

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE  
MINISTRE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE  
SCIENTIFIQUE UNIVERSITE MENTOURI-CONSTANTINE-  
FACULT DES SCIENCES EXACTES

N° d'ordre :.....

Série :.....

# Mémoire

Présenté pour l'obtention du Diplôme de :  
Magister en mathématiques  
OPTION : PROBA-STAT :  
Analyse statistique des séries

## THEME

# LES MODELES DE PROCESSUS LATENTS A TEMPS DISCRET

Par  
BOUZERAIEB HAYET

### **Devant le jury**

*Président : S.Boughaba M.C Université Mentouri Constantine*  
*Rapporteur : M.Bousseboua M.C Université Mentouri Constantine*  
*Examineur : D.Boudaa M.C Université Mentouri Constantine*  
*Examineur : F.Messaci M.C Université Mentouri Constantine*

Soutenu le :10/11/2011

# Remerciements

Avant tout, mes vifs remerciements, je les exprime à Dieu tout puissant.

Je tiens à remercier : Dr Bousseboua, mon encadreur, qui m'a fait bénéficier de son savoir, de ses compétences scientifiques et de sa passion pour la recherche.

Je vous remercie jusqu'au bout de mes idées.

Mes vifs remerciements aux président de jury ainsi que les membres de jury l'honneur qu'ils m'ont fait en acceptant de siéger à mon jury de magistère.

Mes remerciements vont à tous les membres de ma famille à qui je dois beaucoup, sans leurs aides, ce travail n'aurait pu voir le jour.

Merci à tous ceux qui m'ont aidé sans ménager ni leurs temps, ni leurs encouragements, ni leurs savoirs.

Pour finir, je ne voudrais pas oublier mes proches qui m'ont soutenu moralement, sans les nommer explicitement, je les remercie pour leur encouragement.

# Table des matières

<b>1</b>	<b>Modèles de processus latents</b>	<b>5</b>
1.1	Description des modèles de processus latents . . . . .	5
1.2	Le modèle de Rasch . . . . .	7
1.3	Loi de probabilité des observations . . . . .	10
1.3.1	Le modèle de Rasch à effets fixes . . . . .	10
1.3.2	Le modèle de Rasch observable . . . . .	12
<b>2</b>	<b>Les modèles de processus <math>\mathcal{MA}</math> latents</b>	<b>15</b>
2.1	Modèles de processus gaussiens latents . . . . .	16
2.1.1	Les processus Gaussiens . . . . .	16
2.1.2	Modèles de processus gaussiens latents . . . . .	17
2.2	Modèles de processus moyennes mobiles d'ordre 1 latents . . . . .	19
2.2.1	Processus de moyenne mobile d'ordre 1 . . . . .	19
2.2.2	Modèles de processus moyennes mobiles d'ordre 1 latents . . . . .	21
<b>3</b>	<b>Les modèles de processus de Markov latents</b>	<b>23</b>
3.1	Modèles de processus de Markov d'ordre 1 latents . . . . .	24
3.1.1	Processus de Markov d'ordre 1 . . . . .	24
3.1.2	Processus de Markov Rasch Latent . . . . .	25
3.2	Modèles autorégressifs latents . . . . .	29
3.2.1	Modèles autorégressifs latents . . . . .	29

3.2.2	Processus Rasch Latent $\mathcal{AR}(1)$ . . . . .	30
<b>4</b>	<b>Estimation des paramètres du modèle de Rasch latent</b>	<b>33</b>
4.1	Modèles de Rasch à effets fixes . . . . .	34
4.1.1	Maximum de vraisemblance conjointe . . . . .	35
4.1.2	Maximum de vraisemblance conditionnelle . . . . .	36
4.1.3	Estimation des paramètres latents . . . . .	38
4.2	Estimation des paramètres d'un modèle de Rasch observable . . . . .	41
4.2.1	L'algorithme EM . . . . .	42
<b>5</b>	<b>Applications</b>	<b>47</b>
5.1	Bruit blanc gaussien Latent . . . . .	48
5.2	Processus Rasch Latent $\mathcal{MA}(1)$ . . . . .	50
5.3	Processus de Markov Rasch Latent . . . . .	52
5.4	Processus Rasch Latent $\mathcal{AR}(1)$ . . . . .	53
5.5	Annexe 1 . . . . .	55

## Introduction

Certaines variables d'intérêt, telles l'évaluation des connaissances dans le contexte de l'éducation ou la qualité de vie en sciences médicales, ne sont pas accessibles directement à partir des observations. De telles variables sont dites latentes ou encore traits latents. Celles-ci ne sont pas observables ou encore ne sont pas mesurées de façon fiable comme on pourrait le faire pour des grandeurs physiques. Les variables latentes se manifestent à des degrés divers selon les individus, elles sont en fait inhérentes à chaque individu. Dans le contexte de l'éducation par exemple, l'habilité ou l'aptitude dans une discipline, est relative et propre à chaque étudiant. Un moyen d'évaluer l'aptitude ou les connaissances d'un étudiant dans une discipline donnée, est de concevoir un questionnaire formé d'items bien choisis, et d'évaluer ensuite ses connaissances (trait latent) à partir des réponses fournies (observations), à l'aide d'une note ou score. Le questionnaire joue le rôle d'un instrument de mesure qu'on appelle échelle de mesure. Il est formé d'un ensemble d'items. Les items sont presque toujours qualitatifs (dichotomiques ou polytomiques). Il s'agit alors de mesurer des individus selon une variable quantitative non observable qui est la variable latente. A cette fin, on construit une échelle de mesure (questionnaire) qui va permettre de quantifier et d'ordonner une telle variable et ensuite positionner chaque individu sur cette échelle. L'individu est donc assimilé à un sujet sur lequel on souhaite effectuer une mesure.

Les qualités d'une échelle (questionnaire) en tant qu'instrument permettant d'obtenir un indicateur (ou mesure) de la variable d'intérêt sont sa validité et sa fiabilité. L'étude de la validité a pour but de justifier l'utilisation et l'interprétation que l'on fera par la suite des réponses observées. La fiabilité est la précision avec laquelle un questionnaire permet de mesurer la variable latente.

Notre travail ne se préoccupe pas des questionnaires mais de constructions de modèles latents aptes à expliquer le trait latent sous jacent.

Les modèles de processus latents furent introduits dans le contexte de l'analyse des

données longitudinales (Wiggins (1973) [WIG73] et tout récemment par Vermunt et al (1999) [VER99], Bartolucci (2006) [BAR06], Langeheine [LAN94] et Van de Pol 1986 [POL94]), les modèles de processus latents étaient déjà intensivement appliqués dans divers autres domaines (Mesbah (2009) [MES09]). Par exemple, en psychologie, la mesure de l'aptitude (variable latente) de l'individu dans le contexte de l'éducation est cruciale. Dans le domaine de la recherche en Qualité de Vie liée à la Santé, on utilise des questionnaires afin d'évaluer les variations du trait latent aussi bien dans les études transversales (temps fixe) que longitudinales (suivi dans le temps de personnes), rétrospectives où prospectives (Mesbah (2009)[MES09]).

Une des hypothèses de base des modèles de processus latents est l'indépendance locale. Cette hypothèse postule que les variables observées utilisées pour mesurer une variable latente spécifique, sont conditionnellement indépendantes sachant la variable latente correspondante. Cette condition est généralement admise dans tous les modèles latents. Elle est aussi incluse dans le modèle de Rasch (Admane et Mesbah (2006) [ADM06], Fisher and Molenaar (1995) [FIS95]), un des modèles parmi les plus populaires des modèles IRT (Item Response Theory). Ce modèle possède des propriétés intéressantes et permet en outre une écriture explicite de la vraisemblance des observations.

Ce travail comporte deux parties. La première partie consiste en une description des différents modèles et la détermination de la loi jointe des observations. Le premier chapitre comporte une description sommaire des modèles de processus latents en général et des modèles de Rasch latents en particulier ainsi que le calcul des lois marginales des observations sous les hypothèses du modèle de Rasch à effets fixes et à effets aléatoires. Dans un second chapitre, on étudie une classe de modèles de processus latents, lorsque le processus latent est gaussien et lorsque c'est un processus moyenne mobile d'ordre 1 ( $\mathcal{MA}(1)$ ). Quant au troisième chapitre, il est consacré à la classe des modèles de processus de Markov d'ordre 1 latents et la classe des processus autorégressifs d'ordre 1 latents ( $\mathcal{AR}(1)$ ). La deuxième partie

porte sur l'estimation des paramètres de chaque modèle. Dans le chapitre quatre qui correspond au premier de cette partie, on passe en revue les procédures d'estimation des paramètres d'un modèle de processus latent sous les hypothèses du modèle de Rasch à effets fixes et à effets aléatoires. Et dans le chapitre cinq, on applique la méthode d'estimation du maximum de vraisemblance des paramètres d'un modèle à effets aléatoires, en privilégiant une méthode récursive basée sur l'algorithme **EM** (expectation and maximization). À cette fin, les modèles de processus latents décrits ci-dessus sont supposés satisfaire les hypothèses du modèle de Rasch latent. Feddag et Mesbah (2005) [FED05] ont considéré une situation similaire dans un contexte restreint faisant intervenir moins de trois mesures longitudinales seulement et utilisant les méthodes GEE pour estimer les paramètres de difficulté et des variables latentes.





# Chapitre 1

## Modèles de processus latents

### 1.1 Description des modèles de processus latents

Un modèle de processus latent consiste en une trajectoire finie d'un processus multivarié  $\{(\mathbf{X}_i(t), \Theta_i(t)) : 1 \leq i \leq n, 1 \leq t \leq T\}$  où  $\mathbf{X}_i(t) = (X_{i1}(t), \dots, X_{iq}(t))'$  est le processus longitudinal des observations et pour tout  $i : (\Theta_i(t))$  est un processus non observable dit processus latent. Dans la plupart des applications, les  $q$  variables  $X_{i,1}(t), X_{i,2}(t), \dots,$

$X_{i,q}(t)$  sont à chaque occasion  $t$ , les réponses de l'individu  $i$  à un questionnaire dichotomique comportant  $q$ -items. La variable  $X_{i,k}(t)$  représente la réponse à l'instant  $t$  de l'individu  $i$  à l'item  $k$ . Le même questionnaire est soumis, à différentes occasions aux mêmes individus. Une caractéristique de ces modèles est que chaque variable réponse  $X_{i,k}(t)$  ne dépend que de la variable latente  $\Theta_i(t)$  correspondante.

Il s'agit de déterminer la loi du processus latent  $(\Theta_i(t))$  au vue d'une trajectoire finie du processus des observations  $(X_{i,k}(t))$ .

Dans notre contexte le modèle peut être représenté à tout instant  $t$  par

$$\begin{array}{ccccccc} \underbrace{\{X_{1,1}(t) & \dots & X_{1,q}(t)\}} & \cdot & \cdot & \cdot & \underbrace{\{X_{n,1}(t) & \dots & X_{n,q}(t)\}} \\ & & \uparrow & & \cdot & \cdot & \cdot & & \uparrow \\ & & \Theta_1(t) & & \cdot & \cdot & \cdot & & \Theta_n(t) \end{array}$$

Le processus des observations  $(X_{i,k}(t))$  est à valeurs dans  $\{0, 1\}$ . La valeur 1 correspond à une réponse correcte (où positive) du sujet  $i$  à l'item  $k$ . Le terme "correcte" est utilisé dans le contexte de l'éducation, tandis que le terme "positive" peut être utilisé dans un contexte tel que la qualité de vie où plus généralement dans des questionnaires mesurant des traits psychologiques et/ou sociologiques.

La variable latente  $\Theta_i(t)$  ne dépend que de l'individu  $i$ . Elle peut être interprétée comme une mesure de son aptitude dans un contexte éducatif, et plus généralement comme une mesure de sa position sur un trait latent unidimensionnel évaluée par l'instrument de mesure qui est le questionnaire.

On note  $\underline{\mathbf{X}}_i(t) = (X_{i,1}(t), \dots, X_{i,q}(t))'$  le vecteur réponse de l'individu  $i$  à l'instant  $t$  et par  $\pi(x_{i,k}/\theta_i(t))$  la probabilité conditionnelle  $\mathbf{P}(X_{i,k}(t) = x_{i,k}/\Theta_i(t) = \theta_i(t))$  où  $i = 1, \dots, n$ ,  $k = 1, 2, \dots, q$ ,  $t = 1, 2, \dots, T$  et  $x_{i,k} \in \{0, 1\}$  pour tout  $i$  et pour tout  $k$ .

**Proposition 1** *La loi jointe du vecteur réponse  $\underline{\mathbf{X}}_i = (\underline{\mathbf{X}}_i(1), \dots, \underline{\mathbf{X}}_i(T))'$  de l'individu  $i$  s'écrit au vue d'une observation  $\underline{\mathbf{x}}_i = (x_i(1), \dots, x_i(T))'$  :*

$$p(\underline{\mathbf{x}}_i) = \int \dots \int_{\mathbb{R}^T} p(\underline{\mathbf{x}}_i/\underline{\theta}_i) \times g_i(\underline{\theta}_i) d\theta_i(1) \dots d\theta_i(T) \quad (1.1)$$

où  $g_i$  est la densité de probabilité du processus latent individuel  $(\Theta_i(1), \dots, \Theta_i(T))'$ ,  $\underline{\theta}_i = (\theta_i(1), \dots, \theta_i(T))'$  et  $p(\underline{\mathbf{x}}_i(t)/\theta_i(t)) = \mathbf{P}(\underline{\mathbf{X}}_i(t) = \underline{\mathbf{x}}_i(t)/\Theta_i(t) = \theta_i(t))$  et où  $\underline{\mathbf{x}}_i(t) = (x_{i,1}(t), \dots, x_{i,q}(t))'$  est le vecteur réponse de l'individu  $i$ .

### Démonstration

Celle ci dérive directement du fait que la loi de probabilité du vecteur réponse  $\underline{\mathbf{X}}_i$  de l'individu  $i$ , est la marginale de la loi jointe du vecteur  $\underline{\mathbf{X}}_i$  et du vecteur latent  $(\Theta(1), \dots, \Theta(T))'$ . ■

Remarque :

Si le processus latent est à valeurs discrètes, alors la relation (1.1) est dans ce cas

$$p(\underline{\mathbf{x}}_i) = \sum_{\theta_i(1) \in \mathbb{D}} \dots \sum_{\theta_i(T) \in \mathbb{D}} p(\underline{\mathbf{x}}_i / \theta_i) \times g_i(\theta_i) d\theta_i(1) \dots d\theta_i(T)$$

où  $\mathbb{D}$  est l'ensemble des valeurs de  $\Theta_i(t)$  et c'est une partie au plus infinie dénombrable.

Dans le contexte de ce travail, nous supposons que la distribution de probabilité de la variable latente appartient à une famille de lois paramétriques connues. Un des objectifs de ce travail est d'estimer les paramètres inconnus de cette loi à l'aide des observations  $x_{i,k}(t)$ .

A cette fin, l'hypothèse d'indépendance locale des items d'une part, et le choix du modèle de Rasch comme modèle de mesure d'autre part, vont nous permettre de préciser l'écriture des probabilités conditionnelles  $p(\underline{\mathbf{x}}_i(t) / \theta_i(t))$ .

L'indépendance locale des items signifie que les variables réponses  $X_{i,1}(t), \dots, X_{i,q}(t)$  observés de l'individu  $i$ , sont, conditionnellement à leur valeur latente correspondante (inobservée)  $\Theta_i(t)$ , indépendantes.

Cette hypothèse est une des hypothèses du modèle de Rasch, que l'on retrouve aussi dans tous les modèles **IRT** (item response theory) et plus généralement, dans les modèles à variables latentes observables.

Ainsi, l'utilisation de cette hypothèse fondamentale, simplifiera d'abord l'écriture de  $p(\underline{\mathbf{x}}_i(t) / \theta_i(t))$  et ensuite celle de la vraisemblance, qu'elle aidera à factoriser, ce qui facilitera l'estimation des paramètres.

## 1.2 Le modèle de Rasch

De nos jours, dans le contexte des variables latentes et de la Théorie des Réponses aux Items (IRT), le modèle probabiliste de *mesure* le plus populaire est le modèle de Rasch (Fisher et Molenaar (1995) [FIS95]). Un modèle probabiliste de mesure est un modèle pour *la distribution des items observées conditionnellement*

à la variable latente unidimensionnelle. Le modèle de Rasch [RAS62] est très utilisé dans divers domaines psychométriques, dont la recherche en éducation et plus récemment l'analyse de la qualité de la vie liée à la santé (Mesbah (2009) [MES09]). Ce modèle considère les hypothèses suivantes :

( $h_1$ ) : La loi conditionnelle de  $\underline{\mathbf{X}}_i(t)$  par rapport aux variables latentes  $\underline{\Theta}_i(t)$  :  $t = 1, \dots, T$ , ne dépend que de la variable latente correspondante, c'est-à-dire, celle existant au même instant  $t$  où les items ont été observés :

$$P(\underline{\mathbf{X}}_i(t) = \underline{\mathbf{x}}_i(t) / \Theta_i(1) = \theta_i(1), \dots, \Theta_i(T) = \theta_i(T)) = p(\underline{\mathbf{x}}_i(t) / \theta_i(t)) \quad (1.2)$$

où  $p(\underline{\mathbf{x}}_i(t) / \theta_i(t)) = P(\underline{\mathbf{X}}_i(t) = \underline{\mathbf{x}}_i(t) / \Theta_i(t) = \theta_i(t))$  et  $\underline{\mathbf{X}}_i(t)$  le vecteur réponse de l'individu  $i$  lors de la  $t^{ieme}$  consultation.

Cette hypothèse postule que la variable réponse de l'individu  $i$  à l'instant  $t$  ne dépend que de la variable latente  $\Theta_i(t)$ .

( $h_2$ ) : Les vecteurs aléatoires  $\underline{\mathbf{X}}_i(1), \underline{\mathbf{X}}_i(2), \dots, \underline{\mathbf{X}}_i(T)$  sont conditionnellement indépendants sachant le vecteur latent  $(\Theta_i(1) = \theta_i(1), \dots, \Theta_i(T) = \theta_i(T))'$  :

$$p(\underline{\mathbf{x}}_i(1), \dots, \underline{\mathbf{x}}_i(T) / \theta_i(1), \dots, \theta_i(T)) = \prod_{t=1}^T p(\underline{\mathbf{x}}_i(t) / \theta_i(t)) \quad (1.3)$$

où  $p(\underline{\mathbf{x}}_i(1), \dots, \underline{\mathbf{x}}_i(T) / \theta_i(1), \dots, \theta_i(T))$  est la loi jointe de  $\underline{\mathbf{X}}_i(1), \underline{\mathbf{X}}_i(2), \dots, \underline{\mathbf{X}}_i(T)$  conditionnellement à  $\{\Theta_i(1) = \theta_i(1), \dots, \Theta_i(T) = \theta_i(T)\}$ .

Ceci est une hypothèse d'indépendance conditionnelle longitudinale des items observés relativement au processus latent.

( $h_3$ ) : Les variables réponses  $X_{i,1}(t), \dots, X_{i,q}(t)$  sont conditionnellement indépendantes relativement à la variable latente  $\Theta_i(t)$  :

$$p(\underline{\mathbf{x}}_i(t) / \theta_i(t)) = \prod_{k=1}^q \pi(x_{i,k}(t) / \theta_i(t)) \quad (1.4)$$

où  $\pi(x_{i,k}(t) / \theta_i(t)) = P(X_{i,k}(t) = x_{i,k}(t) / \Theta_i(t) = \theta_i(t))$  et  $\underline{\mathbf{x}}_i(t) = (x_1(t), \dots, x_q(t))'$  une observation du vecteur réponse  $\underline{\mathbf{X}}_i(t)$ .

Cette hypothèse correspond à un questionnaire comportant  $q$  items conditionnellement indépendants relativement à la variable latente correspondante. C'est l'hypothèse de l'indépendance locale.

( $h_4$ ) Les probabilités conditionnelles  $\pi(x_{i,k}(t)/\theta_i(t))$  sont données par

$$\pi(1/\theta_i(t), \beta_k) = \frac{\exp(\theta_i(t) - \beta_k)}{1 + \exp((\theta_i(t) - \beta_k))}, \quad (1.5)$$

où  $\theta_i(t)$  caractérise la position de l'individu  $i$  sur le trait latent (l'aptitude de l'individu à mesurer) et  $\beta_k$  un paramètre réel caractérisant la position de l'item  $k$  sur ce même trait latent. Dans le contexte des sciences de l'éducation, ce dernier est souvent interprété comme le niveau de difficulté de l'item (de la question)  $k$ .

L'introduction du paramètre de difficulté  $\beta_k$  nous amène à noter cette loi conditionnelle  $\pi(x_{i,k}(t)/\theta_i(t), \beta_k)$ .

La formule (1.5) est connue comme étant le modèle de Rasch. Nous dirons que le processus latent est *Rasch observable* et les hypothèses  $h_1 - h_4$  ci-dessus définissent le modèle de Rasch à effets aléatoires.

### **Le modèle de Rasch à effets fixes**

Si pour tout  $t$  et pour tout  $i$ , les variables latentes  $\Theta_i(t)$  sont presque sûrement constante, alors l'hypothèse  $h_3$  devient

( $h'_3$ ) : Les variables réponses  $X_{i,1}(t), \dots, X_{i,q}(t)$  sont indépendantes.

Les hypothèses ( $h_1$ ), ( $h_2$ ), ( $h'_3$ ) et ( $h_4$ ) définissent le modèle de Rasch à effets fixes.

Ce modèle appartient à la famille des modèles linéaires à effets fixes et composé d'un paramètre effet groupe  $\underline{\beta} = (\beta_1, \dots, \beta_q)'$  et d'une composante effet individuel  $\underline{\theta} = (\theta_1, \dots, \theta_n)'$ .

Remarque :

L'hypothèse d'indépendance locale dans le modèle de Rasch à effets fixes, est en fait une indépendance globale, au sens que les variables réponses  $X_{i,k}(t)$  et  $X_{i,k}(t-1)$  sont indépendantes.

Les réponses des individus au questionnaire sont évidemment indépendantes, ce

qui revient à dire que les processus latents  $(\Theta_i(t) : 1 \leq t \leq T), i = \overline{1, n}$  sont indépendants. Cette hypothèse fait qu'il est suffisant de décrire pour commencer le modèle et les lois marginales des observations pour un seul et même individu. A cette fin et par commodité, on omettra dans cette première partie l'indice relatif à l'individu.

### Notations

On convient de noter

$\underline{\mathbf{X}}(t) = (X_1(t), \dots, X_q(t))'$  le vecteur réponse d'un individu anonyme  $i$  à l'instant  $t$  et par  $\pi(x_{i,k}/\theta_i(t), \beta_k) = p(x_{i,k}/\theta_i(t), \beta_k)$  la probabilité conditionnelle  $\mathbf{P}(X_{i,k}(t) = x_{i,k}/\Theta_i(t) = \theta_i(t), \beta_k)$  et  $\underline{\mathbf{X}} = (\underline{\mathbf{X}}(1), \dots, \underline{\mathbf{X}}(T))'$  où  $\underline{\mathbf{X}}(t) = (X_1(t), \dots, X_q(t))'$  le vecteur réponse lors de la  $t^{ieme}$  opportunité,  $\underline{\Theta} = (\Theta(1), \dots, \Theta(T))'$  le vecteur latent correspondant et  $\underline{\beta} = (\beta(1), \dots, \beta(q))'$  le vecteur des paramètres de difficulté des items.

## 1.3 Loi de probabilité des observations

### 1.3.1 Le modèle de Rasch à effets fixes

**Proposition 2** *Sous les hypothèses  $h_1, h_2$  et  $h'_3$  ci-dessus, la loi jointe des vecteurs réponses  $\underline{\mathbf{X}}(1), \dots, \underline{\mathbf{X}}(T)$  au vue d'une observation  $\underline{\mathbf{x}}(1), \dots, \underline{\mathbf{x}}(T)$  durant la période  $\{1, \dots, T\}$  s'écrit :*

$$p(\underline{\mathbf{x}}(1), \dots, \underline{\mathbf{x}}(T)/\underline{\theta}) = \prod_{t=1}^T [1 - \alpha_1(\theta(t))]^q \left[ \frac{\alpha_1(\theta(t))}{1 - \alpha_1(\theta(t))} \right]^{R(t)} \quad (1.6)$$

où  $\alpha_1(\theta(t)) = \pi(1/\theta(t))$ ,  $R(t) = \sum_{k=1}^q x_k(t)$  est le score (nombre de réponses correctes) d'un individu lors de la  $t^{ieme}$  opportunité et  $\underline{\theta} = (\theta(1), \dots, \theta(T))'$ .

### Démonstration

En regardant la loi du vecteur réponse  $(\underline{\mathbf{X}}(1), \dots, \underline{\mathbf{X}}(T))'$  comme la marginale de la loi jointe de  $(\underline{\mathbf{X}}(1), \dots, \underline{\mathbf{X}}(T))'$  et de  $(\Theta(1), \dots, \Theta(T))'$  et compte tenu de ce que les variables latentes  $\Theta(t)$  sont presque sûrement constantes, alors la loi marginale de

$\underline{\mathbf{X}}(1), \dots, \underline{\mathbf{X}}(T)$  au vue d'une observation  $\underline{\mathbf{x}}(1), \dots, \underline{\mathbf{x}}(T)$ , s'écrit

$$p(\underline{\mathbf{x}}(1), \dots, \underline{\mathbf{x}}(T)) = p(\underline{\mathbf{x}}(1), \dots, \underline{\mathbf{x}}(T)/\underline{\theta}).$$

Et tenant compte des hypothèses  $h_1$  et  $h_2$ , cette dernière expression devient

$$p(\underline{\mathbf{x}}(1), \dots, \underline{\mathbf{x}}(T)/\underline{\theta}) = \prod_{t=1}^T p(\underline{\mathbf{x}}(t)/\underline{\theta}(t)). \quad (1.7)$$

En appliquant ensuite l'hypothèse  $h'_3$  de l'indépendance locale, chacune des probabilités  $p(\underline{\mathbf{x}}(t)/\underline{\theta}(t))$  de (1.7) s'écrit

$$p(\underline{\mathbf{x}}(t)/\underline{\theta}(t)) = \prod_{k=1}^q \pi(\mathbf{x}_k(t)/\theta(t)).$$

Et comme

$$\pi(x_k(t)/\theta(t)) = [1 - \alpha_1(\theta(t))] \left[ \frac{\alpha_1(\theta(t))}{1 - \alpha_1(\theta(t))} \right]^{x_k(t)}$$

(hypothèse  $h_4$ ) et reportant ensuite ces résultats dans (1.7), alors la relation (1.6) pour un seul individu en découle. ■

La dépendance de l'effet individuel  $\theta$  du temps  $t$  est légitime, car on doit s'attendre à ce que le paramètre individuel  $\theta$  s'améliore au cours des différentes opportunités.

**Corollaire 1** *Sous les hypothèses  $h_1, h_2, h'_3$  et  $h_4$  du modèle de Rasch à effets fixes, la loi jointe des vecteurs réponses  $\underline{\mathbf{X}}(1), \dots, \underline{\mathbf{X}}(T)$  au vue d'une observation  $\underline{\mathbf{x}}(1), \dots, \underline{\mathbf{x}}(T)$  durant la période  $\{1, \dots, T\}$ , s'écrit :*

$$p(\underline{\mathbf{x}}(1), \dots, \underline{\mathbf{x}}(T)/\underline{\theta}, \underline{\beta}) = \frac{\exp(\sum_{t=1}^T R(t)\theta(t) - \sum_{k=1}^q \beta_k S_k)}{\prod_{t=1}^T \prod_{k=1}^q [1 + \exp((\theta(t) - \beta_k))]} \quad (1.8)$$

où  $R(t)$  est le score obtenu lors de la  $t^{\text{ième}}$  consultation et  $S_k = \sum_{t=1}^T x_k(t)$  le nombre de réponses correctes à l'item  $k$  au cours des  $T$ -opportunités.

### Démonstration

En suivant le développement de la démonstration de la proposition 2 ci-dessus et remplaçant chacune des probabilités  $\pi(x_k(t)/\theta(t))$  par son équivalent  $\frac{\exp(x_k(t)(\theta(t) - \beta_k))}{1 + \exp((\theta(t) - \beta_k))}$  de l'hypothèse  $h_4$ , la relation (1.8) en découle. ■

### 1.3.2 Le modèle de Rasch observable

Soit  $\{\underline{\mathbf{x}}(t) : 1 \leq t \leq T\}$  une trajectoire finie du processus des observations  $\{\underline{\mathbf{X}}(t) : 1 \leq t \leq T\}$  relative à un seul individu et où  $\underline{\mathbf{X}}(t) = (X_1(t), \dots, X_q(t))'$  est le vecteur des variables réponses lors de la consultation  $t$  et  $(\Theta(t) : 1 \leq t \leq T)$  le processus latent correspondant. On a le résultat suivant :



**Proposition 3** *Sous les hypothèses  $h_1, h_2$  et  $h_3$  ci-dessus, la loi jointe des observations d'un individu quelconque durant la période  $\{1, \dots, T\}$ , s'écrit :*

$$p(\underline{\mathbf{x}}(1), \dots, \underline{\mathbf{x}}(T)) = \int \dots \int_{\mathbf{R}^T} \prod_{t=1}^T [1 - \alpha_1(\theta(t))]^q \left[ \frac{\alpha_1(\theta(t))}{1 - \alpha_1(\theta(t))} \right]^{R(t)} \cdot g(\underline{\theta}) d\underline{\theta} \quad (1.9)$$

où  $\alpha_1(\theta(t)) = \pi(1/\theta(t))$ ,  $R(t) = \sum_{k=1}^q x_k(t)$  le score (nombre de réponses correctes) de l'individu lors de la  $t^{\text{ième}}$  opportunité et  $g$  la densité du vecteur latent  $(\Theta(1), \dots, \Theta(T))'$ ,  $\underline{\theta} = (\theta(1), \dots, \theta(T))$  et  $d\underline{\theta} = d\theta(1) \dots d\theta(T)$ .

### Démonstration

En regardant la loi jointe des vecteurs  $(\underline{\mathbf{X}}(1), \dots, \underline{\mathbf{X}}(T))$  comme la marginale de la loi jointe de  $(\underline{\mathbf{X}}(1), \dots, \underline{\mathbf{X}}(T))$  et de  $(\Theta(1), \dots, \Theta(T))$  et en tenant compte des hypothèses  $h_1$  et  $h_2$ , nous pouvons écrire :

$$p((\underline{\mathbf{x}}(1), \dots, \underline{\mathbf{x}}(T))) = \int \dots \int_{\mathbf{R}^T} \prod_{t=1}^T p(\underline{\mathbf{x}}(t)/\theta(t)) g(\theta(1), \dots, \theta(T)) d\theta(1) \dots d\theta(T).$$

Pour conclure, il s'agit de calculer la probabilité conditionnelle  $p(\underline{\mathbf{x}}(t)/\theta(t))$ . Aussi, en utilisant l'hypothèse  $h_3$  de l'indépendance locale et le fait que :

$$\pi(x_k(t)/\theta_t) = (1 - \alpha_1(\theta(t))) \left[ \frac{\alpha_1(\theta(t))}{1 - \alpha_1(\theta(t))} \right]^{x_k(t)}$$

fait que cette probabilité conditionnelle peut être mise sous la forme

$$p(\underline{\mathbf{x}}(t)/\theta(t)) = [1 - \alpha_1(\theta(t))]^q \left[ \frac{\alpha_1(\theta(t))}{1 - \alpha_1(\theta(t))} \right]^{R(t)},$$

d'où le résultat. ■

La probabilité  $p(\underline{\mathbf{x}}(1), \dots, \underline{\mathbf{x}}(T))$  dépend des probabilités de transition  $\alpha_1(\theta(t))$  qui constituent autant de paramètres à estimer afin d'identifier totalement le modèle. Cependant, si l'on suppose que ce modèle satisfait l'hypothèse  $h_4$  (modèle de Rasch), alors le nombre de paramètres peut être substantiellement réduit et la probabilité du vecteur des observations relatif à un seul individu devient relativement plus simple.

**Corollaire 2** *Si le modèle  $((\underline{\mathbf{X}}(t), \Theta(t)) : 1 \leq t \leq T)$  est un modèle de Rasch observable, alors la loi de probabilité des vecteurs d'observation  $\underline{\mathbf{X}}(1), \dots, \underline{\mathbf{X}}(T)$  au*

vue d'une observation  $\underline{\mathbf{x}}(1), \dots, \underline{\mathbf{x}}(T)$  d'un seul et même individu durant la période  $\{1, \dots, T\}$ , est donnée par

$$p(\underline{\mathbf{x}}(1), \dots, \underline{\mathbf{x}}(T)) = \int \dots \int_{\mathbb{R}^T} \frac{\exp(\sum_{t=1}^T R(t)\theta_t - \sum_{k=1}^q \beta_k S_k)}{\prod_{t=1}^T \prod_{k=1}^q [1 + \exp((\theta_t - \beta_k))]} g(\underline{\theta}) d\underline{\theta} \quad (1.10)$$

où  $g$  est la densité du vecteur latent,  $R(t)$  est le score,  $S_k = \sum_{t=1}^T x_k(t)$  le nombre de réponses correctes à l'item  $k$  au cours des  $T$ -opportunités,  $\underline{\theta} = (\theta(1), \dots, \theta(T))$  et  $d\underline{\theta} = d\theta(1) \dots d\theta(T)$ .

### Démonstration

Comme dans la démonstration de la proposition 3 ci-dessus, on a

$$p(\underline{\mathbf{x}}(1), \dots, \underline{\mathbf{x}}(T)) = \int \dots \int_{\mathbb{R}^T} \prod_{t=1}^T p(\underline{\mathbf{x}}(t)/\theta(t), \beta_k) g(\underline{\theta}) d\theta(1) \dots d\theta(T) \quad (1.11)$$

où  $\underline{\theta} = (\theta(1), \dots, \theta(T))$ .

Il suffit alors, moyennant l'hypothèse  $h_3$ , d'écrire chaque probabilité  $p(\underline{\mathbf{x}}(t)/\theta(t))$  sous la forme

$$p(\underline{\mathbf{x}}(t)/\theta(t)) = \prod_{k=1}^q \pi(x_k(t)/\theta(t))$$

et de remplacer ensuite chacune des probabilités  $\pi(x_k(t)/\theta(t), \beta_k)$  par son équivalent du modèle de Rasch (hypothèse  $h_4$ ). Ceci fait que chaque probabilité  $p(\underline{\mathbf{x}}(t)/\theta(t), \beta_k)$  devient

$$p(\underline{\mathbf{x}}(t)/\theta(t), \beta_k) = \frac{\exp(\sum_{t=1}^T R(t)\theta(t) - \sum_{k=1}^q \beta_k S_k)}{\prod_{t=1}^T \prod_{k=1}^q [1 + \exp((\theta(t) - \beta_k))]} \quad (1.12)$$

d'où l'on déduit ensuite en reportant ceci dans la relation (1.11), le résultat énoncé dans le corollaire. ■

## Chapitre 2

# Les modèles de processus $\mathcal{MA}$ latents

Dans ce chapitre, nous considérons deux classes de processus latents : celle des processus gaussiens latents et celle des processus moyennes mobiles d'ordre 1 ( $\mathcal{MA}(1)$ ) latents.

## 2.1 Modèles de processus gaussiens latents

### 2.1.1 Les processus Gaussiens

**Définition 1** Un processus  $(X_t : t \geq 1)$  à valeurs dans  $\mathbb{R}$ , est dit un processus gaussien, si toutes les combinaisons linéaires finies du processus  $(X_t : t \geq 1)$  suivent une loi normale.

Ainsi, pour tout  $n \geq 1$  et  $t_1, t_2, \dots, t_n \in \mathbb{N}^*$ , la loi de la variable aléatoire  $\sum_{j=1}^n \alpha_j X_{t_j}$  est gaussienne et ce pour tous  $\alpha_1, \dots, \alpha_n \in \mathbb{R}$ . Autrement dit, tous les vecteurs finis dimensionnelles  $(X_{t_1}, X_{t_2}, \dots, X_{t_n})$  sont gaussiens. Or, nous savons (cf. [BOU04]) que la loi jointe  $P_{X_{t_1}, X_{t_2}, \dots, X_{t_n}}$  est entièrement caractérisée par sa fonction caractéristique  $\varphi_{t_1, \dots, t_n}$  définie par :

$$\varphi_{t_1, \dots, t_n}(\theta) = \exp\left(i \cdot \langle \theta, m_{\mathbf{t}} \rangle - \frac{1}{2} \theta' \Lambda_{\mathbf{t}} \theta\right)$$

où  $\Lambda_{\mathbf{t}}$  est la matrice de covariance du vecteur  $(X_{t_1}, X_{t_2}, \dots, X_{t_n})'$ ,  $m_{\mathbf{t}}$  sa fonction moyenne et  $\mathbf{t} = (t_1, \dots, t_n)$ .

Si de plus  $\Lambda_{\mathbf{t}}$  est inversible, alors le vecteur  $(X_{t_1}, X_{t_2}, \dots, X_{t_p})'$  admet une densité de probabilité  $f$  donnée par

$$f(x_1, \dots, x_p) = \frac{1}{\sqrt{(2\pi)^p} \cdot \sqrt{|\Lambda_{\mathbf{t}}|}} \exp\left\{-\frac{1}{2} \underline{\mathbf{x}}' \Lambda_{\mathbf{t}}^{-1} \underline{\mathbf{x}}\right\} \quad (2.1)$$

où  $\underline{\mathbf{x}} = (x_1, x_2, \dots, x_p)'$  et  $|\Lambda_{\mathbf{t}}|$  le jacobien de la matrice de covariance  $\Lambda_{\mathbf{t}}$

**Définition 2** On appelle processus de bruit blanc, un processus  $(\varepsilon(t) : t \in \mathbb{Z})$  du second ordre, centré et deux à deux non corrélés.

La fonction de covariance  $\gamma$  d'un bruit blanc  $(\varepsilon(t) : t \in \mathbb{Z})$  s'écrit

$$\gamma(t) = \begin{cases} \sigma^2 & \text{si } t = 0 \\ 0 & \text{si } t \neq 0 \end{cases}$$

Et si  $n \geq 1$  et  $t_1, \dots, t_n$  une suite croissante d'entiers, alors la matrice de covariance  $\Lambda(n)$  du vecteur  $(\varepsilon(t_1), \dots, \varepsilon(t_n))'$  est diagonale :  $\Lambda(n) = \sigma^2 \cdot \mathbf{I}_n$  où  $\mathbf{I}_n$  est la matrice unité d'ordre  $n$ . Ceci nous permet de déduire le résultat presque évident suivant :

**Proposition 4** *Un bruit blanc  $(\varepsilon(t) : t \in \mathbb{Z})$  est gaussien si et seulement si, ses marginales unidimensionnelles sont gaussiennes, auquel cas tout vecteur  $(\varepsilon(t_1), \dots, \varepsilon(t_n))'$   $n$  dimensionnel admet une densité de probabilité  $f$  donnée par*

$$f(x_1, \dots, x_n) = \frac{1}{\sqrt{2\pi}^n \cdot \sigma^n} \exp\left\{-\frac{1}{\sigma^2} \sum_{i=1}^n x_i^2\right\}.$$

Enfin nous rappelons le résultat qui est d'ailleurs aisé à démontrer :

**Proposition 5** *Un processus gaussien est stationnaire au second ordre si et seulement si, sa fonction moyenne est constante et sa fonction de covariance est invariante par translation dans le temps.*

Il est tout aussi facile de voir que tout bruit blanc de variance constante est stationnaire et il en est encore de tout bruit blanc gaussien de variance constante. Un tel bruit blanc gaussien forme une suite de variables aléatoires indépendantes et équidistribuées selon une loi normale centrée et de variance  $\sigma^2$ .

### 2.1.2 Modèles de processus gaussiens latents

Ce modèle considère le processus latent  $(\Theta(t) : 1 \leq t \leq T)$  gaussien centré et de matrice de covariance  $\Lambda$  supposée inversible. Le vecteur latent  $(\Theta(1), \dots, \Theta(T))'$  est gaussien et admet une densité de probabilité  $g$  donnée par

$$g(\theta(1), \dots, \theta(T)) = \frac{1}{\sqrt{(2\pi)^T \sqrt{|\Lambda|}}} \exp\left\{-\frac{1}{2} \underline{\theta}' \Lambda^{-1} \underline{\theta}(t)\right\} \quad (2.2)$$

où  $\underline{\theta} = (\theta(1), \dots, \theta(T))'$ .

Nous déduisons directement des résultats du chapitre 1, la loi marginale des observations pour un seul individu.

**Corollaire 3** *Sous les hypothèses  $h_1, h_2$  et  $h_3$  ci-dessus, la loi de probabilité du vecteur des observations  $\underline{\mathbf{X}} = (\underline{\mathbf{X}}(1), \dots, \underline{\mathbf{X}}(T))$  au vue d'une trajectoire  $\underline{\mathbf{x}} = (\underline{\mathbf{x}}(1), \dots, \underline{\mathbf{x}}(T))$  et où  $\underline{\mathbf{x}}(t)$  est le vecteur réponse de l'individu durant toute la période  $1, 2, \dots, T$ , est donnée par*

$$p(\underline{\mathbf{x}}) = \int \dots \int_{\mathbb{R}^T} \prod_{t=1}^T [1 - \alpha_0(\theta(t))]^q \left[ \frac{\alpha_0(\theta(t))}{1 - \alpha_0(\theta(t))} \right]^{R(t)} g(\underline{\theta}) d\underline{\theta}$$

où  $\alpha_0(\theta(t)) = \pi(1/\theta(t))$ ,  $g(\underline{\theta}) = g(\theta(1), \dots, \theta(T))$  la densité du vecteur latent donnée par (2.2) et  $R(t) = \sum_{k=1}^q x_k(t)$  le score obtenu à la  $t^{\text{ieme}}$  occasion.

**Corollaire 4** Si le modèle  $((\underline{\mathbf{X}}(t), \Theta(t)) : 1 \leq t \leq T)$  est un modèle de Rasch observable, alors la loi de probabilité des vecteurs d'observation  $\underline{\mathbf{X}}(1), \dots, \underline{\mathbf{X}}(T)$  au vue d'une observation  $\underline{\mathbf{x}}(1), \dots, \underline{\mathbf{x}}(T)$  d'un seul et même individu durant la période  $\{1, \dots, T\}$ , est donnée par

$$p(\underline{\mathbf{x}}(1), \dots, \underline{\mathbf{x}}(T)/\underline{\theta}, \underline{\beta}) = \int \dots \int_{\mathbb{R}^T} \frac{\exp(\sum_{t=1}^T R(t)\theta(t) - \sum_{k=1}^q \beta_k S_k)}{\prod_{t=1}^T \prod_{k=1}^q [1 + \exp((\theta(t) - \beta_k))]} g(\underline{\theta}) d\underline{\theta} \quad (2.3)$$

où  $g$  est la densité du vecteur latent,  $R(t)$  le score,  $S_k = \sum_{t=1}^T x_k(t)$  le nombre de réponses correctes à l'item  $k$  au cours des  $T$ -opportunités,  $\underline{\theta} = (\theta(1), \dots, \theta(T))$ ,  $\underline{\beta} = (\beta(1), \dots, \beta(q))$  et  $d\underline{\theta} = d\theta(1) \dots d\theta(T)$ .

Ces deux résultats découlent directement de la proposition 3 et du corollaire 2 du chapitre 1.

Remarque :

Si la matrice de covariance  $\Lambda$  du vecteur latent  $\underline{\theta} = (\Theta(1), \dots, \Theta(T))'$  est diagonale ce qui correspond à l'indépendance des variables latentes  $\Theta(j), \dots, \Theta(T)$ , alors la loi marginale des observations s'écrit dans ce cas

$$\begin{aligned} p(\underline{\mathbf{x}}(1), \dots, \underline{\mathbf{x}}(T)/\underline{\theta}, \underline{\beta}) &= \frac{1}{\sigma^T \sqrt{(2\pi)^T}} \int \dots \int_{\mathbb{R}^T} \frac{1}{\prod_{t=1}^T \prod_{k=1}^q [1 + \exp((\theta(t) - \beta_k))]} \\ &\times \exp\left\{ \sum_{t=1}^T R(t)\theta(t) - \sum_{k=1}^q \beta_k S_k - \frac{1}{2\sigma^2} \sum_{t=1}^T \theta^2(t) \right\} d\theta(1) \dots d\theta(T) \end{aligned}$$

## 2.2 Modèles de processus moyennes mobiles d'ordre 1 latents

### 2.2.1 Processus de moyenne mobile d'ordre 1

**Définition 3** On appelle processus moyenne mobile d'ordre 1 un processus  $(X(t) : t \geq 1)$  solution de l'équation au différence d'ordre 1 :

$$X(t) = a\varepsilon(t-1) + \varepsilon(t) \quad (2.4)$$

où  $\varepsilon(t) : t \geq 0$  est un bruit blanc fort (suite de variables aléatoires indépendantes), de variance  $\sigma^2$  et  $a$  une constante réelle.

On note  $\mathcal{MA}(1)$  la classe des processus moyennes mobiles d'ordre 1. Ces processus sont stationnaires au second ordre, centrés et de fonction de covariance  $\gamma(\cdot)$  donnée par

$$\gamma(t) = \begin{cases} (1+a^2)\sigma^2 & \text{si } t = 0 \\ a\sigma^2 & \text{si } |t| = 1 \\ 0 & \text{si } |t| \geq 2 \end{cases} \quad (2.5)$$

où  $\delta_i(j)$  est le symbole de Kronecker. D'autre part, une condition suffisante pour assurer l'inversibilité du processus  $\mathcal{MA}(1)$  est que le polynôme de la moyenne mobile  $q(z) = 1 + az$  n'ait pas de zéro à l'intérieur du disque unité (cf. [BRO87]), ceci nécessite que  $|a| < 1$ . Cette condition sera admise pour toute la suite.

### Lois marginales finies dimensionnelles

On fait remarquer d'abord qu'un processus de la classe  $\mathcal{MA}(1)$  est 1-dépendant, au sens que pour tout  $n \geq 2$ , les variables  $(X(s) : s \leq t)$  et  $(X(s+n) : s \geq t)$  sont indépendantes. Aussi, la loi jointe d'une suite finie de variables  $(X(t_1), X(t_2), \dots, X(t_p))$  telle que les écarts temporels  $|t_j - t_{j-1}| > 1$ , coïncide avec la loi produit  $\bigotimes_{1 \leq j \leq n} P_{X_{t_j}}$  des lois marginales  $P_{X_{t_j}}, 1 \leq j \leq n$ .

Concernant la loi jointe des variables  $X_1, X_2, \dots, X_n$ , on a le résultat suivant :

**Proposition 6** Si le bruit blanc  $(\varepsilon(t) : t \geq 0)$  est gaussien, alors le processus moyenne mobile  $(X(t) : t \geq 1)$  définie par l'équation (2.4) est gaussien centré et de

fonction de covariance  $\gamma_X(\cdot)$  donnée par (2.5). En outre la loi jointe des variables  $X_1, X_2, \dots, X_n$  admet une densité de probabilité  $f$  donnée par

$$f(x_1, x_2, \dots, x_n) = \frac{1}{\sqrt{(2\pi)^n \sigma^n \sqrt{|\Delta_n|}}} \exp\left\{-\frac{1}{2\sigma^2} \mathbf{x}' [\Gamma \cdot \Gamma']^{-1} \mathbf{x}\right\}$$

où  $\Gamma$  est une matrice triangulaire supérieure dont les coefficients  $\gamma_{i,j}$  sont donnés

$$\text{par } \gamma_{i,j} = \begin{cases} a & \text{si } i = j \\ 1 & \text{si } j = i + 1 \\ 0 & \text{si } |i - j| \geq 2 \end{cases}$$

et

$$\Delta_n = \frac{1 - \lambda + a^2 [(1 + a^2) - |1 - a^2|]^n}{a^2} + \lambda \frac{[(1 + a^2) + |1 - a^2|]^n}{2^n}$$

où  $\lambda$  une constante réelle.

**Démonstration** La première assertion de la proposition résulte d'une part, du fait que le processus  $(X(t))$  est 1-dépendant et du fait que le produit de gaussiennes est gaussienne, et d'autre part du fait que le vecteur  $(X_1, X_2, \dots, X_n)'$  étant le transformé linéaire du vecteur gaussien  $(\varepsilon_0, \varepsilon_1, \dots, \varepsilon_n)'$ , est donc un vecteur gaussien. Aussi, ceci confirme que le processus  $(X_t)$  est gaussien centré et de fonction de covariance la fonction  $\gamma(\cdot)$  décrite dans la relation (2.5). Il est aisé d'écrire la densité du vecteur  $(X_{t_1}, X_{t_2}, \dots, X_{t_n})'$  lorsque les écarts  $|t_{j+1} - t_j| \geq 2$ , comme produit de densité gaussiennes. Cette loi peut être regardée aussi comme la marginale de la loi du vecteur  $(X_{t_1}, X_{t_1+1}, \dots, X_{t_2-1}, X_{t_2}, X_{t_2+1}, \dots, X_{t_3-1}, X_{t_3}, \dots, X_{t_{n-1}-1}, X_{t_{n-1}}, \dots, X_{t_n})'$ . Aussi, est il suffisant de déterminer dans les deux cas la densité si elle existe du vecteur  $(X_1, X_2, \dots, X_n)'$ . Ce vecteur est le transformé linéaire du vecteur  $(\varepsilon_0, \varepsilon_1, \dots, \varepsilon_n)'$ , par une matrice triangulaire supérieure  $\Gamma$  à  $(n$  lignes et  $n + 1)$  colonnes et dont les éléments  $\gamma_{i,j}$  sont donnés par

$$\gamma_{i,j} = \begin{cases} a & \text{si } i = j \\ 1 & \text{si } j = i + 1 \\ 0 & \text{si } |i - j| \geq 2 \end{cases}$$

Il s'ensuit que le le vecteur  $(X_1, X_2, \dots, X_n)'$  est gaussien centré et de matrice de covariance  $\sigma^2 \Gamma \cdot \Gamma'$  dont les coefficients sont donnés par



$$\delta_{i,j} = \begin{cases} \sigma^2(1+a^2) & \text{si } i = j \\ \sigma^2 a & \text{si } j = i+1 \text{ ou } j = i-1 \\ 0 & \text{si } |i-j| \geq 2 \end{cases}$$

Le déterminant  $\Delta_n$  de cette matrice satisfait à une constante multiplicative près, une équation de récurrence linéaire du second ordre (cf. COX [COXxxx])

$$\Delta_n = (1+a^2)\Delta_{n-1} - a^2\Delta_{n-2}.$$

Le polynôme caractéristique  $p(\Delta)$  de cette équation admet deux zéros réels

$$\Delta_1 = \frac{(1+a^2) - |1-a^2|}{2} \text{ et } \Delta_2 = \frac{(1+a^2) + |1-a^2|}{2}$$

et donc la valeur du déterminant  $\Delta_n$  est

$$\Delta_n = \lambda\Delta_1^n + \mu\Delta_2^n$$

avec  $\lambda$  et  $\mu$  sont deux constantes déterminées par les conditions initiales

$$\begin{cases} \lambda\Delta_1 + \mu\Delta_2 = 1 + a^2 \\ \lambda\Delta_1^2 + \mu\Delta_2^2 = (1+a^2)^2 - a^2 \end{cases}$$

Comme le processus  $(X(t) : t \geq 1)$  est inversible, alors ce déterminant est non nul, ce qui assure que la matrice de covariance  $\sigma^2\Gamma.\Gamma'$  est inversible, et donc le vecteur  $(X_1, X_2, \dots, X_n)'$  admet une densité de probabilité  $f$  telle que énoncée dans la proposition. ■

### 2.2.2 Modèles de processus moyennes mobiles d'ordre 1 latents

Ce modèle considère le processus latent  $\{\Theta(t) : 1 \leq t \leq T\}$  un processus de la classe  $\mathcal{MA}(1)$  inversible :

$$\Theta(t) = a\varepsilon(t-1) + \varepsilon(t)$$

où  $(\varepsilon(t))$  est un bruit blanc gaussien de variance  $\sigma^2$  et  $a$  une constante réelle choisie de sorte que la moyenne mobile  $(\Theta(t))$  est inversible.

Le vecteur latent  $\underline{\Theta} = (\Theta(1), \dots, \Theta_T)'$  admet une densité de probabilité donnée par

$$g(\theta(1), \dots, \theta(T)) = \frac{1}{\sqrt{(2\pi)^n \sigma^T \sqrt{|\Delta_T|}}} \exp\left(-\frac{1}{2\sigma^2} \underline{\theta}' \Delta_T^{-1} \underline{\theta}\right) \quad (2.6)$$

où  $\sigma^2 \Delta_T$  est la matrice de variance covariance du vecteur latent  $\underline{\Theta}$  et où  $\Delta_T$  est la matrice définie par  $\Delta_T = \Gamma \Gamma'$  avec  $\Gamma$  étant la matrice triangulaire supérieure définie comme dans la section ci-dessus, et  $\underline{\theta} = (\theta(1), \dots, \theta(T))'$ .

Comme pour le cas gaussien latent ci-dessus, nous déterminons la loi marginale des observations sous les hypothèses du modèle de Rasch observable.

**Corollaire 5** *Sous les hypothèses  $h_1, h_2$  et  $h_3$  ci-dessus, la loi de probabilité du vecteur des observations  $\underline{\mathbf{X}} = (\underline{\mathbf{X}}(1), \dots, \underline{\mathbf{X}}(T))$  au vue d'une trajectoire  $\underline{\mathbf{x}} = (\underline{\mathbf{x}}(1), \dots, \underline{\mathbf{x}}(T))$  et où  $\underline{\mathbf{x}}(t)$  est le vecteur réponse de l'individu dlors de la  $t^{ieme}$  opportunité, est donnée par*

$$p(\underline{\mathbf{x}}) = \int \dots \int_{\mathbb{R}^T} \prod_{t=1}^T [1 - \alpha_0(\theta(t))]^q \left[ \frac{\alpha_0(\theta(t))}{1 - \alpha_0(\theta(t))} \right]^{R(t)} g(\underline{\theta}) d\underline{\theta}$$

où  $\alpha_0(\theta(t)) = \pi(1/\theta(t))$ ,  $g(\underline{\theta}) = g(\theta(1), \dots, \theta(T))$  la densité du vecteur latent donnée par (2.6) et  $R(t) = \sum_{k=1}^q x_k(t)$  le score obtenu à la  $t^{ieme}$  occasion.

**Corollaire 6** *Si le modèle  $((\underline{\mathbf{X}}(t), \Theta(t)) : 1 \leq t \leq T)$  est un modèle de Rasch observable et si le processus latent  $\{\Theta(t) : 1 \leq t \leq T\}$  est un processus moyenne mobile d'ordre 1 inversible, alors la loi de probabilité des vecteurs d'observation  $\underline{\mathbf{X}}(1), \dots, \underline{\mathbf{X}}(T)$  au vue d'une observation  $\underline{\mathbf{x}}(1), \dots, \underline{\mathbf{x}}(T)$  d'un seul et même individu durant la période  $\{1, \dots, T\}$ , est donnée par*

$$p(\underline{\mathbf{x}}(1), \dots, \underline{\mathbf{x}}(T) / \underline{\theta}, \underline{\beta}) = \int \dots \int_{\mathbb{R}^T} \frac{\exp(\sum_{t=1}^T R(t)\theta(t) - \sum_{k=1}^q \beta_k S_k)}{\prod_{t=1}^T \prod_{k=1}^q [1 + \exp((\theta(t) - \beta_k))]} g(\underline{\theta}) d\underline{\theta} \quad (2.7)$$

où  $g$  est la densité du vecteur latent donnée par (2.6),  $R(t)$  le score,  $S_k = \sum_{t=1}^T x_k(t)$  le nombre de réponses correctes à l'item  $k$  au cours des  $T$ -opportunités,  $\underline{\theta} = (\theta(1), \dots, \theta(T))$ ,  $\underline{\beta} = (\beta(1), \dots, \beta(q))$  et  $d\underline{\theta} = d\theta(1) \dots d\theta(T)$ .

## Chapitre 3

# Les modèles de processus de Markov latents

Dans cette section, nous considérons deux classes de processus de Markov latents. Il s'agit de la classe des processus de Markov d'ordre 1 latents et de la classe des processus autorégressifs d'ordre 1 notées  $(\mathcal{AR}(1))$  latent.

## 3.1 Modèles de processus de Markov d'ordre 1 latents

### 3.1.1 Processus de Markov d'ordre 1

Nous rappelons pour commencer quelques notions élémentaires sur les processus de Markov et nous renvoyons aux références classiques et en particulier ceux de Feller [FEL58].

**Définition 4** Soit  $(X_t : t \in \mathbb{Z})$  un processus aléatoires à valeurs dans une partie  $\mathbb{E}$  de  $\mathbb{R}$ . On dit que le processus  $(X_t)$  est un processus de Markov d'ordre 1, si l'état du processus à l'instant  $t + 1$  sachant les états antérieurs à l'instant  $t + 1$  ne dépend que de l'état immédiatement antérieur, c'est à-dire de l'état du processus à l'instant  $t$ .

Ainsi, si  $(X_t)$  est un processus de Markov d'ordre 1, alors pour tout borélien  $B$  et pour toute suite  $\{x_s : s \leq t\} \subset \mathbb{E}$ , on a

$$P(X_{t+1} \in B / X_s = x_s : s \leq t) = P(X_{t+1} \in B / X_t = x_t) \quad (3.1)$$

La probabilité conditionnelle  $P(X_{t+1} \in B / X_t = x_t)$  est appelée une probabilité de transition sur  $\mathbb{E} \times \mathcal{B}$ .

Un processus de Markov à valeurs dans un ensemble fini est dit une chaîne de Markov d'ordre 1. Les probabilités de transition notées  $p_{i,j}(t)$  forment une matrice  $\mathbf{P} = (p_{i,j}(t) : 1 \leq i, j \leq n)$  dite matrice des probabilités de transition.

**Définition 5** Un processus de Markov d'ordre 1  $(X_t : t \in \mathbb{Z})$  est dit homogène si pour tout  $t$  :

$$P(X_{t+1} \in B / X_t = x) = P(X_1 \in B / X_0 = x)$$

On déduit en utilisant la formule de Bayes, qu'un processus de Markov homogène d'ordre 1 est stationnaire au second ordre si les variables aléatoires  $X_t$  ont même loi de probabilité.

La loi commune  $P_0$  est appelée la loi initiale et la chaîne de Markov est dite invariante si la loi initiale  $P_0$  est solution de l'équation

$$P_0 = P_0 P$$

où  $P$  est la matrice des probabilités de transition.

Soit  $(X_t : t \in \mathbb{Z})$  une chaîne de Markov homogène d'ordre 1. Si pour toute réalisation  $\omega$  du hasard, la probabilité de transition  $P(X_1/X_0(\omega))$  est absolument continue par rapport à la mesure de Lebesgue, on peut écrire

$$P(X_1 \in B/X_0 = x) = \int_B f^x(y)dy$$

où  $f^x(\cdot)$  est la densité de probabilité de la loi  $P(X_1 \in B/X_0 = x)$ .

La densité de probabilité conditionnelle  $f^x(\cdot)$  est appelée un noyau de transition.

**Définition 6** *Un processus de Markov homogène d'ordre 1, de probabilité de transition absolument continue par rapport à la mesure de Lebesgue dans  $(\mathbb{R}, \mathcal{B})$ , est dite irréductible si  $f^x(y) > 0$  pour tous réels  $x$  et  $y$ , et où  $f^{(\cdot)}(\cdot)$  est le noyau de transition du processus de Markov.*

**Définition 7** *Un état  $x$  d'une chaîne de Markov est dit périodique s'il existe un entier strictement positif  $m(x)$  tel que  $m(x) = \min\{m \in \mathbb{N}^* : p_{x,x}^m > 0\}$ . Un état d'une chaîne de Markov d'ordre 1 est dit apériodique s'il n'est pas périodique. Et la chaîne est dite apériodique si tous ses états ne sont pas périodiques.*

Enfin on termine ce bref rappel par l'analogie du théorème de Bayes.

**Proposition 7** *Soit  $(X_t : t \in \mathbb{Z})$  une chaîne de Markov homogène d'ordre 1, de loi invariante  $f_0$ . Si la probabilité de transition  $P(X_1/X_0(\omega))$  admet une densité de probabilité  $f$  alors la loi jointe de  $X_0, X_1, \dots, X_n$  admet une densité de probabilité  $f$  donnée par*

$$f(x_0, x_1, \dots, x_n) = f_0(x_0) \prod_{k=1}^n f^{x_{k-1}}(x_k) \quad (3.2)$$

où  $f^x(\cdot)$  est la densité de probabilité conditionnelle de la loi de probabilité de transition  $P'(X_1 = \cdot / X_0 = x)$ .

### 3.1.2 Processus de Markov Rasch Latent

Ce modèle considère le processus latent  $(\Theta(t) : 1 \leq t \leq T)$  comme étant un processus de Markov d'ordre 1. On suppose le processus  $(\Theta(t) : 1 \leq t \leq T)$  à support réel, et de loi invariante gaussienne centrée et de variance  $\sigma^2$  et possédant un noyau de transition  $\mathcal{N}$  gaussien de variance  $\sigma^2$ .

**Proposition 8** *Si le processus  $\{\Theta(t) : 1 \leq t \leq T\}$  est à support réel, de loi invariante gaussienne centrée et de variance  $\sigma^2$ , avec pour noyau de transition  $\mathcal{N}$  gaussien et de variance  $\sigma^2$ , alors le vecteur  $(\Theta(1), \dots, \Theta(T))$  est gaussien et a une densité de probabilité  $g$  donnée par*

$$g(\theta(1), \dots, \theta(T)) = \frac{1}{\sigma^T \sqrt{(2\pi)^T}} \exp\left\{-\frac{1}{2\sigma^2}[\theta(1)^2 + \sum_{t=2}^T (\theta(t) - \theta(t-1))^2]\right\} \quad (3.3)$$

**Démonstration**

Par conditionnement, la loi jointe des variables  $\Theta(1), \dots, \Theta(T)$  se déduit facilement :

$$g(\theta(1), \dots, \theta(T)) = \lambda(\theta(1)) \prod_{t=2}^T \mathcal{N}(\theta(t-1), \theta(t)) \quad (3.4)$$

où  $\lambda$  est la loi invariante du processus  $\{\Theta(t) : 1 \leq t \leq T\}$ . Et d'ici nous en déduisons l'expression de  $g$ . ■

**Remarque :**

Si on suppose la chaîne  $\Theta_1, \dots, \Theta_T$  à support fini  $\theta_1, \dots, \theta_m$ , alors la matrice des probabilités de transition  $\mathbf{P}$  a pour coefficients

$$\pi(\theta_u(t)/\theta_v(t-1)) = P(\Theta(t) = \theta_u(t)/\Theta(t-1) = \theta_v(t-1)), 1 \leq u, v \leq m.$$

Et si de plus, on suppose que la chaîne  $(\Theta(t) : 1 \leq t \leq T)$  est homogène, irréductible et possède une loi invariante, alors la matrice des probabilités de transition est indépendante de  $t$  et les lois de probabilité du vecteur  $(\Theta(t) : 1 \leq t \leq T)$  peuvent être écrites

$$P(\Theta(1) = \theta_{i_1}, \dots, \Theta(T) = \theta_{i_T}) = \lambda(\theta_{i_1}) \prod_{t=2}^T \pi(\theta_{i_t}/\theta_{i_{t-1}}) \quad (3.5)$$

où  $i_t \in \{1, \dots, m\}$ .

Dans ce qui suit, on suppose la chaîne de Markov  $(\Theta(t))$  à support réel, homogène et irréductible.

Le résultat suivant spécifie la loi des observations d'un individu. Il résulte directement de la proposition 1, en remplaçant dans (1.9) la densité  $g$  par son expression

donnée dans (3.3).

**Corollaire 7** *Sous les hypothèses de la proposition 8 ci-dessus et sous les hypothèses  $h_1, h_2$  et  $h_3$  ci-dessus, la loi de  $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_T)$  peut être écrite au vue d'une trajectoire  $\mathbf{x} = (\mathbf{x}(1), \dots, \mathbf{x}(T))$  :*

$$p(\mathbf{x}) = \frac{1}{\sigma^T \sqrt{(2\pi)^T}} \int \dots \int_{\mathbf{R}^T} \prod_{t=1}^T [1 - \alpha_0(\theta(t))]^q \left[ \frac{\alpha_0(\theta(t))}{1 - \alpha_0(\theta(t))} \right]^{R(t)} h(\underline{\theta}) d\underline{\theta} \quad (3.6)$$

où  $\alpha_0(\theta(t)) = \pi(1/\theta(t))$ ,  $R(t) = \sum_{k=1}^q x_k(t)$  le score obtenu lors de la  $t^{ieme}$  occasion et

$$h(\underline{\theta}) = h(\theta(1), \dots, \theta(T)) = \exp\left\{-\frac{1}{2\sigma^2} [\theta(1)^2 + \sum_{t=2}^T (\theta(t) - \theta(t-1))^2]\right\}.$$

La probabilité  $p(\mathbf{x}(1), \dots, \mathbf{x}(T))$  dépend des probabilités de transition  $\pi(x_k(t)/\theta(t))$ , aussi, si on suppose de plus que ce modèle est un modèle de Rasch, alors on déduit des corollaires 1 et 2 ci-dessus, le résultat suivant :

**Corollaire 8** *Sous les hypothèses de la proposition 8 ci-dessus et sous les hypothèses du modèle de Rasch latent, la loi  $p(\mathbf{x}(1), \dots, \mathbf{x}(T))$  des observations relatives à un seul individu au cours de la période  $\{1, \dots, T\}$ , est de la forme*

$$p(\mathbf{x}(1), \dots, \mathbf{x}(T)) = \frac{1}{\sigma^T \sqrt{(2\pi)^T}} \int \dots \int \phi(\mathbf{x}, \theta, \beta) h(\theta) d\theta(1) \dots d\theta(T) \quad (3.7)$$

où

$$\phi(\mathbf{x}, \theta, \beta) = \frac{\exp(\sum_{t=1}^T \theta(t) \cdot R(t) - \sum_{k=1}^q \beta_k S_k)}{\prod_{t=1}^T \prod_{k=1}^q [1 + \exp(\theta(t) - \beta_k)]},$$

$S_k = \sum_{t=1}^T x_k(t)$  et  $\theta = (\theta(1), \dots, \theta(T))$ ,  $\beta = (\beta_1, \dots, \beta_q)$ .

**Remarque :**

Si la chaîne est à support fini  $\{\theta_1, \dots, \theta_m\}$ , alors la loi (1.13) devient

$$p(\mathbf{x}) = \sum_{i_1=1}^m \dots \sum_{i_T=1}^m \prod_{t=1}^T \lambda(\theta_{i_1}) \cdot \pi(\theta_{i_t}/\theta_{i_{t-1}}) \cdot [1 - \alpha_0(\theta_{i_t})]^q \cdot \left[ \frac{\alpha_0(\theta_{i_t})}{1 - \alpha_0(\theta_{i_t})} \right]^{R(t)} \quad (3.8)$$

où  $\alpha_0(\theta_{i_t}) = \pi(1/\theta_{i_t})$  et  $r(t) = \sum_{k=1}^q x_k(t)$  le score obtenu à la  $t^{th}$  occasion. Et dans un modèle de Rasch observable, cette probabilité s'écrit

$$p(\mathbf{x}) = \sum_{i_1=1}^m \dots \sum_{i_T=1}^m \prod_{t=1}^T \lambda(\theta_{i_1}) \cdot \pi(\theta_{i_t}/\theta_{i_{t-1}}) \frac{\exp\{\sum_{t=1}^T \theta_t \cdot R(t) - \sum_{k=1}^q \beta_k \cdot S_k\}}{\prod_{t=1}^T \prod_{k=1}^q [1 + \exp(\theta_{i_t} - \beta_k)]} \quad (3.9)$$



## 3.2 Modèles autorégressifs latents

### 3.2.1 Modèles autorégressifs latents

On suppose que le processus latent  $(\Theta(t) : 1 \leq t \leq T)$  est un processus autorégressif d'ordre 1 :

$$\Theta(t) = \rho\Theta(t-1) + \varepsilon(t) \quad (3.10)$$

où  $(\varepsilon(t))$  est un bruit blanc gaussien de variance  $\sigma^2$  et  $\rho$  une constante réelle. Afin d'assurer la stationnarité à l'ordre 2 du processus latent  $(\Theta(t))$ , il est suffisant de considérer le paramètre  $\rho$  de module strictement inférieur à 1. D'autre part, il est aisé de vérifier que ce processus est gaussien et définit un processus de Markov d'ordre 1. Concernant ce dernier point, on a le résultat suivant.

**Proposition 9** *Si le processus latent  $(\Theta(t))$  est un processus autorégressif d'ordre 1, alors ce processus est un processus de Markov d'ordre 1. De plus, si  $|\rho| < 1$ , ce processus est stationnaire au second ordre, de loi initiale la loi normale centrée et de variance  $\frac{\sigma^2}{1-\rho^2}$  et la loi conditionnelle de  $\Theta(t)$  sachant  $\Theta(t-1) = \theta_0$  est gaussienne de moyenne  $\rho\theta_0$  et de variance  $\sigma^2$ .*

#### Démonstration

La loi jointe des variables  $(\Theta(1), \dots, \Theta(T))$  se déduit facilement par conditionnement, et on trouve

$$g(\theta_1, \dots, \theta_T) = f_0(\theta_1) \prod_{t=2}^T f^{\theta_{t-1}}(\theta_t) \quad (3.11)$$

qui fournit

$$g(\theta_1, \dots, \theta_T) = \frac{\sqrt{1-\rho^2}}{\sqrt{(2\pi)^T \sigma^T}} \exp\left\{-\frac{1}{2\sigma^2}[(1-\rho^2)\theta_1^2 + \sum_{t=2}^T (\theta_t - \rho\theta_{t-1})^2]\right\} \quad (3.12)$$

où  $f_0$  est la densité de la loi initiale de la chaîne  $(\Theta(t))$  et  $f^{\theta_{t-1}}(\theta_t)$  la densité conditionnelle de  $\Theta(t)$  sachant  $\Theta(t-1)$ . On fait remarquer que le caractère markovien du processus latent assure que  $\Theta(t)$  est conditionnellement indépendant de  $\Theta(t-2), \dots, \Theta(1)$  sachant  $\Theta(t-1)$ . ■

### 3.2.2 Processus Rasch Latent $\mathcal{AR}(1)$

Comme pour le cas markovien, nous déduisons le résultat suivant :

**Corollaire 9** *Sous les hypothèses que les processus latent  $\Theta(1), \dots, \Theta(T)$  est un  $\mathcal{AR}(1)$  causal et si le modèle satisfait les hypothèses  $h_1, h_2$  et  $h_3$  ci-dessus, alors la loi jointe des observations  $(\mathbf{X}(1), \dots, \mathbf{X}(T))$  s'écrit au vue d'une trajectoire  $\underline{\mathbf{x}} = (\underline{\mathbf{x}}(1), \dots, \underline{\mathbf{x}}(T))$*

$$p(\underline{\mathbf{x}}, \underline{\theta}, \rho) = \frac{\sqrt{1-\rho^2}}{\sqrt{(2\pi)^T \cdot \sigma^T}} \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \psi(\underline{\mathbf{x}}, \underline{\theta}, \rho) \times H(\underline{\mathbf{x}}, \underline{\theta}, \rho) d\theta(T) \dots d\theta(1)$$

où

$$\begin{aligned} \psi(\underline{\mathbf{x}}, \underline{\theta}, \rho) &= (1 - \alpha_0(\theta(1)))^q e^{-\frac{1}{2\sigma^2}(1-\rho^2)\theta(1)^2} \left( \frac{\alpha_0(\theta(1))}{1-\alpha_0(\theta(1))} \right)^{R(1)} \\ H(\underline{\mathbf{x}}, \underline{\theta}, \rho) &= \prod_{t=2}^T (1 - \alpha_0(\theta(t)))^q \left( \frac{\alpha_0(\theta(t))}{1-\alpha_0(\theta(t))} \right)^{R(t)} \cdot e^{-\frac{1}{2\sigma^2}[(\theta(t)-\rho\theta(t-1))^2]}, \\ \alpha_0(\theta(t)) &= \pi(1/\theta(t)) \text{ et } R(t) = \sum_{k=1}^q x_k(t). \end{aligned}$$

#### Démonstration

Il suffit d'écrire

$$p(\underline{\mathbf{x}}, \underline{\theta}, \rho) = \int \dots \int_{\mathbf{R}^T} p(\underline{\mathbf{x}}/\underline{\theta}) g(\theta(1), \dots, \theta(T)) d\theta(1) \dots d\theta(T) \quad (3.13)$$

et tenant compte des hypothèses  $h_1$  et  $h_3$ , cette probabilité devient

$$p(\underline{\mathbf{x}}, \underline{\theta}, \rho) = \int \dots \int_{\mathbf{R}^T} \prod_{t=1}^T p(\underline{\mathbf{x}}(t)/\theta(t)) g(\theta(1), \dots, \theta(T)) d\theta(1) \dots d\theta(T) \quad (3.14)$$

ce qui donne

$$p(\underline{\mathbf{x}}, \underline{\theta}, \rho) = \int \dots \int_{\mathbf{R}^T} \prod_{t=1}^T \prod_{k=1}^q \pi(x_k(t)/\theta(t)) g(\theta(1), \dots, \theta(T)) d\theta(1) \dots d\theta(T) \quad (3.15)$$

D'où compte tenu de la définition de la densité  $g$  du vecteur latent et de ce que  $\pi(x_k(t)/\theta(t)) = [\alpha_0(\theta(t))]^{x_k(t)} [1 - \alpha_0(\theta(t))]^{1-x_k(t)}$ , on obtient l'expression de cette loi. ■

La loi de probabilité  $p(\underline{\mathbf{x}}(1), \dots, \underline{\mathbf{x}}(T))$  dépend des probabilités de transition  $\pi(x_k(t)/\theta(t))$ , aussi si l'on suppose que ce modèle est un modèle de Rasch observable, alors on a

**Corollaire 10** *Si les probabilités conditionnelles  $\pi(x_k(t)/\theta(t))$  suivent le modèle de Rasch, alors la loi du modèle pour un seul individu, s'écrit*

$$p(\underline{\mathbf{x}}) = \frac{\sqrt{1-\rho^2}}{\sqrt{(2\pi)^T \sigma^T}} \int_{-\infty}^{+\infty} \frac{\exp\{-\frac{1}{2\sigma^2}(1-\rho^2)\theta(1)^2 + \theta(1).R(1) - S(1)\}}{\prod_{k=1}^q (1 + \exp(\theta(1) - \beta_k))} \times F(\underline{\mathbf{x}}) d\theta(1) \quad (3.16)$$

où

$$F(\underline{\mathbf{x}}) = \prod_{t=T}^2 \int_{-\infty}^{+\infty} \frac{\exp\{-\frac{1}{2\sigma^2}(\theta(t) - \rho\theta(t-1))^2 + \theta(t).R(t) - S(t)\}}{\prod_{k=1}^q (1 + \exp(\theta(t) - \beta_k))} d\theta(T) d\theta(T-1) \dots d\theta(2)$$

$$\text{et } S(t) = \sum_{k=1}^q x_k(t) \beta_k.$$

### Démonstration

Il suffit de remplacer dans la relation (3.13) les probabilités de transition  $\pi(x_k(t)/\theta(t)) = [\alpha_0(\theta(t))]^{x_k(t)} [1 - \alpha_0(\theta(t))]^{1-x_k(t)}$  par leurs équivalents de Rasch et de déduire ensuite le résultat demandé. ■



## Chapitre 4

# Estimation des paramètres du modèle de Rasch latent

Dans ce chapitre, nous décrivons différentes approches de l'estimation des paramètres du modèle de Rasch latent. Les différentes approches s'appuient toutes sur la méthode du maximum de vraisemblance. Le premier est consacré à l'estimation des paramètres du modèle de Rasch Latent à effets fixes et le second à l'estimation des paramètres du modèle de Rasch latent observable.

## 4.1 Modèles de Rasch à effets fixes

La variable latente relative à chaque individu est supposée constante. Le paramètre à estimer est constitué des effets fixes  $(\theta_1, \dots, \theta_n)$  et des paramètres de difficulté  $\beta_1, \dots, \beta_q$  des items. Les paramètres latents  $\theta_i$  sont des fonctions déterministes du temps  $t$ . On s'attend à ce que l'aptitude de l'individu  $i$  s'améliore au fur et à mesure des consultations. On peut supposer par exemple que  $\theta_i = a_i t + b_i$  où  $a_i$  et  $b_i$  sont des paramètres réels relatifs à l'individu  $i$ .

On rappelle que les vecteurs réponses des individus sont indépendants. Aussi, au vu des observations  $(\underline{x}(1), \underline{x}(2), \dots, \underline{x}(T))$  où  $\underline{x}(t) = (\underline{x}_1(t), \dots, \underline{x}_n(t))$  et où  $\underline{x}_i(t) = (x_{i,1}(t), \dots, x_{i,q}(t))$  est le vecteur réponse de l'individu  $i$  à l'instant  $t$ , et compte tenu des hypothèses de Rasch à effets fixes, la fonction de vraisemblance s'écrit

$$L(\underline{\mathbf{x}}(1), \dots, \underline{\mathbf{x}}(T); \underline{\theta}, \underline{\beta}) = \frac{\exp\{\sum_{i=1}^n \langle \underline{\theta}_i, \underline{R}_i \rangle - \langle \underline{\beta}, \underline{S} \rangle\}}{\prod_{i=1}^n \prod_{t=1}^T \prod_{k=1}^q [1 + \exp((\theta_i(t) - \beta_k))]} \quad (4.1)$$

où  $\underline{R}_i = (R_i(1), \dots, R_i(T))'$  et  $R_i(t)$  la statistique score de l'individu  $i$  lors de la  $t^{\text{ième}}$  consultation,  $\underline{S} = (S_1, \dots, S_q)'$  où  $S_k = \sum_{i=1}^n S_{i,k}$  et  $S_{i,k} = \sum_{t=1}^T x_{i,k}(t)$  le nombre de réponses correctes à l'item  $k$  de l'individu  $i$  au cours des  $T$  opportunités.

Il apparaît au vu de la relation (3.1) que la statistique  $\underline{R} = (\underline{R}_1, \dots, \underline{R}_n)'$  et  $\underline{S} = (\underline{S}_1, \dots, \underline{S}_q)'$  est une statistique exhaustive pour le paramètre  $(\underline{\theta}, \underline{\beta})$ .

Le paramètre à effet fixe  $\underline{\theta}$  est dans un espace de dimension égale à la taille de l'échantillon (le nombre des individus consultés). Aussi, un estimateur de ce paramètre par une des méthodes d'estimation ne peut être consistant. L'estimation conjointe du paramètre  $(\underline{\theta}, \underline{\beta})$  par la méthode du maximum de vraisemblance appelée **méthode d'estimation du maximum de vraisemblance conjointe** fournit des estimateurs difficiles à calculer et qui nécessitent des méthodes itératives et de plus les estimateurs obtenus ne sont pas consistants. Aussi, on adopte la démarche suivante. On estime par la méthode du maximum de vraisemblance le vecteur  $\underline{\beta}$  seulement en prenant la vraisemblance conditionnelle par rapport à la statistique exhaustive  $\underline{R}$  des scores afin d'éliminer les paramètres à effets fixes regardés comme

paramètres nuisibles et ensuite on estime le paramètre à effets fixes  $\underline{\theta}$  par la méthode du maximum de vraisemblance, en remplaçant le paramètre item  $\underline{\beta}$  par son estimation obtenue lors de l'étape précédente.

#### 4.1.1 Maximum de vraisemblance conjointe

Dans cette situation le paramètre à effet fixe est regardé comme une variable presque sûrement constante. La fonction du log de vraisemblance s'écrit

$$l(\underline{x}; \underline{\theta}, \underline{\beta}) = \sum_{i=1}^n \sum_{t=1}^T \theta_i(t) R_i(t) - \sum_{i=1}^n \sum_{k=1}^q \beta_k S_{i,k} - \sum_{i=1}^n \sum_{t=1}^T \sum_{k=1}^q \ln[1 + \exp((\theta_i(t) - \beta_k))] \quad (4.2)$$

Une estimation du maximum de vraisemblance du paramètre  $(\underline{\theta}, \underline{\beta})$  est solution du système d'équations

$$\left\{ \begin{array}{l} \sum_{k=1}^q \frac{e^{\theta_i(t) - \beta_k}}{1 + e^{\theta_i(t) - \beta_k}} = R_i(t) \\ i = 1, \dots, n, \text{ et } t = 1, \dots, T, \\ \text{et} \\ \sum_{i=1}^n \sum_{t=1}^T \frac{\exp((\theta_i(t) - \beta_k))}{1 + e^{\theta_i(t) - \beta_k}} = \sum_{i=1}^n S_{i,k} \\ k = 1, \dots, q \text{ et } i = 1, \dots, n. \end{array} \right.$$

Pour que le modèle soit identifiable, on impose en général la contrainte  $\sum_{k=1}^q \beta_k = 0$ . Ce système ne peut être résolu analytiquement et on recourt le plus souvent à des méthodes numériques itératives tel l'algorithme de Newton - Raphson.

Comme signalé plus haut l'estimateur des effets fixes  $\underline{\theta}$  ne peut être consistant à cause de sa dimension (il y a autant de paramètres que d'individus) alors que le nombre d'items reste constant. Concernant le vecteur des paramètres de difficulté, on a le résultat suivant dû à Ghosh [GHO98].

**Théorème 1** Soit  $\{(\mathbf{X}_{i,k}(t), \Theta_i) : 1 \leq i \leq n, 1 \leq k \leq q, 1 \leq t \leq T\}$  un modèle de rasch à effets fixes. L'estimateur du maximum de vraisemblance  $\hat{\beta} = (\hat{\beta}_1, \dots, \hat{\beta}_q)$  du paramètre de difficulté  $\beta = (\beta_1, \dots, \beta_q)$ , est un estimateur non consistant du vecteur des paramètres de difficulté.

### 4.1.2 Maximum de vraisemblance conditionnelle

Un moyen de résoudre le problème de la consistance de l'estimateur du paramètre de difficulté  $\underline{\beta}$  compte tenu de l'exhaustivité de la statistique  $\underline{R} = (\underline{R}_1, \dots, \underline{R}_n)$ , est de considérer la distribution conditionnelle des observations sachant la statistique exhaustive  $\underline{R}$ . Cette probabilité conditionnelle est indépendante du paramètre à effet fixe  $\underline{\theta}$  et il s'agit alors d'estimer le paramètre de difficulté  $\underline{\beta}$  par la méthode du maximum de vraisemblance.

On commence par calculer la distribution conditionnelle des observations  $\underline{x}$  relativement à la statistique des scores  $\underline{R} = (\underline{R}_1, \dots, \underline{R}_n)$ .

**Proposition 10** *Sous les hypothèses du modèle de Rasch à effets fixes et au vue de l'observation  $\underline{x} = (\underline{x}(1), \underline{x}(2), \dots, \underline{x}(T))$  où  $\underline{x}(t) = (\underline{x}_1(t), \dots, \underline{x}_n(t))$  et  $\underline{x}_i(t)$  le vecteur réponse de l'individu  $i$  au questionnaire lors de la  $t^{\text{ième}}$  consultation, la fonction du log de vraisemblance conditionnelle des observations sachant le vecteur  $(\underline{R}_1 = r_1, \dots, \underline{R}_n = r_n)$ , est donnée par*

$$l_c(\underline{x}, \underline{\theta}, \underline{\beta} \mid r_1, \dots, r_n) = - \sum_{i=1}^n \sum_{k=1}^q \beta_k s_{i,k} - \sum_{i=1}^n \sum_{t=1}^T \log(\gamma_q(r_i(t))) \quad (4.3)$$

où

$$\gamma_q(r_i(t)) = \sum_{(y_{i,k}(t)) : \sum_{k=1}^q y_{i,k}(t) = r_i(t)} \prod_{k: y_{i,k}(t)=1} e^{-\beta_k}$$

est la fonction symétrique élémentaire. De plus, les estimateurs du maximum de la vraisemblance conditionnelle, sont solutions des équations

$$\begin{cases} e^{-\beta_k} \sum_{i=1}^n \sum_{t=1}^T \frac{\gamma_{q-1}(r_i(t)-1)}{\gamma_q(r_i(t))} = \sum_{i=1}^n s_{i,k} \\ k = 1, \dots, q. \end{cases} \quad (4.4)$$

#### Démonstration

En conditionnant par rapport aux scores observés  $(\underline{R}_1 = r_1, \dots, \underline{R}_n = r_n)$ , la vraisemblance conditionnelle ne dépend que du paramètre  $\underline{\beta}$  et s'écrit

$$\begin{aligned} L_c(\underline{x}, \underline{\beta} \mid r_1, \dots, r_n) &= P(\underline{X} = \underline{x} \mid \underline{R}_1 = r_1, \dots, \underline{R}_n = r_n) \\ &= \prod_{i=1}^n P(\underline{X}_i = \underline{x}_i \mid \underline{R}_i = r_i) \end{aligned}$$



car les vecteurs réponses relatifs à chaque individus sont indépendants. Comme  $\underline{X}_i = (\underline{X}_i(1), \dots, \underline{X}_i(T))'$  et  $\underline{X}_i(t) = (X_{i,1}(t), \dots, X_{i,q}(t))'$  pour tout  $t = 1, \dots, T$ , alors on peut écrire pour tout  $i$  :

$$\begin{aligned} P(\underline{X}_i = \underline{x}_i / \underline{R}_i = \underline{r}_i) &= \frac{P(\underline{X}_i = \underline{x}_i, \underline{R}_i = \underline{r}_i / \underline{\theta}_i, \underline{\beta})}{P(\underline{R}_i = \underline{r}_i)} \\ &= \prod_{t=1}^T \frac{P(\underline{X}_i(t) = \underline{x}_i(t), R_i(t) = r_i(t) / \underline{\theta}_i(t), \underline{\beta})}{P(R_i(t) = r_i(t))} \end{aligned}$$

et comme l'événement  $\{\underline{X}_i(t) = \underline{x}_i(t) : \sum_{k=1}^q x_{i,k}(t) = r_i(t)\} \subset \{R_i(t) = r_i(t)\}$ , alors cette probabilité devient

$$P(\underline{X}_i = \underline{x}_i / \underline{R}_i = \underline{r}_i) = \prod_{t=1}^T \frac{P(\underline{X}_i(t) = \underline{x}_i(t) / \underline{\theta}_i(t), \underline{\beta})}{P(R_i(t) = r_i(t))} \quad (4.5)$$

De plus, compte tenu de l'indépendance des observations (hypothèse de l'indépendance locale,) la probabilité du numérateur est égale à

$$P(\underline{X}_i(t) = \underline{x}_i(t) / \underline{\theta}_i(t), \underline{\beta}) = \frac{\exp\{r_i(t)\theta_i - \sum_{k=1}^q \beta_k s_{i,k}\}}{\prod_{k=1}^q (1 + e^{\theta_i(t) - \beta_k})}$$

et celle figurant au dénominateur, s'écrit quant à elle

$$P(R_i(t) = r_i(t)) = \sum_{(y_{i,k}(t)) / \sum_{k=1}^q y_{i,k}(t) = r_i(t)} \frac{\exp\{r_i(t)\theta_i - \sum_{k=1}^q \beta_k y_{i,k}(t)\}}{\prod_{k=1}^q (1 + e^{\theta_i(t) - \beta_k})}$$

En reportant ces expressions dans (3.5) et moyennant quelques manipulations, cette dernière relation devient

$$P(\underline{X}_i(t) = \underline{x}_i(t) / \underline{R}_i(t) = \underline{r}_i(t)) = \frac{\exp\{-\sum_{k=1}^q \beta_k s_{i,k}\}}{\sum_{(y_{i,k}(t)) / \sum_{k=1}^q y_{i,k}(t) = r_i(t)} \exp\{-\sum_{k=1}^q \beta_k y_{i,k}(t)\}}.$$

Par suite, le logarithme de la vraisemblance conditionnelle  $L_c(\underline{x}, \underline{\theta}, \underline{\beta} / \underline{r}_1, \dots, \underline{r}_n)$ , revêt la forme suivante :

$$\begin{aligned} \ln L_c(\underline{x}, \underline{\theta}, \underline{\beta} / \underline{r}_1, \dots, \underline{r}_n) &= - \sum_{i=1}^n \sum_{k=1}^q \beta_k s_{i,k} \\ &\quad - \sum_{i=1}^n \sum_{t=1}^T \log \left\{ \sum_{(y_{i,k}(t)) / \sum_{k=1}^q y_{i,k}(t) = r_i(t)} \exp\left\{- \sum_{k: y_{i,k}(t)=1} \beta_k\right\} \right\} \end{aligned}$$

et que l'on peut encore écrire sous la forme suivante

$$l_c(\underline{x}, \underline{\theta}, \underline{\beta} \mid r_1, \dots, r_n) = - \sum_{i=1}^n \sum_{k=1}^q \beta_k s_{i,k} - \sum_{i=1}^n \sum_{t=1}^T \log(\gamma_q(r_i(t)))$$

où

$$\gamma_q(r_i(t)) = \sum_{(y_{i,k}(t)) : \sum_{k=1}^q y_{i,k}(t) = r_i(t)} \prod_{k: y_{i,k}(t)=1} e^{-\beta_k}$$

la fonction symétrique élémentaire.

En dérivant par rapport à  $\beta_k$ , on obtient

$$\frac{\partial}{\partial \beta_k} l_c(\underline{x}, \underline{\theta}, \underline{\beta} \mid r_1, \dots, r_n) = - \sum_{i=1}^n s_{i,k} - \sum_{i=1}^n \sum_{t=1}^T \frac{1}{\gamma_q(r_i(t))} \frac{\partial}{\partial \beta_k} (\gamma_q(r_i(t)))$$

où

$$\begin{aligned} \frac{\partial}{\partial \beta_k} (\gamma_q(r_i(t))) &= \frac{\partial}{\partial \beta_k} \left( \sum_{(y_{i,k}(t)) : \sum_{k=1}^q y_{i,k}(t) = r_i(t)} \prod_{k: y_{i,k}(t)=1} e^{-\beta_k} \right) \\ &= -e_k^{-\beta} \left\{ \sum_{(y_{i,l}(t)) : \sum_{l=1, l \neq k}^{q-1} y_{i,l}(t) = r_i(t) - 1} \prod_{l: y_{i,l}(t)=1} e^{-\beta_l} \right\} \\ &= -e^{-\beta_k} \gamma_{q-1}(r_i(t) - 1) \end{aligned}$$

d'où nous déduisons les équations (3.4) de la fonction du log de vraisemblance. ■

Une procédure numérique est nécessaire pour le calcul de ses estimateurs et qui nécessite le calcul à chaque étape des fonctions symétriques  $\gamma_q(r_i(t))$ .

Concernant la consistance de l'estimateur du maximum de vraisemblance conditionnelle  $\hat{\underline{\beta}}$ , on a le résultat suivant dû à Planzagl (1993)[PLA93].

**Théorème 2** *L'estimateur du maximum de vraisemblance conditionnelle  $\hat{\underline{\beta}}$  du paramètre de difficulté  $\underline{\beta}$  est un estimateur consistant du vecteur  $\underline{\beta}$ .*

### 4.1.3 Estimation des paramètres latents

On estime par la méthode du maximum de vraisemblance les paramètres à effets fixes  $\underline{\theta}$  en supposant connu les paramètres de difficulté  $\underline{\beta}$  (prendre par exemple pour

valeur de ces paramètres leurs estimés obtenues par la méthode du maximum de vraisemblance conditionnelle).

On rappelle que le log de la fonction de vraisemblance s'écrit (relation (3.2) ci-dessus) :

$$l(\underline{x}; \underline{\theta}, \underline{\beta}) = \sum_{i=1}^n \sum_{t=1}^T \theta_i(t) R_i(t) - \sum_{i=1}^n \sum_{k=1}^q \beta_k S_{i,k} - \sum_{i=1}^n \sum_{t=1}^T \sum_{k=1}^q \ln[1 + \exp((\theta_i(t) - \beta_k))]$$

Les estimateurs du maximum de vraisemblance des paramètres latents, sont solutions des équations

$$\begin{cases} \sum_{t=1}^T \sum_{k=1}^q \frac{e^{\theta_i(t) - \beta_k}}{1 + e^{\theta_i(t) - \beta_k}} = \sum_{t=1}^T R_i(t) \\ i = 1, \dots, n. \end{cases}$$

Cette équation a une solution unique sauf dans les cas de scores nuls ou de scores parfait égale à 1. L'unicité vient de ce que la fonction  $K(\theta) = \sum_{k=1}^q \frac{e^{\theta_i(t) - \beta_k}}{1 + e^{\theta_i(t) - \beta_k}}$  est continue et strictement croissante, nulle à  $-\infty$  et tend vers une limite finie lorsque  $\theta \rightarrow +\infty$ .

Pour calculer la valeur de  $\hat{\theta}_i(t)$ , on utilise l'algorithme de Raphson - Newton :

$$\widehat{\theta}_i^{p+1} = \widehat{\theta}_i^p - \frac{\partial^2}{\partial \theta_i^2} \{ [l(\underline{x}; \underline{\theta}, \underline{\beta})]^{-1} \}_{\theta_i = \widehat{\theta}_i^p} \times \frac{\partial}{\partial \theta_i} \{ [l(\underline{x}; \underline{\theta}, \underline{\beta})]^{-1} \}_{\theta_i = \widehat{\theta}_i^p}$$

expression de laquelle on a volontairement omis l'instant  $t$ .

On trouve à chaque étape  $p$  :

$$\widehat{\theta}_i^{p+1} = \widehat{\theta}_i^p + \frac{r_i(t) - \sum_{k=1}^q \frac{e^{\widehat{\theta}_i^p - \beta_k}}{1 + e^{\widehat{\theta}_i^p - \beta_k}}}{\sum_{k=1}^q \frac{e^{\widehat{\theta}_i^p - \beta_k}}{1 + e^{\widehat{\theta}_i^p - \beta_k}} \cdot 2}.$$

Concernant les propriétés asymptotiques de ces estimateurs, on a le résultat suivant (Lord 1993 [LOR.93]).

**Théorème 3** *On a*

$$var(\widehat{\theta}_i(t)) = \frac{1}{I(\widehat{\theta}_i(t))} + o\left(\frac{1}{q}\right)$$

et le biais de cet estimateur est donné par

$$b(\widehat{\theta}_i(t)) = \frac{\mathbf{J}(\widehat{\theta}_i(t))}{2\mathbf{I}^2(\widehat{\theta}_i(t))} + O\left(\frac{1}{q}\right)$$

où  $\mathbf{I}(\widehat{\theta}_i(t)) = \sum_{k=1}^q \pi_{i,k}(t)(1-\pi_{i,k}(t))$  est la matrice de Fischer,  $\mathbf{J}_i(\widehat{\theta}(t)) = \sum_{k=1}^q \frac{\partial^2 \pi_{i,k}(t)}{\partial \theta_i^2(t)}$  et  $\pi_{i,k}(t) = \frac{e^{\theta_i(t) - \beta_k}}{1 + e^{\theta_i(t) - \beta_k}}$ .

Le résultat suivant (Clauet 1990 [CLA90]) fournit la distribution asymptotique de l'estimateur du maximum de vraisemblance du paramètre individuel  $\theta_i(t)$ .

**Théorème 4** *Sous certaines conditions de régularité, on a*

$$\sqrt{\mathbf{I}(\widehat{\theta}_i(t))}(\widehat{\theta}_i(t) - \theta_i(t)) \xrightarrow{loi} \mathcal{N}(0, 1)$$

## 4.2 Estimation des paramètres d'un modèle de Rasch observable

Les variables latentes étant ici aléatoires et les processus latents individuels  $\underline{\Theta}_1, \dots, \underline{\Theta}_n$  sont supposés indépendants et satisfont les hypothèses du modèle de Rasch observable. De plus, nous supposons que ces processus latents sont gaussiens centrés et de variance respective  $\sigma_i^2$ . En outre, pour assurer l'identifiabilité du modèle, on suppose que la somme  $\sum_{k=1}^n \beta_k = 0$ . On note  $\eta$  le vecteur des paramètres du modèle et dont la structure dépend de la nature du modèle que l'on spécifiera au chapitre suivant pour chacun des modèles envisagés. L'objectif est d'estimer le paramètre  $\eta$  par la méthode du maximum de vraisemblance.

### La vraisemblance marginale

Le modèle de Rasch latent peut être considéré comme un cas particulier de modèles plus généraux pour l'analyse des données incomplètes. Le vecteur des observations ( $\underline{\mathbf{x}}_i$ ) peut être considéré comme incomplet, du fait que le trait latent ne soit pas observé. Ce vecteur incomplet peut être considéré comme une fonction observable du vecteur ( $\underline{\mathbf{x}}_i, \theta_i$ ) des données (partiellement observées) complètes. La fonction de vraisemblance des données incomplètes appelée **vraisemblance marginale**, s'écrit :

$$p(\underline{\mathbf{x}}; \eta) = \prod_{i=1}^n \int \dots \int_{\mathbf{R}^T} p(\underline{\mathbf{x}}_i / \underline{\theta}_i; \eta_i) g_i(\theta_i(1), \dots, \theta_i(T)) d\theta_i(1) \dots d\theta_i(T) \quad (4.6)$$

où  $\underline{\theta}_i = (\theta_i(1), \dots, \theta_i(T))'$ ,  $g_i$  la densité de probabilité de  $\underline{\Theta}_i = (\Theta_i(1), \dots, \Theta_i(T))'$ ,  $\underline{\mathbf{x}} = ((\underline{\mathbf{x}}_1(t), \dots, \underline{\mathbf{x}}_n(t))' : 1 \leq t \leq T)$  une réalisation du processus d'observation  $\underline{\mathbf{X}} = \{(\underline{\mathbf{X}}_1(t), \dots, \underline{\mathbf{X}}_n(t))' : 1 \leq t \leq T\}$  et où  $\underline{\mathbf{x}}_i(t) = (\underline{\mathbf{x}}_{i,1}(t), \dots, \underline{\mathbf{x}}_{i,q}(t))'$  le vecteur réponse de l'individu  $i$  lors de la  $t^{ieme}$  opportunité.

En utilisant les hypothèses caractérisant le modèle de Rasch, la vraisemblance marginale s'écrit ici (cf. (1.10) chap. 1§3) :

$$p(\eta; \underline{\mathbf{x}}) = \prod_{i=1}^n \int \dots \int_{\mathbf{R}^T} p(\underline{\mathbf{x}}_i / \underline{\eta}_i, \underline{\beta}) g_i(\theta_i(1), \dots, \theta_i(T)) d\theta_i(1) \dots d\theta_i(T) \quad (4.7)$$

où la probabilité conditionnelle  $p(\underline{\mathbf{x}}_i/\eta_i, \underline{\beta})$  est ici donnée par

$$p(\underline{\mathbf{x}}_i/\eta_i, \underline{\beta}) = \frac{\exp(\sum_{t=1}^T R_i(t)\theta_i(t) - \sum_{k=1}^q \beta_k S_{i,k})}{\prod_{t=1}^T \prod_{k=1}^q [1 + \exp(\theta_i(t) - \beta_k)]} \quad (4.8)$$

La méthode qui consiste à estimer le paramètre  $\eta$  en maximisant cette vraisemblance marginale est difficile à mettre en oeuvre. Aussi, parmi les approches communément utilisées pour maximiser cette fonction, celle basée sur l'application de l'algorithme E.M est généralement privilégiée. D'autres approches telles l'approche SEM (stochastic EM) ou GEE (équations d'estimation généralisées) sont utilisées comme alternatives à l'approche EM. Une autre approche basée sur la vraisemblance composée (cf. Renard [RENxx]), a été initié récemment par Feddag (2009) [FED09].

Les intégrales multiples figurant dans l'expression de la vraisemblance marginale ne sont pas faciles à calculer, aussi utilise-t-on l'approche des quadratures de Gauss - Hermite (cf. annexe 1) pour obtenir une approximation plus facile à calculer. Ci-dessous, nous décrivons l'algorithme EM et clôturerons cette section par quelques observations sur les approches SEM et GEE.

#### 4.2.1 L'algorithme EM

L'algorithme EM (cf. Dempstair [DEM77]) consiste à maximiser itérativement, non pas la log-vraisemblance des données incomplètes, mais la moyenne conditionnelle de la log-vraisemblance des données complètes par rapport aux observations. En fait, il s'agit de maximiser l'expression suivante de la fonction log-vraisemblance :

$$\mathbf{Q}(\eta/\eta^{(p)}) = \mathbf{E}\{\log[f(\underline{\mathbf{x}}, \underline{\theta}; \eta)]/\underline{\mathbf{x}}, \eta^{(p)}\}$$

où

$$f(\underline{\mathbf{x}}, \underline{\theta}, \eta) = \prod_{i=1}^n p(\underline{\mathbf{x}}_i/\theta_i; \eta_i) g_i(\theta_i, \eta_i),$$

et  $\underline{\theta} = (\theta_1, \dots, \theta_n)'$  avec  $\theta_i = (\theta_i(1), \dots, \theta_i(T))'$ ,  $\underline{\mathbf{x}} = (\underline{\mathbf{x}}_1, \dots, \underline{\mathbf{x}}_n)'$ ,  $\underline{\mathbf{x}}_i = (\underline{\mathbf{x}}_i(1), \dots, \underline{\mathbf{x}}_i(T))'$  et  $\underline{\mathbf{x}}_i(t) = (x_{i,1}(t), \dots, x_{i,q}(t))'$  le vecteur réponse de l'individu  $i$  lors de la  $t^{ieme}$  opportunité.

Plus précisément, à l'étape  $p+1$ , étant donné la valeur  $\eta^{(p)}$  de l'estimation obtenue à l'étape  $p$ , l'algorithme est comme suit :

**Étape E** (Espérance) :

Cette étape consiste à calculer l'espérance conditionnelle de la log-vraisemblance complète sachant les valeurs courantes des estimateurs et le vecteur des observations  $\underline{\mathbf{x}}$

$$\mathbf{Q}(\eta/\eta^{(p)}) = \mathbf{E}(\log f(\underline{\mathbf{x}}, \underline{\theta}; \eta) / \underline{\mathbf{x}}, \eta^{(p)}). \quad (4.9)$$

Nous avons le résultat suivant :

**Proposition 11** *Sous les hypothèses du modèle de Rasch observable, la moyenne conditionnelle  $\mathbf{Q}(\eta/\eta^{(p)})$  est donnée par :*

$$\begin{aligned} \mathbf{Q}(\eta/\eta^{(p)}) &= \sum_{i=1}^n \int \dots \int_{\mathbf{R}^T} \log\{g_i(\underline{\theta}_i, \eta_i)\} \times q_i(\underline{\theta}_i/\underline{\mathbf{x}}_i, \eta^{(p)}) d\theta_i(1) \dots d\theta_i(T) \quad (4.10) \\ &+ \sum_{i=1}^n \int \dots \int_{\mathbf{R}^T} \left\{ \sum_{t=1}^T R_i(t)\theta_i(t) - \sum_{k=1}^q \beta_k S_{i,k} - \sum_{t=1}^T \sum_{k=1}^q \log[1 + e^{\theta_i(t) - \beta_k}] \right\} \\ &\times q_i(\underline{\theta}_i/\underline{\mathbf{x}}_i, \eta^{(p)}) d\theta_i(1) \dots d\theta_i(T) \end{aligned}$$

où

$$q_i(\underline{\theta}_i/\underline{\mathbf{x}}_i, \eta^{(p)}) \propto p(\underline{\mathbf{x}}_i, \eta^{(p)} / \theta_i(1), \dots, \theta_i(T)) \times g_i(\theta_i(1), \dots, \theta_i(T) / \eta^{(p)}) \quad (4.11)$$

et où

$$p(\underline{\mathbf{x}}_i, \eta^{(p)} / \theta_i(1), \dots, \theta_i(T)) = \frac{\exp(\sum_{t=1}^T R_i(t)\theta_i(t) - \sum_{k=1}^q \beta_k^{(p)} S_{i,k})}{\prod_{t=1}^T \prod_{k=1}^q [1 + \exp(\theta_i(t) - \beta_k^{(p)})]}$$

et  $g_i(\theta_i(1), \dots, \theta_i(T) / \eta^{(p)})$  la valeur de cette densité lorsque la valeur du paramètre  $\eta$  est  $\eta^{(p)}$ .

### Démonstration

On a

$$\begin{aligned} \mathbf{Q}(\eta/\eta^{(p)}) &= \sum_{i=1}^n \int \dots \int_{\mathbf{R}^T} [\log\{g_i(\underline{\theta}_i, \eta_i)\} + \log\{p(\underline{\mathbf{x}}_i, \underline{\theta}_i; \eta_i/\eta^{(p)})\}] \\ &\cdot q_i(\underline{\theta}_i/\underline{\mathbf{x}}_i, \eta^{(p)}) d\theta_i(1) \dots d\theta_i(T) \end{aligned}$$

où  $q_i(\underline{\theta}_i/\underline{\mathbf{x}}_i, \eta^{(p)})$  est la densité conditionnelle du vecteur latent  $\underline{\Theta}_i$  sachant le vecteur des observations  $\underline{\mathbf{X}}_i$ .

Par la formule classique de Bayes, cette densité peut être mise sous la forme

$$q_i(\underline{\theta}_i/\underline{\mathbf{x}}_i, \eta^{(p)}) = \frac{p(\underline{\mathbf{x}}_i/\underline{\theta}_i, \eta^{(p)})}{p(\underline{\mathbf{x}}_i/\eta^{(p)})} \times g(\underline{\theta}_i/\eta^{(p)}).$$

Et comme

$$\begin{aligned} p(\underline{\mathbf{x}}_i/\eta^{(p)}) &= \int \dots \int_{\mathbb{R}^T} p(\underline{\mathbf{x}}_i/\underline{\theta}_i, \eta^{(p)}) \\ &\quad \times g(\theta_i(1), \dots, \theta_i(T), \eta^{(p)}) d\theta_i(1) \dots d\theta_i(T) \end{aligned}$$

il en résulte que  $q_i(\underline{\theta}_i/\underline{\mathbf{x}}_i, \eta^{(p)})$  coïncide à une constante près avec

$$p(\underline{\mathbf{x}}_i/\underline{\theta}_i, \eta^{(p)}) \times g(\underline{\theta}_i/\eta^{(p)})$$

et où  $p(\underline{\mathbf{x}}_i/\underline{\theta}_i, \eta^{(p)})$  est donnée par la relation (3.8) en remplaçant le paramètre  $\eta$  par son estimé  $\eta^{(p)}$ .

De plus, comme

$$\begin{aligned} \log\{p(\underline{\mathbf{x}}, \underline{\theta}, \eta/\underline{\mathbf{x}}, \eta^{(p)})\} &= \sum_{i=1}^n \sum_{t=1}^T \sum_{k=1}^q \int \dots \int_{\mathbb{R}^T} \log[\pi(\mathbf{x}_{i,k}(t)/\theta_i(t), \eta_i^{(p)})] \\ &\quad \cdot q_i(\underline{\theta}_i(t)/\underline{\mathbf{x}}_i, \eta^{(p)}) d\theta_i(1) \dots d\theta_i(T) \end{aligned}$$

et que

$$\pi(\mathbf{x}_{i,k}(t)/\theta_i(t), \eta_i^{(p)}) = \frac{e^{x_{i,k}(t)(\theta_i(t)-\beta_k)}}{1 + e^{\theta_i(t)-\beta_k}}$$

le résultat de la proposition en découle. ■

On remarquera que la densité  $q_i$  est en fait la densité de la loi à postériori du paramètre sachant les observations et la densité  $g_i$  est la loi à priori.

### Étape M (Maximisation)

Cette deuxième étape consiste à trouver la valeur  $\eta^{(p+1)}$  qui réalise le maximum de la quantité  $Q(\eta/\eta^{(p)})$  :

$$Q(\eta^{(p+1)}) = \arg \max_{\eta} Q(\eta/\eta^{(p)}). \quad (4.12)$$



Cette étape exige une écriture explicite des densités  $g_i$  des vecteurs latents  $\underline{\Theta}_i$  et de la probabilité  $q_i$  fournit par la relation (4.10) afin de faire ressortir la dépendance de  $Q(\eta/\eta^{(p)})$  comme une fonction des paramètres du modèle.

Pour calculer la valeur  $\eta^{(p+1)}$ , on applique l'algorithme de Raphson - Newton et les intégrales sont approchées numériquement en utilisant les formules des quadratures de Gauss-Hermite.

Les étapes **E** and **M** sont alternativement répétées jusqu'à ce que la différence

$$Q(\eta^{(p+1)}) - Q(\eta^{(p)})$$

soit inférieure à une certaine quantité fixée à priori.

### Remarque

Pas moins de trois algorithmes sont utilisés pour trouver une approximation de l'estimateur du maximum de vraisemblance : les quadratures de Gauss Hermite pour approximer les intégrales, l'algorithme EM qui nécessite à son tour l'utilisation de l'algorithme de Raphson - Newton pour maximiser la vraisemblance approchée.

D'autres approches d'estimation sont utilisées et permettent de contourner certaines difficultés rencontrées dans l'approche **EM**. Il s'agit de

1. L'approche **SEM Stochastic EM (SEM) (Hanon 2000) [HAN00]**, basée sur la méthode du maximum de vraisemblance. Elle permet de contourner l'écueil du calcul des intégrales multiples et d'éviter en conséquent l'algorithme de Gauss - Hermite. Cette approche consiste à utiliser la vraisemblance jointe en complétant les données manquantes relatives aux paramètres individuels  $\theta_i$  et ce en simulant ces dernières suivant leur loi respective.

La procédure de simulation repose sur l'algorithme de métropolis [xxx].

2. L'approche (**GEE (Liang et Zeger 1986) [ZEG86]**) (méthode des équations d'estimation généralisées) qui constitue une alternative à l'estimation par la méthode du maximum de vraisemblance. C'est une extension de l'approche

quasi-vraisemblance de Wedderburn, 1974 [WED74]) aux mesures répétées. Cette méthode nécessite la connaissance des deux premiers moments des variables réponses et une relation fonctionnelle entre la variance et la moyenne. Cette dernière relation est vérifiée par les modèles linéaires généralisés.

Pour plus de détails concernant ces différentes approches nous renvoyons à l'excellent papier de Adname et Mesbah [ADN06].

## Chapitre 5

# Applications

Nous appliquons l'approche **EM** à chacun des modèles étudiés dans les chapitres précédents. Chacune de ces classes de modèles est supposée satisfaire les hypothèses du modèle de Rasch. On suppose que les processus latents  $(\Theta_i(t) : 1 \leq t \leq T)$ ,  $i = \overline{1, n}$ , sont indépendants et normalement distribués de variance respective  $\sigma_i^2$ .

## 5.1 Bruit blanc gaussien Latent

Dans ce cas le paramètre  $\eta = (\beta, \sigma^2)$  avec  $\beta = (\beta_1, \dots, \beta_q)'$  le vecteur des paramètres de difficulté et  $\sigma^2 = (\sigma_1^2, \dots, \sigma_n^2)'$  et où  $\sigma_i^2$  est la variance du bruit blanc gaussien  $\Theta_i$ .

On rappelle que la densité  $g_i$  est dans ce cas donnée par

$$g_i(\theta_i(1), \dots, \theta_i(T)) = \frac{1}{\sqrt{(2\pi)^T \cdot \sigma_i^T}} \exp\left\{-\frac{1}{\sigma_i^2} \sum_{t=1}^T \theta_i(t)^2\right\}.$$

Le résultat suivant qui découle directement des relations (4.10) et (4.11) ci-dessus, nous permet de préciser la moyenne conditionnelle  $\mathbf{Q}(\eta/\eta^{(p)})$ .

**Corollaire 11** *Sous les hypothèses du modèle de Rasch, la moyenne conditionnelle  $\mathbf{Q}(\eta/\eta^{(p)})$  de la log-vraisemblance complète sachant la valeur courante de l'estimateur  $\eta^{(p)}$  est donnée par*

$$\begin{aligned} \mathbf{Q}(\eta/\eta^{(p)}) &= \sum_{i=1}^n \int \dots \int_{\mathbf{R}^T} \left\{ -\frac{T}{2} \log(2\pi) - \frac{T}{2} \log(\sigma_i^2) - \frac{1}{2\sigma_i^2} \sum_{t=1}^T \theta_i^2(t) \right\} \\ &\quad \times q_i(\underline{\theta}_i/\underline{\mathbf{x}}_i, \eta^{(p)}) d\theta_i(1) \dots d\theta_i(T) \\ &+ \sum_{i=1}^n \int \dots \int_{\mathbf{R}^T} \left\{ \sum_{t=1}^T R_i(t) \theta_i(t) - \sum_{k=1}^q \beta_k S_{i,k} - \sum_{t=1}^T \sum_{k=1}^q \log(1 + e^{\theta_i(t) - \beta_k}) \right\} \\ &\quad \times q_i(\underline{\theta}_i/\underline{\mathbf{x}}_i, \eta^{(p)}) d\theta_i(1) \dots d\theta_i(T) \end{aligned} \quad (5.1)$$

où

$$\begin{aligned} q_i(\underline{\theta}_i/\underline{\mathbf{x}}_i, \eta^{(p)}) &\propto \frac{1}{\sqrt{(2\pi)^T (\sigma_i^{(p)})^T}} \\ &\quad \exp\left\{ \sum_{t=1}^T \theta_i(t) \cdot R_i(t) - \sum_{k=1}^q \beta_k^{(p)} S_{i,k} - \frac{1}{2(\sigma_i^{(p)})^2} \sum_{t=1}^T \theta_i^2(t) \right\} \\ &\quad \times \frac{1}{\prod_{t=1}^T \prod_{k=1}^q [1 + \exp(\theta_i(t) - \beta_k^{(p)})]} \end{aligned}$$

La maximisation par rapport aux  $\sigma_i^2$  concerne uniquement le deuxième et troisième termes de la première somme sur  $i$  et la maximisation par rapport aux  $\beta_k$  fait appel uniquement au second et troisième terme de la seconde somme sur  $i$ . Ceci nous amène à calculer

$$\sigma_i^{2(p+1)} = \frac{1}{T} \int \dots \int \sum_{t=1}^T \theta_i(t)^2 q_i(\underline{\theta}_i/\underline{\mathbf{x}}, \beta^{(p)}, \sigma^{2(p)}) d\theta_i(1) \dots d\theta_i(T)$$

pour tout  $i$  et à résoudre les  $q$  équations suivantes :

$$\sum_{i=1}^n \sum_{t=1}^T \int \dots \int \frac{1}{1+\exp((\beta_k^{(p+1)} - \theta_i(t)))} q_i(\underline{\theta}_i / \underline{\mathbf{x}}, \eta^{(p)}) d\theta_i(1) \dots d\theta_i(T) = \sum_{i=1}^n S_{i,k}$$

pour tout  $k : 1 \leq k \leq q$ .

Et les intégrales sont de nouveau approchées numériquement en utilisant l'algorithme de Raphson-Newton et les quadratures de Gauss-Hermite.

## 5.2 Processus Rasch Latent $\mathcal{MA}(1)$

Dans ce cas, on rappelle que le paramètre est  $\eta = (\underline{\beta}, \underline{\alpha}, \sigma^2)$  avec  $\beta = (\beta_1, \dots, \beta_q)$  le vecteur des paramètres de difficulté,  $\alpha = (\alpha_1, \dots, \alpha_n)$  le vecteur des paramètres des processus moyennes mobiles latents ( $\underline{\Theta}_i : i = \overline{1, n}$  et  $\sigma^2 = (\sigma_1^2, \dots, \sigma_n^2)$ ). Précisons pour commencer que la densité du vecteur latent  $(\Theta_i(1), \dots, \Theta_i(T))'$  peut être réécrite sous la forme

$$g_i(\theta_i(1), \dots, \theta_i(T)) = \frac{\phi_1(\underline{\theta}_i, \underline{\alpha}_i)}{\sqrt{(2\pi)^T \sigma_i^T}} \exp\left(-\frac{1}{2\sigma_i^2} \phi_2(\underline{\theta}_i, \underline{\alpha}_i)\right)$$

où  $\phi_1(\underline{\theta}_i, \underline{\alpha}_i)$  est une fonction strictement positive et de classe  $\mathcal{C}^1$  sur  $\mathbb{R}^T \times \mathbb{R}_+^n$  et  $\phi_2(\underline{\theta}_i, \underline{\alpha}_i)$  une forme quadratique strictement positive.

**Corollaire 12** *Sous les hypothèses du modèle de Rasch, la moyenne conditionnelle  $\mathbf{Q}(\eta/\eta^{(p)})$  de la log-vraisemblance complète sachant la valeur courante de l'estimateur  $\eta^{(p)}$  est donnée par*

$$\begin{aligned} \mathbf{Q}(\eta/\eta^{(p)}) &= \sum_{i=1}^n \int \dots \int_{\mathbf{R}^T} \left\{ -\frac{T}{2} \log(2\pi) - \frac{T}{2} \log(\sigma_i^2) + \log(\phi_1(\underline{\theta}_i, \underline{\alpha}_i)) - \frac{1}{2\sigma_i^2} \phi_2(\underline{\theta}_i, \underline{\alpha}_i) \right\} \\ &\quad \times q_i(\underline{\theta}_i/\underline{\mathbf{x}}_i, \eta^{(p)}) d\theta_i(1) \dots d\theta_i(T) \\ &+ \sum_{i=1}^n \int \dots \int_{\mathbf{R}^T} \left\{ \sum_{t=1}^T R_i(t) \theta_i(t) - \sum_{k=1}^q \beta_k S_{i,k} - \sum_{t=1}^T \sum_{k=1}^q \log(1 + e^{\theta_i(t) - \beta_k}) \right\} \\ &\quad \times q_i(\underline{\theta}_i/\underline{\mathbf{x}}_i, \eta^{(p)}) d\theta_i(1) \dots d\theta_i(T) \end{aligned} \quad (5.2)$$

où

$$\begin{aligned} q_i(\underline{\theta}_i/\underline{\mathbf{x}}_i, \eta^{(p)}) &\propto \frac{\phi_1(\underline{\theta}_i, \underline{\alpha}_i^{(p)})}{\sqrt{(2\pi)^T (\sigma_i^{(p)})^T}} \\ &\quad \times \frac{\exp\left\{ \sum_{t=1}^T \theta_i(t) \cdot R_i(t) - \sum_{k=1}^q \beta_k^{(p)} S_{i,k} - \frac{1}{2(\sigma_i^{(p)})^2} \phi_2(\underline{\theta}_i, \underline{\alpha}_i^{(p)}) \right\}}{\prod_{t=1}^T \prod_{k=1}^q [1 + \exp(\theta_i(t) - \beta_k^{(p)})]} \end{aligned}$$

La maximisation par rapport aux  $\sigma_i^2$  concerne uniquement le deuxième et le quatrième terme de la première somme de cette expression et la maximisation par rapport aux  $\alpha_i$  porte sur le troisième et quatrième terme de cette même expression. Ceci nous amène à résoudre le système d'équations suivant, pour  $i = 1, \dots, n$  :

$$(\sigma_i^{(p+1)})^2 = \frac{1}{T} \int \dots \int \phi_2(\underline{\theta}_i, \underline{\alpha}_i) q_i(\underline{\theta}_i / \underline{\mathbf{x}}, \eta^{(p)}) d\underline{\theta}_i,$$

et

$$\int \dots \int \{2(\sigma_i^{(p+1)})^2 \frac{\partial \phi_1(\underline{\theta}_i, \underline{\alpha}_i)}{\partial \alpha_i} - \phi_1(\underline{\theta}_i, \underline{\alpha}_i) \frac{\partial \phi_2(\underline{\theta}_i, \underline{\alpha}_i)}{\partial \alpha_i}\} q_i(\underline{\theta}_i / \underline{\mathbf{x}}, \eta^{(p)}) d\underline{\theta}_i = 0$$

La maximisation par rapport aux  $\beta_k$  concerne uniquement la double somme sur  $i$  et  $t$  du dernier terme de cette expression. Ceci conduit à résoudre les  $q$  équations suivantes, pour  $k = 1, \dots, q$  :

$$\sum_{i=1}^n \sum_{t=1}^T x_{i,k} = \sum_{i=1}^n \sum_{t=1}^T \int \dots \int [1 + \exp((\beta_k^{(p+1)} - \theta_i(t)))]^{-1} q_i(\underline{\theta}_i / \underline{\mathbf{x}}, \eta^{(p)}) d\underline{\theta}_i$$

Dans les deux cas, les intégrales sont de nouveau approchées numériquement en utilisant des quadratures de Gauss-Hermite.

### 5.3 Processus de Markov Rasch Latent

Dans ce cas le paramètre  $\eta = (\beta, \sigma^2)$  avec  $\beta = (\beta_1, \dots, \beta_q)'$  le vecteur des paramètres de difficulté et  $\sigma^2 = (\sigma_1^2, \dots, \sigma_n^2)'$ , avec  $\sigma_i^2$  la variance du processus de Markov  $\Theta_i$ .

L'espérance conditionnelle de la log-vraisemblance complète sachant la valeur courante de l'estimateur  $\eta^{(p)}$  s'écrit dans ce cas

**Corollaire 13** *Sous les hypothèses du modèle de Rasch, la moyenne conditionnelle  $\mathbf{Q}(\eta/\eta^{(p)})$  de la log-vraisemblance complète sachant la valeur courante de l'estimateur  $\eta^{(p)}$  est donnée par*

$$\mathbf{Q}(\eta/\eta^{(p)}) = -\frac{T}{2} \sum_{i=1}^n \log(2\pi\sigma_i^2) + H_1 + H_2 \quad (5.3)$$

où

$$H_1 = -\sum_{i=1}^n \frac{1}{2\sigma_i^2} \int \dots \int [\theta_i^2(1) + \sum_{t=2}^T (\theta_i(t) - \theta_i(t-1))^2] q_i(\underline{\theta}_i/\underline{\mathbf{x}}, \beta^{(p)}, \sigma^{2(p)}) d\underline{\theta}_i,$$

$$H_2 = \sum_{i=1}^n \sum_{t=1}^T \sum_{k=1}^q \int \dots \int \log \left\{ \frac{\exp(x_{i,k}(t)(\theta_i(t) - \beta_k))}{1 + \exp((\theta_i(t) - \beta_k))} \right\} q_i(\underline{\theta}_i/\underline{\mathbf{x}}, \eta^{(p)}) d\theta_i(1) \dots d\theta_i(T)$$

et

$$q_i(\underline{\theta}_i/\underline{\mathbf{x}}, \eta^{(p)}) \propto \frac{\exp\{\sum_{t=1}^T \theta_i(t) \cdot R_i(t) - \sum_{k=1}^q \beta_k \cdot S_i(k) - \frac{1}{2\sigma_i^2} [\theta_i^2(1) + \sum_{t=2}^T (\theta_i(t) - \theta_i(t-1))^2]\}}{\prod_{t=1}^T \prod_{k=1}^q [1 + \exp(\theta_i(t) - \beta_k)]}.$$

La maximisation par rapport aux  $\sigma_i^2$  concerne uniquement les deux premiers termes de cette dernière expression et la maximisation par rapport aux  $\beta_k$  fait appel uniquement à la double somme sur  $i$  et  $t$  du dernier terme de cette même expression. Ceci nous amène à calculer

$$\sigma_i^{2(p+1)} = \frac{1}{T} \int \dots \int [\theta_i^2(1) + \sum_{t=2}^T (\theta_i(t) - \theta_i(t-1))^2] \pi_i(\underline{\theta}_i/\underline{\mathbf{x}}, \beta^{(p)}, \sigma^{2(p)}) d\theta_i(1) \dots d\theta_i(T)$$

pour tout  $i$  et à résoudre les  $q$  équations suivantes : pour tout  $k$  tel que  $1 \leq k \leq q$ ,

$$\sum_{i=1}^n \sum_{t=1}^T x_{i,k}(t) = \sum_{i=1}^n \sum_{t=1}^T \int \dots \int [1 + \exp((\beta_k - \theta_i(t)))]^{-1} \pi_i(\underline{\theta}_i/\underline{\mathbf{x}}, \eta^{(p)}) d\underline{\theta}_i.$$

Et les intégrales sont de nouveau approchées numériquement en utilisant l'algorithme de Raphson-Newton et les approximations par des quadratures de Gauss-Hermite.



## 5.4 Processus Rasch Latent $\mathcal{AR}(1)$

Dans ce cas, on rappelle que le paramètre est  $\eta = (\beta, \rho, \sigma^2)$  with  $\beta = (\beta_1, \dots, \beta_q)$ ,  $\rho = (\rho_1, \dots, \rho_n)$  and  $\sigma^2 = (\sigma_1^2, \dots, \sigma_n^2)$ .

**Corollaire 14** *Sous les hypothèses du modèle de Rasch, la moyenne conditionnelle  $\mathbf{Q}(\eta/\eta^{(p)})$  de la log-vraisemblance complète sachant la valeur courante de l'estimateur  $\eta^{(p)}$  est donnée par*

$$\mathbf{Q}(\theta/\eta^{(p)}) = -\frac{T}{2} \sum_{i=1}^n \log(2\pi\sigma_i^2) + \frac{1}{2} \sum_{i=1}^n \log(1 - \rho_i^2) + G_1 + G_2 \quad (5.4)$$

où

$$G_1 = -\sum_{i=1}^n \frac{1}{2\sigma_i^2} \int \dots \int \varphi(\underline{\theta}_i, \underline{\rho}_i) \mathbf{q}_i(\underline{\theta}_i/\underline{\mathbf{x}}_i, \eta^{(p)}) d\theta_i(1) \dots d\theta_i(T)$$

$$G_2 = +\sum_{i=1}^n \sum_{t=1}^T \sum_{k=1}^q \int \dots \int \psi(\underline{\mathbf{x}}_i, \underline{\theta}_i) \mathbf{q}_i(\underline{\theta}_i/\underline{\mathbf{x}}_i, \eta^{(p)}) d\theta_i(1) \dots d\theta_i(T)$$

avec

$$\varphi(\underline{\theta}_i, \underline{\rho}_i) = (1 - \rho_i^2) \theta_1^2 + \sum_{t=2}^T (\theta_i(t) - \rho_i \theta_i(t-1))^2,$$

$$\psi(\underline{\mathbf{x}}_i, \underline{\theta}_i) = \ln \frac{\exp(x_{i,k}(t)(\theta_i(t) - \beta_k))}{1 + \exp((\theta_i(t) - \beta_k))} \text{ et } q_i(\underline{\theta}_i/\underline{\mathbf{x}}_i, \eta^{(p)}) \propto C$$

et où

$$C = \frac{\exp\{\sum_{t=1}^T \theta_i(t) \cdot r_i(t) - \sum_{k=1}^q \beta_k^{(p)} \cdot r_i(k) - \frac{1}{2\sigma_i^{2(p)}} [(1 - \rho_i^{2(p)}) \theta_i^2(1) + \sum_{t=2}^T (\theta_i(t) - \rho_i^{(p)} \theta_i(t-1))^2]\}}{\prod_{t=1}^T \prod_{k=1}^q [1 + \exp(\theta_i(t) - \beta_k^{(p)})]}.$$

### Démonstration

L'espérance conditionnelle de la log-vraisemblance complète prend la forme (1.32)

avec  $\mathbf{q}_i(\underline{\theta}_i/\underline{\mathbf{x}}_i, \eta^{(p)}) = q_i(\underline{\theta}_i/\underline{\mathbf{x}}_i, \eta^{(p)})$ , et nous voyons que  $\mathbf{q}_i(\underline{\theta}_i/\underline{\mathbf{x}}_i, \eta^{(p)}) \propto C$ . ■

La maximisation par rapport aux  $\sigma_i^2$  concerne uniquement le premier et le troisième terme de cette dernière expression et la maximisation par rapport aux  $\rho_i$  porte sur le second et le troisième terme de cette même expression. Ceci nous amène à résoudre le système d'équations suivant, pour  $i = 1, \dots, n$  :

$$\sigma_i^{2(p+1)} = \frac{1}{T} \sum_{i=1}^n \int \dots \int [(1 - \rho_i^2) \theta_i^2(1) + \sum_{t=2}^T (\theta_i(t) - \rho_i \theta_i(t-1))^2] \cdot \pi(\underline{\theta}_i/\underline{\mathbf{x}}_i, \eta^{(p)}) d\underline{\theta}_i,$$

et

$$\frac{\rho_i}{1 - \rho_i^2} = \frac{1}{\sigma^2} \int \dots \int [\rho_i \theta_i^2(1) + \sum_{t=2}^T \theta_i(t) - 1(\theta_i(t) - \rho_i \theta_i(t-1))] \pi(\underline{\theta}_i/\underline{\mathbf{x}}_i, \beta^{(p)}, \rho^{(p)}, \sigma^{2(p)}) d\underline{\theta}_i$$

La maximisation par rapport aux  $\beta_k$  concerne uniquement la double somme sur  $i$  et  $t$  du dernier terme de cette expression. Ceci conduit à résoudre les  $q$  équations

suivantes, pour  $k = 1, \dots, q$  :

$$\sum_{i=1}^n \sum_{t=1}^T x_{i,k} = \sum_{i=1}^n \sum_{t=1}^T \int \dots \int [1 + \exp((\beta_k - \theta_i(t)))]^{-1} \pi_{\mathbf{i}}(\underline{(\theta_i)}/\underline{\mathbf{x}}, \underline{\eta}^{(p)}) d\underline{\theta}_i.$$

Dans les deux cas, les intégrales sont de nouveau approchées numériquement en utilisant des quadratures de Gauss-Hermite.

## 5.5 Annexe 1

### L'approche des quadratures de Gauss-Hermite

Cette approche consiste à approximer les intégrales simples de la forme  $\int_{-\infty}^{\infty} f(x)e^{-\frac{x^2}{2}} dx$  par une somme finie  $\sum_{i=1}^m f(t_i)v_i$ , où  $x_i$  et  $w_i$  sont les noeuds et les poids de la quadrature de Gauss que l'on peut retrouver dans Abramowitz et Stegun (1974), ou encore peuvent être calculés dans le package statmod du logiciel  $\mathcal{R}$ . Lorsque l'intégrale est une intégrale multiple de la forme

$$\varphi(\underline{\mu}, \Lambda) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f(x_1, \dots, x_T) e^{[-\frac{1}{2}(\underline{x}-\underline{\mu})'\Lambda^{-1}(\underline{x}-\underline{\mu})]} dx_1 \dots dx_T$$

où  $\Lambda$  est la matrice de variance covariance supposée régulière,  $\underline{\mu}$  le vecteur moyen de cette loi et  $\underline{x} = (x_1, \dots, x_T)$  une variable dans  $\mathbb{R}^T$ , on se ramène au cas précédent, en écrivant cette intégrale comme le produit de  $T$  intégrales simples généralisées. A cette fin, on envisage deux cas :

1. La matrice de covariance  $\Lambda$  est diagonale.

Alors moyennant une transformant de la variable en une variable centrée réduite, on applique l'approche décrite ci-dessus à chacune des intégrales du produit par une somme, ce qui permet d'approcher l'intégrale multiple par la somme

$$\sum_{i_1=1}^{m_1} \dots \sum_{i_T=1}^{m_T} f(x_{i_1}, \dots, x_{i_T}) v_{i_1} \dots v_{i_T},$$

où  $x_{i_k}$  et  $w_{i_k} : 1 \leq k \leq T$  sont les noeuds et les poids de la quadrature de Gauss.

2. La matrice de covariance  $\Lambda$  est non diagonale.

On se ramène au cas précédent moyennant la transformation standard  $\underline{y} = \mathbf{C}\underline{z}$  où  $\underline{z} = (\frac{x_1-\mu_1}{\sigma_1^2}, \dots, \frac{x_T-\mu_T}{\sigma_T^2})'$  est un vecteur normal centré réduit, et  $\mathbf{C}$  est la matrice triangulaire inférieure de Choleski telle que  $\Lambda = \mathbf{C}\mathbf{C}'$ . Dans ce cas, l'intégrale devient avec des notations évidentes

$$\varphi(\underline{z}, \underline{\mu}, \mathbf{C}) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \psi(z_1, \dots, z_T) e^{[-\frac{1}{2} \sum_{k=1}^T z_k^2]} dz_1 \dots dz_T$$

et d'où l'on déduit l'approximation décrite ci-dessus.

# Bibliographie

- [ADM06] O. Admane et M. Mesbah 2006. Estimation des paramètres du modèle Rasch dichotomique. Annales de L'ISUP. Vol 1-Fas 1-2 p,3-30. (2006)
- [BAR06] F. Bartolucci 2006. Likelihood inference for a class of latent Markov models under linear hypotheses on the transition probabilities. Journal of the royal Statistical Society, Series B, vol. 68, p, 155-178.
- [BOU04] M. Boussebouda (2004). Eléments de la théorie des probabilités. Tome 2. Ed. OPU
- [BRO87] Brockwell and Davis (1987). Times Series. Theory and Method. Springer Verlag.
- [COX62] D.R. Cox and (1972). The statistical analysis of series Events. Chapman and Hall London
- [DEM77] A. P. Dempster, N. M. Laird, D. S. Rubin 1977. Maximum likelihood from incomplete data via the EM algorithm. Journal of royal statistical Society, Ser. B.,39, 1-38.
- [FED05] M. L. Feddag, et M. Mesbah 2005. Generalized estimating equations for longitudinal mixed Rasch model. Journal of Statistical Planning and Inference, vol. 129, p, 159-179. (2005)
- [FED02] M. L. Feddag, M. GRAMA et M. Mesbah 2002. Estimation des paramètres d'un modèles de Rasch mixte par la méthode GEE. Application en qualité de vie. Journal de la société Française de statistique. 143, 169-175.

- [FED09] M. L. Feddag 2009. Pairwise likelihood for the Longitudinal mixed Rasch model. *Computational Statistics and Data Analysis*, 53, 1027 - 1037.
- [FEL..] W. Feller (19..). An introduction to probability theory and its applications. Tome 1 et 2. Ed. Wiley
- [FIS95] Fisher, G.H. and Molenaar, I.W. Rasch models, Foundations, Recent Developments and Applications. Springer-Verlag, New York. (1995)
- [GHO98] M. Ghosh (1995). Inconsistent maximum likelihood estimators for the Rasch Model. *Statistics and probability letters.*, 23,165-170.
- [HMN00] A. Hanon, 2000. Modèle de Rasch et validation de questionnaire de Qualité de Vie. thse de doctorat. Université de Bretagne Sud, Vannes, France.
- [LAN94] Langeheine, R. and Van de Pol, F. Discrete-time mixed Markov latent class models. In : Dale, A., Davies, R.B. (eds) *Analysing social and political change. A casebook of methods.* London :Sage. p, 170-197. (1994)
- [LIA86] Liang. K. Y. et Zeger S. L. 1986. Longitudinal data analysis using generalized linear models. *Biometrika.* 73, 13-22.
- [LOR93] F. M LORD (1983). Unbiased estimators of ability parameters, of their variance and of their parallel forms reliability. *Psychometrika*, 48(2), 233-245.
- [KLA90] K. C KLAUER (1990). Asymptotic properties of the ML estimator of ability parameter with items parameters are known. *Methodika*, 4, 23-36.
- [MES09] Mesbah, M. Statistical Quality of Life. In : Balakrishnan, N., Campbell, B.R., Vidakovic, B., Kotz, S., Johnson, N.L. (eds) *Methods and Applications of Statistics in the Life and Health Sciences.* Chapter 73. Wiley. New York. (2009)
- [PLA93] J. Planzagl (1993). On the consistency of the conditional maximum likelihood estimators. *Ann. Ins.Statist. Math.* 45, 703-719.
- [POL86] Van de Pol, F., de Leeuw, J. A latent Markov model to correct for measurement error. *Sociological Methods and Research*, 15, 118-141. (1986)

- [RAS60] G. Rasch 1960. Probabilistic Models for some intelligence and attainment test. Cicago, The University of Chicago Press.
- [VER86] Vermunt, J.R., Langeheine, R., Böckenholt, U. Discrete-time discrete-state latent Markov models with time-constant and time-varying covariates. *Journal of Educational and Behavioral Statistics*, 24, pp. 179-207. (1999)
- [WED74] R. W. M Weddeburn 1974. Quasi likelihood function, Generalized linear models and the Gauss-Newton method. *Biometrika* 48, 439-447.
- [WIG73] Wiggins, L.M. Panel analysis. Amsterdam ; Elsevier. (1973)