigma_{\underline{Z}_p}^+ \Sigma_{\underline{Z}_p, \underline{Y}_n}$ , on a :

$$(a) \quad \underline{U}_{m+n} | \underline{Z}_p \sim \left[ \begin{pmatrix} \underline{m}_{X|Z} \\ \underline{m}_{Y|Z} \end{pmatrix}; \begin{pmatrix} C_{X|Z} & C_{U|Z} \\ C_{U|Z}^t & C_{Y|Z} \end{pmatrix} \right],$$

$$(b) \quad \underline{X}_m | \underline{V}_{n+p} \sim$$

$$\left[ \underline{m}_{X|Z} + C_{U|Z} C_{Y|Z}^+ (\underline{Y}_n - \underline{m}_{Y|Z}); C_{X|Z} - C_{U|Z} C_{Y|Z}^+ C_{U|Z}^t \right],$$

(c) quand  $\underline{X}_m$  et  $\underline{Y}_n$ , conditionnellement à  $\underline{Z}_p$ , ne sont pas corrélés et que l'une des espérances conditionnelles

$$E[\underline{X}_m | \sigma(\underline{Z}_p)], \quad E[\underline{Y}_n | \sigma(\underline{Z}_p)]$$

est une fonction affine de  $\underline{Z}_p$ , alors

$$\underline{X}_m | \underline{V}_{n+p} = \underline{X}_m | \underline{Z}_p \sim [\underline{m}_{X|Z}; C_{X|Z}].$$

6. Si,  $A$  et  $B$  étant des matrices et  $\underline{c}$  un vecteur de dimensions appropriées,

$$\begin{aligned} \underline{Y}_n | \underline{Z}_p &\sim [\underline{m}_{Y|Z}; C_{Y|Z}], \\ \underline{X}_m | \underline{V}_{n+p} &\sim [A\underline{Y}_n + B\underline{Z}_p + \underline{c}; C_{X|V}], \end{aligned}$$

alors

$$(a) \quad \underline{U}_{m+n} | \underline{Z}_p \sim$$

$$\left[ \begin{pmatrix} A\underline{m}_{Y|Z} + B\underline{Z}_p + \underline{c} \\ \underline{m}_{Y|Z} \end{pmatrix}; \begin{pmatrix} C_{Y|V} + AC_{Y|Z}A^t & AC_{Y|Z} \\ (AC_{Y|Z})^t & C_{Y|Z} \end{pmatrix} \right],$$

(b) et, en particulier,

$$\underline{X}_m | \underline{Z} \sim [A\underline{m}_{Y|Z} + B\underline{Z}_p + \underline{c}; C_{Y|V} + AC_{Y|Z}A^t].$$

On utilise encore les formules suivantes pour des vecteurs  $\underline{a} \in \mathbb{R}^n$ ,  $\underline{b} \in \mathbb{R}^n$  et un nombre  $\alpha \in \mathbb{R}$  :

1. On suppose  $|a_i| > 0$ ,  $1 \leq i \leq n$ , et  $\alpha \neq -\frac{1}{\sum_{k=1}^n a_k}$ .  $A$  est une matrice diagonale dont les éléments diagonaux sont  $a_1, \dots, a_n$ , et  $\underline{1}_n$  est l'élément de  $\mathbb{R}^n$  dont toutes les composantes valent 1. Alors

$$(A + \alpha \underline{a} \underline{a}^t)^{-1} = A^{-1} - \frac{\alpha}{1 + \alpha \sum_{k=1}^n a_k} \underline{1}_n \underline{1}_n^t.$$

- 2.

$$\langle \underline{a}, (A + \alpha \underline{a} \underline{a}^t)^{-1} \underline{b} \rangle_{\mathbb{R}^n} = \frac{\sum_{k=1}^n b_k}{1 + \alpha \sum_{k=1}^n a_k}.$$

### 9.2.1 Les équations d'estimation

La recherche des paramètres produisant le maximum de la vraisemblance se fait en partie à partir des équations que l'on obtient dans ce qui suit. Tout d'abord,

$$\frac{\partial}{\partial \underline{\alpha}} \mathcal{L}_{N, \underline{\theta}}(\underline{\alpha}, \underline{\beta}, \underline{\gamma}, \tau^2, \eta^2) = \frac{\partial}{\partial \underline{\alpha}} \mathcal{L}_{N| \underline{\theta}}(\underline{\alpha}).$$

Si l'on fixe alors les indices  $j_0 \in [1 : p]$ ,  $k_0 \in [1 : m]$ ,  $l_0 \in [1 : a]$ , on obtient que

$$\begin{aligned} D_{j_0, k_0, l_0} &:= \frac{\partial}{\partial \alpha_{k_0}^{(j_0, l_0)}} \mathcal{L}_{N| \underline{\theta}}(\underline{\alpha}) \\ &= \sum_{i=1}^{n_{j_0}} \left\{ N_i^{(j_0, k_0)} u_i^{(j_0, l_0)} - \theta_i^{(j_0)} a_i^{(j_0, k_0)} u_i^{(j_0, l_0)} \right\} \\ &= \sum_{i=1}^{n_{j_0}} u_i^{(j_0, l_0)} \left\{ N_i^{(j_0, k_0)} - \theta_i^{(j_0)} a_i^{(j_0, k_0)} \right\}. \end{aligned}$$

Les dérivées partielles correspondant à  $\underline{\alpha}_1^{(1)}$  sont respectivement

$$D_{1,1,1}, \dots, D_{1,1,a},$$

qui correspondent aux sommes successives

$$\begin{aligned} &\sum_{i=1}^{n_1} u_i^{(1,1)} \left\{ N_i^{(1,1)} - \dots \right\}, \\ &\vdots \\ &\sum_{i=1}^{n_1} u_i^{(1,a)} \left\{ N_i^{(1,1)} - \dots \right\}. \end{aligned}$$

Cela conduit à un premier produit matriciel :

$$\begin{bmatrix} u_1^{(1,1)} & \cdots & u_{n_1}^{(1,1)} \\ \vdots & & \vdots \\ u_1^{(1,a)} & \cdots & u_{n_1}^{(1,a)} \end{bmatrix} \times \begin{bmatrix} N_1^{(1,1)} \\ \vdots \\ N_{n_1}^{(1,1)} \end{bmatrix}.$$

Les dérivées partielles relatives à  $\underline{\alpha}_m^{(1)}$  conduisent à

$$\begin{bmatrix} u_1^{(1,1)} & \cdots & u_{n_1}^{(1,1)} \\ \vdots & & \vdots \\ u_1^{(1,a)} & \cdots & u_{n_1}^{(1,a)} \end{bmatrix} \times \begin{bmatrix} N_1^{(1,m)} \\ \vdots \\ N_{n_1}^{(1,m)} \end{bmatrix},$$

celles relatives à  $\underline{\alpha}_1^{(p)}$  produisent

$$\begin{bmatrix} u_1^{(p,1)} & \cdots & u_{n_p}^{(p,1)} \\ \vdots & & \vdots \\ u_1^{(p,a)} & \cdots & u_{n_p}^{(p,a)} \end{bmatrix} \times \begin{bmatrix} N_1^{(p,1)} \\ \vdots \\ N_{n_p}^{(p,1)} \end{bmatrix},$$

et celles relatives à  $\underline{\alpha}_m^{(p)}$  produisent

$$\begin{bmatrix} u_1^{(p,1)} & \cdots & u_{n_p}^{(p,1)} \\ \vdots & & \vdots \\ u_1^{(p,a)} & \cdots & u_{n_p}^{(p,a)} \end{bmatrix} \times \begin{bmatrix} N_1^{(p,m)} \\ \vdots \\ N_{n_p}^{(p,m)} \end{bmatrix}.$$

Si l'on pose

$$\underline{u}_1^{(1)} = \begin{bmatrix} u_1^{(1,1)} \\ \vdots \\ u_1^{(1,a)} \end{bmatrix}, \dots, \underline{u}_{n_1}^{(1)} = \begin{bmatrix} u_{n_1}^{(1,1)} \\ \vdots \\ u_{n_1}^{(1,a)} \end{bmatrix},$$

puis  $U_1 =$

$$\left[ \begin{array}{c|c|c|c|c|c} \underline{u}_1^{(1)} & \underline{0} & \underline{0} & \cdots & \underline{0} & \underline{0} & \underline{u}_2^{(1)} & \underline{0} & \underline{0} & \cdots & \underline{0} & \underline{0} & \cdots & \underline{0} & \underline{0} \\ \underline{0} & \underline{u}_1^{(1)} & \underline{0} & \cdots & \underline{0} & \underline{0} & \underline{0} & \underline{u}_2^{(1)} & \underline{0} & \cdots & \underline{0} & \underline{0} & \cdots & \underline{0} & \underline{0} \\ \vdots & \vdots \\ \underline{0} & \underline{0} & \underline{0} & \cdots & \underline{0} & \underline{u}_1^{(1)} & \underline{0} & \underline{0} & \underline{0} & \cdots & \underline{0} & \underline{u}_2^{(1)} & \cdots & \underline{0} & \underline{0} & \underline{u}_{n_1}^{(1)} \end{array} \right],$$

et que l'on définit de même  $U_2, \dots, U_p$ , puis

$$U = \begin{bmatrix} U_1 & O_{am,mn_2} & O_{am,mn_3} & \cdots & O_{am,mn_{p-1}} & O_{am,mn_p} \\ O_{am,mn_1} & U_2 & O_{am,mn_3} & \cdots & O_{am,mn_{p-1}} & O_{am,mn_p} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ O_{am,mn_1} & O_{am,mn_2} & O_{am,mn_3} & \cdots & O_{am,mn_{p-1}} & U_p \end{bmatrix},$$

le produit  $U_1 \underline{N}^{(1)}$  contient le premier terme des dérivées partielles, par rapport à  $\underline{\alpha}^{(1)}$ , de la vraisemblance, et  $U \underline{N}$  le premier terme de toutes ces dérivées. Le second terme a la forme  $U \underline{X}$ , où  $\underline{X}$  s'obtient comme suit. On forme la matrice

$$A_1 = \begin{bmatrix} \underline{a}_1^{(1)} & \underline{0} & \underline{0} & \cdots & \underline{0} & \underline{0} \\ \underline{0} & \underline{a}_2^{(1)} & \underline{0} & \cdots & \underline{0} & \underline{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \underline{0} & \underline{0} & \underline{0} & \cdots & \underline{0} & \underline{a}_{n_1}^{(1)} \end{bmatrix},$$

puis de manière similaire les matrices  $A_2, \dots, A_p$ , et enfin la matrice

$$A = \begin{bmatrix} A_1 & O_{mn_1, n_2} & O_{mn_1, n_2} & \cdots & O_{mn_1, n_{p-1}} & O_{mn_p, n_p} \\ O_{mn_2, n_1} & A_2 & O_{mn_2, n_3} & \cdots & O_{mn_2, n_{p-1}} & O_{mn_2, n_p} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ O_{mn_p, n_1} & O_{mn_p, n_2} & O_{mn_p, n_3} & \cdots & O_{mn_p, n_{p-1}} & A_p \end{bmatrix}.$$

On a  $\underline{X} = A \underline{\theta}_{[-0]}$ . Finalement,

$$\frac{\partial}{\partial \underline{\alpha}} \mathcal{L}_{\underline{N}, \underline{\theta}}(\underline{\alpha}, \underline{\beta}, \underline{\gamma}, \tau^2, \eta^2) = U [\underline{N} - A \underline{\theta}_{[-0]}].$$

Comme ci-dessus, tout d'abord,

$$\frac{\partial}{\partial \underline{\beta}} \mathcal{L}_{\underline{N}, \underline{\theta}}(\underline{\alpha}, \underline{\beta}, \underline{\gamma}, \tau^2, \eta^2) = \frac{\partial}{\partial \underline{\beta}} \mathcal{L}_{\underline{\theta} | \underline{\theta}_0}(\underline{\beta}, \eta^2).$$

Puis, si l'on fixe  $j_0 \in [1 : p]$  et  $k_0 \in [1 : b]$ , on obtient que

$$\begin{aligned} D_{j_0, k_0} &:= \frac{\partial}{\partial \beta^{(j_0, k_0)}} \mathcal{L}_{\underline{\theta} | \underline{\theta}_0}(\underline{\beta}, \eta^2) \\ &= \phi \sum_{i=1}^{n_{j_0}} \Delta v_i^{(j_0, k_0)} \left\{ \frac{\theta_i^{(j_0)}}{b_i^{(j_0)}} - \theta_{i-1}^{(j_0)} \right\}. \end{aligned}$$

Pour  $1 \leq j \leq p$ , on introduit alors les matrices suivantes :  $B_j$  est la matrice diagonale dont les éléments diagonaux sont respectivement  $b_1^{(j)}, \dots, b_{n_j}^{(j)}$ ,

$$\tilde{B}_j = \begin{bmatrix} -b_1^{(j)} & 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & -b_2^{(j)} & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & b_3^{(j)} & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & -b_{n_j}^{(j)} & 1 \end{bmatrix},$$

$$\Delta V_j = \left[ \Delta \underline{v}_1^{(j)} \mid \cdots \mid \Delta \underline{v}_{n_j}^{(j)} \right].$$

Soit  $\underline{D}_j$  le vecteur de composantes respectives  $D_{j,1}, \dots, D_{j,b}$ . On a que :

$$\underline{D}_j = \Delta V_j B_j^{-1} \tilde{B}_j \underline{\theta}^{(j)}.$$

Soient maintenant les matrices suivantes :

$$B = \begin{bmatrix} B_1 & O_{n_1, n_1} & O_{n_1, n_2} & \cdots & O_{n_1, n_{p-1}} & O_{n_1, n_p} \\ O_{n_2, n_1} & B_2 & O_{n_2, n_3} & \cdots & O_{n_2, n_{p-1}} & O_{n_2, n_p} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ O_{n_p, n_1} & O_{n_p, n_2} & O_{n_p, n_3} & \cdots & O_{n_p, n_{p-1}} & B_p \end{bmatrix},$$

$\Delta V$  et  $\tilde{B}$  construites de manière analogue. On a alors

$$\frac{\partial}{\partial \underline{\beta}} \mathcal{L}_{\underline{N}, \underline{\theta}}(\underline{\alpha}, \underline{\beta}, \underline{\gamma}, \tau^2, \eta^2) = \Delta V B^{-1} \tilde{B} \underline{\theta}.$$

Finalement

$$\frac{\partial}{\partial \underline{\gamma}} \mathcal{L}_{\underline{N}, \underline{\theta}}(\underline{\alpha}, \underline{\beta}, \underline{\gamma}, \tau^2, \eta^2) = \frac{\partial}{\partial \underline{\gamma}} \mathcal{L}_{\underline{\theta}_0}(\underline{\gamma}, \tau^2),$$

et, si l'on fixe  $k_0 \in [1 : c]$ , on obtient que

$$\begin{aligned} D_{k_0} &:= \frac{\partial}{\partial \gamma_{k_0}} \mathcal{L}_{\underline{\theta}_0}(\underline{\gamma}, \tau^2) \\ &= \psi \sum_{j=1}^p w_{k_0}^{(j)} \left[ \frac{\theta_0^{(j)}}{g_j} - 1 \right]. \end{aligned}$$

Si alors  $\underline{g}$  est le vecteur de composantes  $g_1, \dots, g_p$ ,  $G$  la matrice diagonale dont les éléments diagonaux sont les composantes de  $\underline{g}$ , et qu'enfin

$$W = \left[ \underline{w}^{(1)} \mid \cdots \mid \underline{w}^{(p)} \right],$$

on obtient

$$\frac{\partial}{\partial \underline{\gamma}} \mathcal{L}_{\underline{N}, \underline{\theta}}(\underline{\alpha}, \underline{\beta}, \underline{\gamma}, \tau^2, \eta^2) = W G^{-1} [\underline{\theta}_0 - \underline{g}].$$

Si maintenant

$$\begin{aligned} \underline{\eta} &= \begin{bmatrix} \underline{\alpha} \\ \underline{\beta} \\ \underline{\gamma} \end{bmatrix}, \\ \underline{F}_1(\underline{\eta}) &= U \left[ \underline{N} - A \underline{\theta}_{[-0]} \right], \end{aligned}$$

$$\begin{aligned}
\underline{E}_2(\underline{\eta}) &= \Delta V B^{-1} \tilde{B} \underline{\theta}, \\
\underline{E}_3(\underline{\eta}) &= W G^{-1} [\underline{\theta}_0 - \underline{g}], \\
\underline{E}(\underline{\eta}) &= \begin{bmatrix} \underline{E}_1(\underline{\eta}) \\ \underline{E}_2(\underline{\eta}) \\ \underline{E}_3(\underline{\eta}) \end{bmatrix},
\end{aligned}$$

les équations d'estimation sont données par l'expression  $\underline{E}(\underline{\eta}) = \underline{0}$ .

### 9.2.2 Une formule de dérivation

Voici la formule de dérivation du produit d'une matrice de fonctions agissant sur un vecteur dont les composantes sont elles-mêmes des fonctions. Soit  $\underline{u}$  le vecteur de composantes  $u_1, \dots, u_m$ . Soit  $A(\underline{u})$  une matrice de taille  $(n, p)$  de composantes  $a_{i,j}(\underline{u})$ . Soit enfin  $\underline{f}(\underline{u})$  le vecteur de composantes  $f_1(\underline{u}), \dots, f_p(\underline{u})$ . On pose

$$\underline{h}(\underline{u}) = A(\underline{u}) \underline{f}(\underline{u})$$

et on a besoin d'une expression pour  $\frac{\partial \underline{h}}{\partial \underline{u}}$ . On a tout d'abord que

$$\frac{\partial \underline{h}}{\partial \underline{u}} = \begin{bmatrix} \frac{\partial h_1}{\partial u_1} & \dots & \frac{\partial h_1}{\partial u_m} \\ \vdots & \dots & \vdots \\ \frac{\partial h_n}{\partial u_1} & \dots & \frac{\partial h_n}{\partial u_m} \end{bmatrix},$$

et que

$$\begin{bmatrix} \frac{\partial h_i}{\partial u_1} \\ \vdots \\ \frac{\partial h_i}{\partial u_m} \end{bmatrix} = \begin{bmatrix} \frac{\partial a_{i,1}}{\partial u_1} & \dots & \frac{\partial a_{i,p}}{\partial u_1} \\ \vdots & \dots & \vdots \\ \frac{\partial a_{i,1}}{\partial u_m} & \dots & \frac{\partial a_{i,p}}{\partial u_m} \end{bmatrix} \underline{f} + \begin{bmatrix} \frac{\partial f_1}{\partial u_1} & \dots & \frac{\partial f_p}{\partial u_1} \\ \vdots & \dots & \vdots \\ \frac{\partial f_1}{\partial u_m} & \dots & \frac{\partial f_p}{\partial u_m} \end{bmatrix} \begin{bmatrix} a_{i,1} \\ \vdots \\ a_{i,p} \end{bmatrix},$$

si bien que, si  $\underline{a}_i$  est le vecteur dont le transposé  $\underline{a}_i^t$  est la ligne numéro  $i$  de  $A$ , alors

$$\left[ \frac{\partial h_i}{\partial u_1} \mid \dots \mid \frac{\partial h_i}{\partial u_m} \right] = \underline{f}^t \times \frac{\partial \underline{a}_i}{\partial \underline{u}} + \underline{a}_i^t \times \frac{\partial \underline{f}}{\partial \underline{u}}.$$

Par suite

$$\frac{\partial \underline{h}}{\partial \underline{u}} = \begin{bmatrix} \underline{f}^t \times \frac{\partial \underline{a}_1}{\partial \underline{u}} \\ \vdots \\ \underline{f}^t \times \frac{\partial \underline{a}_p}{\partial \underline{u}} \end{bmatrix} + A \times \frac{\partial \underline{f}}{\partial \underline{u}} := \dot{A} + A \times \frac{\partial \underline{f}}{\partial \underline{u}}.$$