



HAL
open science

Construction d'échelles d'items unidimensionnelles en qualité de vie

Jean-Benoit Hardouin

► **To cite this version:**

Jean-Benoit Hardouin. Construction d'échelles d'items unidimensionnelles en qualité de vie. Sciences du Vivant [q-bio]. Université René Descartes - Paris V, 2005. Français. NNT : . tel-00011754

HAL Id: tel-00011754

<https://theses.hal.science/tel-00011754>

Submitted on 4 Mar 2006

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

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

UNIVERSITÉ RENÉ DESCARTES - PARIS V
Centre Universitaire des Saints Pères
UFR Biomédicale

Thèse en vue de l'obtention du grade de DOCTEUR
de l'Université RENÉ DESCARTES - PARIS V

Discipline : Santé Publique

Spécialité : Biostatistique

CONSTRUCTION D'ÉCHELLES D'ITEMS
UNIDIMENSIONNELLES EN QUALITÉ DE VIE

Jean-Benoit Hardouin

Soutenance le 14 novembre 2005 devant le jury composé de :

M.	Jean-Louis AUGET	Rapporteur - Professeur Université de Nantes
M.	Michel CHAVANCE	Examineur - Directeur de recherche INSERM Villejuif
M.	Paul DEHEUELS	Examineur - Professeur Université Paris VI
Mme	Catherine HUBER	Présidente du jury - Professeur Université Paris V
M.	Mounir MESBAH	Directeur de thèse - Professeur Université Paris VI
M.	El Mostafa QANNARI	Rapporteur - Professeur ENITIAA, Nantes

*A Caroline,
et à mes trois enfants, Orane, Théophile et Philémon*

Table des matières

Table des figures	ix
Table des tableaux	xii
Remerciements	xiii
Introduction	1
Plan de cette thèse	3
Glossaire	5
I L'analyse statistique de la qualité de vie	7
1 Notions et notations	9
1.1 Notions	9
1.1.1 Trait latent	9
1.1.2 Questionnaires, profils, tests, items	9
1.1.3 Réponses positives/correctes et négatives/incorrectes	9
1.1.4 Mesures indirectes, scores	10
1.1.5 Dimensions, traits latents, échelles, sous-échelles	10
1.1.6 Psychométrie, théorie classique, théorie de réponse aux items	10
1.2 Notations	10
1.2.1 Notations mathématiques générales	10
1.2.2 Notations spécifiques	11
1.3 Les modèles classiques	12
1.3.1 Principes	12
1.3.2 Définition	12
1.3.3 La fiabilité	12
1.3.4 Le coefficient alpha de Cronbach	12
1.4 Synthèse	13
2 La Théorie de Réponse aux items (Item Response Theory - IRT)	15
2.1 Définition générale	15
2.2 Les fonctions de réponse aux items (Items Response Functions - IRF) et les courbes caractéristiques des items (Items Characteristic Curves - ICC)	15

2.3	Hypothèses fondamentales	15
2.3.1	L'unidimensionnalité	15
2.3.2	La monotonie	16
2.3.3	L'indépendance locale	16
2.4	Les traces des items	16
2.4.1	Principes généraux : les traces classiques	16
2.4.2	Traces particulières : les traces cumulées	16
2.4.3	Détection des mauvais items à l'aide des traces	17
2.4.4	Adaptation des traces : les traces en fonction du sous-score	17
2.5	IRT paramétrique et IRT non-paramétrique	17
2.5.1	L'IRT non paramétrique	17
2.5.2	L'IRT paramétrique	17
2.6	Le modèle de Rasch (RM ou 1-PLM)	18
2.6.1	Modélisation	18
2.6.2	Courbes caractéristiques des items dans le modèle de Rasch	18
2.6.3	Interprétation des paramètres de difficulté des items	18
2.6.4	Paramètres fixes et paramètres aléatoires	19
2.6.5	Représentation graphique	19
2.6.6	Le modèle de Rasch comme un modèle linéaire généralisé (mixte)	19
2.6.7	Propriétés spécifiques	21
2.6.8	L'estimation des paramètres	23
2.6.9	Tester le modèle de Rasch	31
2.7	Les autres modèles de l'IRT pour items dichotomiques	38
2.7.1	Le modèle logistique à un paramètre (One Parameter Logistic Model) (OPLM)	38
2.7.2	Courbes caractéristiques des items dans l'OPLM	40
2.7.3	Le modèle de Birnbaum (2-PLM)	40
2.7.4	Autres modèles	41
2.8	Les modèles de l'IRT pour items polytomiques	42
2.8.1	L'extension au cadre polytomique du modèle de Rasch	42
2.8.2	Le modèle d'échelle de classement (Rating Scale Model) (RSM)	42
2.8.3	Le modèle du crédit partiel (Partial Credit Model) (PCM)	43
2.8.4	Lien entre le RSM et le PCM	43
2.8.5	Estimation du PCM et du RSM	44
2.9	Synthèse	44
3	L'IRT multidimensionnelle	45
3.1	Définition	45
3.2	Extension des hypothèses fondamentales de l'IRT	45
3.2.1	La dimension	45
3.2.2	La monotonie	45
3.2.3	L'indépendance locale	46
3.3	Les structures simples (SS) et approximativement simple (ASS)	46
3.4	Les extensions multidimensionnelles du modèle de Rasch	46
3.4.1	Le modèle multidimensionnel de Rasch	46
3.4.2	Le modèle à trait latent à plusieurs composantes (multicomponent latent trait model)	47

3.4.3	Le modèle multidimensionnel polytomique à trait latent (multidimensional polytomous latent trait model) (MPLT)	47
3.4.4	Le modèle logistique multinomial à coefficients aléatoires multidimensionnel (Multidimensional Random Coefficients Multinomial Logit Model) (MRCML)	60
3.5	Autres modèles de l'IRT multidimensionnelle	60
3.5.1	Les extensions des modèles unidimensionnels	60
3.5.2	Les extensions longitudinales	61
3.6	Synthèse	61
4	Une proposition : le modèle (polytomique) multidimensionnel de Rasch marginalement exhaustif (M(P)MSRM)	63
4.1	Un nouveau modèle	63
4.2	L'exhaustivité marginale	64
4.3	Modèles marginaux	64
4.4	Définition	64
4.5	Proposition	64
4.6	Représentation graphique	65
4.7	Estimation	66
4.8	Interprétation des paramètres de difficulté des items	67
4.9	Nécessité d'une structure simple	67
4.10	Extension au cadre polytomique	69
4.11	Application sur trois échelles du Sickness Impact Profil (SIP)	70
4.12	Synthèse	72
II	La sélection d'échelles unidimensionnelles d'items	73
5	Les méthodes classiques de sélection d'items	75
5.1	But de ce chapitre	75
5.2	La procédure pas à pas du alpha de Cronbach	75
5.3	L'analyse factorielle exploratoire	76
5.3.1	Principes de l'analyse en facteurs communs et spécifiques	76
5.3.2	L'analyse en composantes principales (ACP)	78
5.3.3	Le cas des variables qualitatives	78
5.3.4	L'analyse en facteurs d'information complète (full information item factor analysis)	80
5.3.5	La procédure Varclus	80
5.3.6	La procédure de classification autour de variables latentes - CLV	81
5.3.7	L'analyse parallèle modifiée révisée (Revised Modified Parallel Analysis)	82
5.4	Synthèse	83
6	Les procédures existantes en Théorie de Réponse aux Items non paramétrique	85
6.1	La procédure HCA/CCPROX	85
6.1.1	Définition générale	85
6.1.2	Les mesures de proximité entre les items	85

6.1.3	Les mesures de proximité entre les ensembles d'items	87
6.1.4	Résultats	88
6.2	La Mokken Scale Procedure (MSP)	89
6.2.1	Le Modèle Monotonement Homogène de Mokken (Mokken Monotone Homogeneous Model) (MMHM)	89
6.2.2	Les erreurs de Guttman	89
6.2.3	Les coefficients d'échelles de Loevinger	90
6.2.4	Lien entre les coefficients d'échelles de Loevinger et le MMHM	93
6.2.5	Algorithme	93
6.2.6	Résultats	94
6.3	La procédure DETECT	95
6.3.1	L'indice DETECT et les indices Iss et R	95
6.3.2	Estimation pratique des indices	96
6.3.3	Algorithme de la procédure DETECT	96
6.3.4	Résultats	97
6.4	Comparaison de HCA/CCPROX, MSP et DETECT	98
6.4.1	Étude de simulations	98
6.4.2	Résultats	98
6.5	Synthèse	99
7	Propositions : La classification d'items en sous-échelles basée sur l'adéquation à un modèle	101
7.1	Introduction	101
7.2	La procédure "BackRasch", procédure pas à pas descendante basée sur le modèle de Rasch	102
7.2.1	Principes	102
7.2.2	Algorithme	102
7.3	La procédure "Raschfit"	102
7.3.1	Principes	102
7.3.2	Représentation graphique des deux modèles testés à chaque étape	104
7.3.3	Algorithme	104
7.3.4	Améliorations sur le temps de calcul : la procédure Raschfit-Fast	107
7.3.5	Extensions à d'autres modèles	110
7.4	Étude de simulations	110
7.4.1	Les paramètres de simulations	111
7.4.2	Résultats	113
7.5	Applications	120
7.5.1	Application sur le Profil d'impact des maladies (Sickness Impact Profil) (SIP)	120
7.5.2	Application dans un autre champ : prédire le risque de récurrence de la tentative suicidaire par un score simple	124
7.6	Synthèse	129

III	La Théorie de Réponse aux Items sous SAS et Stata	131
8	Analyse de l'existant et nouveaux macros-programmes pour l'IRT	133
8.1	L'IRT et les logiciels généralistes	133
8.2	SAS	134
8.2.1	Version de base	134
8.2.2	Ajouts	134
8.3	Stata	134
8.3.1	Version de base	134
8.3.2	Ajouts	135
8.4	Théorie classique et estimations des modèles de l'IRT sous SAS	135
8.4.1	Le macro-programme %AnaQol	135
8.5	Théorie classique et estimations des modèles de l'IRT sous Stata	137
8.5.1	Le programme traces	137
8.5.2	Le programme raschtest	137
8.5.3	Le programme mmsrm	140
8.5.4	Le programme geekel2d	141
8.6	Les procédures de sélection d'échelles unidimensionnelles sous SAS	142
8.6.1	Le macro-programme %BackRasch	142
8.6.2	Les macro-programmes %LoevH et %MSP	142
8.6.3	Le macro-programme %DETECT	143
8.7	Les procédures de sélection d'échelles unidimensionnelles sous Stata	144
8.7.1	Le programme raschfit	144
8.7.2	Le programme backrasch	145
8.7.3	Les programmes loevh et mSP	146
8.7.4	Le programme hcaccprox	147
8.7.5	Le programme detect	147
8.7.6	Le programme simirt	148
8.8	Le site web AnaQol et le projet FreeIRT	149
8.8.1	AnaQol	149
8.8.2	Le projet FreeIRT	149
	Conclusion et perspectives	151
	Bibliographie	152
IV	Annexes	163
A	Comparaison du M2-PLM et du 5-PAM	165
B	Résultats des simulations	169
B.1	La procédure Raschfit	170
B.2	La procédure Raschfit-Fast	171
B.3	La procédure BackRasch	172
B.4	La procédure MSP (c=0.3)	173
B.5	La procédure MSP (c=0.2)	174
B.6	La procédure HCA/CCPROX	175

B.7	La procédure AFCS-Varimax	176
B.8	La procédure ACP-Varimax	177
C	Sorties des programmes	179
C.1	Les données	179
C.2	Le macro-programme SAS %AnaQol	179
C.3	Le programme Stata traces	181
C.4	Le programme Stata raschtest	182
C.5	Le programme Stata mmsrm	187
C.6	Le programme Stata geekel2d	188
C.7	Le macro-programme SAS %BackRasch	189
C.8	Les macro-programmes SAS %LoevH et %MSP	192
C.9	Le macro-programme SAS %DETECT	192
C.10	Le programme Stata raschfit	194
C.11	Le programme Stata backrasch	195
C.12	Les programmes Stata loevh et misp	197
C.13	Le programme Stata hcaccprox	200
C.14	Le programme Stata detect	200

Table des figures

2.1	Courbes caractéristiques de 7 items vérifiant un modèle de Rasch	18
2.2	Représentation graphique du modèle de Rasch	19
2.3	Représentation graphique de la propriété d'exhaustivité du score sur le trait latent dans le cadre du modèle de Rasch	22
2.4	Courbes caractéristiques de 7 items vérifiant un OPLM	40
2.5	Courbes caractéristiques de 7 items vérifiant un modèle de Birnbaum	41
2.6	Courbes caractéristiques d'un item à 4 modalités suivant un RSM avec les paramètres $\delta_j = -2.0$, $\tau_2 = 1.0$ et $\tau_3 = 1.5$	43
3.1	Représentation graphique du Multidimensional Polytomous Latent Trait Model	48
3.2	Représentation graphique de la propriété d'exhaustivité du vecteur des sous-scores sur le trait latent multidimensionnel dans le cadre du MPLT	49
4.1	Représentation graphique du modèle multidimensionnel marginalement exhaustif de Rasch	65
4.2	Représentation graphique de la propriété d'exhaustivité marginale dans le cadre du MMSRM	66
7.1	Modèle unidimensionnel avec trois items déjà sélectionnés et un nouvel item	104
7.2	Modèle bi-dimensionnel avec trois items déjà sélectionnés et un nouvel item	104
7.3	Modèle "bi-dimensionnel" avec trois items déjà sélectionnés et un nouvel item dans le cadre de la version rapide de la procédure "Raschfit"	108
7.4	Dendrogramme issue de la classification des items en fonction du nombre de classements dans un même groupe par les 11 procédures	122
7.5	Représentation graphique du MMSRM obtenu avec les 19 items	127
7.6	Représentation graphique du MMSRM avec les deux principaux traits latents et la variable "risque de récidence"	127
C.1	Procédure pas à pas du alpha de Cronbach sur les items eb1 à eb9	180
C.2	Trace de l'item c1	182
C.3	Trace de l'item c4	182
C.4	Traces logistiques des items c1 à c9	183
C.5	Courbes caractéristiques des items c1 et c4 sous le modèle de Rasch	186
C.6	Graphique de l'information des items c1 à c9 sous le modèle de Rasch	186

Liste des tableaux

4.1	Énoncés des items des trois sous-échelles du SIP choisies	71
4.2	Estimations des paramètres du MMSRM pour les trois sous-échelles du SIP . .	72
5.1	Tableau de contingence entre les items dichotomiques X_1 et X_2	79
6.1	Tableau de contingence entre les items dichotomiques X_j et X_k	85
6.2	Tableau de contingence entre les items dichotomiques X_j et X_k pour les individus ayant un score $S = s$	86
6.3	Tableau de contingence entre les items dichotomiques X_j et X_k	89
6.4	Tableau de contingence théorique entre les items dichotomiques X_j et X_k pour maximiser la covariance	91
7.1	Exemple de résultats types, avec deux ensembles de 7 items simulés reliés à deux traits latents distincts	114
7.2	Exemple de résultats type, avec deux ensembles de 7 items simulés reliés au même trait latent	115
7.3	Résultats agrégés des simulations avec le M2-PLM et une structure simple (Taux de résultats de type 1 ou 2/Taux de résultats de type 4 (Taux de résultats de type 5))	116
7.4	Résultats agrégés des simulations avec le M2-PLM et une structure approximativement simple (Taux de résultats de type 1 ou 2/Taux de résultats de type 4 (Taux de résultats de type 5))	117
7.5	Résultats agrégés des simulations avec le M5-PAM (Taux de résultats de type 1 ou 2/Taux de résultats de type 4 (Taux de résultats de type 5))	117
7.6	Temps de calculs moyens, minimum et maximum en secondes et écart-type pour chaque procédure sur les simulations en fonction du nombre d'items dans chaque dimension	118
7.7	Matrice des coefficients kappa de concordance entre les méthodes	119
7.8	Caractéristiques de chacune des méthodes de sélection d'items	121
7.9	Partition synthétique des 28 items des sous-échelles "communication", "mobilité" et "comportement émotionnelle"	123
7.10	Variables regroupées avec la procédure Raschfit	126
7.11	Indicateurs de validité des scores proposés	129
8.1	Macro-programmes de Christensen et Bjorner	134
8.2	Options du macro-programmes SAS % AnaQol	136
8.3	Options du programmes Stata traces	138
8.4	Options du programme Stata raschtest	139

8.5	Options du programme Stata mmsrm	141
8.6	Options du programme Stata geekel2d	141
8.7	Options du macro-programme SAS %BackRasch	143
8.8	Options du macro-programme SAS %MSP	143
8.9	Options du macro-programme SAS %DETECT	144
8.10	Options du programme Stata raschfit	145
8.11	Options du programme Stata backrasch	145
8.12	Options du programme Stata loevh	146
8.13	Options du programme Stata mmp	147
8.14	Options du programme Stata hcaccprox	148
8.15	Options du programme Stata detect	148
8.16	Options du programme Stata simirt	149
B.1	Résultat des 360 simulations avec la procédure Raschfit	170
B.2	Résultat des 360 simulations avec la procédure Raschfit-Fast	171
B.3	Résultat des 360 simulations avec la procédure BackRasch	172
B.4	Résultat des 360 simulations avec la procédure MSP (c=0.3)	173
B.5	Résultat des 360 simulations avec la procédure MSP (c=0.2)	174
B.6	Résultat des 360 simulations avec la procédure HCA/CCPROX	175
B.7	Résultat des 360 simulations avec la procédure AFCS-Varimax	176
B.8	Résultat des 360 simulations avec la procédure ACP-Varimax	177

Remerciements

Mes remerciements vont bien entendu à ceux qui m'ont aidé et soutenu durant l'ensemble de cette thèse, à mes deux rapporteurs, les Professeurs Jean-Louis Auget et El Mostafa Qanari, qui ont accepté de lire précautionneusement ce document, et aux autres membres du jury, les Professeurs Catherine Huber, Michel Chavance et Paul Deheuvels.

Je remercie particulièrement le Professeur Mounir Mesbah qui m'a encadré pour ce travail pour son implication et ses remarques.

Ensuite je veux remercier Caroline qui m'a supporté durant cette période, et mes trois enfants, Orane, Théophile et Philémon qui ont été globalement sages quand j'en avais besoin. A eux quatre, je souhaiterais dédier ce travail.

Je remercie aussi toutes les personnes qui, directement ou indirectement m'ont soutenu, parfois sans doute inconsciemment : Mohand-Larbi, Laurent, Maina, Myren, Maxime, Mathilde, Xavier, Cécile, Jean-Michel, Jean-François, Antoine... Un merci particulier à Véronique Sébille et Alessio Fusco en ce sens, qui ont montré un intérêt particulier pour mon travail.

Un merci aussi tout particulier à mes parents qui ont relu ce document pour la chasse à la faute et à Mohand-Larbi qui m'a fait l'honneur de ses commentaires.

Plus prosaïquement, merci à l'Observatoire Régional de la Santé du Centre qui a financé certaines missions dans le cadre de ce travail.

Je salue en outre toutes les personnes qui m'ont contacté dans le cadre du site du Projet FreeIRT, pour leurs remarques et commentaires vis-à-vis des programmes qui y ont été présentés, parfois pour des remerciements, parfois pour me proposer de l'aide pour aller plus loin (en particulier, merci aux anglophones qui m'ont toujours, avec gentillesse, proposé d'améliorer mon anglais), en particulier Ronán Conroy, mais aussi Stefan Thau, Jeroen Weesie, Nick Cox, et Karl Bang Christensen.

Introduction

L'analyse statistique de la qualité de vie est devenue depuis une dizaine d'année primordiale dans le cadre de la recherche clinique. En effet, il est de plus en plus courant de voir, pour de nouveaux traitements, une équivalence avec les traitements de référence en terme de taux de guérison ou d'allongement de la durée de survie : on cherche alors à réduire par les nouveaux traitements les effets secondaires. La mesure de la qualité de vie des patients peut alors apparaître comme un véritable outil de référence pour mesurer l'impact du traitement ressenti par le patient sur son quotidien.

De nombreuses difficultés résident cependant dans la mesure d'une telle variable, non directement mesurable. Mais on peut désormais s'appuyer sur une littérature consistante datant du début du 20^e siècle pour la mesure de tels traits latents (avec Spearman dès 1904 [137] et les premiers travaux formels sur la mesure de l'intelligence) et de plus de vingt ans dans l'application concrète en mesure de la qualité de vie.

Il semble cependant que si l'aspect "confirmatoire" servant à valider des outils de mesure de la qualité de vie a été souvent traité, il n'en est pas de même pour les aspects "exploratoires" servant à construire de tels outils, dont les premiers travaux sont plus récents (années 90). La raison à cet état de fait tient sans doute à l'importance des experts (psychologues, psychiatres, cliniciens) dans la construction de questionnaire de qualité de vie. Le statisticien est ainsi appelé plutôt pour valider des outils déjà construits, mais moins souvent durant la phase de construction de tels outils. Pourtant, il peut y avoir un intérêt à construire des outils à partir de leurs propriétés psychométriques, ce qui peut limiter ensuite certains problèmes de validation. Bien entendus, l'apport du statisticien ne doit être que dans le cadre d'une aide à la décision, et il semble difficile de construire du début à la fin un outil de mesure de la qualité de vie seulement en fonction des qualités psychométriques recherchées. A ce titre, les outils exploratoires doivent pouvoir permettre l'introduction d'*a priori*, en particulier la définition d'un noyau d'items servant de base à la construction d'une échelle unidimensionnelle.

Dans un tel cadre exploratoire et psychométrique, on peut s'interroger sur les propriétés que doivent vérifier une "bonne" échelle de mesure de la qualité de vie. L'unidimensionnalité est aujourd'hui communément admise : une échelle doit mesurer un concept unique. La psychométrie moderne impose en outre, dans le cadre de la théorie de réponse aux items (IRT) deux autres propriétés : l'indépendance locale et la monotonie. Enfin, la simplicité d'utilisation de l'instrument de mesure est une notion importante. A ce titre, rechercher un ensemble d'items vérifiant un modèle de Rasch peut être une solution puisque ce modèle est sous-jacent à toute notion de score simple (non pondéré). Dans un cadre clinique, le fait qu'un outil soit particulièrement simple d'utilisation est un atout considérable, puisque ces outils

s'adressent alors à des personnes généralement peu au fait de considérations psychométriques, et hostiles à tout instrument nécessitant des calculs, aussi simples soient ils. La notion de score non pondéré (revenant à compter un nombre d'items que le patient vérifiera) reste en général en dessous du seuil de tolérance des praticiens pour l'utilisation de tels outils.

Ce travail s'inscrit donc dans le cadre du modèle de Rasch, et plus spécifiquement de la sélection, à partir d'un ensemble d'items, de sous-ensembles d'items vérifiant un modèle de Rasch. Contrairement au cadre causal où l'on cherche le meilleur modèle pour s'adapter aux données (ici les réponses aux items), on va chercher les ensembles d'items qui, selon le modèle de Rasch, donc à l'aide de scores non pondérés, permettent de mesurer le mieux possible les concepts sous-jacents mesurés par l'ensemble des items, quitte à parfois rejeter les items qui ne permettent pas d'aboutir à une meilleure qualité de la mesure (à moins d'utiliser des modèles plus complexes, donc au détriment de la simplicité d'utilisation des échelles de mesure).

Une seconde partie du travail concernera le développement d'outils psychométriques sous deux logiciels statistiques généralistes : SAS et Stata. Plusieurs programmes pour effectuer des modélisations ou de la sélection d'items seront présentés.

Plan de cette thèse

Dans le premier chapitre, nous expliciterons un certain nombre de termes, dont la définition n'est pas toujours unanime dans la littérature, aussi, ce chapitre ne doit pas être considéré comme un dictionnaire de définitions précises, mais comme une présentation détaillée de termes peu connus, qui seront utilisés dans le reste du document. Ce chapitre fera ensuite état de l'ensemble des notations utilisées par la suite. Il se conclura sur une présentation succincte des modèles dits "classiques" en analyse de qualité de vie.

Le second chapitre présentera rapidement les hypothèses fondamentales de la théorie de réponse aux items (IRT) et les modèles de cette théorie les plus couramment utilisés dans le domaine de la qualité de vie, en particulier le modèle de Rasch et ses extensions pour les items dichotomiques (OPLM) ou polytomiques (Partial Credit Model et Rating Scale Model).

Le troisième chapitre concernera l'extension de l'IRT au champ multidimensionnel, c'est-à-dire à la prise en compte simultanée de plusieurs traits latents. Les extensions de la littérature du modèle de Rasch dans ce champ sont présentées et, pour un de ces modèles, l'estimation des paramètres par les Equations d'Estimation Généralisées (GEE) sera proposée [64]. Un nouveau modèle sera proposé au quatrième chapitre : le modèle (polytomique) multidimensionnel de Rasch marginalement exhaustif (M(P)MSRM). Ce modèle a fait l'objet de la principale partie d'une publication parue dans *Communication in Statistics - Theory and Methods* [68].

Les deux chapitres suivants s'attacheront à expliquer un certain nombre de procédures issues de la théorie classique (chapitre 5) ou de l'IRT (chapitre 6) permettant de trouver une partition d'items telle que chaque sous-échelle formée vérifie un certain nombre de propriétés (unidimensionnalité, fiabilité du score pour l'analyse classique, vérification des hypothèses fondamentales pour l'IRT).

Dans ces six premiers chapitres (hors chapitre 4), nous nous attacherons bien sûr à faire une revue des thèmes traités, mais nous détaillerons aussi, de façon plus systématique, les outils utilisés dans la partie originale de ce travail, à savoir le chapitre 4 et l'ensemble des deux derniers chapitres.

Le chapitre 7 définira deux méthodes de classification des items en se basant sur l'adéquation à un modèle de Rasch. La première méthode est une adaptation au champ de l'IRT des procédures pas à pas descendantes (Backward). La seconde méthode est basée sur une procédure ascendante et a fait l'objet d'une partie dans la publication déjà citée [68] ainsi que deux communications orales dans des congrès internationaux (23^e conférence annuelle de la " Society for Multivariate Analysis in the Behavioural Sciences (SMABS)" [56], 2nd Euro-Japanese

Workshop on Stochastic Modelling for Finance, Insurance, Production and Reliability [57]). Une amélioration particulièrement efficace en terme de réduction du temps de calcul a donné lieu à une autre conférence (Workshop en l'honneur de Marvin Zelen [67]) et à un article en cours d'écriture [69]. Ce chapitre s'achèvera par une étude de simulations comparant les différentes méthodes de sélection des items en sous-échelles (présentée principalement lors du Workshop "Environment and Health Related Quality of Life : Statistical Perspectives" [66], et par deux exemples sur des données réelles : le premier concerne l'analyse de trois sous-échelles issues du "Sickness Impact Profil", présentée lors de l' International Conference on Statistics in Health Sciences [60]), l' autre concerne un travail dans un cadre différent, où la procédure a été utilisée dans le but de construire un score simple (non pondéré) pour prédire le risque de récurrence de la tentative suicidaire chez des patients reçus dans un service d'urgence. Cette dernière étude a donné lieu à deux conférences (séminaire de Probabilité et Statistique de l'Université de Montpellier 2 et de l'ENSAM [59], XXXVèmes Journées du Groupement d'Etude sur la Prévention du Suicide (GEPS) [119]) et à une publication dans la Revue Française de psychiatrie et de psychologie médicale [120].

Enfin, le dernier chapitre présentera un travail d'écriture de macro-programmes sous les logiciels généralistes SAS et Stata afin d'analyser les échelles de qualité de vie sous ces logiciels sans faire appel à des logiciels spécifiques. Ce travail concerne principalement l'estimation des paramètres des modèles de l' IRT et les méthodes de sélection d'items en sous-échelles. Le premier travail en ce sens est décrit en [65] et a été présenté lors d'un atelier INSERM [79] puis lors d'une formation de la Société Française de Statistique [58]. Un des principaux macro-programmes sous SAS, %AnaQol, a donné lieu à un article soumis à "Communications in Statistics - Simulation and Computation" [70], et un des principaux modules sous Stata a donné lieu à un article soumis à "The Stata Journal" [63]. Pour certains autres programmes, un manuel a été écrit ([62] et [61]).

Glossaire

Abbréviation	Signification	Référence
2-PLM	Two Parameters Logistic Model : Modèle logistique à deux paramètres (modèle de Birnbaum)	2.7.3
5-PAM	Five Parameters Acceleration Model : Modèle d'accélération à cinq paramètres	2.7.4
ACP	Analyse en Composantes Principales	5.3.2
AFCS	Analyse en Facteurs Communs et Spécifiques	5.3.1
AIC	Akaike Information Criterion : Critère d'information d'Akaike	7.3.3.4
ASS	Approximate Simple Structure : Structure approximativement simple	3.3
CL	Complete Linkage : Lien complet	6.1.3
CLV	Classification around Latent Variables : classification autour de variables latentes	5.3.6
CML	Conditional Maximum Likelihood : Maximum de vraisemblance conditionnelle	2.6.8.3
DETECT	Dimensionality Evaluation To Enumerate Contributing Traits : Evaluation de la dimensionalité par l'énumération des traits contributifs	6.3
EM	Expectation-Maximisation : Estimation-Maximisation	2.6.8.6
GEE	Generalized Estimating Equations : Equations d'estimation généralisées	2.6.8.5
GLM	Generalized Linear Models : Modèles linéaires généralisés	2.6.6.1
GLMM	Generalized Linear Mixed Models : Modèles linéaires généralisés mixtes	2.6.6.3
GMRM	Generalized Multidimensional Rasch Model : Modèle de Rasch multidimensionnel généralisé	3.4.3.7
HCACCPROX	Hierarchical Cluster Analysis / Conditional Proximity Measures : Classification hiérarchique avec mesures de proximité conditionnelles	6.1
ICC	Item Characteristic Curve : Courbe caractéristique de l'item	2.2
IRF	Item Response Function : Fonction de réponse à l'item	2.2
IRT	Item Response Theory : Théorie de réponse à l'item	2
(...)		

Abbréviation	Signification	Référence
JML	Joint Maximum Likelihood : Maximum de vraisemblance jointe	2.6.8.2
M2-PLM	Multidimensional Two Parameters Logistic Model : Modèle logistique à deux paramètres multidimensionnel	7.4.1.2
M5-PAM	Multidimensional Five Parameters Acceleration Model : Modèle d'accélération à cinq paramètres multidimensionnel	7.4.1.2
MML	Marginal Maximum Likelihood : Maximum de vraisemblance marginale	2.6.8.4
MMHM	Mokken Monotone Homogeneous Model : Modèle monotone homogène de Mokken	6.2.1
MMSRM	Multidimensional Marginally Sufficient Rasch Model : Modèle multidimensionnel marginalement exhaustif de Rasch	4.5
MPMSRM	Multidimensional Polytomous Marginally Sufficient Rasch Model : Modèle polytomique multidimensionnel marginalement exhaustif de Rasch	4.10
MPLT	Multidimensional Polytomous Latent Trait Model : Modèle multidimensionnel polytomique à trait latent	3.4.3
MRCML	Multidimensional Random Coefficients Multinomial Logit Model : Modèle logistique multinomial à coefficients aléatoires multidimensionnel	3.4.4
MSP	Mokken Scale Procedure : Procédure d'échelle de Mokken	6.2
OPLM	One Parameter Logistic Model : Modèle logistique à un paramètre	2.7.1
PCM	Partial Credit Model : Modèle de crédit partiel	2.8.3
RM	Rasch Model : Modèle de Rasch	2.6
RSM	Rating Scale Model : Modèle d'échelle de classement	2.8.2
SEM	Stochastic Expectation-Maximisation : Estimation-Maximisation Stochastique	2.6.8.6
SIP	Sickness Impact Profil : Profile d'impact des maladies	7.5.1
SL	Single Linkage : lien unique	6.1.3
SS	Simple Structure : Structure simple	3.3
UPGMA	Unweighted Pair-Group Method of Averages : Méthode des moyennes par paire non pondérée	6.1.3
WPGMA	Weighted Pair-Group Method of Averages : Méthode des moyennes par paire pondérée	6.1.3

Première partie

L'analyse statistique de la qualité de
vie

Chapitre 1

Notions et notations

1.1 Notions

1.1.1 Trait latent

Mesurer la qualité de vie est un exercice délicat, car la qualité de vie, comme d'autres variables (état de santé, intelligence, aptitude, connaissance...) n'est pas une variable qui peut être directement mesurée. On appelle généralement ces variables des *traits latents* lesquels "font référence à une particularité humaine (...) présente chez chacun et qui se manifeste à des degrés divers selon les individus" [54].

1.1.2 Questionnaires, profils, tests, items

La mesure de variables latentes s'effectue à l'aide de *questionnaires*, appelés aussi dans le domaine de la qualité de vie des *profils* (en général quand ces questionnaires mesurent plusieurs dimensions, comme par exemple le "Sickness Impact Profil", ou le "Profil de qualité de vie subjective") ou des *tests* (par exemple "test de QI"). Ces questionnaires sont composés d'*items* (variables) à réponses le plus souvent binaires ("oui"/"non", "d'accord"/"pas d'accord"...) ou ordinales ("pas du tout"/"peu"/"en partie"/"complètement", "jamais"/"quelquefois"/"souvent"/"toujours" ...). On note aussi plus rarement des items à réponses analogiques (règles graduées ou non), discrètes (nombre entre 0 et 10 par exemple) ou continues (réponse libre comprise entre deux valeurs), mais il n'en sera pas question dans ce document.

1.1.3 Réponses positives/correctes et négatives/incorrectes

C'est dans le domaine des sciences de l'éducation et de la psychologie (mesure de l'intelligence, mesure de la connaissance...) que se sont développées principalement les méthodes d'analyse de traits latents. Elles ont ensuite été appliquées en recherche clinique au cours des années 80. De cette origine, subsistent des termes de vocabulaire mal adaptés au domaine spécifique de la qualité de vie, notamment l'appellation des modalités de réponses aux items comme *positives/correctes* ou *négatives/incorrectes* : à chaque item, on associe une modalité *négative* (ou *incorrecte*) définie comme celle étant la moins favorable à une bonne qualité de vie (ou à un bon état de santé) ; les autres modalités sont appelées *positives* (ou *correctes*), et correspondent aux modalités les plus favorables eu égard à la qualité de vie. On note que dans le cas d'items à réponses binaires, il existe pour chaque item une seule modalité positive.

1.1.4 Mesures indirectes, scores

Pour mesurer les variables latentes, on utilise des méthodes de *mesures indirectes*, basées sur des observations réelles (réponses aux items). Cette inférence passe par la construction d'un modèle établissant un lien entre les traits latents et les observations. Le modèle ainsi construit aboutit à la création d'un (ou plusieurs) *score(s)*, pondéré(s) ou non, avec des pondérations connues ou estimées, le(s)quel(s) reflète(nt) le(s) niveau(x) des individus par rapport au(x) trait(s) latent(s) étudié(s).

1.1.5 Dimensions, traits latents, échelles, sous-échelles

Les notions de *dimension*, *échelles* et *sous-échelles* restent floues dans la littérature. Nous utiliserons dans ce document le terme *dimension* pour parler du concept mesuré, celui d'*échelle* comme l'ensemble des items d'un questionnaire mesurant un (ou plusieurs) *trait(s) latent(s)*, et celui de *sous-échelle* comme le sous-ensemble des items d'un questionnaire mesurant un seul trait latent donné.

En outre, dans le cadre d'un questionnaire unidimensionnel, il y aura confusion entre les termes *questionnaire*, *d'échelle* et *sous-échelle*. Ce sera notamment le cas dans le chapitre 2 de ce document.

1.1.6 Psychométrie, théorie classique, théorie de réponse aux items

Il est courant de différencier, dans le champ de la mesure de traits latents psychologiques (*Psychométrie*), la *théorie classique* et la *théorie de réponse aux items* (IRT) [théorie moderne]. Les modèles de la théorie classique décomposent le score observé en un vrai score et une erreur, et modélise les observations par des modèles linéaires. Les modèles de l'IRT entrent dans le cadre des modèles linéaires généralisés et décrivent une probabilité de réponse par rapport au trait latent, généralement via un lien logistique ou probit (plus rarement logarithmique).

1.2 Notations

1.2.1 Notations mathématiques générales

Dans ce document, \mathbb{N} et \mathbb{R} représenteront respectivement l'ensemble des entiers naturels et l'ensemble des réels.

Sauf exceptions explicitement indiquées, les scalaires seront notés en minuscules maigres (a, b, α, \dots), les vecteurs en minuscules grasses ($\mathbf{a}, \boldsymbol{\theta}, \dots$) et les matrices en majuscules grasses ($\mathbf{A}, \boldsymbol{\Theta}, \dots$). Les constantes, comme celles définies par exemple au paragraphe 1.2.2, seront notées en majuscules maigres.

\mathbf{A}' représente la transposée de la matrice \mathbf{A} , \mathbf{A}^{-1} son inverse (si \mathbf{A} est inversible) et \mathbf{A}^- une inverse généralisée.

Les variables aléatoires seront représentées par des majuscules (X, Y, \dots). Les réalisations de ces variables seront notées par la même lettre en minuscule et indexées (x_1, y_n, \dots). $E(X)$ et $V(X)$ sont respectivement l'espérance et la variance de X . $Cov(X, Y)$ et $Corr(X, Y)$ sont

respectivement la covariance et la corrélation entre les variables aléatoires X et Y .

Les estimations seront notées par un chapeau : \hat{x} , $\hat{\delta}$...

Le vecteur $\mathbf{x} = (x_1, \dots, x_l, \dots, x_L)'$ sera notée $\mathbf{x} = (x_l)_{l=1, \dots, L}$, tandis que $\mathbf{x} = (x_1, \dots, x_l, \dots, x_L)$ sera notée $\mathbf{x} = (x_l)_{l=1, \dots, L}'$.

\mathbf{I}_d représente la matrice identité d'ordre d et $\mathbf{1}_d$ représente le vecteur unité de taille d . $\mathbf{0}$ représente, selon le contexte, une matrice ou un vecteur composé de 0.

$x = \mathbb{1}_I(y)$ signifie que $x = 1$ si $y \in I$ et $x = 0$ sinon.

1.2.2 Notations spécifiques

La taille du questionnaire (nombre d'items de l'échelle) sera notée J . Les items sont indexés par $j = 1, \dots, J$. La dimension du questionnaire (nombre de sous-échelles) est notée Q , et chaque dimension est indexée par $q = 1, \dots, Q$. La taille (nombre d'items) de la q^e sous-échelle est notée par J_q .

Le questionnaire est passé à un échantillon de N individus, chaque individu étant indexé par $n = 1, \dots, N$ ¹.

Les variables aléatoires de réponses des individus aux items sont notées X_j , $j = 1, \dots, J$. La réponse du n^e individu au j^e item est notée x_{nj} . m_j représente le nombre de modalités positives de l'item j ($m_j = 1$ si le j^e item est dichotomique).

Les Q traits latents sont notés Θ_q , et leurs valeurs individuelles pour le n^e individu sont $\theta_n = (\theta_{n1}, \dots, \theta_{nq}, \dots, \theta_{nQ})'$. Dans un cadre unidimensionnel ($Q = 1$), on notera Θ le trait latent et θ_n sa valeur individuelle pour le n^e individu.

Pour chaque item, la modalité négative sera codée 0. Dans le cadre dichotomique, la modalité positive de chaque item est codée 1, et dans un cadre polytomique, les modalités positives seront codées par des entiers supérieurs ou égaux à 1 : on considèrera que les modalités sont ordonnées de la modalité la moins favorable à un bon niveau sur le trait latent (codage le plus petit, en général 1) à la modalité la plus favorable à un bon niveau sur le trait latent (le plus souvent codée m_j pour le j^e item). En général, bien que ce ne soit pas la règle absolue, plusieurs modalités d'un même item ne sont pas codées de la même manière.

Le score est la variable aléatoire $S = \sum_{j=1}^J X_j$. Cette variable prend pour valeur $s_n = \sum_{j=1}^J x_{nj}$ pour l'individu n .

M représente la valeur maximale possible du score $M = \sum_{j=1}^J m_j$, la valeur minimale du score étant 0.

¹volontairement, nous avons choisi de ne pas indexer les individus par i comme cela est généralement fait, dans un soucis de cohérence

D'autres notations seront introduites au fur et à mesure du texte.

1.3 Les modèles classiques

1.3.1 Principes

La théorie classique [95] est basée sur l'analyse du score observé considéré comme une estimation d'un vrai score représentant la variable latente à mesurer [91].

1.3.2 Définition

En théorie classique [95], on suppose un lien linéaire entre le score (on parlera de score pour la réponse à chaque item) observé X_{nj} et un "vrai" score τ_{nj} défini comme l'espérance de X_{nj} . Le modèle sous-jacent est donc un modèle linéaire à effets aléatoires :

$$X_{nj} = \tau_{nj} + \epsilon_{nj} \quad (1.1)$$

où ϵ_{nj} est l'erreur de mesure, avec $E(\epsilon_{nj}) = 0$ et $Cov(\tau_{nj}, \epsilon_{nj}) = 0$.

1.3.3 La fiabilité

La fiabilité d'une mesure (d'un score) est définie par le rapport entre la variance du "vrai" score τ_{nj} (que l'on notera $\sigma_{\tau_j}^2$) et celle du score observé X_{nj} (que l'on notera $\sigma_{X_j}^2$). On la note pour l'item X_j :

$$\rho_j^2 = \frac{\sigma_{\tau_j}^2}{\sigma_{X_j}^2} \quad (1.2)$$

1.3.4 Le coefficient alpha de Cronbach

Le coefficient alpha de Cronbach est l'un des indices le plus souvent retrouvé dans les analyses d'échelles de qualité de vie. C'est un indice qui permet de mesurer la fiabilité d'un modèle classique, c'est-à-dire la fiabilité quand on utilise le score S_n pour mesurer le vrai score τ_n .

On suppose que le score S_n est la somme de J variables aléatoires X_{nj} , $j = 1, \dots, J$ telles que $\rho_X^2 = Corr(X_{nj}, X_{nl})$, $\forall j \neq l$ et ρ_X^2 fixé², alors la fiabilité du score S_n s'écrit [54] :

$$\rho_S^2 = \frac{J\rho_X^2}{1 + (J-1)\rho_X^2} \quad (1.3)$$

Cette égalité est connue sous le nom de formule de Spearman-Brown. Le coefficient alpha de Cronbach [24] est une estimation de ρ_S^2 . Dans le cas de la normalité des variables X_{nj} , $j = 1, \dots, J$, l'estimateur du maximum de vraisemblance du alpha de Cronbach est :

$$\alpha = \hat{\rho}_S^2 = \frac{J\hat{r}}{1 + (J-1)\hat{r}} \quad (1.4)$$

²on parle alors d'un modèle parallèle

où \hat{r} est une estimation de la corrélation moyenne entre les variables X_j , $j = 1, \dots, J$ prises deux à deux.

$$\hat{r} = \frac{2}{(J-1)J} \sum_{j=1}^J \sum_{k=j+1}^J \text{Corr}(X_j, X_k) \quad (1.5)$$

1.4 Synthèse

Ce premier chapitre introductif a permis de préciser les notions élémentaires de la psychométrie. Le chapitre suivant présentera la théorie de réponse aux items (IRT) et les principaux modèles de cette théorie utilisés dans le cadre de l'analyse de la qualité de vie.

Chapitre 2

La Théorie de Réponse aux items (Item Response Theory - IRT)

2.1 Définition générale

La théorie de réponse aux items (Items Response Theory - IRT) [114] [147] est basée sur une relation entre les réponses à des items mesurant un trait latent et ce trait latent. *A contrario* de l'analyse classique, la relation entre le score observé et le trait latent n'est plus (forcément) linéaire.

L'IRT repose sur trois hypothèses fondamentales : l'unidimensionnalité, la monotonie et l'indépendance globale.

Dans cette partie, et sauf mention contraire, l'IRT fera référence à l'IRT paramétrique telle que définie à la section 2.5.2

2.2 Les fonctions de réponse aux items (Items Response Functions - IRF) et les courbes caractéristiques des items (Items Characteristic Curves - ICC)

Les fonctions de réponses aux items (IRF) [147] sont les fonctions $P(X_{nj} = x_{nj} | \theta_n, \nu_j)$ où ν_j est un vecteur de paramètres caractérisant l'item j . En IRT, on modélise les IRF en fonction du (ou des) trait(s) latent(s) et des paramètres associés aux items.

Les courbes caractéristiques des items (ICC)[14] sont les représentations graphiques des IRF en fonction du trait latent.

2.3 Hypothèses fondamentales

2.3.1 L'unidimensionnalité

L'unidimensionnalité signifie que le trait latent sous-jacent à un ensemble d'items (à travers leur fonction réponse) est un même et unique scalaire. La réalisation du trait latent pour

l'individu n est notée θ_n . Un tel ensemble d'items est dit, par abus de langage, unidimensionnel. Nous verrons au chapitre 3 que cette hypothèse peut être dépassée quand le trait latent sous-jacent aux items est vectoriel : on dit alors que les items sont multidimensionnels.

2.3.2 La monotonie

La monotonie concerne les fonctions réponses des items. En IRT, les IRF doivent être non décroissantes en Θ , ou en d'autres termes, on doit avoir :

$$\frac{dP(X_{nj} = x_{nj}/\theta_n, \boldsymbol{\nu}_j)}{d\theta_n} \geq 0 \quad (2.1)$$

2.3.3 L'indépendance locale

L'indépendance locale signifie que conditionnellement au trait latent, les réponses aux différents items sont indépendantes. Pour un individu donné, ou plus généralement pour l'ensemble des individus ayant la même valeur pour le trait latent, les réponses aux différents items sont indépendantes les unes des autres (les variables X_{nj} et X_{nk} , avec $j \neq k$, les indices de deux items différents, sont indépendantes, $\forall n = 1, \dots, N$).

$$P(\mathbf{X}_n = \mathbf{x}_n/\theta_n, \boldsymbol{\nu}) = \prod_{j=1}^J P(X_{nj} = x_{nj}/\theta_n, \boldsymbol{\nu}_j), \quad (2.2)$$

avec $\mathbf{X}_n = (X_{nj})_{j=1, \dots, J}$, $\mathbf{x}_n = (x_{nj})_{j=1, \dots, J}$ et $\boldsymbol{\nu} = (\boldsymbol{\nu}_j)_{j=1, \dots, J}$.

2.4 Les traces des items

2.4.1 Principes généraux : les traces classiques

Les traces des items sont des représentations graphiques simplifiées des courbes caractéristiques des items. On représente, pour un item donné, et pour chaque valeur du score de 0 à M la probabilité qu'un individu ayant ce score ait répondu à une des modalités positives de cet item. La trace d'un item j correspondant à sa k^e modalité représente donc $P(X_j = k)$ en fonction du score S . Dans une échelle bien construite constituée d'items dichotomiques, la trace doit être une fonction non décroissante : plus les individus ont un score élevé, et plus la probabilité de répondre positivement à chacun des items doit être élevée (hypothèse de monotonie). Dans un cadre polytomique, l'analyse est plus délicate : en effet, les individus ayant de très hautes valeurs sur le trait latent ont tendance à plus souvent répondre aux modalités les plus hautes, au détriment des modalités moyennes. Pour ces modalités moyennes, on assiste généralement à une tendance croissante de la trace, puis à une tendance décroissante.

2.4.2 Traces particulières : les traces cumulées

Pour une analyse plus simple, les traces cumulées peuvent être utilisées pour les items polytomiques. On ne représente alors plus la probabilité de répondre à une modalité particulière d'un item, mais à la probabilité de répondre "au moins" à chacune des modalités d'un item. La trace correspondant à sa k^e modalité représente donc $P(X_j \geq k)$ en fonction du score S . Toutes les traces cumulées doivent être non décroissantes.

2.4.3 Détection des mauvais items à l'aide des traces

Lorsque la trace d'un item dichotomique ou une trace cumulée d'un item polytomique est plate, sur la totalité du spectre du score, un item mal intégré à l'échelle est à suspecter. En effet, cela signifie que la probabilité de répondre positivement à cet item est indépendant de la valeur du score, donc que l'item mesure un autre concept que celui mesuré par le score.

Il arrive régulièrement d'avoir des items pour lesquels, à partir d'une certaine valeur du score, tous les individus répondent correctement à cet item, ou au contraire, pour lesquels, en dessous d'une certaine valeur du score, aucun individu ne répond correctement à cet item. On observe alors dans ces deux cas des parties complètement plates des traces. Dans le premier cas, on parle d'un effet plafond, et dans le second cas d'un effet plancher. Cela signifie que, dans le premier cas, l'item est sensiblement plus facile que le reste des items, et dans le second cas, que l'item est sensiblement plus difficile que les autres. Ces items sont le plus souvent inutiles car mal adaptés à la population étudiée.

2.4.4 Adaptation des traces : les traces en fonction du sous-score

Lorsque qu'un item ne mesure pas le même concept que les autres items utilisés pour calculer un score, les traces sont toujours légèrement croissantes, puisque l'item visé est utilisé pour calculer le score (donc la réponse à l'item et le score sont légèrement corrélés positivement). Ce phénomène est d'autant plus important que le nombre d'items est restreint. Pour pallier à ce problème, on peut utiliser le sous-score (rest-score), c'est-à-dire le score calculé sans l'item donné (sous-score par rapport à l'item k $S_k = \sum_{j=1, j \neq k}^J X_j = S - X_k$), pour effectuer la représentation des traces : dans ce cas, la trace d'un item mesurant un autre concept que celui mesuré par le reste des items est plate.

2.5 IRT paramétrique et IRT non-paramétrique

2.5.1 L'IRT non paramétrique

L'IRT non paramétrique a démarré au début des années 70 avec les travaux de Mokken [107] mais s'est principalement développée dans le courant des années 90 (voir par exemple [108], [115], [140], [71], [155]...). Elle vise à vérifier les hypothèses fondamentales de l'IRT [116]. Dans ce contexte, on ne s'intéresse pas à la caractérisation des items et des individus par le biais de paramètres. Ce champ a été notamment utilisé dans le cadre de la sélection d'items en échelles vérifiant les hypothèses fondamentales de l'IRT (voir chapitre 6).

2.5.2 L'IRT paramétrique

L'IRT paramétrique représente la forme originelle de l'IRT : les fonctions de réponse aux items sont modélisées en fonction de paramètres caractérisant les items et les individus, et tels que les modèles permettent de vérifier les hypothèses fondamentales de l'IRT.

2.6 Le modèle de Rasch (RM ou 1-PLM)

2.6.1 Modélisation

Le modèle de Rasch [126] est un des modèles les plus simples de l'IRT puisque chaque item est caractérisé par un paramètre unique appelé le paramètre de difficulté de l'item. Une fonction de lien logistique relie les réponses de chaque item à un prédicteur linéaire des paramètres caractérisant les capacités des individus (le trait latent θ_n) et les paramètres de difficulté des items (δ_j). Le modèle de Rasch est un modèle unidimensionnel (le trait latent est un scalaire). Ce modèle est défini par la propriété d'indépendance locale et par ses IRF (croissantes par construction) :

$$P(X_{nj} = x_{nj}/\theta_n, \delta_j) = \frac{\exp(x_{nj}(\theta_n - \delta_j))}{1 + \exp(\theta_n - \delta_j)}$$

2.6.2 Courbes caractéristiques des items dans le modèle de Rasch

Lorsque l'on représente les courbes caractéristiques d'un ensemble d'items vérifiant un modèle de Rasch, on obtient des courbes logistiques non sécantes comme présentées à la figure 2.1.



FIG. 2.1 – Courbes caractéristiques de 7 items vérifiant un modèle de Rasch

2.6.3 Interprétation des paramètres de difficulté des items

Les paramètres de difficultés des items s'interprètent facilement : pour un individu donné (θ_n fixé), la fonction $P(X_{nj} = x_{nj}/\theta_n, \delta_j)$ est décroissante en δ_j , ce qui peut s'interpréter par le fait que plus la valeur de δ_j est grande, et plus la probabilité qu'un individu donné réponde positivement à cet item est petite (c'est-à-dire plus cet item est difficile).

La valeur du paramètre δ_j représente la valeur du trait latent qu'il faut atteindre pour avoir une chance sur deux de répondre positivement à l'item j : les capacités des individus et

les difficultés des items sont donc représentés dans la même échelle. Géométriquement, sur la courbe caractéristique d'un item, la difficulté correspond à l'abscisse du point de l'IRF ayant une ordonnée de 0.5.

2.6.4 Paramètres fixes et paramètres aléatoires

Il existe deux manières de traiter le modèle de Rasch. La plus ancienne consiste à considérer les paramètres d'aptitude des individus (θ_n) comme des paramètres fixes. Le modèle est alors composé de N paramètres d'aptitude et de J paramètres de difficulté. Nous verrons ensuite qu'il n'existe en fait que $(J - 1)$ paramètres d'aptitude distincts pouvant être estimés de façon consistante.

Une seconde manière de traiter les paramètres d'aptitude des individus est de considérer que ces paramètres sont des réalisations d'une variable aléatoire Θ : on suppose alors la famille de distribution de cette variable aléatoire et l'on estime les $k = \text{card}(\nu)$ paramètres ν de cette fonction de distribution $G(\Theta/\nu)$. En général, on suppose que cette variable aléatoire est distribuée selon une loi normale, aussi l'on estime la moyenne et la variance de cette variable ($k = 2$), et le modèle est alors composé de $(J + 2)$ paramètres.

2.6.5 Représentation graphique

Kreiner [90] propose de définir le modèle de Rasch à effets fixes comme un modèle graphique. La figure 2.2 montre un exemple d'une telle représentation pour cinq items. On note que les flèches représentent le sens d'une relation de dépendance causale (le trait latent explique la réponse à chaque item). En outre, il n'y a pas de liaison (arrête) entre les items, ce qui signifie donc que les réponses aux items sont indépendantes conditionnellement au trait latent (hypothèse d'indépendance locale) [et aux autres items].

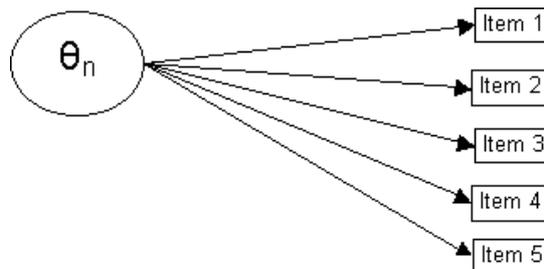


FIG. 2.2 – Représentation graphique du modèle de Rasch

2.6.6 Le modèle de Rasch comme un modèle linéaire généralisé (mixte)

2.6.6.1 Les modèles linéaires généralisés - GLM

Soient pour chaque observation i ($i = 1, \dots, N$) une variable réponse notée y_i , un vecteur de covariables noté \mathbf{x}_i et un vecteur de paramètres inconnus noté β .

Les modèles linéaires généralisés sont définis par [100] :

1. les observations sont mutuellement indépendantes,
2. la densité de la variable réponse y_i appartient à la famille exponentielle (dont la densité est donnée ci dessous - équation (2.3)),
3. l'espérance $\mu_i = E(y_i)$ est reliée au prédicteur linéaire $\eta_i = \mathbf{u}_i' \boldsymbol{\beta}$ avec $\mathbf{u}_i = u(\mathbf{x}_i)$ un vecteur connu et la fonction $h(\cdot)$ bijective appelée fonction lien, par la relation $\mu_i = h(\eta_i)$.

La densité d'une variable y_i est de la famille exponentielle si, en posant θ_i un paramètre canonique, ϕ un paramètre de dispersion, w_i un poids associé à la réalisation de y_i , $d_i(\phi) = \frac{\phi}{w_i}$ et $a(\cdot)$ et $c(\cdot)$ des fonctions spécifiques à chaque distribution, on peut écrire la densité de y_i comme :

$$f(y_i; \theta_i) = \exp \left(\frac{y_i \theta_i - a(\theta_i)}{d_i(\phi)} + c(y_i; \phi) \right) \quad (2.3)$$

2.6.6.2 Le modèle de Rasch à effets fixes

Dans le cadre général du modèle de Rasch, les observations x_{nj} sont par définition indépendantes, de par la propriété d'indépendance locale.

Dans le cas du modèle de Rasch à effet fixe, on peut montrer qu'il appartient aux GLM en montrant que la densité de la variable réponse x_{nj} est de la forme exponentielle, en posant $\theta_{nj} = \log \left(\frac{p_{nj}}{1-p_{nj}} \right)$ avec $p_{nj} = \frac{\exp(\theta_n - \delta_j)}{1 + \exp(\theta_n - \delta_j)}$, $a(\theta_{nj}) = \log(1 + \exp(\theta_n - \delta_j))$, $w_{nj} = 1$, $\phi = 1$ et $c(x_{nj}, \phi) = 0$.

La fonction lien $h(\cdot)$ est donc la fonction logit et le modèle est un modèle logistique.

En posant les vecteurs $\mathbf{u}_{\mathbf{n}j}' = (\underbrace{0, \dots, 0, -1, 0, \dots, 0}_J \text{ paramètres } \delta_j, \underbrace{0, \dots, 0, 1, 0, \dots, 0}_N \text{ paramètres } \theta_n)$, $\boldsymbol{\delta} = (\delta_j)_{j=1, \dots, J}$, $\boldsymbol{\theta} = (\theta_n)_{n=1, \dots, N}$ et $\boldsymbol{\beta}' = (\boldsymbol{\delta}', \boldsymbol{\theta}')$, on retrouve le modèle de Rasch à effets fixes.

2.6.6.3 Les modèles linéaires généralisés mixtes - GLMM

Les GLMM (GLM à effets aléatoires) sont une extensions des GLM où des effets aléatoires sont introduits dans le prédicteur linéaire η_i , de telle sorte que, conditionnellement aux effets aléatoires, le modèle est défini comme un GLM [33]. Le vecteur $\boldsymbol{\beta}$ ne contient alors plus que les effets fixes.

Soit \mathbf{b} un vecteur aléatoire réel non observé appelé vecteur des effets aléatoires décomposé en Q parties $\mathbf{b} = (\mathbf{b}_{\mathbf{q}})_{q=1, \dots, Q}$ avec $\mathbf{b}_{\mathbf{q}}$ définissant le q^e trait latent ($\mathbf{b}_{\mathbf{q}} = (\mathbf{b}_{i\mathbf{q}})_{i=1, \dots, N}$). Soit un vecteur \mathbf{z}_i connu permettant de lier l'observation i au q^e trait latent. Alors le prédicteur linéaire η_i s'écrit :

$$\eta_i = \mathbf{u}_i' \boldsymbol{\beta} + \mathbf{z}_i' \mathbf{b} \quad (2.4)$$

Par analogie avec les GLM, on peut définir un GLMM de la manière suivante :

1. les observations sont mutuellement indépendantes,
2. conditionnellement aux effets aléatoires, la densité de la variable réponse y_i appartient à la famille exponentielle (définie à l'équation (2.3)),

3. l'espérance $\mu_i = E(y_i/\mathbf{b})$ est reliée au prédicteur linéaire η_i par la relation $\mu_i = h(\eta_i)$,
4. les erreurs sont indépendantes des effets aléatoires,
5. la loi du vecteur \mathbf{b} est en général supposée comme une distribution multinormale centrée de matrice de variance-covariance Σ inconnue.

2.6.6.4 Le modèle de Rasch à effets aléatoires

Dans le cas du modèle de Rasch à effets, on peut montrer qu'il appartient aux GLMM en montrant que conditionnellement au trait latent, la densité de la variable réponse x_{nj} est de la forme exponentielle. Dans ce cas, $Q = 1$ et le vecteur des effets aléatoires est défini comme $\mathbf{b} = (\theta_n)_{n=1, \dots, N}$. On pose $\boldsymbol{\beta} = \boldsymbol{\delta}$, $\mathbf{u}_{\mathbf{n}j} = (0, \dots, 0, -1, 0, \dots, 0)$ de dimension J et $\mathbf{z}_{\mathbf{n}j} = (0, \dots, 0, 1, 0, \dots, 0)$ de dimension N .

La fonction lien $h(\cdot)$ est donc la fonction logit et le modèle est un modèle logistique mixte.

2.6.7 Propriétés spécifiques

Le modèle de Rasch est un modèle simple à interpréter. Cependant, les hypothèses sous-jacentes qu'il implique (ICC non sécantes par exemple), fait qu'il est parfois difficile de trouver un ensemble d'items qui vérifient un modèle de Rasch. De nombreux autres modèles plus souples ont été proposés mais aucun n'a connu le succès du modèle de Rasch. Deux raisons peuvent expliquer ce phénomène : la première est qu'un modèle plus souple, ou qui exige des hypothèses moins spécifiques, offre l'avantage de mieux s'adapter à un ensemble d'items existant, mais pose des problèmes d'estimation des paramètres, de mesure de l'adéquation et d'interprétation ; la seconde tient sans doute à deux propriétés importantes et spécifiques à une famille de modèles proches du modèle de Rasch : l'exhaustivité du score sur le trait latent, et la "specific objectivity".

Ceci a amené certains psychométriciens, dont Molenaar [114, page 5], à estimer qu'il est souvent préférable, pour mesurer un phénomène latent, de rechercher un ensemble d'items qui vérifient un modèle de Rasch avec ce trait latent, plutôt que de trouver un modèle plus complexe qui a une bonne adéquation à un ensemble d'items donnés. En effet, on cherche plutôt en psychométrie à définir l'ensemble des items qui va permettre de mesurer le mieux possible le trait latent, c'est-à-dire un ensemble d'items vérifiant, au minimum, les hypothèses fondamentales de l'IRT. La simplicité d'interprétation (et donc d'utilisation) du modèle de Rasch, fait qu'il est par conséquent plus intéressant de trouver un bon ensemble d'items vérifiant un tel modèle, quitte à jeter certains items, que de trouver un modèle plus complexe ayant de moins bonnes propriétés de mesure mais s'ajustant mieux à l'ensemble des données.

2.6.7.1 Exhaustivité du score sur le trait latent

L'exhaustivité du score sur le trait latent signifie que, sous le modèle de Rasch, toute l'information disponible sur la valeur du trait latent d'un individu n est contenu dans la statistique exhaustive $S_n = \sum_{j=1}^J X_{nj}$ (score). On peut vérifier simplement cette propriété en montrant que sous l'hypothèse d'indépendance locale, la densité conjointe des variables

$X_{n1}, \dots, X_{nj}, \dots, X_{nJ}$ (que l'on notera $f(X_{n1}, \dots, X_{nj}, \dots, X_{nJ}/\theta_n, \delta)$) appartient à la famille exponentielle :

$$\begin{aligned} f(X_{n1} = x_{n1}, \dots, X_{nj} = x_{nj}, \dots, X_{nJ} = x_{nJ}/\theta_n, \delta) &= \prod_{j=1}^J \frac{\exp[x_{nj}(\theta_n - \delta_j)]}{1 + \exp(\theta_n - \delta_j)} \\ &= \frac{\exp\left[s_n \theta_n - \sum_{j=1}^J x_{nj} \delta_j\right]}{\prod_{n=1}^N \prod_{j=1}^J 1 + \exp(\theta_n - \delta_j)} \quad (2.5) \end{aligned}$$

Par le théorème de factorisation de la famille exponentielle (voir par exemple [30], ou sous le nom de théorème de Darmais [135]), on montre formellement que la statistique S_n est exhaustive du trait latent θ_n . Le modèle de Rasch n'est pas le seul modèle à vérifier cette propriété (si l'on considère le score comme pouvant être pondéré) : les modèles IRT vérifiant une telle propriété sont dits "de la famille du modèle de Rasch" (Rasch family models) [36]. Cependant, le modèle de Rasch est le seul modèle de l'IRT où cette propriété s'applique, pour des items dichotomiques, au score simple (non pondéré) [9].

La figure 2.3 représente la propriété d'exhaustivité du score sur le trait latent dans le cadre d'une représentation graphique : dans cette figure, on note que toute l'information disponible sur l'ensemble des items est contenue dans le score (aucun lien direct entre les items et le trait latent), et que désormais, conditionnellement au score, les réponses aux items ne sont plus indépendantes.

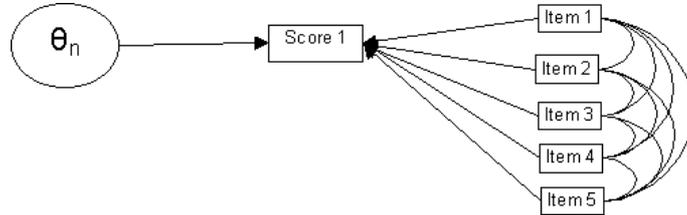


FIG. 2.3 – Représentation graphique de la propriété d'exhaustivité du score sur le trait latent dans le cadre du modèle de Rasch

2.6.7.2 L'objectivité spécifique (Specific Objectivity)

L'objectivité spécifique est une propriété du modèle de Rasch définie et nommée ainsi par Rasch [128]. Elle repose sur l'idée qu'il est possible, au sein du modèle de Rasch, de comparer "objectivement" deux individus, c'est-à-dire indépendamment de l'instrument de mesure (le questionnaire). Un exemple typique de cette probabilité est de montrer que la probabilité qu'entre deux individus n_1 et n_2 donnés, n_1 ait répondu positivement à l'item j et n_2 ait répondu négativement au même item, sachant qu'un seul des deux a répondu positivement à cet item, ne dépend pas du paramètre de difficulté de l'item (δ_j) :

$$P(X_{n_1j} = 1, X_{n_2j} = 0/\theta_{n_1}, \theta_{n_2}, \delta_j, X_{n_1j} + X_{n_2j} = 1) = \frac{\exp(\theta_{n_1})}{\exp(\theta_{n_1}) + \exp(\theta_{n_2})} \quad (2.6)$$

Cet exemple peut être généralisé à J items.

Une des conséquences de cette propriété est que pour des individus ne présentant pas le même niveau sur le trait latent, les difficultés des items restent les mêmes (en terme d'ordre et d'écart entre les valeurs des paramètres).

Fisher (voir une revue sur ce thème en [36]) a montré que le modèle de Rasch est le seul modèle vérifiant les hypothèses de l'IRT à respecter cette propriété.

2.6.8 L'estimation des paramètres

2.6.8.1 Identifiabilité des paramètres du modèle de Rasch

- Le modèle de Rasch est un modèle linéaire généralisé qui associe à la variable réponse X_{nj} :
- les J paramètres δ_j et les N paramètres θ_n dans le cadre du modèle de Rasch à effets fixes,
 - les J paramètres δ_j et les k paramètres composant $\boldsymbol{\nu}$ dans le cadre du modèle de Rasch à effets aléatoires.

Dans le cas du modèles de Rasch à effets fixes, la matrice $U = (\mathbf{u}_{nj}')_{n=1,\dots,N,j=1,\dots,J}$ (voir les définitions à la section 2.6.6.2) est de rang $(J + N - 1)$ tandis que le vecteur $\boldsymbol{\beta}$ est de dimension $(J + N)$, ceci justifie la non identifiabilité de ce modèle. L'ajout d'une contrainte d'identifiabilité permet d'identifier le modèle [113].

En général, la contrainte d'identifiabilité utilisée est celle de la nullité de la moyenne des paramètres de difficulté :

$$\sum_{j=1}^J \delta_j = 0$$

C'est notamment le cas pour le logiciel RSP [46]. On trouve aussi parfois comme contrainte la nullité d'un paramètre de difficulté particulier ($\delta_k = 0$).

Dans le cadre du modèle de Rasch à effets aléatoires, le modèle est identifiable si l'on considère la distribution du trait latent comme centrée. Dans l'autre cas, le modèle revient à un modèle équivalent où la moyenne du trait latent peut être considéré comme une constante, et l'on doit alors poser une contrainte d'identifiabilité sur les paramètres de difficultés des items (du même type que dans le cadre du modèle de Rasch à effets fixes).

2.6.8.2 La méthode du maximum de vraisemblance jointe

La méthode du maximum de vraisemblance jointe (Joint Maximum Likelihood - JML) est la méthode d'estimation la plus naturelle dans le cas où le trait latent est considéré comme un ensemble de paramètres fixes. Elle consiste à trouver les estimateurs qui maximisent la vraisemblance jointe :

$$L_J(\boldsymbol{\delta}, \boldsymbol{\theta} / \mathbf{x}) = \prod_{n=1}^N \prod_{j=1}^J P(X_{nj} = x_{nj} / \theta_n, \delta_j) \quad (2.7)$$

avec $\boldsymbol{\delta} = (\delta_j)_{j=1,\dots,J}$ et $\boldsymbol{\theta} = (\theta_n)_{n=1,\dots,N}$.

En dérivant le logarithme de cette quantité par rapport à θ_n ($n = 1, \dots, N$) et δ_j ($j = 1, \dots, J$), on obtient les équations du maximum de vraisemblance suivante :

$$\begin{cases} s_n = \sum_{j=1}^J \frac{\exp(\theta_n - \delta_j)}{1 + \exp(\theta_n - \delta_j)} \\ t_j = \sum_{n=1}^N \frac{\exp(\theta_n - \delta_j)}{1 + \exp(\theta_n - \delta_j)} \end{cases}, n = 1, \dots, N; j = 1, \dots, J \quad (2.8)$$

Les estimations des paramètres de difficulté du modèle obtenues par cette méthode sont cependant non consistants [43]. Andersen [7] montre cependant qu'elles le sont à condition que $N \rightarrow +\infty$, $J \rightarrow +\infty$ et $N/J \rightarrow +\infty$. Aussi cette méthode est à proscrire, en particulier lorsque le nombre d'items est important, ou lorsque le nombre d'individus est petit.

2.6.8.3 La méthode du maximum de vraisemblance conditionnelle

La méthode du maximum de vraisemblance conditionnelle (Conditional Maximum Likelihood - CML) permet de résoudre le problème de l'inconsistance des estimateurs du JML lorsque l'on considère le trait latent comme un ensemble de paramètres fixes. Le principe de cette méthode est alors de s'affranchir du trait latent en utilisant la vraisemblance des observations conditionnellement au score :

$$\begin{aligned} L_{Cn}(\boldsymbol{\delta}, \theta_n / \mathbf{x}_n, s_n) &= P(\mathbf{X}_n = \mathbf{x}_n / \theta_n, \boldsymbol{\delta}, S_n = s_n) \\ &= \frac{P(\mathbf{X}_n = \mathbf{x}_n / \theta_n, \boldsymbol{\delta})}{P(S_n = s_n / \theta_n, \boldsymbol{\delta})} \\ &= \prod_{j=1}^J \frac{\exp[x_{nj}(\theta_n - \delta_j)]}{1 + \exp(\theta_n - \delta_j)} \times \left[\sum_{\mathbf{y} \in \Omega / \sum_{j=1}^J y_j = s_n} \prod_{j=1}^J P(X_{nj} = y_j / \theta_n, \delta_j) \right]^{-1} \\ &= \frac{\exp(\theta_n s_n) \exp(-\sum_{j=1}^J x_{nj} \delta_j)}{\prod_{j=1}^J [1 + \exp(\theta_n - \delta_j)]} \times \frac{\prod_{j=1}^J [1 + \exp(\theta_n - \delta_j)]}{\exp(\theta_n s_n) \sum_{\mathbf{y} \in \Omega / \sum_{j=1}^J y_j = s_n} \exp(-\sum_{j=1}^J y_j \delta_j)} \\ &= \frac{\exp\left(-\sum_{j=1}^J x_{nj} \delta_j\right)}{\gamma_{s_n}(\boldsymbol{\delta})} \\ &= L_{Cn}(\boldsymbol{\delta} / \mathbf{x}_n, s_n) \end{aligned} \quad (2.9)$$

où $\mathbf{s} = (s_n)_{n=1,\dots,N}$, Ω est l'ensemble des 2^J vecteurs possible de dimension J composés de 0 et de 1 et $\mathbf{y} = (y_j)_{j=1,\dots,J}$ est un vecteur quelconque de Ω . La fonction symétrique gamma est définie comme :

$$\gamma_s(\boldsymbol{\delta}) = \sum_{\mathbf{y} \in \Omega / \sum_{j=1}^J y_j = s} \exp\left(-\sum_{j=1}^J y_j \delta_j\right) \quad (2.10)$$

Le trait latent disparaît alors dans cette vraisemblance conditionnelle. On note que dans ce cas, seuls les individus n'ayant pas un score nul ou parfait ($s_n \neq 0$ et $s_n \neq J$) apportent de l'information (dans le cas d'un score nul ou parfait, la vraisemblance associée à un individu est égale à 1). Seuls ces individus peuvent donc être utilisés dans le processus de maximisation

de la vraisemblance.

Alors :

$$\begin{aligned}
 L_C(\boldsymbol{\delta}/\mathbf{x}, \mathbf{s}) &= \prod_{\substack{n=1 \\ s_n \neq 0, s_n \neq J}}^N L_{Cn}(\boldsymbol{\delta}/\mathbf{x}_n, s_n) \\
 &= \frac{\prod_{n=1/s_n \neq 0, s_n \neq J}^N \exp\left(-\sum_{j=1}^J x_{nj} \delta_j\right)}{\prod_{s=1}^{J-1} [\gamma_s(\boldsymbol{\delta})]^{n_s}} \quad (2.11)
 \end{aligned}$$

avec n_s le nombre d'individus tels que $S_n = s$. On montre que

$$\frac{d\gamma_s(\boldsymbol{\delta})}{d\delta_j} = -\exp(-\delta_j) \gamma_{s-1}(\boldsymbol{\delta}_j^*) \quad (2.12)$$

Avec $\boldsymbol{\delta}_j^* = (\delta_1, \dots, \delta_{j-1}, \delta_{j+1}, \dots, \delta_J)$. Donc en dérivant $L_C(\boldsymbol{\delta}/\mathbf{x}_n, \mathbf{s})$ par rapport aux δ_j , on obtient les équations du maximum de vraisemblance conditionnelle :

$$t_j = \sum_{n=1/s_n \neq 0, s_n \neq J}^N \frac{\exp(-\delta_j) \gamma_{s_n-1}(\boldsymbol{\delta}_j^*)}{\gamma_{s_n}(\boldsymbol{\delta})} = \sum_{s=1}^{J-1} n_s \frac{\exp(-\delta_j) \gamma_{s-1}(\boldsymbol{\delta}_j^*)}{\gamma_s(\boldsymbol{\delta})}, \quad j = 1, \dots, J \quad (2.13)$$

Les estimations obtenues sont finies et uniques dès lors que la matrice des réponses est bien conditionnée [38]. De plus, les estimations sont consistantes quand $N \rightarrow +\infty$ à J fixé [7] [122].

2.6.8.4 La méthode du maximum de vraisemblance marginale

La méthode du maximum de vraisemblance marginale (Marginal Maximum Likelihood - MML) permet d'estimer les paramètres du modèle de Rasch lorsque le trait latent est considéré comme une variable aléatoire. La vraisemblance marginale est obtenue en intégrant la vraisemblance par la fonction de distribution du trait latent $G(\Theta/\boldsymbol{\nu})$, où $\boldsymbol{\nu}$ est un ensemble de paramètres définissant la fonction $G(\cdot)$.

$$\begin{aligned}
 L_M(\boldsymbol{\delta}, \boldsymbol{\nu}/\mathbf{x}) &= \prod_{n=1}^N \int_{\mathbb{R}} \prod_{j=1}^J P(X_{nj} = x_{nj}/\theta, \delta_j) G(\theta/\boldsymbol{\nu}) d\theta \\
 &= \prod_{n=1}^N L_{Mn}(\boldsymbol{\delta}, \boldsymbol{\nu}/\mathbf{x}_n) \quad (2.14)
 \end{aligned}$$

On peut alors estimer de façon consistante et conjointement les paramètres de difficulté des items et les paramètres $\boldsymbol{\nu}$. Il est courant de supposer la fonction de distribution du trait latent comme une loi normale centrée de variance σ^2 [113]. Cela résulte sans doute du fait que le modèle de Rasch à effets mixtes et un GLMM et que le choix de la distribution multinormale pour les effets aléatoires est très fréquent dans ce domaine (voir quasi-systématique) car ce choix permet le rapprochement entre les GLMM et les GLM hétéroscédastiques [33]. Dans le cas du modèle de Rasch à effets mixtes, il est en outre commode de centrer la distribution de

la variable aléatoire, ce qui évite des problème d'indétermination (voir section 2.6.8.1).

L'estimation se fait à l'aide d'un algorithme de maximisation classique (par exemple Newton-Raphson). A chaque étape, les intégrales sont estimées à l'aide de quadratures de Gauss-Hermite. Ces quadratures permettent d'estimer des intégrales du type :

$$\int_{\mathbb{R}} f(x) \exp(-x^2) dx \approx \sum_{u=1}^K w_u f(A_u) \quad (2.15)$$

Où les u sont les indices des K noeuds de quadratures A_u pondérés par les poids w_u . Ces quadratures sont donc employées quand la variable d'intégration est distribuée normalement.

Les noeuds A_u , $u = 1, \dots, K$ sont les K racines des polynômes orthogonaux d'Hermite de degré K [28] définis par [51] :

$$P_K(x) = (-1)^K \exp(x^2) \frac{d^K \exp(-x^2)}{dx^K} \quad (2.16)$$

Le poids w_u associé au noeud A_u vaut alors [51] :

$$w_u = \frac{(K-1)! \sqrt{2\pi}}{K [P_{K-1}(A_u)]^2} \quad (2.17)$$

Des tables existent pour retrouver ces noeuds et poids associés [28] : voir par exemple l'ouvrage de Stroud et Secrest [141].

On note que lorsque la variable latente est distribuée selon une loi normale centrée de variance σ^2 , on procède à un changement de variable. Alors on utilise les noeuds transformés $A_u^* = \sqrt{2}\sigma A_u$, et en posant $f_n(\mathbf{x}_n/y, \boldsymbol{\delta}) = \prod_{j=1}^J P(X_{nj} = x_{nj}/y, \delta_j)$, on obtient :

$$\begin{aligned} L_{Mn}(\boldsymbol{\delta}, \sigma^2/\mathbf{x}) &\approx \frac{1}{\sqrt{\pi}} \sum_{u=1}^K w_u f_n(\mathbf{x}_n/A_u^*, \boldsymbol{\delta}) \\ &\approx \frac{1}{\sqrt{\pi}} \sum_{u=1}^K w_u f_n(\mathbf{x}_n/\sqrt{2}\sigma A_u, \boldsymbol{\delta}) \end{aligned} \quad (2.18)$$

Ainsi, on peut estimer la log-vraisemblance $l_M(\boldsymbol{\delta}, \sigma^2/\mathbf{x}) = \log L_M(\boldsymbol{\delta}, \sigma^2/\mathbf{x})$ par

$$\begin{aligned} l_M(\boldsymbol{\delta}, \boldsymbol{\nu}/\mathbf{x}) &\approx \log \prod_{n=1}^N \frac{1}{\sqrt{\pi}} \sum_{u=1}^K w_u f_n(\mathbf{x}_n/\sqrt{2}\sigma A_u, \boldsymbol{\delta}) \\ &\approx -\frac{N}{2} \log \pi + \sum_{n=1}^N \log \sum_{u=1}^K w_u f_n(\mathbf{x}_n/\sqrt{2}\sigma A_u, \boldsymbol{\delta}) \end{aligned} \quad (2.19)$$

On obtient alors les equations du maximum de vraisemblance marginale par

$$\begin{cases} \frac{dl_M(\boldsymbol{\delta}, \boldsymbol{\nu}/\mathbf{x})}{d\delta_j} \approx \sum_{n=1}^N \sum_{u=1}^K c_{nu}(\boldsymbol{\delta}, \sigma^2) \frac{df_n(\mathbf{x}_n/\sqrt{2}\sigma A_u, \boldsymbol{\delta})}{d\delta_j} = 0 \quad \forall j = 1, \dots, J \\ \frac{dl_M(\boldsymbol{\delta}, \boldsymbol{\nu}/\mathbf{x})}{d\sigma^2} \approx \sum_{n=1}^N \sum_{u=1}^K c_{nu}(\boldsymbol{\delta}, \sigma^2) \frac{df_n(\mathbf{x}_n/\sqrt{2}\sigma A_u, \boldsymbol{\delta})}{d\sigma^2} = 0 \end{cases} \quad (2.20)$$

avec $c_{nu}(\boldsymbol{\delta}, \sigma^2) = \frac{w_u}{\sum_{u=1}^K w_u f_n(\mathbf{x}_n / \sqrt{2}\sigma A_u, \boldsymbol{\delta})}$.

D'autres auteurs comme De Leeuw et Verhelst [26] ont proposé de ne pas définir de façon paramétrique la fonction $G(\cdot)$ (modèle semi paramétrique), ce qui est asymptotiquement équivalent à la méthode du maximum de vraisemblance conditionnelle [41].

Enfin, Zwinderman [156] a montré que les estimations par MML peuvent être biaisées si la fonction $G(\cdot)$ est mal spécifiée.

2.6.8.5 Les équations d'estimations généralisées - GEE

Les équations d'estimations généralisées (dénotées GEE) ont été développées en 1986 par Liang et Zeger [92] dans un cadre longitudinal. L'application des GEE dans le cadre des modèles linéaires généralisés nécessite l'estimation des deux premiers moments de la variable réponse. Dans le cadre de modèles à effets aléatoires, ces moments peuvent être approximés [154]. Cependant, pour estimer les paramètres de la distribution des effets aléatoires, il faut pouvoir disposer de ces moments jusqu'à l'ordre 4. Sutradhar et Rao [142] proposent des approximations de ces moments. Feddag, Grama et Mesbah [35] montrent que les estimations des paramètres du modèle de Rasch obtenus avec seulement les deux premiers moments sont de bonne qualité comparées à celles obtenues avec les quatre premiers moments. C'est cette approche qui est présentée par la suite.

Dans cette approche proposée par Feddag, Grama et Mesbah [35], [34], la vraisemblance marginale est linéarisée par une approximation de Taylor. Il est alors facile d'intégrer cette expression par rapport à la densité du trait latent (supposée normale), puis d'estimer les premiers moments des variables réponses [142].

a- Approximation de la vraisemblance et estimation des moments

Posons $L_n(\theta_n, \boldsymbol{\delta} / \mathbf{x}_n)$ la vraisemblance associée à l'individu n :

$$L_n(\theta_n, \boldsymbol{\delta} / \mathbf{x}_n) = \prod_{j=1}^J P(X_{nj} = x_{nj} / \theta_n, \delta_j) = \prod_{j=1}^J \frac{\exp [x_{nj} (\theta_n - \delta_j)]}{1 + \exp (\theta_n - \delta_j)} \quad (2.21)$$

On peut approximer $P(X_{nj} = x_{nj} / \theta_n, \delta_j)$ par une approximation de Taylor de degrés 4 autour de $\theta_n = 0$:

$$\begin{aligned} P(X_{nj} = x_{nj} / \theta_n, \delta_j) &= P(X_{nj} = x_{nj} / 0, \delta_j) + \left. \frac{dP(X_{nj} = x_{nj} / \theta_n, \delta_j)}{d\theta_n} \right|_{\theta_n=0} \theta_n \\ &+ \frac{1}{2} \left. \frac{d^2 P(X_{nj} = x_{nj} / \theta_n, \delta_j)}{d\theta_n^2} \right|_{\theta_n=0} \theta_n^2 \\ &+ \frac{1}{6} \left. \frac{d^3 P(X_{nj} = x_{nj} / \theta_n, \delta_j)}{d\theta_n^3} \right|_{\theta_n=0} \theta_n^3 \\ &+ \frac{1}{24} \left. \frac{d^4 P(X_{nj} = x_{nj} / \theta_n, \delta_j)}{d\theta_n^4} \right|_{\theta_n=0} \theta_n^4 + O(|\theta_n|^5) \end{aligned} \quad (2.22)$$

Notons $l_{nj}^{(0)}$ la fonction :

$$l_{nj}^{(0)} = \frac{\exp(\theta_n - \delta_j)}{1 + \exp(\theta_n - \delta_j)} \quad (2.23)$$

Et notons $l_{nj}^{(k)}$ la k^e dérivée de cette fonction par rapport à θ_n :

$$l_{nj}^{(k)} = \frac{d^k l_{nj}^{(0)}}{d\theta_{nj}^k} \quad (2.24)$$

Et $l_{nj}^{*(k)}$ la valeur de $l_{nj}^{(k)}$ en $\theta_n = 0$:

$$l_{nj}^{*(k)} = \left. \frac{d^k l_{nj}^{(0)}}{d\theta_{nj}^k} \right|_{\theta_n=0} \quad (2.25)$$

Notons $h_{nj} = P(X_{nj} = x_{nj}/\theta_n, \delta_j)$. Par dérivations successives, on montre que :

$$\frac{dP(X_{nj} = x_{nj}/\theta_n, \delta_j)}{d\theta_n} = h_{nj} A_{nj} \quad (2.26)$$

$$\frac{d^2 P(X_{nj} = x_{nj}/\theta_n, \delta_j)}{d\theta_n^2} = h_{nj} Q_{nj} \quad (2.27)$$

$$\frac{d^3 P(X_{nj} = x_{nj}/\theta_n, \delta_j)}{d\theta_n^3} = h_{nj} R_{nj} \quad (2.28)$$

$$\frac{d^4 P(X_{nj} = x_{nj}/\theta_n, \delta_j)}{d\theta_n^4} = h_{nj} S_{nj} \quad (2.29)$$

Avec $A_{nj} = x_{nj} - l_{nj}^{(0)}$, $C_{nj} = -l_{nj}^{(1)}$, $D_{nj} = -l_{nj}^{(2)}$, $E_{nj} = -l_{nj}^{(3)}$

$$Q_{nj} = A_{nj}^2 - C_{nj}$$

$$R_{nj} = A_{nj}^3 - 3A_{nj}C_{nj} - D_{nj}$$

$$S_{nj} = A_{nj}^4 - 6A_{nj}^2C_{nj} - A_{nj}D_{nj} + 3C_{nj}^2 - E_{nj}$$

Soient h_{nj}^* , A_{nj}^* , C_{nj}^* , D_{nj}^* , E_{nj}^* , Q_{nj}^* , R_{nj}^* , S_{nj}^* les valeurs respectives de h_{nj} , A_{nj} , C_{nj} , D_{nj} , E_{nj} , Q_{nj} , R_{nj} , S_{nj} pour $\theta_n = 0$, alors :

$$P(X_{nj} = x_{nj}/\theta_n, \delta_j) = h_{nj}^* \left\{ 1 + A_{nj}^* \theta_n + \frac{1}{2} Q_{nj}^* \theta_n^2 + \frac{1}{6} R_{nj}^* \theta_n^3 + \frac{1}{24} S_{nj}^* \theta_n^4 \right\} + O(|\theta_n|^5) \quad (2.30)$$

Cette approximation est une fonction linéaire du trait latent θ_n . En intégrant sur la distribution du trait latent (supposée être une loi normale centrée de variance σ^2), on obtient :

$$P(X_{nj} = x_{nj}/\sigma^2, \delta_j) = h_{nj}^* \int_{\mathbb{R}} \left\{ 1 + A_{nj}^* \theta + \frac{1}{2} Q_{nj}^* \theta^2 + \frac{1}{6} R_{nj}^* \theta^3 + \frac{1}{24} S_{nj}^* \theta^4 \right\} G(\theta/\sigma^2) d\theta + O(|\theta_n|^5) \quad (2.31)$$

Or, l'on sait que si Y est une fonction aléatoire distribuée selon une loi normale centrée de variance σ^2 notée $G(Y/\sigma^2)$, alors :

$$\int_{\mathbb{R}} Y G(Y/\sigma^2) dY = 0 \quad (2.32)$$

$$\int_{\mathbb{R}} Y^2 G(Y/\sigma^2) dY = \sigma^2 \quad (2.33)$$

$$\int_{\mathbb{R}} Y^3 G(Y/\sigma^2) dY = 0 \quad (2.34)$$

$$\int_{\mathbb{R}} Y^4 G(Y/\sigma^2) dY = 3\sigma^4 \quad (2.35)$$

Aussi

$$P(X_{nj} = x_{nj}/\sigma^2, \delta_j) = h_{nj}^* \left\{ 1 + \frac{\sigma^2}{2} Q_{nj}^* + \frac{\sigma^4}{8} S_{nj}^* \right\} + O(|\theta_n|^5) \quad (2.36)$$

Et, en utilisant l'hypothèse d'indépendance locale :

$$P(X_{nj} = x_{nj}, X_{nk} = x_{nk}/\sigma^2, \delta_j, \delta_k) = P(X_{nj} = x_{nj}/\sigma^2, \delta_j) P(X_{nk} = x_{nk}/\sigma^2, \delta_k) \quad (2.37)$$

Si $x_{nj} = 1$, on montre que $h_{nj}^* = l_j^{*(1)}$, $A_{nj}^* = \frac{l_j^{*(2)}}{l_j^{*(1)}}$, $Q_{nj}^* = \frac{l_j^{*(3)}}{l_j^{*(1)}}$, $R_{nj}^* = \frac{l_j^{*(4)}}{l_j^{*(1)}}$ et $S_{nj}^* = \frac{l_j^{*(5)}}{l_j^{*(1)}}$

Donc les moments d'ordre 1 valent :

$$\begin{aligned} \mu_j &= P(X_{nj} = 1/\sigma^2, \delta_j) \\ &= l_j^{*(1)} + \frac{\sigma^2}{2} l_j^{*(3)} + \frac{\sigma^4}{8} l_j^{*(5)} + O(|\theta_n|^5) \end{aligned} \quad (2.38)$$

Et les moments d'ordre 2 valent :

$$\begin{aligned} \sigma_{jj} &= \mu_j(1 - \mu_j) \\ &= l_j^{*(2)} + \frac{\sigma^2}{2} \left(l_j^{*(3)} - 2l_j^{*(1)} l_j^{*(3)} \right) + \frac{\sigma^4}{8} \left(l_j^{*(5)} - 2l_j^{*(1)} l_j^{*(5)} - 2l_j^{*(3)} l_j^{*(3)} \right) + O(|\theta_n|^5) \end{aligned} \quad (2.39)$$

$$\begin{aligned} \sigma_{jk} &= P(X_{nj} = 1, X_{nk} = 1/\sigma^2, \delta_j, \delta_k) - \mu_j \mu_k \\ &= \sigma^2 l_j^{*(2)} l_k^{*(2)} + \frac{\sigma^4}{2} \left(l_j^{*(2)} l_k^{*(4)} + l_j^{*(3)} l_k^{*(3)} + l_j^{*(4)} l_k^{*(2)} \right) + O(|\theta_n|^5) \end{aligned} \quad (2.40)$$

b- Estimation par GEE

Soient $\boldsymbol{\mu} = (\mu_j)_{j=1, \dots, J}$ l'espérance du vecteur \mathbf{X}_n , $s_{n,jl} = (x_{nj} - \mu_j)(x_{nl} - \mu_l)$ le moment empirique d'ordre 2 de \mathbf{X}_n , $\mathbf{s}_n = (s_{n,jl})'_{1 \leq j < l \leq J}$ et $\boldsymbol{\eta} = E(\mathbf{s}_n) = (\sigma_{jl})_{1 \leq j < l \leq J}$.

Posons $\boldsymbol{\xi}_n = (x_n, s_n)'$, $n = 1, \dots, N$ le vecteur des observations dont l'espérance et la variance sont données respectivement par $\boldsymbol{\phi} = (\boldsymbol{\mu}, \boldsymbol{\eta})'$ et $\mathbf{V} = \begin{pmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{12} & \mathbf{V}_{22} \end{pmatrix}$ où $\mathbf{V}_{11} = \text{Var}(\mathbf{X}_n) = (\sigma_{jl})_{j,l=1, \dots, J}$, $\mathbf{V}_{12} = \text{Cov}(\mathbf{X}_n, \mathbf{S}_n)$ et $\mathbf{V}_{22} = \text{Var}(\mathbf{S}_n)$.

Alors les équations d'estimations généralisées (GEE) pour les différents moments du modèle sont de la forme :

$$\mathbf{D}' \mathbf{V}^{-1} \sum_{n=1}^N (\boldsymbol{\xi}_n - \boldsymbol{\phi}) = \mathbf{0}, \quad (2.41)$$

avec

$$\mathbf{D} = \begin{pmatrix} \frac{d\boldsymbol{\mu}}{d\boldsymbol{\delta}} & \frac{d\boldsymbol{\mu}}{d\sigma^2} \\ \frac{d\boldsymbol{\eta}}{d\boldsymbol{\delta}} & \frac{d\boldsymbol{\eta}}{d\sigma^2} \end{pmatrix} \quad (2.42)$$

Les GEE sont en général résolues par un algorithme de Fisher-scoring, ce qui permet d'obtenir l'estimation des paramètres $(\boldsymbol{\delta}, \sigma^2)$. A l'étape $k + 1$, ces estimations sont calculées par :

$$\begin{pmatrix} \hat{\boldsymbol{\delta}}^{(k+1)} \\ \hat{\sigma}^{2(k+1)} \end{pmatrix} = \begin{pmatrix} \hat{\boldsymbol{\delta}}^{(k)} \\ \hat{\sigma}^{2(k)} \end{pmatrix} + \frac{1}{N} \left(\hat{\mathbf{D}}^{(k)'} \hat{\mathbf{V}}^{(k)-1} \hat{\mathbf{D}}^{(k)} \right)^{-1} \hat{\mathbf{D}}^{(k)'} \hat{\mathbf{V}}^{(k)-1} \sum_{n=1}^N \left(\hat{\boldsymbol{\xi}}_n^{(k)} - \hat{\boldsymbol{\phi}}^{(k)} \right) \quad (2.43)$$

où $\hat{\mathbf{D}}^{(k)}$, $\hat{\mathbf{V}}^{(k)}$ et $\hat{\boldsymbol{\phi}}^{(k)}$ sont les valeurs respectives de \mathbf{D} , \mathbf{V} et $\boldsymbol{\phi}$ au point $(\hat{\boldsymbol{\delta}}^{(k)}, \hat{\sigma}^{2(k)})'$.

La matrice de variance-covariance \mathbf{W} des estimations $(\boldsymbol{\delta}, \sigma^2)'$ est obtenue par l'estimateur sandwich [33] :

$$\mathbf{W} = \frac{1}{N^2} \left(\hat{\mathbf{D}} \hat{\mathbf{V}}^{-1} \hat{\mathbf{D}} \right)^{-1} \left(\hat{\mathbf{D}} \hat{\mathbf{V}}^{-1} \sum_{n=1}^N \left(\hat{\boldsymbol{\xi}}_n - \hat{\boldsymbol{\phi}} \right) \left(\hat{\boldsymbol{\xi}}_n' - \hat{\boldsymbol{\phi}}' \right) \hat{\mathbf{V}}^{-1} \hat{\mathbf{D}} \right) \left(\hat{\mathbf{D}} \hat{\mathbf{V}}^{-1} \hat{\mathbf{D}} \right)^{-1} \quad (2.44)$$

où $\hat{\mathbf{D}}$, $\hat{\mathbf{V}}$ et $\hat{\boldsymbol{\phi}}$ sont les valeurs respectives de \mathbf{D} , \mathbf{V} et $\boldsymbol{\phi}$ au point $(\hat{\boldsymbol{\delta}}, \hat{\sigma}^2)'$.

2.6.8.6 Autres méthodes : algorithmes EM, SEM

Cette revue des méthodes d'estimations sous le modèle de Rasch ne serait pas complète sans parler de l'algorithme EM (Expectation-Maximisation) proposé par Dempster, Laird et Rubin [27] et implémenté, entre autre, dans le logiciel RSP [46] pour le modèle de Rasch à effets aléatoires. Cet algorithme ne résoud cependant pas le problème primordial de la méthode du maximum de vraisemblance marginale, à savoir l'estimation d'intégrales (par quadratures de Gauss-Hermite) à chaque étape de l'algorithme. L'estimation se fait par exemple avec un algorithme de Newton-Raphson [46]. Hamon [54] [55] propose d'utiliser un algorithme SEM (Stochastic Expectation and Maximisation) afin de contourner ce problème.

2.6.8.7 Estimation des valeurs individuelles du trait latent

Suivant la méthode d'estimation utilisée, les valeurs individuelles du trait latent peuvent être obtenues. Dans le cadre du JML, ces estimations sont directement obtenues. Dans les autres méthodes, on peut les estimer conditionnellement aux estimations des paramètres de difficulté des items obtenues.

Lorsque le trait latent est considéré comme un ensemble de paramètres fixes, on maximise alors la quantité [73] :

$$L_{MLE}(\theta_n / \hat{\boldsymbol{\delta}}, \mathbf{x}_n) = \prod_{j=1}^J P(X_{nj} = x_{nj} / \theta_n, \hat{\delta}_j), \quad n = 1, \dots, N \quad (2.45)$$

avec $\hat{\delta}_j$ l'estimateur de δ_j , $j = 1, \dots, J$ obtenu par CML.

Néanmoins, ces estimations, appelées estimateurs du maximum de vraisemblance (Maximum Likelihood Estimators - MLE) sont biaisées, en particulier si J est petit [73]. De plus, il n'est pas possible d'estimer ces valeurs pour les individus ayant un score nul ou parfait. Il est possible, par analogie avec le modèle de Rasch à effet aléatoire, de considérer une sorte de distribution du trait latent notée $g(\theta)$, en posant :

$$g(\theta) = \sqrt{\sum_{j=1}^J \left(\frac{\exp(\theta - \hat{\delta}_j)}{1 + \exp(\theta - \hat{\delta}_j)} \right) \left(1 - \frac{\exp(\theta - \hat{\delta}_j)}{1 + \exp(\theta - \hat{\delta}_j)} \right)} \quad (2.46)$$

En maximisant la quantité

$$L_{WLE}(\theta_n / \hat{\boldsymbol{\delta}}, \mathbf{x}_n) = \prod_{j=1}^J P(X_{nj} = x_{nj} / \theta_n; \hat{\delta}_j) g(\theta_n) \quad (2.47)$$

on obtient les estimations du maximum de vraisemblance pondérées (Weighted Likelihood Estimator - WLE), qui sont non biaisées et permettent d'obtenir des estimations pour le trait latent même pour les scores nuls ou parfaits [73].

Lorsque le trait latent est considéré comme une variable aléatoire, on peut maximiser la quantité précédente (estimation classique), ou bien utiliser l'estimation modale de Bayes (Bayes Modal Estimators - BME) en utilisant la loi *a priori* du trait latent $G(\Theta / \hat{\sigma}^2)$ et en maximisant (voir [73] [33] [77]) :

$$L_{BME}(\theta_n / \hat{\boldsymbol{\delta}}, \hat{\sigma}^2, \mathbf{x}_n) = \prod_{j=1}^J P(X_{nj} = x_{nj} / \theta_n; \hat{\delta}_j) G(\theta_n / \hat{\sigma}^2) \quad (2.48)$$

avec $\hat{\delta}_j$ l'estimateur de δ_j , $j = 1, \dots, J$ et $\hat{\sigma}^2$ l'estimateur de σ^2 obtenus par MML ou GEE. Notons que ce sont les estimations obtenues avec la procédure NLMIXED de SAS [77].

Une autre façon d'estimer les valeurs du trait latent lorsque celui-ci est considéré comme une variable aléatoire est d'utiliser les estimateurs attendus *a posteriori* de Bayes (Bayes Expected A Posteriori Estimators - EAP), en maximisant les quantités [73] :

$$L_{EAP}(\theta_n / \hat{\boldsymbol{\delta}}, \hat{\sigma}^2, \mathbf{x}_n) = \frac{\int_{\mathbb{R}} \theta_n \prod_{j=1}^J \frac{\exp(x_{nj}(\theta_n - \hat{\delta}_j))}{\exp(\theta_n - \hat{\delta}_j)} G(\theta / \hat{\sigma}^2) d\theta_n}{\int_{\mathbb{R}} \prod_{j=1}^J \frac{\exp(x_{nj}(\theta_n - \hat{\delta}_j))}{\exp(\theta_n - \hat{\delta}_j)} G(\theta / \hat{\sigma}^2) d\theta_n} \quad (2.49)$$

Notons que ces estimations sont celles obtenues avec le module **gllapred** de Stata [123].

2.6.9 Tester le modèle de Rasch

Divers tests ont été proposés pour tester l'adéquation de données à un modèle de Rasch [48], ou pour tester la vérification de certaines hypothèses de l'IRT. Concernant les tests d'adéquation, on peut les regrouper en trois grandes familles : les tests globaux, les tests du premier ordre et les tests du second ordre. Si l'ensemble de ces tests permettent de tester l'adéquation au modèle de Rasch, cette typologie repose sur le fait que certains tests sont plus

sensibles que d'autre à la violation de certaines propriétés du modèle de Rasch. Nous avons en outre intégré à cette partie les indices OUTFIT et INFIT qui permettent de détecter des individus ou items abérants, le test U qui permet de tester l'égalité des pentes des ICC des items et le split test qui permet de vérifier la propriété d'objectivité spécifique.

2.6.9.1 Les tests globaux

Les tests globaux sont les tests qui permettent de tester l'adéquation globale des données à un modèle de Rasch. Ces tests sont en outre sensibles au fait que les estimations des paramètres soient différents dans différents sous-échantillons. Le principal test global est connu sous le nom de test Z du rapport de vraisemblance d'Andersen [8]. Il repose sur le fait que les estimations des paramètres de difficulté du modèle de Rasch doivent être similaires pour des populations ayant des niveaux différents sur le trait latent (principe d'objectivité spécifique). La population est alors séparée en fonction de la valeur du score s en G groupes (un groupe correspondant à un intervalle donnée du score, hors score nul ou parfait). La statistique de test est alors :

$$Z = -2 \log \left(\max_{\boldsymbol{\delta}} L_C(\boldsymbol{\delta}/\mathbf{x}, \mathbf{s}) \right) + 2 \log \left(\sum_{g=1}^G \max_{\boldsymbol{\delta}_g} L_C(\boldsymbol{\delta}_g/\mathbf{x}_g) \right) \quad (2.50)$$

avec $L_C(\boldsymbol{\delta}/\mathbf{x}, \mathbf{s})$ définie par l'équation (2.11) et $L_C(\boldsymbol{\delta}_g/\mathbf{x}_g)$ la vraisemblance conditionnelle obtenue avec les individus du groupe g .

Cette statistique suit sous l'hypothèse nulle d'une bonne adéquation au modèle de Rasch une loi de Chi-deux avec $(J-1)(G-1)$ degrés de liberté [8].

Martin-Löf [97] propose un autre test en scindant l'ensemble des J items en deux groupes de J_1 et J_2 items. La vraisemblance conditionnelle estimée dans le groupe k est notée $L_C^{(k)}(\boldsymbol{\delta}_k/\mathbf{x}_k)$. Le score obtenu par l'individu n pour les items du groupe k est noté s_{nk} . n_s est le nombre d'individus tels que $s_n = k$ et $n_{s_1 s_2}$ est le nombre d'individus tels que $s_{n1} = s_1$ et $s_{n2} = s_2$. Ainsi la statistique de test vaut :

$$\begin{aligned} ML = & 2 \left(\sum_{s_1=0}^{J_1} \sum_{s_2=0}^{J_2} n_{s_1 s_2} \log \left(\frac{n_{s_1 s_2}}{N} \right) - \sum_{s=0}^J n_s \left(\frac{n_s}{N} \right) - \log(L_C(\boldsymbol{\delta}/\mathbf{x}, \mathbf{s})) \right. \\ & \left. + \log \left(L_C^{(1)}(\boldsymbol{\delta}_1/\mathbf{x}_1) \right) + \log \left(L_C^{(2)}(\boldsymbol{\delta}_2/\mathbf{x}_2) \right) \right) \end{aligned} \quad (2.51)$$

Cette statistique suit sous l'hypothèse nulle une loi de Chi-deux avec $(J_1 J_2 - 1)$ degrés de liberté [97].

Outre ces deux tests globaux les plus utilisés dans le cadre du modèle de Rasch, Glas et Verhelst [48] font une revue de l'ensemble des tests utilisés sous le modèle de Rasch.

2.6.9.2 Les tests du premier ordre

Les tests du premier ordre sont une famille de tests permettant de tester l'adéquation des données à un modèle de Rasch et qui sont particulièrement sensibles aux violations de la monotonie, de "parallélisme" des ICC et de l'exhaustivité du score sur le trait latent. Les

tests les plus connus dans cette famille sont les tests Y (Wright-Panchapakesan), R_{1C} , R_{1M} et Q_1 (van den Wollenberg). Ces tests comparent, pour chaque valeur du score, les nombres attendus et observés d'individus ayant répondu positivement à chacun des items.

On note n_{sj} le nombre d'individus ayant un score $s_n = s$ et ayant répondu positivement à l'item j ($X_{nj} = 1$) et \hat{n}_{sj} ce nombre attendu sous l'hypothèse du modèle de Rasch. Les scores $s = 1, \dots, J - 1$ sont regroupés en G groupes avec I_g l'ensemble des scores s du groupe g , $g = 1, \dots, J - 1$. Soient $d_{sj} = n_{sj} - \hat{n}_{sj}$, $\mathbf{d}_{1s} = (d_{sj})_{j=1, \dots, J}$ et $\mathbf{d}_{1g} = (\sum_{s \in I_g} d_{sj})_{j=1, \dots, J}$.

Wright et Panchapakesan [151] utilisent les approximations $\hat{n}_{sj} = n_s \hat{\pi}_{sj}$ avec $\hat{\pi}_{sj} = P(X_{nj} = 1 / \hat{\theta}_n, S_n = s, \hat{\boldsymbol{\delta}})$, et posent :

$$y_{gj} = \frac{\sum_{s \in I_g} d_{sj}}{\sqrt{\sum_{s \in I_g} n_s \hat{\pi}_{sj} (1 - \hat{\pi}_{sj})}}, \quad (2.52)$$

puis proposent les statistiques :

$$Y_j = \sum_{g=1}^G y_{gj}^2, \quad (2.53)$$

qui suivent sous l'hypothèse de monotonie de la fonction réponse de l'item j une loi du chi-deux avec $G - 1$ degrés de liberté. La statistique $Y = \sum_{j=1}^J Y_j$ suit sous l'hypothèse de monotonie de l'ensemble des IRF une loi du chi-deux à $(J - 1)(G - 1)$ degrés de liberté. Van den Wollenberg [145] souligne cependant des erreurs de constructions dans les statistiques Y dues à l'estimation des π_{sj} .

Martin-Löf [97] et Glas [45] proposent des statistiques équivalentes. Nous détaillerons ici celles de Glas dénommées statistiques R_{1C} . Comparées au test Y , les π_{sj} sont estimées par :

$$\hat{\pi}_{sj} = \frac{\exp(-\hat{\delta}_j) \gamma_{s-1}(\hat{\boldsymbol{\delta}}_j^*)}{\gamma_s(\hat{\boldsymbol{\delta}})}, \quad (2.54)$$

avec $\boldsymbol{\delta}_j^* = (\delta_l)_{l=1, \dots, j-1, j+1, \dots, J}$.

Démonstration.

Posons $I = 1, \dots, j, \dots, J$ l'ensemble des items, $I^* = 1, \dots, j - 1, j + 1, \dots, J$, $\mathbf{X}_{\mathbf{n}}^* = (X_{nl})_{l \in I^*}$, $\mathbf{x}_{\mathbf{n}}^* = (x_{nl})_{l \in I^*}$ et $s_n^* = \sum_{l \in I^*} x_{nl}$. D'après l'équation (2.9), on a

$$P(\mathbf{X}_{\mathbf{n}} = \mathbf{x}_{\mathbf{n}} / \theta_n, \boldsymbol{\delta}) = \frac{\exp(-\sum_{l \in I} x_{nl} \delta_l)}{\gamma_{s_n}(\boldsymbol{\delta})} \quad (2.55)$$

Or

$$\begin{aligned} \pi_{sj} &= P(X_{nj} = 1 / \mathbf{X}_{\mathbf{n}}^* = \mathbf{x}_{\mathbf{n}}^*, \theta_n, \boldsymbol{\delta}, S_n = s) \\ &= \frac{P(X_{nj} = 1, \mathbf{X}_{\mathbf{n}}^* = \mathbf{x}_{\mathbf{n}}^* / \theta_n, \boldsymbol{\delta}, S_n = s)}{P(\mathbf{X}_{\mathbf{n}}^* = \mathbf{x}_{\mathbf{n}}^* / \theta_n, \boldsymbol{\delta}_j^*, S_n^* = s - 1)} \\ &= \frac{\exp(-\sum_{l \in I} x_{nl} \delta_l)}{\gamma_s(\boldsymbol{\delta})} \left[\frac{\exp(-\sum_{l \in I^*} x_{nl} \delta_l)}{\gamma_{s-1}(\boldsymbol{\delta}^*)} \right]^{-1} \\ &= \frac{\exp(-\delta_j) \gamma_{s-1}(\boldsymbol{\delta}_j^*)}{\gamma_s(\boldsymbol{\delta})} \end{aligned} \quad (2.56)$$

□

En outre, une matrice $\mathbf{V}_{1\mathbf{g}}$ de pondérations des $\mathbf{d}_{1\mathbf{g}}$ est utilisée. La statistique de test

$$R_{1Cg} = \mathbf{d}_{1\mathbf{g}}' \mathbf{V}_{1\mathbf{g}}^{-1} \mathbf{d}_{1\mathbf{g}} \quad (2.57)$$

suit sous l'hypothèse de monotonie des IRF une loi du chi-deux à $(J - 1)$ degrés de liberté, et la statistique $R_{1C} = \sum_{g=1}^G R_{1Cg}$ suit sous l'hypothèse de monotonie de l'ensemble des IRF une loi du chi-deux à $(J - 1)(G - 1)$ degrés de liberté.

Les pondérations utilisés dans les matrices $\mathbf{V}_{1\mathbf{g}}$ sont $e_{gjj} = \sum_{s \in I_g} n_s \hat{\pi}_{sj} = \sum_{s \in I_g} \hat{n}_{sj}$ pour les éléments diagonaux de la matrice et

$$e_{gjk} = \sum_{s \in I_g} n_s \frac{\exp(\hat{\delta}_j) \exp(\hat{\delta}_k) \gamma_{s-2}(\hat{\delta}_{j,k}^*)}{\gamma_s(\hat{\delta})} \quad (2.58)$$

pour les éléments non diagonaux avec $\hat{\delta}_{j,k}^* = (\delta_l)_{l=1, \dots, J, l \neq j, k \neq k}$.

La contribution de l'item j à la statistique R_{1C} peut être évaluée en calculant

$$R_{1Cj} = \sum_{g=1}^G \mathbf{V}_{\mathbf{g}}^{-1/2} \mathbf{d}_{\mathbf{g}} \quad (2.59)$$

avec $\mathbf{V}_{\mathbf{g}}^{-1/2}$ la décomposition de Cholesky de $\mathbf{V}_{\mathbf{g}}$ ($\mathbf{V}_{\mathbf{g}}^{-1/2'} \mathbf{V}_{\mathbf{g}}^{-1/2} = \mathbf{V}_{\mathbf{g}}$). Cette quantité suit, sous l'hypothèse d'une adéquation correcte de l'item j au modèle de Rasch une loi du chi-deux à $G - 1$ degrés de liberté.

En outre, une approximation des statistiques R_{1C} a été proposée par van den Wollenberg [146]. R_{1Cj} peut être approximée par Q_{1j} telle que :

$$Q_{1j} = \sum_{g=1}^G \left\{ \frac{\left[\sum_{s \in I_g} n_{sj} - \hat{n}_{sj} \right]^2}{\sum_{s \in I_g} \hat{n}_{sj}} + \frac{\left[\sum_{s \in I_g} n_{sj} - \hat{n}_{sj} \right]^2}{\sum_{s \in I_g} n_s - \hat{n}_{sj}} \right\} = \sum_{g=1}^G \frac{\sum_{s \in I_g} n_s d_{sj}^2}{\sum_{s \in I_g} \hat{n}_{sj} (n_s - \hat{n}_{sj})} \quad (2.60)$$

et R_{1C} est approximée par Q_1 :

$$Q_1 = \frac{J-1}{J} \sum_{j=1}^J Q_{1j} \quad (2.61)$$

La statistique R_{1C} a été étendue à la statistique R_{1M} dans le cadre du modèle de Rasch à effet aléatoire [45]. Dans ce cas, le nombre de réponses positives à l'item j attendus sous le modèle de Rasch est :

$$\hat{N}_{gj} = N \sum_{s \in I_g} \exp(\hat{\delta}_j) \gamma_{s-1}(\hat{\delta}_j^*) \int_{\mathbb{R}} \frac{\exp(s\theta)}{\prod_{j=1}^J (1 + \exp(\theta - \hat{\delta}_j))} G(\theta/\hat{\sigma}^2) d\theta \quad (2.62)$$

Les matrices de pondérations $\mathbf{V}_{\mathbf{g}}$ contiennent les éléments $e_{gjj} = \hat{N}_{gj}$ sur la diagonale et

$$\hat{e}_{gjk} = N \sum_{s \in I_g} \exp(\hat{\delta}_j) \exp(\hat{\delta}_k) \gamma_{s-1}(\hat{\delta}_{j,k}^*) \int_{\mathbb{R}} \frac{\exp(s\theta)}{\prod_{j=1}^J (1 + \exp(\theta - \hat{\delta}_j))} G(\theta/\hat{\sigma}^2) d\theta \quad (2.63)$$

en dehors des diagonales.

Posons

$$\hat{n}_0 = N \int_{\mathbb{R}} \frac{1}{\prod_{j=1}^J (1 + \exp(\theta - \hat{\delta}_j))} G(\theta/\hat{\sigma}^2) d\theta \quad (2.64)$$

et

$$\hat{n}_J = N \int_{\mathbb{R}} \frac{\exp(J\theta - \sum_{j=1}^J \hat{\delta}_j)}{\prod_{j=1}^J (1 + \exp(\theta - \hat{\delta}_j))} G(\theta/\hat{\sigma}^2) d\theta \quad (2.65)$$

Alors, la statistique R_{1M} vaut

$$R_{1M} = c_0 + \sum_{g=1}^G \mathbf{d}_{\mathbf{g}j}' \mathbf{V}_{\mathbf{1g}}^{-1} \mathbf{d}_{\mathbf{g}j} + c_J \quad (2.66)$$

avec $c_0 = \frac{(n_0 - \hat{n}_0)^2}{\hat{n}_0}$ et $c_J = \frac{(n_J - \hat{n}_J)^2}{\hat{n}_J}$. R_{1M} suit sous l'hypothèse de monotonie des items une loi du chi-deux à $(J-1)^2 - 1$ degrés de liberté.

2.6.9.3 Les tests du second ordre

Les tests du second ordre sont une famille de tests permettant de tester l'adéquation des données à un modèle de Rasch et qui sont sensibles à la violation des hypothèses d'unidimensionnalité du trait latent et d'indépendance locales des réponses aux items. Les tests les plus connus dans cette famille sont les tests R_{2C} , R_{2M} et Q_2 (van den Wollenberg). Ces tests comparent, pour chaque valeur du score, les nombres attendus et observés d'individus ayant répondu positivement à chaque paire possible d'items.

On note n_{1j} le nombre d'individus ayant un score $s_n = 1$ et ayant répondu positivement à l'item j ($X_{nj} = 1$) et \hat{n}_{1j} ce nombre attendu sous l'hypothèse du modèle de Rasch. Soient $d_{1j} = n_{1j} - \hat{n}_{1j}$, \mathbf{d}_1 le vecteur des $\frac{d_{1j}}{\sqrt{n}}$, $j = 1 \dots J$.

On note n_{2jl} le nombre d'items ayant un score $s_n > 1$ et ayant répondu positivement aux items j et l ($X_{nj} = 1$ et $X_{nl} = 1$) et \hat{n}_{2jl} ce nombre attendu sous l'hypothèse du modèle de Rasch. Soient $d_{2jl} = n_{2jl} - \hat{n}_{2jl}$, $\sqrt{\mathbf{d}_2}/\sqrt{n}$ le vecteur des d_{2jl} , $j = 1 \dots J$, $l = j + 1, \dots, J$. \mathbf{V}_1 et \mathbf{V}_2 sont deux matrices de pondération.

La statistique R_{2Cj} [113] est donnée par :

$$R_{2C} = \mathbf{d}_2' \mathbf{V}_2^{-1} \mathbf{d}_2 + \mathbf{d}_1' \mathbf{V}_1^{-1} \mathbf{d}_1 \quad (2.67)$$

Sous l'hypothèse d'unidimensionnalité, R_{2C} suit une loi du chi-deux à $\frac{J(J-1)}{2}$ degrés de liberté.

Cette statistique peut être étendue au modèle de Rasch à effets aléatoires [48]. La statistique Q_2 proposée par van den Wollenberg [146] est une estimation de R_{2C} . On note alors n_{sjl}

le nombre d'individus ayant un score $s_n = s$, $s > 2$ et ayant répondu positivement aux items j et l et \hat{n}_{sjl} l'estimation de ce nombre sous le modèle de Rasch :

$$\hat{n}_{sjl} = n_s \frac{\exp(-\hat{\delta}_j) \exp(-\hat{\delta}_l) \gamma_{s-2}(\hat{\boldsymbol{\delta}}_{j,l}^*)}{\gamma_s(\hat{\boldsymbol{\delta}})} \quad (2.68)$$

Soient $d^2 = (n_{sjl} - \hat{n}_{sjl})^2$ et $q_{2sjl} = \frac{d^2}{\hat{n}_{sjl}} + \frac{d^2}{\hat{n}_{sj} - \hat{n}_{sjl}} + \frac{d^2}{\hat{n}_{sl} - \hat{n}_{sjl}} + \frac{d^2}{n_s - \hat{n}_{sjl} - \hat{n}_{sj} + \hat{n}_{sjl} - \hat{n}_{sl} + \hat{n}_{sjl}}$, alors la statistique donnée par :

$$Q_2 = \frac{J-3}{J-1} \sum_{s=1}^{J-1} \sum_{j=1}^J \sum_{l=j+1}^J q_{2sjl} \quad (2.69)$$

suit une loi de chi-deux à $\frac{J(J-1)(J-3)}{2}$ degrés de liberté sous l'hypothèse d'unidimensionnalité .

Dans le cas de regroupement de scores en G groupes, les degrés de liberté de Q_2 deviennent $\frac{JG(J-3)}{2}$

2.6.9.4 Détecter les observations abérantes : les indices OUTFIT et INFIT

Deux indices [93] très utilisés permettent de détecter les observations abérantes : OUTFIT et INFIT.

Ces indices sont basées sur les résidus

$$r_{nj} = x_{nj} - \hat{\pi}_{s_n j} \quad (2.70)$$

Pour l'item j , ces deux indices sont calculés par

$$OUTFIT_j = \frac{1}{N} \sum_{n=1}^N \frac{r_{nj}^2}{\hat{\pi}_{s_n j} (1 - \hat{\pi}_{s_n j})} \quad (2.71)$$

$$INFIT_j = \frac{1}{N} \frac{\sum_{n=1}^N r_{nj}^2}{\sum_{n=1}^N \hat{\pi}_{s_n j} (1 - \hat{\pi}_{s_n j})} \quad (2.72)$$

Pour l'individu n , ces deux indices sont calculés par

$$OUTFIT_n = \frac{1}{N} \sum_{j=1}^J \frac{r_{nj}^2}{\hat{\pi}_{s_n j} (1 - \hat{\pi}_{s_n j})} \quad (2.73)$$

$$INFIT_n = \frac{1}{N} \frac{\sum_{j=1}^J r_{nj}^2}{\sum_{j=1}^J \hat{\pi}_{s_n j} (1 - \hat{\pi}_{s_n j})} \quad (2.74)$$

OUTFIT permet de détecter des observations abérantes du type un individu avec un score élevé qui n'a pas répondu positivement à un item facile (ou inversement, un individu avec un score faible qui a répondu positivement à un item difficile).

INFIT permet de détecter des observations ayant un profil de réponses incohérent, du type un individu avec un score moyen qui a répondu positivement aux items les plus difficiles et négativement aux items les plus faciles.

Les valeurs de ces deux indices doivent se trouver dans l'intervalle $[0.6; 1.4]$ (selon les sources, on trouve parfois des valeurs légèrement différentes).

2.6.9.5 Tester l'égalité des pentes des ICC : le test U

Afin de tester l'égalité des pentes des ICC, Molenaar [111] a développé le test U dans le cadre d'une estimation par CML.

L'échantillon est découpé en 3 parties en fonction du score des items. Une première partie concerne les individus avec un score au plus égal à un seuil c_1 , une seconde concerne ceux qui ont un score supérieur ou égal à c_2 . La troisième partie concerne les autres individus. On impose

$$\sum_{s=1}^{c_1} n_s > 25\%N \quad (2.75)$$

et

$$\sum_{s=c_2}^{J-1} n_s > 25\%N \quad (2.76)$$

La statistique U est égale pour l'item j ($j = 1, \dots, J$) à :

$$U_j = \frac{z_1 - z_2}{\sqrt{c_1 + J - c_2}} \quad (2.77)$$

avec

$$z_1 = \sum_{s=1}^{c_1} \frac{\pi_{sj} - \hat{\pi}_{sj}}{\sqrt{n_s \hat{\pi}_{sj} (1 - \hat{\pi}_{sj})}} \quad (2.78)$$

et

$$z_2 = \sum_{s=c_2}^{J-1} \frac{\pi_{sj} - \hat{\pi}_{sj}}{\sqrt{n_s \hat{\pi}_{sj} (1 - \hat{\pi}_{sj})}} \quad (2.79)$$

La statistique U_j suit sous l'hypothèse d'égalité de la pente de l'ICC de l'item j à la moyenne des pentes des autres items du modèle une loi normale centrée réduite : une statistique significativement négative signifie que la pente de l'item j est trop forte (plus forte que la moyenne des autres items) et inversement.

2.6.9.6 Le test de cission de l'échantillon (split test)

Ce test est proposé par Molenaar [111] afin de détecter les items qui ne vérifieraient pas le principe d'objectivité spécifique. L'échantillon est scindé en deux en fonction des réponses (positives ou négatives) à un item donné et dans chaque sous-échantillon, on estime par CML les paramètres de difficulté des items (on les centrera pour faciliter les comparaisons). On représente alors dans un graphique les estimations obtenues dans un des deux groupes en fonction de celles obtenues dans l'autre groupe.

Des écarts importants entre ces représentations et la diagonale montrent que le principe d'objectivité spécifique est bafoué. On s'intéressera notamment aux items ayant une difficulté plus importante pour le groupe ayant répondu positivement à l'item ayant servi à scinder l'échantillon.

Les différences entre les deux séries d'estimations peuvent être testées par un test de rapport de vraisemblance Z de Andersen (voir section 2.6.9.1).

2.7 Les autres modèles de l'IRT pour items dichotomiques

2.7.1 Le modèle logistique à un paramètre (One Parameter Logistic Model) (OPLM)

2.7.1.1 Définition

Le modèle logistique à un paramètre (OPLM) [40] est une extrapolation du modèle de Rasch où l'on ne considère plus les ICC comme non sécantes. Dans ce modèle, des paramètres $B_j > 0$, $j = 1, \dots, J$, connus (supposés) *a priori* permettent d'étendre le modèle de Rasch. Ces paramètres sont fixés subjectivement par des experts (en donnant plus ou moins d'importance à chaque item), ou peuvent être choisis comme des valeurs proches des estimations des pouvoirs discriminants obtenus dans un modèle logistique à deux paramètres (voir section 2.7.3).

Outre qu'il vérifie les propriétés de l'IRT, l'OPLM est défini par les IRF suivantes :

$$P(X_{nj} = x_{nj} / \theta_n, B_j, \delta_j) = \frac{\exp [B_j x_{nj} (\theta_n - \delta_j)]}{1 + \exp [B_j (\theta_n - \delta_j)]} \quad (2.80)$$

Les paramètres (poids) B_j sont appelés pouvoirs discriminants (ou pouvoirs discriminatoires - discriminating power) : en effet, plus leur valeur est élevée, et plus l'item concerné permet de séparer (discriminer) deux groupes d'individus par rapport à la valeur de leur trait latent, en utilisant une valeur seuil de ce trait latent correspondant à la valeur du paramètre de difficulté de l'item.

En termes géométriques, le fait que les items n'aient pas tous la même valeur pour les pseudos-paramètres B_j se traduit par des pentes différentes des ICC à leur point d'inflexion (lorsque $\theta_n = \delta_j$). Pour obtenir un modèle tel que ces pentes soient égales pour tous les items, il faut fixer les paramètres B_j à une valeur fixée B : le modèle devient alors un modèle équivalent au modèle de Rasch.

Démonstration.

La fonction réponse de l'item j vaut :

$$P_{nj} = P(X_{nj} = 1 / B_j, \theta_n, \delta_j) = \frac{\exp [B_j (\theta_n - \delta_j)]}{1 + \exp [B_j (\theta_n - \delta_j)]} \quad (2.81)$$

Soit $u_{nj} = B_j (\theta_n - \delta_j)$.

La pente de l'ICC en θ est donnée par :

$$\left. \frac{dP_{nj}}{d\theta_n} \right|_{\theta_n=\theta} = \left. \frac{B_j \exp(u_{nj})}{(1 + \exp(u_{nj}))} \right|_{\theta_n=\theta} \quad (2.82)$$

Cette pente est maximisée lorsque $\frac{d^2 P_{nj}}{d\theta_n^2} = 0$. Or

$$\frac{d^2 P_{nj}}{d\theta_n^2} = B_j^2 \frac{\exp(u_{nj}) (1 - \exp(2u_{nj}))}{(1 + \exp(u_{nj}))^4} \quad (2.83)$$

Donc la seule solution à $\frac{d^2 P_{nj}}{d\theta_n^2} = 0$ est

$$\begin{aligned} 1 - \exp(2u_{nj}) &= 0 \\ u_{nj} &= 0 \\ \theta_n &= \delta_j \end{aligned} \quad (2.84)$$

Donc la pente de l'ICC est maximale en $\theta_n = \delta_j$ et vaut

$$\left. \frac{dP_{nj}}{d\theta_n} \right|_{\theta_n=\delta_j} = \frac{B_j}{4} \quad (2.85)$$

Ainsi pour que deux items j et k aient la même pente maximale, il faut que $B_j = B_k$ et pour que tous les items aient la même pente maximale, il faut que $B_j = B \forall j = 1, \dots, J$. Dans ce cas, en posant $\theta_n^* = B\theta_n$ et $\delta_j^* = B\delta_j$, la fonction réponse d'un item j s'écrit :

$$P(X_{nj} = 1/B, \theta_n, \delta_j) = \frac{\exp[B(\theta_n - \delta_j)]}{1 + \exp[B(\theta_n - \delta_j)]} = \frac{\exp(\theta_n^* - \delta_j^*)}{1 + \exp(\theta_n^* - \delta_j^*)} \quad (2.86)$$

On obtient alors un modèle équivalent au modèle de Rasch. \square

2.7.1.2 Propriétés

De la même manière que pour le modèle de Rasch, on montre que le score pondéré $\tilde{S}_n = \sum_{j=1}^J B_j X_{nj}$ est une statistique exhaustive du trait latent θ_n . Dans un souci de simplification pour l'utilisation de tels scores pondérés, et pour réduire le nombre de scores possibles, les auteurs proposent de restreindre l'espace des pseudos-paramètres B_j à l'ensemble des premiers entiers naturels. Dans ce contexte, il existe alors $(\sum_{j=1}^J B_j + 1)$ scores possibles différents.

2.7.1.3 Estimation

Comme pour le modèle de Rasch, l'OPLM n'est pas identifiable (une contrainte d'identifiabilité est nécessaire) et l'estimation conjointe des paramètres de difficulté et des valeurs individuelles du trait latent est, dans le cas général, inconsistante.

Etant donné la propriété d'exhaustivité du score pondéré \tilde{S}_n sur le trait latent, il est possible d'utiliser la méthode du maximum de vraisemblance conditionnelle (conditionnellement au score pondéré \tilde{S}_n) qui rend les estimations consistantes. Il est en outre possible d'utiliser, en considérant le trait latent comme une variable aléatoire, la méthode du maximum de vraisemblance marginale, ou d'autres méthodes d'estimations (GEE, algorithme EM...).

2.7.2 Courbes caractéristiques des items dans l'OPLM

Lorsque l'on représente les courbes caractéristiques d'un ensemble d'items vérifiant un OPLM, on obtient des courbes logistiques sécantes comme présenté à la figure 2.4. Seuls les items ayant les mêmes pouvoirs discriminants ont des pentes au point d'inflexion des courbes caractéristiques identiques.



FIG. 2.4 – Courbes caractéristiques de 7 items vérifiant un OPLM

2.7.3 Le modèle de Birnbaum (2-PLM)

Le modèle de Birnbaum [15] est un modèle très connu en psychométrie, malgré des problèmes réels d'estimation des paramètres de ce modèle. Ce modèle extrapole le OPLM, en considérant les pouvoirs discriminants comme des paramètres inconnus. Nous noterons dans ce modèle le pouvoir discriminant du j^e item α_j (qui s'interprètent comme les paramètres B_j sous l'OPLM). Outre qu'il vérifie les hypothèses fondamentales de l'IRT, le modèle de Birnbaum est défini par ses IRF :

$$P(X_{nj} = x_{nj}/\theta_n, \alpha_j, \delta_j) = \frac{\exp[\alpha_j x_{nj} (\theta_n - \delta_j)]}{1 + \exp[\alpha_j (\theta_n - \delta_j)]} \quad (2.87)$$

Dans ce modèle, il n'existe plus de scores connus *a priori* (avant l'estimation des paramètres) qui soit exhaustif du trait latent. A ce titre, le modèle de Birnbaum n'est pas considéré comme une extension du modèle de Rasch, et la méthode du maximum de vraisemblance conditionnelle n'est plus envisageable.

On trouve souvent ce modèle sous la forme équivalente :

$$P(X_{nj} = x_{nj}/\theta_n, \alpha_j, \delta_j) = \frac{\exp[1.7\alpha_j x_{nj} (\theta_n - \delta_j)]}{1 + \exp[1.7\alpha_j (\theta_n - \delta_j)]} \quad (2.88)$$

Le coefficient 1.7 étant une approximation de $c^{-1} = \frac{\pi}{\sqrt{3.41}}$ [82, page 120]. Ce coefficient est utilisé pour approcher la fonction logistique $\text{logit}(x)$ par la fonction probit $\Phi(x)$. En effet on montre [18] que :

$$\forall u(x) \in \mathbb{R}, \text{logit}(u(x)) \approx \Phi(c.u(x)) \quad (2.89)$$

Lorsque l'on représente les courbes caractéristiques d'un ensemble d'items vérifiant un modèle de Birnbaum, on obtient des courbes logistiques sécantes comme présenté à la figure 2.5. A la différence de l'OPLM, les pentes des courbes caractéristiques à leur point d'inflexion sont toutes différentes.

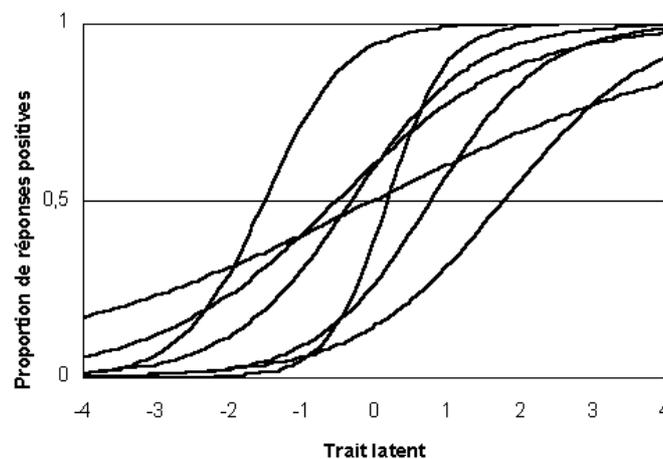


FIG. 2.5 – Courbes caractéristiques de 7 items vérifiant un modèle de Birnbaum

2.7.4 Autres modèles

En IRT, un certain nombre d'autres modèles existent dans un cadre dichotomique unidimensionnel. Cependant, la plupart d'entre eux ont été définis en sciences de l'éducation (mesure de connaissances, de l'intelligence...) et considèrent un paramètre de "guessing" (réponse positive par hasard). Nous nous sommes restreint à l'analyse de la qualité de vie dans ce travail, or, dans ce domaine, on ne considère pas la possibilité de "réponses par hasard", puisque les réponses aux items correspondent à un état clinique du patient, et qu'il n'existe pas d'enjeu lors du processus de réponse. Nous ne ferons donc que citer certains de ces modèles : le 3-PLM (modèle logistique à trois paramètres), le 4-PLM (modèle logistique à 4 paramètres, faisant état d'une réponse négative par hasard) ou le 5-PAM (modèle d'accélération à 5 paramètres - 5-Parameters acceleration model - qui introduit un paramètre d'accélération permettant de considérer des ICC non symétriques)...

Nous pourrions cependant citer quelques modèles non étudiés dans ce travail qui sont parfois utilisés en qualité de vie : le Linear Logistic Test Model (LLTM) [47] par exemple, qui permet de décomposer la difficulté de chaque item pour modéliser le processus de réponse, en

fonction des opérations cognitives sollicitées ; ou le Log-linear Rasch Model (LLRM) [85] [84] [87] qui permet d'expliquer la réponse à l'item par des covariables (sexe, âge...).

Enfin, toute une littérature existe autour des modèles probit : la fonction de lien logistique est alors remplacée par une fonction probit. C'est le cas des "normal ogive models" à un ou deux paramètres [95].

2.8 Les modèles de l'IRT pour items polytomiques

2.8.1 L'extension au cadre polytomique du modèle de Rasch

Si le modèle de Rasch est un modèle intéressant en IRT, et de ce fait très utilisé, il ne permet de prendre en considération que des items à réponses dichotomiques. Or, en qualité de vie, il est courant d'avoir recours à des questionnaires composés d'items à réponses qualitatives, et le plus souvent ordinales. Des extensions au cadre polytomique du modèle de Rasch ont été proposées en ce sens. Rasch a lui-même proposé un développement polytomique du modèle de Rasch [127], qui est décrit à la section 3.4.1. Ce modèle est peu utilisé en pratique, car c'est un modèle très restrictif. En revanche, dans les années 70, Andrich [13] [10] propose le Rating Scale Model (RSM), basé sur des résultats d'Andersen [9]. Masters [98] propose au début des années 80 le Partial Credit Model (PCM). Ces deux modèles restent à ce jour les principales extensions au cadre polytomique du modèle de Rasch.

Ces deux modèles pour items à réponses ordinales sont considérés comme des extensions du modèle de Rasch car ils conservent tous les deux la propriété d'exhaustivité du score sur le trait latent. Dans ce cadre, il existe encore pour chaque item une réponse négative codée 0, choisie comme la plus défavorable eu égard au concept mesuré, et les autres modalités, appelées réponses positives, sont ordonnées et prennent pour valeur $1, \dots, m_j$ (pour l'item j).

2.8.2 Le modèle d'échelle de classement (Rating Scale Model) (RSM)

Le RSM [12] [11] est le plus simple des deux modèles : tous les items ont le même nombre de modalités $m_j = m \forall j = 1, \dots, J$. L'item j est modélisé par un paramètre de difficulté δ_j . En outre, pour chaque réponse positive h codée au delà de 1, un paramètre d'ajustement τ_h est posé (ce paramètre n'étant pas différent entre les items).

La fonction réponse correspondant à la modalité positive h de l'item j dans le cadre du RSM s'écrit :

$$P(X_{nj} = h/\theta_n, \delta_j, \tau_2, \dots, \tau_m) = \frac{\exp\left(h\theta_n - \delta_j - \sum_{p=2}^h \tau_p\right)}{\sum_{l=0}^m \exp\left(h\theta_n - \delta_j - \sum_{p=2}^l \tau_p\right)} \quad (2.90)$$

Le paramètre δ_j représente la difficulté globale de l'item : plus sa valeur est élevée, et plus il est difficile pour un individu donné d'avoir une réponse positive à l'item j . Sa valeur correspond au niveau du trait latent qu'il faut atteindre pour avoir autant de chance de répondre négativement à l'item que de répondre à la première modalité (notée "1").

Plus la valeur d'un paramètre τ_h est élevée et plus il est difficile de répondre à une modalité au moins égale à h pour les individus ayant une réponse au moins égale à $h - 1$: il s'agit donc de la difficulté de la réponse h pour les individus ayant une réponse au moins égale à $h - 1$. La valeur τ_h représente la valeur du trait latent qu'il faut atteindre pour avoir autant de chance de répondre à la modalité $h - 1$ qu'à la modalité h .

Lorsque l'on représente les courbes caractéristiques d'un items vérifiant un RSM, on obtient une figure comme celle présentée à la figure 2.6.

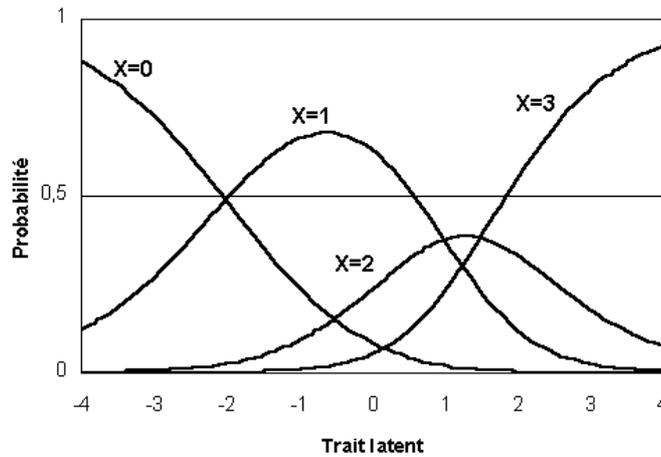


FIG. 2.6 – Courbes caractéristiques d'un item à 4 modalités suivant un RSM avec les paramètres $\delta_j = -2.0$, $\tau_2 = 1.0$ et $\tau_3 = 1.5$

2.8.3 Le modèle du crédit partiel (Partial Credit Model) (PCM)

Le PCM [99] [11] introduit quant-à lui un paramètre δ_{jh} pour chaque modalité h positive de chaque item j . Les items n'ont alors pas obligatoirement le même nombre de modalités.

La fonction réponse correspondant à la modalité positive h de l'item j dans le cadre du PCM s'écrit :

$$P(X_{nj} = h/\theta_n, \delta_{j1}, \dots, \delta_{jm_j}) = \frac{\exp \left[h \left(\theta_n - \sum_{p=1}^h \delta_{jp} \right) \right]}{\sum_{l=0}^{m_j} \exp \left(h \theta_n - \sum_{p=1}^l \delta_{jp} \right)} \quad (2.91)$$

Les paramètres du PCM s'interprètent de la même manière que ceux du RSM.

2.8.4 Lien entre le RSM et le PCM

On peut considérer le RSM comme un cas particulier du PCM où les items auraient le même nombre de modalités m et où pour tout $j = 1, \dots, J$:

$$\begin{cases} \delta_{j1} = \delta_j \\ \delta_{jh} = \tau_h, \quad \forall h = 2, \dots, m \end{cases} \quad (2.92)$$

2.8.5 Estimation du PCM et du RSM

Etant donné la propriété d'exhaustivité du score sur le trait latent, il est possible d'estimer de façon consistante les paramètres des deux modèles par la méthode du maximum de vraisemblance conditionnelle. L'algorithme est par exemple présenté en [11].

En considérant, en parallèle du modèle de Rasch, le trait latent comme une variable aléatoire, il est possible d'estimer les paramètres des deux modèles par la méthode du maximum de vraisemblance marginale [11].

2.9 Synthèse

Dans ce chapitre, nous avons posé les hypothèses fondamentales de l'IRT et présenté les principaux modèles unidimensionnels de l'IRT utilisés dans le champ de la qualité de vie. Dans la suite, nous allons présenter l'extension de l'IRT au champ multidimensionnel, et les modèles existants qui en découlent.

Chapitre 3

L'IRT multidimensionnelle

3.1 Définition

L'IRT multidimensionnelle est une extension naturelle, mais néanmoins récente de l'IRT : les principales contributions dans ce domaine ont été proposées dans les années 90.

L'IRT multidimensionnelle est encore dans une phase de développement intense, et il n'existe pas de typologie bien fixée des modèles et de leurs appellations. Peu de modèles bien définis ont été retrouvés, et il existe plutôt un certain nombre d'extrapolations de modèles unidimensionnels utilisés et retrouvés (voir par exemple [130], [32], [150], [129]).

3.2 Extension des hypothèses fondamentales de l'IRT

3.2.1 La dimension

En IRT multidimensionnelle, il n'y a bien entendu plus d'hypothèse d'unidimensionnalité. En revanche, la dimension Q , c'est-à-dire le nombre de traits latents permettant d'expliquer les réponses à l'ensemble des items, est fixée (connue) *a priori*. Junker [83] définit en outre que $Q \ll J$. Zhang et Stout [155] définissent la dimension comme le nombre de traits latents nécessaire pour que soient vérifiées l'indépendance locale et la monotonie.

Il existe donc Q traits latents Θ_q , $q = 1, \dots, Q$, et chaque individu est caractérisé par un vecteur Q -dimensionnel $\theta_n = (\theta_{nq})_{q=1, \dots, Q}$ appelé trait latent multidimensionnel.

3.2.2 La monotonie

En IRT multidimensionnelle, l'hypothèse de monotonie est étendue aux Q traits latents, c'est-à-dire que les fonctions de réponses aux items sont non décroissantes pour chaque trait latent, tous les autres traits latents étant par ailleurs fixés. Analytiquement, cela s'écrit :

$$\frac{dP(X_{nj} = 1/\theta_n; \nu_j)}{d\theta_{nq}} \geq 0, \quad \forall q = 1, \dots, Q, \quad (3.1)$$

avec ν_j un vecteur de paramètres caractérisant l'item j .

3.2.3 L'indépendance locale

En IRT multidimensionnelle, l'hypothèse d'indépendance locale est étendue aux Q traits latents, ou au trait latent multidimensionnel. On a donc pour tout j, j' tels que $1 \leq j \neq j' \leq J$ les variables aléatoires X_{nj} et $X_{nj'}$ indépendantes conditionnellement à θ_n . Ainsi

$$P(\mathbf{X}_n = \mathbf{x}_n / \theta_n; \boldsymbol{\nu}) = \prod_{j=1}^J P(X_{nj} = x_{nj} / \theta_n; \nu_j), \quad (3.2)$$

avec $\boldsymbol{\nu} = (\nu_j)_{j=1, \dots, J}$.

3.3 Les structures simples (SS) et approximativement simple (ASS)

Une structure simple [44] est la structure sous-jacente à un ensemble de J items formant une partition de ces items en Q groupes tels que les réponses de tous les items d'un même groupe sont explicitées par un trait latent spécifique, et ne sont pas expliquées par les autres traits latents. Dans une structure approximativement simple, les réponses aux items d'un même groupe sont expliquées principalement par un trait latent spécifique, et de manière moins sensible par les autres traits latents. Zhang et Stout [155] parlent dans ce cas de groupes d'items suffisamment séparés ("sufficiently separated"). Ils supposent qu'une structure simple est une structure souhaitable mais peu envisageable en pratique, et qu'une structure approximativement simple est une approche suffisante d'une structure simple en pratique.

3.4 Les extensions multidimensionnelles du modèle de Rasch

3.4.1 Le modèle multidimensionnel de Rasch

Le modèle multidimensionnel de Rasch est sans doute le premier modèle de l'IRT multidimensionnelle qui a été proposé. Il est proposé par Rasch lui-même [127], et explicité par exemple par Fisher [39] et par Kelderman et Rijkes [87].

Ce modèle permet de traiter les réponses d'un ensemble de J items polytomiques à M modalités positives. Dans ce modèle, la dimension (le nombre de traits latents) est fixée à $Q = M$. Chaque réponse positive d'un item donné est associée à un et un seul trait latent, tel que il ne puisse y avoir, pour un même item, deux réponses associées à un même trait latent. Les réponses aux items sont codées de manière à ce que toutes les réponses codées de la même manière soient associées au même trait latent. Soit h un codage particulier d'une réponse ($h = 1, \dots, M$), alors, à chaque réponse h d'un item, on associe le h^e trait latent Θ_h . Le modèle s'écrit donc :

$$P(X_{nj} = h / \theta_n, \delta_j) = \frac{\exp(\theta_{nh} - \delta_j)}{1 + \exp(\theta_{nh} - \delta_j)} \quad (3.3)$$

On note que si les items sont dichotomiques, on retrouve le modèle de Rasch classique.

Ce modèle très particulier est difficile à mettre en pratique, puisque les modalités de chaque item doivent être étudiées de manière à correspondre chacune à la mesure d'un trait latent particulier. Dans un certain nombre de domaines, et notamment dans le domaine de l'étude de la qualité de vie, de tels items semblent très difficiles à obtenir : les différentes modalités d'un item dans ce champ correspondent plutôt à différents niveaux sur la variable latente. Aucune application concrète de ce modèle n'a été retrouvée dans la littérature¹. Enfin, ce modèle n'est utilisable que dans un cadre "confirmatoire" où l'on chercherait à vérifier que les liens supposés entre chaque modalité de chaque item et les différents traits latents sont corrects, mais pas dans un cadre "exploratoire" où l'on rechercherait les liens entre les items (ou les modalités des items) et les traits latents. Nous ne développerons pas plus l'étude de ce modèle.

3.4.2 Le modèle à trait latent à plusieurs composantes (multicomponent latent trait model)

Ce modèle élaboré par Whitely [150] pour analyser des réponses à des items dichotomiques faisant appel à plusieurs caractéristiques a été présenté par Stegelmann [139] comme une extension multidimensionnelle du modèle de Rasch, puisqu'il préserve la propriété d'objectivité spécifique. Le modèle suppose qu'il existe Q composantes du trait latent, et qu'il existe une "variable" (mot proposé par l'auteur) B prenant pour valeurs b_{jq} tels que :

$$b_{jq} = \begin{cases} 1, & \text{si une réponse correcte à l'item } j \text{ est influencée par la } q^e \text{ composante du trait latent} \\ 0, & \text{sinon} \end{cases} \quad (3.4)$$

La fonction de réponse à l'item j s'écrit alors :

$$P(X_{nj} = x_{nj} / \boldsymbol{\theta}_n, \boldsymbol{\delta}_j) = \frac{\left[\prod_{q=1}^Q \exp(b_{jq}(\theta_{nq} - \delta_{jq})) \right]^{x_{nj}}}{1 + \prod_{q=1}^Q \exp(b_{jq}(\theta_{nq} - \delta_{jq}))} \quad (3.5)$$

Stegelman [139] montre que le vecteur $\mathbf{s}_n = (s_{nq})_{q=1, \dots, Q}$ des sous-scores $s_{nq} = \sum_{j=1}^J b_{jq} x_{nj}$ est une statistique exhaustive du trait latent multidimensionnel $\boldsymbol{\theta}_n$. En outre, il propose une méthode d'estimation des paramètres du modèle basée sur une extrapolation du maximum de vraisemblance conditionnelle (voir la section 3.4.3) et un test d'adéquation globale basée sur l'extension du test de rapport de vraisemblance Z de Andersen.

3.4.3 Le modèle multidimensionnel polytomique à trait latent (multidimensional polytomous latent trait model) (MPLT)

3.4.3.1 Définition et notations

Le MPLT extrapole le modèle précédent de deux manières : d'une part, le modèle est étendu aux items polytomiques, d'autre part, les liens entre les items et les traits latents, définis par la "variable" B , peuvent être quantifiés à la manière de l'OPLM, c'est-à-dire que l'on peut donner plus ou moins de poids à la "variable" B pour représenter l'impact plus ou moins important de chaque item dans la mesure de chaque trait latent (voir section 2.7.1).

¹on peut cependant trouver un certain nombre d'exemples dans les magazines de sociétés concernant des tests psychologiques vérifiant, ou supposés vérifiant, ce type de modèle

Ce modèle est défini par Kelderman et Rijkens [87] [86]. Les auteurs remplacent le terme "variable" pour définir B par "poids".

A chaque item j ayant m_j modalités positives est associé $Q(m_j + 1)$ poids B_{jqx} , avec $q = 1, \dots, Q$ et $x = 0, \dots, m_j$. Ces poids valent 0 si la réponse x de l'item j ne fait pas appel au q^e trait latent, sinon ce poids est défini par un nombre positif. Ils sont choisis subjectivement par des experts, en fonction des liens supposés entre chaque item et chaque trait latent, et de la "force" supposée de chacune de ces relations.

A chaque modalité x d'un item j est associé Q paramètre δ_{jqx} , $q = 1, \dots, Q$ définissant la difficulté de la réponse x de l'item j par rapport au q^e trait latent.

Le multidimensional polytomous latent trait model s'écrit alors :

$$P(X_{nj} = x / \boldsymbol{\theta}_n, \boldsymbol{\delta}_j) = \frac{\exp\left(\sum_{q=1}^Q B_{jqx}(\theta_{nq} - \delta_{jqx})\right)}{\sum_{y=0}^{m_j} \exp\left(\sum_{q=1}^Q B_{jqy}(\theta_{nq} - \delta_{jqy})\right)} \quad (3.6)$$

Les auteurs montrent que les paramètres δ_{jqx} ne sont pas identifiables, en revanche, les paramètres $\delta_{jx} = \sum_{q=1}^Q B_{jqx} \delta_{jqx}$ le sont. Aussi, le modèle est généralement réécrit [74] :

$$P(X_{nj} = x / \boldsymbol{\theta}_n, \boldsymbol{\delta}_j) = \frac{\exp\left(\sum_{q=1}^Q B_{jqx} \theta_{nq} - \delta_{jx}\right)}{\sum_{y=0}^{m_j} \exp\left(\sum_{q=1}^Q B_{jqy} \theta_{nq} - \delta_{jy}\right)} \quad (3.7)$$

3.4.3.2 Représentation graphique

Par extrapolation de la représentation graphique proposée à la section 2.6.5 (page 19), la figure 3.1 propose une représentation graphique du MPLT.

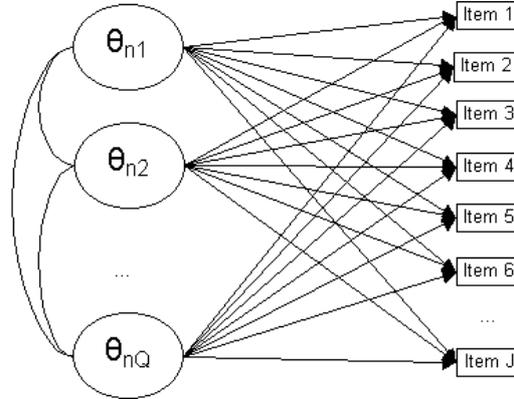


FIG. 3.1 – Représentation graphique du Multidimensional Polytomous Latent Trait Model

3.4.3.3 L'exhaustivité du vecteur des sous-scores sur le trait latent multidimensionnel

En développant l'équation (3.7) sous l'hypothèse d'indépendance locale, en posant $s_{nq}^* = \sum_{j=1}^J B_{jqx_{nj}}$, on obtient :

$$\begin{aligned}
 P(\mathbf{X}_n = \mathbf{x}_n / \boldsymbol{\theta}_n, \boldsymbol{\delta}) &= \prod_{j=1}^J P(X_{nj} = x_{nj} / \boldsymbol{\theta}_n, \boldsymbol{\delta}_j) \\
 &= \prod_{j=1}^J \frac{\exp\left(\sum_{q=1}^Q B_{jqx_{nj}} \theta_{nq} - \delta_{jx_{nj}}\right)}{\sum_{y=0}^{m_j} \exp\left(\sum_{q=1}^Q B_{jqy} \theta_{nq} - \delta_{jy}\right)} \\
 &= \frac{\exp\left(\sum_{j=1}^J \left[\sum_{q=1}^Q B_{jqx_{nj}} \theta_{nq} - \delta_{jx_{nj}}\right]\right)}{\prod_{j=1}^J \sum_{y=0}^{m_j} \exp\left(\sum_{q=1}^Q B_{jqy} \theta_{nq} - \delta_{jy}\right)} \\
 &= \frac{\exp\left(\sum_{q=1}^Q s_{nq}^* \theta_{nq} - \sum_{j=1}^J \delta_{jx_{nj}}\right)}{\prod_{j=1}^J c(\boldsymbol{\theta}_n, \boldsymbol{\delta}_j)} \tag{3.8}
 \end{aligned}$$

avec $c(\boldsymbol{\theta}_n, \boldsymbol{\delta}_j) = \sum_{y=0}^{m_j} \exp\left(\sum_{q=1}^Q B_{jqy} \theta_{nq} - \delta_{jy}\right)$

Ainsi la densité jointe des X_{nj} , $j = 1, \dots, J$ fait partie de la fonction exponentielle, et par le théorème de factorisation, le vecteur $\mathbf{s}_n^* = (s_{nq}^*)_{q=1, \dots, Q}$ est exhaustif du vecteur $\boldsymbol{\theta}_n$.

Par analogie avec le OPLM (voir section 2.7.1), les auteurs restreignent les poids B_{jqx} aux premiers entiers naturels afin de réduire le nombre de vecteurs \mathbf{s}_n^* possibles.

La figure 3.2 représente graphiquement la propriété d'exhaustivité du vecteur des sous-scores \mathbf{s}_n^* sur le trait latent multidimensionnel $\boldsymbol{\theta}_n$ dans le cadre du MPLT

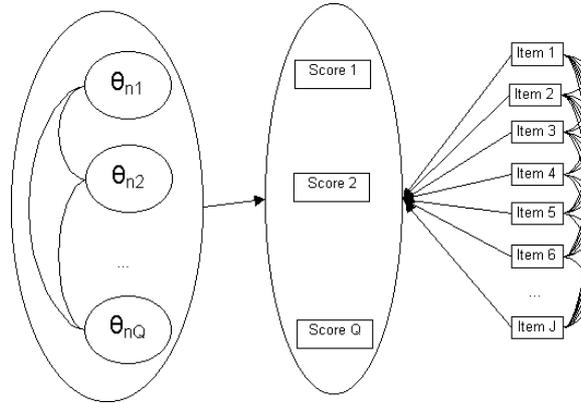


FIG. 3.2 – Représentation graphique de la propriété d'exhaustivité du vecteur des sous-scores sur le trait latent multidimensionnel dans le cadre du MPLT

3.4.3.4 Estimation par CML et MML

En posant, par analogie avec l'équation (2.9) :

$$\begin{aligned} P(\mathbf{S}_n^* = \mathbf{s}_n^*/\boldsymbol{\theta}_n, \boldsymbol{\delta}) &= \sum_{\mathbf{y} \in \Omega/\mathbf{s}_y = \mathbf{s}_n^*} P(\mathbf{X}_n = \mathbf{y}/\boldsymbol{\theta}_n, \boldsymbol{\delta}) \\ &= \gamma(\mathbf{s}_n^*, \boldsymbol{\delta}) \exp \left(\sum_{q=1}^Q \theta_{nq} s_{nq}^* \right) \prod_{j=1}^J [c(\boldsymbol{\theta}_n, \boldsymbol{\delta}_j)]^{-1}, \end{aligned} \quad (3.9)$$

avec Ω l'ensemble des 2^J vecteurs possibles de dimension J composés de 0 et de 1, $\mathbf{y} = (y_j)_{j=1, \dots, J}$ un vecteur quelconque de Ω , $\mathbf{s}_y = (\sum_{j=1}^J B_{jq} y_j)_{q=1, \dots, Q}$ et

$$\gamma(\mathbf{s}_n^*, \boldsymbol{\delta}) = \sum_{\mathbf{y} \in \Omega/\mathbf{s}_y = \mathbf{s}_n^*} \exp(-\delta_{jy_j}) \quad (3.10)$$

La vraisemblance conditionnelle des réponses de l'individu n s'écrit alors :

$$\begin{aligned} P(\mathbf{X}_n = \mathbf{x}_n/\mathbf{S}_n^* = \mathbf{s}_n^*, \boldsymbol{\theta}_n, \boldsymbol{\delta}) &= \frac{P(\mathbf{X}_n = \mathbf{x}_n/\boldsymbol{\theta}_n, \boldsymbol{\delta})}{P(\mathbf{S}_n^* = \mathbf{s}_n^*/\boldsymbol{\theta}_n, \boldsymbol{\delta})} \\ &= \frac{\exp \left(-\sum_{j=1}^J \delta_{jx_{nj}} \right)}{\gamma(\mathbf{s}_n^*, \boldsymbol{\delta})} \\ &= P(\mathbf{X}_n = \mathbf{x}_n/\mathbf{S}_n^* = \mathbf{s}_n^*, \boldsymbol{\delta}) \end{aligned} \quad (3.11)$$

Cette vraisemblance conditionnelle est indépendante du trait latent multidimensionnel. Il est alors possible d'estimer de façon consistante les paramètres de difficulté $\boldsymbol{\delta} = (\delta_{jx})_{j=1, \dots, J; x=0, \dots, m_j}$ en maximisant :

$$L_C(\boldsymbol{\delta}/\mathbf{x}, \mathbf{s}^*) = \prod_{n=1}^N P(\mathbf{X}_n = \mathbf{x}_n/\mathbf{S}_n^* = \mathbf{s}_n^*, \boldsymbol{\delta}) \quad (3.12)$$

Pour résoudre les équations de vraisemblance, il faut poser $Q(J+1)$ contraintes d'identifiabilité : on retrouve systématiquement les $J \cdot Q$ contraintes $B_{jq0} = 0, \forall j = 1, \dots, J$ et $q = 1, \dots, Q$ qui sont les extrapolations multidimensionnelles du fait qu'il existe, pour chaque item, une réponse négative ; Kelderman et Rijkes propose les Q autres contraintes suivantes : $\sum_{j=1}^J \sum_{x=1/\exists B_{jqx} \neq 0}^{m_j} \delta_{jx} = 0, \forall q = 1, \dots, Q$ qui est l'extrapolation multidimensionnelle de la contrainte de nullité de la moyenne des paramètres de difficultés des items dans les modèles de l'IRT. Ces contraintes sont notamment utilisées dans les logiciels LOGIMO et ConQuest qui permettent d'estimer de tels modèles [74] [88] [152].

Une autre manière d'obtenir des estimations consistantes consiste à considérer le trait latent comme une variable aléatoire multidimensionnelle notée $G(\cdot)$ (en général, on suppose cette variable distribuée selon une loi multinormale de paramètres $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$: dans ce cas, les Q dernières contraintes sont généralement $\boldsymbol{\mu} = \mathbf{0}$), et à intégrer la vraisemblance sur l'ensemble des valeurs possible de $\boldsymbol{\theta}$, c'est-à-dire sur \mathbb{R}^Q . On obtient alors la vraisemblance marginale donnée par :

$$L_M(\boldsymbol{\Sigma}, \boldsymbol{\delta}) = \prod_{n=1}^N \int_{\mathbb{R}^Q} P(\mathbf{X}_n = \mathbf{x}_n/\boldsymbol{\theta}; \boldsymbol{\delta}) G(\boldsymbol{\theta}/\boldsymbol{\Sigma}) d\boldsymbol{\theta} \quad (3.13)$$

Tout comme dans le cadre unidimensionnel, les estimations obtenues par cette méthode sont biaisées si la fonction $G(\cdot)$ est mal définie [2].

3.4.3.5 Cadre dichotomique

Dans un cadre dichotomique, il n'existe plus que Q poids non systématiquement nuls pour chaque item. On notera ces poids associés à la réponse positive ($x = 1$) B_{jq} , $q = 1, \dots, Q$. Parallèlement, nous noterons la difficulté de la réponse positive de l'item j relative à la q^e dimension δ_{jq} et nous noterons que pour chaque item, il n'existe plus qu'une seule quantité estimable $\sum_{q=1}^Q B_{jq}\delta_{jq} = \delta_j$.

Le modèle s'écrit alors :

$$P(X_{nj} = x/\theta_n, \delta_j) = \frac{\exp\left(x\left(\sum_{q=1}^Q B_{jq}\theta_{nq} - \delta_j\right)\right)}{1 + \exp\left(\sum_{q=1}^Q B_{jq}\theta_{nq} - \delta_j\right)} \quad (3.14)$$

3.4.3.6 Approche GEE pour le MPLT dans le cas dichotomique

a- Approximation de la vraisemblance marginale

Nous proposons ici d'estimer les paramètres d'un MPLT par l'approche GEE. Le développement ci-dessous est inspiré du même développement réalisé sous le modèle de Rasch par Feddag, Grama et Mesbah [35] (et présenté à la section 2.6.8.5).

La fonction de réponse à un item dichotomique j dans le cadre du MPLT s'écrit :

$$P(X_{nj} = x_{nj}/\theta_n; \mathbf{B}, \delta_j) = \frac{\exp\left(x_{nj}\left(\sum_{q=1}^Q B_{jq}\theta_{nq} - \delta_j\right)\right)}{1 + \exp\left(\sum_{q=1}^Q B_{jq}\theta_{nq} - \delta_j\right)} \quad (3.15)$$

On note $h_{nj}(y)$ la fonction :

$$h_{nj}(y) = \frac{\exp(x_{nj}y)}{1 + \exp(y)} = \exp[x_{nj}y - \log(1 + \exp(y))] \quad (3.16)$$

Soit $g_{nj}(\theta_n) = \sum_{q=1}^Q B_{jq}\theta_{nq} - \delta_j$, alors

$$P(X_{nj} = x_{nj}/\theta_n; \mathbf{B}, \delta_j) = h_{nj}(g_{nj}(\theta_n)) \quad (3.17)$$

Soit $l(y)$ la fonction $l(y) = \frac{\exp(y)}{1 + \exp(y)}$. Notons $l_{nj} = l(g_{nj}(\theta_n))$.

On montre que les dérivées d'ordre 1 à 3 de cette fonction sont données par :

$$l_{njq}^{(1)} = \frac{dl_{nj}}{d\theta_{nq}} = B_{jq}C_{nj}, \quad q = 1, \dots, Q, \quad (3.18)$$

avec

$$C_{nj} = \frac{\exp(g_{nj})}{[1 + \exp(g_{nj})]^2}, \quad (3.19)$$

$$l_{njq_1q_2}^{(2)} = \frac{d^2 l_{nj}}{d\theta_{nq_1} d\theta_{nq_2}} = B_{jq_1} B_{jq_2} D_{nj}, \quad q_1, q_2 = 1, \dots, Q, \quad (3.20)$$

avec

$$D_{nj} = \frac{\exp(g_{nj})(1 - \exp(g_{nj}))}{[1 + \exp(g_{nj})]^3}, \quad (3.21)$$

$$l_{njq_1q_2q_3}^{(3)} = \frac{d^3 l_{nj}}{d\theta_{nq_1} d\theta_{nq_2} d\theta_{nq_3}} = B_{jq_1} B_{jq_2} B_{jq_3} E_{nj}, \quad q_1, q_2, q_3 = 1, \dots, Q, \quad (3.22)$$

avec

$$E_{nj} = \frac{\exp(g_{nj})(1 - 4 \exp(g_{nj})) + \exp(2g_{nj})}{[1 + \exp(g_{nj})]^4}. \quad (3.23)$$

On montre aussi que

$$h_{njq} = \frac{dh_{nj}}{d\theta_{nq}} = B_{jq} h_{nj} A_{nj}, \quad q = 1, \dots, Q, \quad (3.24)$$

avec $A_{nj} = x_{nj} - l_{nj}$,

$$h_{njq_1q_2} = \frac{d^2 h_{nj}}{d\theta_{nq_1} d\theta_{nq_2}} = B_{jq_1} B_{jq_2} h_{nj} Q_{nj}, \quad q_1, q_2 = 1, \dots, Q, \quad (3.25)$$

avec $Q_{nj} = A_{nj}^2 - C_{nj}$,

$$h_{njq_1q_2q_3} = \frac{d^3 h_{nj}}{d\theta_{nq_1} d\theta_{nq_2} d\theta_{nq_3}} = B_{jq_1} B_{jq_2} B_{jq_3} h_{nj} R_{nj}, \quad q_1, q_2, q_3 = 1, \dots, Q, \quad (3.26)$$

avec $R_{nj} = A_{nj}^3 - 3A_{nj}C_{nj} - D_{nj}$,

$$h_{njq_1q_2q_3q_4} = \frac{d^4 h_{nj}}{d\theta_{nq_1} d\theta_{nq_2} d\theta_{nq_3} d\theta_{nq_4}} = B_{jq_1} B_{jq_2} B_{jq_3} B_{jq_4} h_{nj} S_{nj}, \quad q_1, q_2, q_3, q_4 = 1, \dots, Q, \quad (3.27)$$

avec $S_{nj} = A_{nj}^4 - 6A_{nj}^2 C_{nj} - A_{nj} D_{nj} + 3C_{nj}^2 - E_{nj}$.

Notons l_j^* , $l_{jq}^{*(1)}$, $l_{jq_1q_2}^{*(2)}$, $l_{jq_1q_2q_3}^{*(3)}$, h_{nj}^* , h_{njq}^* , $h_{njq_1q_2}^*$, $h_{njq_1q_2q_3}^*$, $h_{njq_1q_2q_3q_4}^*$, A_{nj}^* , C_{nj}^* , D_{nj}^* , E_{nj}^* , Q_{nj}^* , R_{nj}^* et S_{nj}^* les valeurs de ces fonctions en $\theta_n = \mathbf{0}$.

L'approximation de Taylor de h_{nj} à l'ordre 4 vaut :

$$h_{nj} = h_{nj}^* \left(1 + A_{nj}^* G_{1nj} + \frac{1}{2} Q_{nj}^* G_{2nj} + \frac{1}{6} R_{nj}^* G_{3nj} + \frac{1}{24} S_{nj}^* G_{4nj} \right) + O(\|\theta_n\|^5), \quad (3.28)$$

avec

$$G_{1nj} = \sum_{q=1}^Q B_{jq} \theta_{nq}, \quad (3.29)$$

$$G_{2nj} = \sum_{q_1=1}^Q \sum_{q_2=1}^Q B_{jq_1} B_{jq_2} \theta_{nq_1} \theta_{nq_2}, \quad (3.30)$$

$$G_{3nj} = \sum_{q_1=1}^Q \sum_{q_2=1}^Q \sum_{q_3=1}^Q B_{jq_1} B_{jq_2} B_{jq_3} \theta_{nq_1} \theta_{nq_2} \theta_{nq_3}, \quad (3.31)$$

$$G_{4nj} = \sum_{q_1=1}^Q \sum_{q_2=1}^Q \sum_{q_3=1}^Q \sum_{q_4=1}^Q B_{jq_1} B_{jq_2} B_{jq_3} B_{jq_4} \theta_{nq_1} \theta_{nq_2} \theta_{nq_3} \theta_{nq_4}, \quad (3.32)$$

Par récurrence, on obtient

$$\begin{aligned} \prod_{j=1}^J h_{nj} &= P(\mathbf{X}_n = \mathbf{x}_n / \boldsymbol{\theta}_n; \mathbf{B}_j, \delta_j) \\ &= \left(\prod_{j=1}^J h_{nj}^* \right) \times \left(1 + \sum_{j=1}^J A_{nj}^* G_{1nj} + \frac{1}{2} \sum_{j=1}^J Q_{nj}^* G_{2nj} + \frac{1}{6} \sum_{j=1}^J R_{nj}^* G_{3nj} + \frac{1}{24} \sum_{j=1}^J S_{nj}^* G_{4nj} \right. \\ &\quad + \sum_{j_1=1}^J \sum_{j_2=j_1+1}^J A_{nj_1}^* A_{nj_2}^* G_{1nj_1} G_{1nj_2} + \frac{1}{2} \sum_{j_1=1}^J \sum_{j_2=1}^J A_{nj_1}^* Q_{nj_2}^* G_{1nj_1} G_{2nj_2} \\ &\quad + \frac{1}{6} \sum_{j_1=1}^J \sum_{j_2=1}^J A_{nj_1}^* R_{nj_2}^* G_{1nj_1} G_{3nj_2} + \frac{1}{4} \sum_{j_1=1}^J \sum_{j_2=j_1+1}^J Q_{nj_1}^* Q_{nj_2}^* G_{2nj_1} G_{2nj_2} \\ &\quad + \sum_{j_1=1}^J \sum_{j_2=j_1+1}^J \sum_{j_3=j_2+1}^J A_{nj_1}^* A_{nj_2}^* A_{nj_3}^* G_{1nj_1} G_{1nj_2} G_{1nj_3} \\ &\quad + \frac{1}{2} \sum_{j_1=1}^J \sum_{j_2=j_1+1}^J \sum_{j_3=1}^J A_{nj_1}^* A_{nj_2}^* Q_{nj_3}^* G_{1nj_1} G_{1nj_2} G_{2nj_3} \\ &\quad \left. + \sum_{j_1=1}^J \sum_{j_2=j_1+1}^J \sum_{j_3=j_2+1}^J \sum_{j_4=j_3+1}^J A_{nj_1}^* A_{nj_2}^* A_{nj_3}^* A_{nj_4}^* G_{1nj_1} G_{1nj_2} G_{1nj_3} G_{1nj_4} \right) \\ &\quad + O(\|\boldsymbol{\theta}_n\|^5) \end{aligned} \quad (3.33)$$

Aussi, la vraisemblance marginale correspondante à un individu n est

$$L_n(\mathbf{x}_n, \boldsymbol{\delta}, \boldsymbol{\Sigma}) = \int_{\mathbb{R}^Q} P(\mathbf{X}_n = \mathbf{x}_n / \boldsymbol{\theta}_n; \mathbf{B}_j, \boldsymbol{\delta}) G(\boldsymbol{\theta}_n, \boldsymbol{\Sigma}) d\boldsymbol{\theta}_n, \quad (3.34)$$

avec $G(\boldsymbol{\theta}_n, \boldsymbol{\Sigma})$ la fonction de répartition de la loi normale centrée de matrice de variance covariance

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma(\theta)_{11} & \dots & \sigma(\theta)_{1q} & \dots & \sigma(\theta)_{1Q} \\ \dots & \dots & \dots & \dots & \dots \\ \sigma(\theta)_{q1} & \dots & \sigma(\theta)_{qq} & \dots & \sigma(\theta)_{qQ} \\ \dots & \dots & \dots & \dots & \dots \\ \sigma(\theta)_{Q1} & \dots & \sigma(\theta)_{Qq} & \dots & \sigma(\theta)_{QQ} \end{pmatrix} \quad (3.35)$$

Rappelons que les moments joints d'ordre 1 à 4 de la variable $\boldsymbol{\theta}$ sont donnés par :

$$\int_{\mathbb{R}^Q} \theta_q G(\boldsymbol{\theta}, \boldsymbol{\Sigma}) d\boldsymbol{\theta} = 0, \quad q = 1, \dots, Q, \quad (3.36)$$

$$\int_{\mathbb{R}^Q} \theta_{q_1} \theta_{q_2} G(\boldsymbol{\theta}, \boldsymbol{\Sigma}) d\boldsymbol{\theta} = \sigma(\theta)_{q_1 q_2}, \quad q_1, q_2 = 1, \dots, Q, \quad (3.37)$$

$$\int_{\mathbb{R}^Q} \theta_{q_1} \theta_{q_2} \theta_{q_3} G(\boldsymbol{\theta}, \boldsymbol{\Sigma}) d\boldsymbol{\theta} = 0, \quad q_1, q_2, q_3 = 1, \dots, Q, \quad (3.38)$$

$$\int_{\mathbb{R}^Q} \theta_{q_1} \theta_{q_2} \theta_{q_3} \theta_{q_4} G(\boldsymbol{\theta}, \boldsymbol{\Sigma}) d\boldsymbol{\theta} = \sigma_{(\theta)q_1 q_2} \sigma_{(\theta)q_3 q_4} + \sigma_{(\theta)q_1 q_3} \sigma_{(\theta)q_2 q_4} \\ + \sigma_{(\theta)q_1 q_4} \sigma_{(\theta)q_2 q_3}, \quad q_1, q_2, q_3, q_4 = 1, \dots, Q. \quad (3.39)$$

La vraisemblance marginale correspondante à un individu n peut être approchée par :

$$L_n^*(\mathbf{x}_n, \boldsymbol{\delta}, \boldsymbol{\Sigma}) \approx \left(\prod_{j=1}^J h_{nj}^* \right) \times \int_{\mathbb{R}^2} \left(1 + \frac{1}{2} \sum_{j=1}^J Q_{nj}^* G_{2nj} + \frac{1}{24} \sum_{j=1}^J S_{nj}^* G_{4nj} \right. \\ + \sum_{j_1=1}^J \sum_{j_2=j_1+1}^J A_{nj_1}^* A_{nj_2}^* G_{1nj_1} G_{1nj_2} + \frac{1}{6} \sum_{j_1=1}^J \sum_{j_2=1}^J A_{nj_1}^* R_{nj_2}^* G_{1nj_1} G_{3nj_2} \\ + \frac{1}{4} \sum_{j_1=1}^J \sum_{j_2=j_1+1}^J Q_{nj_1}^* Q_{nj_2}^* G_{2nj_1} G_{2nj_2} \\ + \frac{1}{2} \sum_{j_1=1}^J \sum_{j_2=j_1+1}^J \sum_{j_3=1}^J A_{nj_1}^* A_{nj_2}^* Q_{nj_3}^* G_{1nj_1} G_{1nj_2} G_{2nj_3} \\ \left. + \sum_{j_1=1}^J \sum_{j_2=j_1+1}^J \sum_{j_3=j_2+1}^J \sum_{j_4=j_3+1}^J A_{nj_1}^* A_{nj_2}^* A_{nj_3}^* A_{nj_4}^* G_{1nj_1} G_{1nj_2} G_{1nj_3} G_{1nj_4} \right) G(\boldsymbol{\theta}_n, \boldsymbol{\Sigma}) d\boldsymbol{\theta}_n \quad (3.40)$$

b- Calcul des moments

On vérifie que pour tout j, l, k, m tels que $j, l, k, m = 1, \dots, J$, nous avons :

$$G_{1nj} G_{1nl} = \sum_{q_1=1}^Q \sum_{q_2=1}^Q B_{jq_1} B_{lq_2} \theta_{nq_1} \theta_{nq_2}, \quad (3.41)$$

$$G_{1nj} G_{3nl} = \sum_{q_1=1}^Q \sum_{q_2=1}^Q \sum_{q_3=1}^Q \sum_{q_4=1}^Q B_{jq_1} B_{lq_2} B_{lq_3} B_{lq_4} \theta_{nq_1} \theta_{nq_2} \theta_{nq_3} \theta_{nq_4}, \quad (3.42)$$

$$G_{2nj} G_{2nl} = \sum_{q_1=1}^Q \sum_{q_2=1}^Q \sum_{q_3=1}^Q \sum_{q_4=1}^Q B_{jq_1} B_{jq_2} B_{lq_3} B_{lq_4} \theta_{nq_1} \theta_{nq_2} \theta_{nq_3} \theta_{nq_4}, \quad (3.43)$$

$$G_{1nj} G_{1nl} G_{2nk} = \sum_{q_1=1}^Q \sum_{q_2=1}^Q \sum_{q_3=1}^Q \sum_{q_4=1}^Q B_{jq_1} B_{lq_2} B_{kq_3} B_{kq_4} \theta_{nq_1} \theta_{nq_2} \theta_{nq_3} \theta_{nq_4}, \quad (3.44)$$

$$G_{1nj} G_{1nl} G_{1nk} G_{1nm} = \sum_{q_1=1}^Q \sum_{q_2=1}^Q \sum_{q_3=1}^Q \sum_{q_4=1}^Q B_{jq_1} B_{lq_2} B_{kq_3} B_{mq_4} \theta_{nq_1} \theta_{nq_2} \theta_{nq_3} \theta_{nq_4}. \quad (3.45)$$

Notons

$$H_{2j} = \int_{\mathbb{R}^2} G_{2nj} G(\boldsymbol{\theta}, \boldsymbol{\Sigma}) d\boldsymbol{\theta} = \sum_{q_1=1}^Q \sum_{q_2=1}^Q B_{jq_1} B_{jq_2} \sigma_{(\theta)q_1 q_2}, \quad (3.46)$$

$$H_{2jl} = \int_{\mathbb{R}^2} G_{1nj} G_{1nl} G(\boldsymbol{\theta}, \boldsymbol{\Sigma}) d\boldsymbol{\theta} = \sum_{q_1=1}^Q \sum_{q_2=1}^Q B_{jq_1} B_{lq_2} \sigma(\theta)_{q_1 q_2}, \quad (3.47)$$

$$\begin{aligned} H_{4j} &= \int_{\mathbb{R}^2} G_{4nj} G(\boldsymbol{\theta}, \boldsymbol{\Sigma}) d\boldsymbol{\theta} \\ &= \sum_{q_1=1}^Q \sum_{q_2=1}^Q \sum_{q_3=1}^Q \sum_{q_4=1}^Q B_{jq_1} B_{jq_2} B_{jq_3} B_{jq_4} \left(\sigma(\theta)_{q_1 q_2} \sigma(\theta)_{q_3 q_4} + \sigma(\theta)_{q_1 q_3} \sigma(\theta)_{q_2 q_4} + \sigma(\theta)_{q_1 q_4} \sigma(\theta)_{q_2 q_3} \right), \end{aligned} \quad (3.48)$$

$$\begin{aligned} H_{4j^1 l^3} &= \int_{\mathbb{R}^2} G_{1nj} G_{3nl} G(\boldsymbol{\theta}, \boldsymbol{\Sigma}) d\boldsymbol{\theta} \\ &= \sum_{q_1=1}^Q \sum_{q_2=1}^Q \sum_{q_3=1}^Q \sum_{q_4=1}^Q B_{jq_1} B_{lq_2} B_{lq_3} B_{lq_4} \left(\sigma(\theta)_{q_1 q_2} \sigma(\theta)_{q_3 q_4} + \sigma(\theta)_{q_1 q_3} \sigma(\theta)_{q_2 q_4} + \sigma(\theta)_{q_1 q_4} \sigma(\theta)_{q_2 q_3} \right), \end{aligned} \quad (3.49)$$

$$\begin{aligned} H_{4j^2 l^2} &= \int_{\mathbb{R}^2} G_{2nj} G_{2nl} G(\boldsymbol{\theta}, \boldsymbol{\Sigma}) d\boldsymbol{\theta} \\ &= \sum_{q_1=1}^Q \sum_{q_2=1}^Q \sum_{q_3=1}^Q \sum_{q_4=1}^Q B_{jq_1} B_{jq_2} B_{lq_3} B_{lq_4} \left(\sigma(\theta)_{q_1 q_2} \sigma(\theta)_{q_3 q_4} + \sigma(\theta)_{q_1 q_3} \sigma(\theta)_{q_2 q_4} + \sigma(\theta)_{q_1 q_4} \sigma(\theta)_{q_2 q_3} \right). \end{aligned} \quad (3.50)$$

On montre que

$$H_{2j}^2 = \sum_{q_1=1}^Q \sum_{q_2=1}^Q \sum_{q_3=1}^Q \sum_{q_4=1}^Q B_{jq_1} B_{jq_2} B_{jq_3} B_{jq_4} \sigma(\theta)_{q_1 q_2} \sigma(\theta)_{q_3 q_4} = \frac{H_{4j}}{3}, \quad (3.51)$$

$$H_{2j} H_{2l} = \sum_{q_1=1}^Q \sum_{q_2=1}^Q \sum_{q_3=1}^Q \sum_{q_4=1}^Q B_{jq_1} B_{jq_2} B_{lq_3} B_{lq_4} \sigma(\theta)_{q_1 q_2} \sigma(\theta)_{q_3 q_4}. \quad (3.52)$$

$$\begin{aligned} P(x_{nj}/\mathbf{B}_j, \boldsymbol{\delta}) &= \int_{\mathbb{R}^Q} h_{nj} dG(\boldsymbol{\theta}, \boldsymbol{\Sigma}) \\ &= h_{nj}^* \int_{\mathbb{R}^Q} 1 + A_{nj}^* G_{1n} + \frac{1}{2} Q_{nj}^* G_{2n} + \frac{1}{6} R_{nj}^* G_{3n} + \frac{1}{24} S_{nj}^* G_{4n} dG(\boldsymbol{\theta}, \boldsymbol{\Sigma}) \\ &= h_{nj}^* \left(1 + \frac{1}{2} Q_{nj}^* H_{2j} + \frac{1}{24} S_{nj}^* H_{4j} \right), \end{aligned} \quad (3.53)$$

$$\begin{aligned} P(x_{nj}, x_{nl}/\mathbf{B}_j, \boldsymbol{\delta}) &= h_{nj}^* h_{nl}^* \left(1 + \frac{1}{2} Q_{nj}^* H_{2j} + \frac{1}{2} Q_{nl}^* H_{2l} + \frac{1}{24} S_{nj}^* H_{4j} + \frac{1}{24} S_{nl}^* H_{4l} \right. \\ &\quad \left. A_{nj}^* A_{nl}^* H_{2jl} + \frac{1}{6} A_{nj}^* R_{nl}^* H_{4j^1 l^3} + \frac{1}{6} R_{nj}^* A_{nl}^* H_{4j^3 l^1} + \frac{1}{4} Q_{nj}^* Q_{nl}^* H_{4j^2 l^2} \right). \end{aligned} \quad (3.54)$$

Posons

$$l_j(y) = \frac{1}{1 + \exp(y - \delta_j)}, \quad (3.55)$$

$$l_j^{(k)}(y) = \frac{d^k l_j(y)}{dy^k}, \quad k = 1, \dots, 5, \quad (3.56)$$

et $l_j^*, l_j^{*(k)}$, $k = 1, \dots, 5$ les valeurs de ces fonctions en $y = 0$. En remarquant que pour $x_{nj} = 1$, on a

$$h_{nj}^* = l_j^*, \quad (3.57)$$

$$A_{nj}^* = \frac{l_j^{*(1)}}{l_j^*}, \quad (3.58)$$

$$C_{nj}^* = l_j^{*(1)}, \quad (3.59)$$

$$D_{nj}^* = l_j^{*(2)}, \quad (3.60)$$

$$E_{nj}^* = l_j^{*(3)}, \quad (3.61)$$

$$Q_{nj}^* = \frac{l_j^{*(2)}}{l_j^*}, \quad (3.62)$$

$$R_{nj}^* = \frac{l_j^{*(3)}}{l_j^*}, \quad (3.63)$$

$$S_{nj}^* = \frac{l_j^{*(4)}}{l_j^*}, \quad (3.64)$$

on obtient

$$\begin{aligned} \mu_j &= E(X_{nj}) = P(X_{nj} = 1/\mathbf{B}_j, \boldsymbol{\delta}) \\ &= l_j^* + \frac{H_{2j}}{2} l_j^{*(2)} + \frac{H_{4j}}{24} l_j^{*(4)}, \end{aligned} \quad (3.65)$$

$$\begin{aligned} \sigma_{jj} &= \mu_j(1 - \mu_j) \\ &= l_j^{*(1)} + \frac{H_{2j}}{2} \left(l_j^{*(2)} - 2l_j^* l_j^{*(2)} \right) + \frac{H_{4j}}{24} \left(l_j^{*(4)} - 2l_j^* l_j^{*(4)} - 2l_j^{*(2)} l_j^{*(2)} \right), \end{aligned} \quad (3.66)$$

$$\begin{aligned} \sigma_{jl} &= E(X_{nj} X_{nl}) - E(X_{nj}) E(X_{nl}) = P(X_{nj} = 1, X_{nl} = 1/\mathbf{B}_j, \boldsymbol{\delta}) - \mu_j \mu_l \\ &= H_{2jl} l_j^{*(1)} l_l^{*(1)} + \frac{H_{4j} l^3}{6} l_j^{*(1)} l_l^{*(3)} + \frac{H_{4j} l^3}{6} l_j^{*(3)} l_l^{*(1)} + \frac{H_{4j} l^2 l^2 - H_{2j} H_{2l}}{4} l_j^{*(2)} l_l^{*(2)}. \end{aligned} \quad (3.67)$$

c- Les équations d'estimations généralisées

Posons

$$\boldsymbol{\mu} = E(\mathbf{X}_n) = (\mu_j)_{j=1, \dots, J}, \quad (3.68)$$

$$s_{n,jl} = (x_{nj} - \mu_j)(x_{nl} - \mu_l), \quad (3.69)$$

$$\mathbf{s}_n = (s_{n,jl})_{1 \leq j < l \leq J}^t, \quad (3.70)$$

$$\boldsymbol{\eta} = E(\mathbf{S}_n) = (\sigma_{jl})_{1 \leq j < l \leq J}, \quad (3.71)$$

avec \mathbf{X}_n et \mathbf{S}_n les variables aléatoires dont les réalisations sont notées par \mathbf{x}_n et \mathbf{s}_n .

Soit $\boldsymbol{\phi} = (\boldsymbol{\mu}, \boldsymbol{\eta})'$ et $\boldsymbol{\xi}_n = (x_n, s_n)'$.

Soient

$$\mathbf{V} = \begin{pmatrix} (\sigma_{jl})_{j,l=1\dots J} & COV(\mathbf{x}_n, \mathbf{s}_n) \\ COV(\mathbf{s}_n, \mathbf{x}_n) & COV(\mathbf{s}_n) \end{pmatrix}, \quad (3.72)$$

$$\boldsymbol{\delta} = (\delta_j)_{j=1,\dots,J}, \quad (3.73)$$

$$\boldsymbol{\alpha} = (\sigma_{(\theta)qq})_{q=1,\dots,Q}, \quad (3.74)$$

$$\boldsymbol{\gamma} = (\sigma_{(\theta)q_1q_2})_{1 \leq q_1 < q_2 \leq Q}, \quad (3.75)$$

$$\mathbf{D} = \begin{pmatrix} \left(\frac{d\boldsymbol{\mu}}{d\boldsymbol{\delta}} \right) & \left(\frac{d\boldsymbol{\mu}}{d\boldsymbol{\alpha}} \right) & \left(\frac{d\boldsymbol{\mu}}{d\boldsymbol{\gamma}} \right) \\ \left(\frac{d\boldsymbol{\eta}}{d\boldsymbol{\delta}} \right) & \left(\frac{d\boldsymbol{\eta}}{d\boldsymbol{\alpha}} \right) & \left(\frac{d\boldsymbol{\eta}}{d\boldsymbol{\gamma}} \right) \end{pmatrix}. \quad (3.76)$$

Les GEE sont de la forme :

$$U(\boldsymbol{\delta}, \boldsymbol{\alpha}, \boldsymbol{\gamma}) = \mathbf{D}'\mathbf{V}^{-1} \sum_{n=1}^N (\boldsymbol{\xi}_n - \boldsymbol{\phi}) = 0 \quad (3.77)$$

On les résoud par l'algorithme de Fisher-scoring en utilisant $\hat{\boldsymbol{\delta}}^{(k)}$, $\hat{\boldsymbol{\alpha}}^{(k)}$, $\hat{\boldsymbol{\gamma}}^{(k)}$, $\hat{\mathbf{D}}^{(k)}$, $\hat{\mathbf{V}}^{(k)}$, $\hat{\boldsymbol{\xi}}_n^{(k)}$ et $\hat{\boldsymbol{\phi}}^{(k)}$ les estimations respectives de $\boldsymbol{\delta}$, $\boldsymbol{\alpha}$, $\boldsymbol{\gamma}$, \mathbf{D} , \mathbf{V} , $\boldsymbol{\xi}_n$ et $\boldsymbol{\phi}$ à l'étape k et dont les équations itératives à l'étape $(k+1)$ sont données par :

$$\begin{pmatrix} \hat{\boldsymbol{\delta}}^{(k+1)} \\ \hat{\boldsymbol{\alpha}}^{(k+1)} \\ \hat{\boldsymbol{\gamma}}^{(k+1)} \end{pmatrix} = \begin{pmatrix} \hat{\boldsymbol{\delta}}^{(k)} \\ \hat{\boldsymbol{\alpha}}^{(k)} \\ \hat{\boldsymbol{\gamma}}^{(k)} \end{pmatrix} + \left(\hat{\mathbf{D}}^{(k)'} \hat{\mathbf{V}}^{(k)-1} \hat{\mathbf{D}}^{(k)} \right)^{-1} \hat{\mathbf{D}}^{(k)'} \hat{\mathbf{V}}^{(k)-1} \sum_{n=1}^N (\hat{\boldsymbol{\xi}}_n^{(k)} - \boldsymbol{\phi}^{(k)}) \quad (3.78)$$

d- Estimation finale de la log-vraisemblance marginale

La log-vraisemblance marginale peut être estimée par quadratures de Gauss-Hermite après la dernière itération. Celle-ci est égale à :

$$L_M(\hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\Sigma}}/\mathbf{x}) = \int_{\mathbb{R}} \prod_{n=1}^N P(\mathbf{X}_n = \mathbf{x}_n/\boldsymbol{\theta}, \hat{\boldsymbol{\delta}}) G(\boldsymbol{\theta}/\hat{\boldsymbol{\Sigma}}) d\boldsymbol{\theta} \quad (3.79)$$

On se place dans le cadre bi-dimensionnel, avec $\boldsymbol{\theta} = (\theta_1, \theta_2)$ et $f(\theta_1, \theta_2) = \prod_{n=1}^N P(\mathbf{X}_n = \mathbf{x}_n/\boldsymbol{\theta}, \hat{\boldsymbol{\delta}})$. Soient θ_1^* et θ_2^* telles que $(\theta_1^*, \theta_2^*)'(\theta_1^*, \theta_2^*) = \frac{(\theta_1, \theta_2)' \hat{\boldsymbol{\Sigma}}^{-1} (\theta_1, \theta_2)}{2}$.

On a donc $(\theta_1^*, \theta_2^*) = \frac{1}{\sqrt{2}} \hat{\boldsymbol{\Sigma}}^{-1/2} (\theta_1, \theta_2)$ avec $\hat{\boldsymbol{\Sigma}}^{-1/2}' \hat{\boldsymbol{\Sigma}}^{-1/2} = \hat{\boldsymbol{\Sigma}}^{-1}$ donc $\hat{\boldsymbol{\Sigma}}^{-1/2}$ est la matrice de Cholesky de $\hat{\boldsymbol{\Sigma}}^{-1}$.

On montre que si $\hat{\boldsymbol{\Sigma}} = \begin{pmatrix} \hat{\sigma}_{11} & \hat{\sigma}_{12} \\ \hat{\sigma}_{12} & \hat{\sigma}_{22} \end{pmatrix}$, alors $\hat{\boldsymbol{\Sigma}}^{-1} = \frac{1}{|\hat{\boldsymbol{\Sigma}}|} \begin{pmatrix} \hat{\sigma}_{22} & -\hat{\sigma}_{12} \\ -\hat{\sigma}_{12} & \hat{\sigma}_{11} \end{pmatrix}$,

$$\hat{\boldsymbol{\Sigma}}^{-1/2} = \frac{1}{\sqrt{|\hat{\boldsymbol{\Sigma}}|}} \begin{pmatrix} \sqrt{\hat{\sigma}_{22}} & 0 \\ -\frac{\hat{\sigma}_{12}}{\sqrt{\hat{\sigma}_{22}}} & \sqrt{\hat{\sigma}_{11} - \frac{\hat{\sigma}_{12}^2}{\hat{\sigma}_{22}}} \end{pmatrix}. \quad (3.80)$$

Selon le théorème du changement de variable, on a :

$$\int f(\mathbf{a})d\mathbf{a} = \int f(\mathbf{a}^*)|J(\mathbf{a}/\mathbf{a}^*)|d\mathbf{a}^* \quad (3.81)$$

Avec $J(\mathbf{a}/\mathbf{a}^*)$ la matrice jacobienne de \mathbf{a} en \mathbf{a}^* . On montre que :

$$\theta_1 = \sqrt{\frac{2|\hat{\Sigma}|}{\hat{\sigma}_{22}}}\theta_1^*, \quad (3.82)$$

$$\theta_2 = \sqrt{2\hat{\sigma}_{22}}\theta_2^* + \frac{\hat{\sigma}_{12}}{\sqrt{|\hat{\Sigma}|}}\theta_1 = \sqrt{2\hat{\sigma}_{22}}\theta_2^* + \frac{\hat{\sigma}_{12}\sqrt{2}}{\sqrt{\hat{\sigma}_{22}}}\theta_1^*. \quad (3.83)$$

La matrice jacobienne de $\boldsymbol{\theta}$ en $\boldsymbol{\theta}^*$ vaut :

$$J(\boldsymbol{\theta}/\boldsymbol{\theta}^*) = \begin{pmatrix} \frac{d\theta_1}{d\theta_1^*} & \frac{d\theta_1}{d\theta_2^*} \\ \frac{d\theta_2}{d\theta_1^*} & \frac{d\theta_2}{d\theta_2^*} \end{pmatrix} = \begin{pmatrix} \sqrt{\frac{2|\hat{\Sigma}|}{\hat{\sigma}_{22}}} & 0 \\ \frac{\hat{\sigma}_{12}\sqrt{2}}{\sqrt{\hat{\sigma}_{22}}} & \sqrt{2\hat{\sigma}_{22}} \end{pmatrix}. \quad (3.84)$$

Donc $J(\boldsymbol{\theta}/\boldsymbol{\theta}^*)| = 2\sqrt{|\hat{\Sigma}|}$.

La vraisemblance marginale peut alors être approchée par

$$\begin{aligned} L_M(\hat{\boldsymbol{\delta}}, \hat{\Sigma}) &= \int f(\theta_1, \theta_2)G(\boldsymbol{\theta}/\hat{\Sigma})d\boldsymbol{\theta} \\ &= \int f(\theta_1^*, \theta_2^*)\frac{2\sqrt{|\hat{\Sigma}|}}{2\pi|\hat{\Sigma}|^{1/2}}\exp(-\boldsymbol{\theta}^{*'}\boldsymbol{\theta}^*)d\boldsymbol{\theta}^* \\ &= \int f(\theta_1^*, \theta_2^*)\frac{1}{\pi}\exp(-\boldsymbol{\theta}^{*'}\boldsymbol{\theta}^*)d\boldsymbol{\theta}^* \\ &\approx \frac{1}{\pi}\sum_{u=1}^R\sum_{v=1}^R w_u w_v f(A_u^*, A_v^*), \end{aligned} \quad (3.85)$$

avec R le nombre de points de quadratures utilisés, A_r , $r = 1, \dots, R$ les noeuds de quadrature et w_r , $r = 1, \dots, R$ les poids associés. Les noeuds transformés A_u^* et A_v^* sont calculés ainsi :

$$A_u^* = \sqrt{\frac{\hat{\sigma}_{22}}{2|\hat{\Sigma}|}}A_u, \quad (3.86)$$

$$A_v^* = \left(A_v - \frac{\hat{\sigma}_{12}}{\sqrt{|\hat{\Sigma}|}}A_u \right) / \sqrt{2\hat{\sigma}_{22}}. \quad (3.87)$$

e- Variance des estimations

La matrice de variance-covariance \mathbf{W} des estimations $(\hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\gamma}})'$ est obtenue par l'estimateur sandwich [33] :

$$\mathbf{W} = \frac{1}{N^2} \left(\hat{\mathbf{D}}\hat{\mathbf{V}}^{-1}\hat{\mathbf{D}} \right)^{-1} \left(\hat{\mathbf{D}}\hat{\mathbf{V}}^{-1} \sum_{n=1}^N (\hat{\boldsymbol{\xi}}_n - \hat{\boldsymbol{\phi}})(\hat{\boldsymbol{\xi}}_n - \hat{\boldsymbol{\phi}})'\hat{\mathbf{V}}^{-1}\hat{\mathbf{D}} \right) \left(\hat{\mathbf{D}}\hat{\mathbf{V}}^{-1}\hat{\mathbf{D}} \right)^{-1}, \quad (3.88)$$

où $\hat{\mathbf{D}}$, $\hat{\mathbf{V}}$, $\hat{\boldsymbol{\xi}}_n$ et $\hat{\boldsymbol{\phi}}$ sont les approximations de \mathbf{D} , \mathbf{V} , $\boldsymbol{\xi}_n$ et $\boldsymbol{\phi}$ au point $(\hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\gamma}})'$.

3.4.3.7 Le modèle de Rasch multidimensionnel généralisé (Generalized Multidimensional Rasch Model) (GMRM)

Hoijtink, Rooks et Wilmlink [74] proposent un modèle dichotomique plus spécifique nommé le "Generalized Multidimensional Rasch Model" (GMRM), présenté comme un cas particulier du MPLT (c'est pourquoi nous le présentons dans cette section) et comme une extrapolation multidimensionnelle de l'OPLM. Dans ce modèle, des restrictions plus fortes sont posées sur les poids B_{jq} . Ils considèrent en effet que chaque item doit avoir le même poids global B , c'est-à-dire que $\forall j = 1, \dots, J, \sum_{q=1}^Q B_{jq} = B$. En outre, ils restreignent les poids à $Q + 1$ valeurs : $0, 1, \dots, Q$.

3.4.3.8 Interprétation des paramètres de difficulté des items

Etant donné que les paramètres δ_{jqx} ne sont pas identifiables, on ne peut estimer dans le MPLT que les sommes $\delta_{jx} = \sum_{q=1}^Q B_{jqx} \delta_{jqx}$ ($\delta_j = \sum_{q=1}^Q \delta_{jq}$ dans le cadre dichotomique). Ce problème d'identifiabilité pose des problèmes d'interprétations des paramètres caractérisant les items : à quoi correspond la valeur $\hat{\delta}_{jx}$ ($\hat{\delta}_j$) ? Kelderman et Rijkes [87] éludent le problème en ne proposant pas d'interprétations de ces paramètres et en comparant graphiquement, via les courbes caractéristiques des items, des items dans une étude sur des données réelles. Hoijtink, Rooks et Wilmlink [74] parlent plus largement de ce problème : ils interprètent le paramètre δ_j comme une difficulté moyenne de l'item j sur l'ensemble des Q traits latents. Ils proposent en outre de comparer seulement les difficultés des items j et j' qui auraient des poids identiques sur toutes les dimensions $\forall j, j' = 1, \dots, J; j \neq j', \forall q = 1, \dots, Q, B_{jq} = B_{j'q}$.

3.4.3.9 Lien avec les modèles unidimensionnels

Il est possible de retrouver avec le MPLT certains modèles unidimensionnels en posant des contraintes sur les poids B_{jqx} dans un cadre unidimensionnel, $Q = 1$ et on simplifie B_{jqx} par B_{jx} (et dans un cadre dichotomique B_{jq} par B_j), et θ_{nq} par θ_n . L'équation (3.7) s'écrit dans un cadre unidimensionnel :

$$P(X_{nj} = x/\theta_n, \boldsymbol{\delta}_j) = \frac{\exp(B_{jx}\theta_n - \delta_{jx})}{\sum_{y=0}^{m_j} \exp(B_{jy}\theta_n - \delta_{jy})}, \quad (3.89)$$

et plus spécifiquement avec des items dichotomiques (cf équation (3.14)) :

$$P(X_{nj} = x/\theta_n, \delta_j) = \frac{\exp(x(B_j\theta_n - \delta_j))}{1 + \exp(B_j\theta_n - \delta_j)}. \quad (3.90)$$

L'équation (3.90) représente de fait l'OPLM (voir section 2.7.1, après une transformation des paramètres de difficulté δ_j). En posant les contraintes $B_j = 1, \forall j = 1, \dots, J$, on retrouve le modèle de Rasch.

A partir de l'équation (3.89), en posant les contraintes $B_{jx} = x, \forall j = 1, \dots, J, \forall x = 1, \dots, m_j$, on retrouve le Partial Credit Model (voir section 2.8.3, après une transformation des δ_{jx}).

3.4.4 Le modèle logistique multinomial à coefficients aléatoires multidimensionnel (Multidimensional Random Coefficients Multinomial Logit Model) (MRCML)

Le MRCML a été défini en parallèle du MPLT, dont il est très proche, par Adams, Wilson et Wang [2] : le MRCML décrit matriciellement le même modèle que le MPLT décrit analytiquement. Il a vocation à généraliser les modèles de la famille de Rasch, tant unidimensionnels que multidimensionnels.

Soit un ensemble d'items polytomiques Y_j , $j = 1, \dots, J$ à m_j modalités positives et soit, pour chaque item Y_j un sous-ensemble de variables aléatoires X_{jk} , $k = 0, \dots, m_j$ telles que :

$$X_{jk} = \begin{cases} 1, & \text{si } Y_j = k \\ 0, & \text{sinon} \end{cases} \quad (3.91)$$

Soient $\mathbf{X}_j = (X_{jk})_{k=1, \dots, m_j}$, $\mathbf{X} = (X_j)_{j=1, \dots, J}$ deux vecteurs de tailles respectives m_j et $K = \sum_{j=1}^J m_j$ et \mathbf{x}_{nj} et \mathbf{x}_n les réalisations de ces deux vecteurs aléatoires pour l'individu n .

Soient le vecteur $\boldsymbol{\delta} = (\delta_{jk})_{j=1, \dots, J, k=1, \dots, m_j}$ de dimension K et $\boldsymbol{\Theta} = (\Theta_q)_{q=1, \dots, Q}$ le vecteur aléatoire des Q traits latents dont la réalisation pour l'individu n est notée $\boldsymbol{\theta}_n$.

Soit le vecteur $\mathbf{a}_{jk} = (a_{jkl})_{l=1, \dots, K}$ où $a_{jkl} \neq 0$ si l'on modélise la k^e fonction caractéristique de l'item j en fonction du l^e paramètre de $\boldsymbol{\delta}$ et soit la matrice $\mathbf{A} = (\mathbf{a}_{jk})'_{j=1, \dots, J, k=1, \dots, m_j}$.

Soient enfin le vecteur $\mathbf{B}_{jx} = (B_{jqx})_{q=1, \dots, Q}$, et les matrices $\mathbf{B}_j = (\mathbf{B}_{jx})'_{x=1, \dots, m_j}$ et $\mathbf{B} = (\mathbf{B}_j)_{j=1, \dots, J}$.

Le modèle s'écrit alors :

$$P(\mathbf{X}_n = \mathbf{x}_n / \boldsymbol{\theta}_n; \mathbf{A}, \mathbf{B}, \boldsymbol{\delta}) = \Psi(\boldsymbol{\theta}_n, \boldsymbol{\delta}) \exp [\mathbf{x}_n' (\mathbf{B}\boldsymbol{\theta}_n + \mathbf{A}\boldsymbol{\delta})], \quad (3.92)$$

où

$$\Psi(\boldsymbol{\theta}_n, \boldsymbol{\delta}) = \left\{ \sum_{\mathbf{z} \in \Omega} \exp [\mathbf{z}' (\mathbf{B}\boldsymbol{\theta}_n + \mathbf{A}\boldsymbol{\delta})] \right\}^{-1}, \quad (3.93)$$

avec Ω l'ensemble de tous les vecteurs \mathbf{z} possibles.

Les auteurs du modèle montrent qu'il est possible de considérer le trait latent multidimensionnel soit comme un ensemble de paramètres fixes (et proposent l'estimation par maximum de vraisemblance conditionnelle), soit comme un vecteur aléatoire. Dans ce dernier cadre, ils proposent une estimation des paramètres par l'algorithme EM [2] [3].

3.5 Autres modèles de l'IRT multidimensionnelle

3.5.1 Les extensions des modèles unidimensionnels

Les extensions multidimensionnelles des modèles unidimensionnels sont souvent retrouvées dans la littérature. Ce sont les formes les plus simples des modèles multidimensionnels, mais

nous n'avons pas retrouvé de littérature étudiant de façon précise ces modèles. Ils consistent à extrapoler la référence au trait latent unidimensionnel dans les modèles classiques de l'IRT par une somme pondérée des Q composantes du trait latent multidimensionnel. Les pondérations utilisées correspondent alors à des pouvoirs discriminants reliant chaque item à chaque trait latent. Un exemple courant est celui du modèle de Birnbaum (modèle logistique à deux paramètres) dont la forme unidimensionnelle est :

$$P(X_{nj} = x_{nj}/\theta_n; \alpha_j, \delta_j) = \frac{\exp [x_{nj}\alpha_j (\theta_n - \delta_j)]}{1 + \exp [\alpha_j (\theta_n - \delta_j)]}, \quad (3.94)$$

et dont l'extension multidimensionnelle souvent rencontrée (voir par exemple [131], [1] ou [144]) est

$$P(X_{nj} = x_{nj}/\theta_n; \alpha_j, \delta_j) = \frac{\exp \left[x_{nj} \left(\sum_{q=1}^Q \alpha_{jq} \theta_{nq} - \delta_j^* \right) \right]}{1 + \exp \left(\sum_{q=1}^Q \alpha_{jq} \theta_{nq} - \delta_j^* \right)}, \quad (3.95)$$

avec $\alpha_j = (\alpha_{jq})_{q=1, \dots, Q}$ et δ_j^* correspond à la difficulté moyenne (pondérée par les α_{jq}) de l'item j sur les Q traits latents. Il est ainsi possible d'étendre tous les modèles de l'IRT unidimensionnelle à un cadre multidimensionnel.

3.5.2 Les extensions longitudinales

On trouve parfois la notion de modèles multidimensionnels pour parler de modèles longitudinaux (voir par exemple [31]). Nous ne considérerons pas ces modèles à proprement parler comme des modèles multidimensionnels. En effet, dans les modèles longitudinaux, si l'on s'intéresse aux relations entre les mesures d'un même trait latent à différents instants, le problème ne se pose pas tout à fait de la même façon, même si globalement, les modèles peuvent amplement se ressembler. La principale différence entre les deux approches tient sans doute au fait que dans un modèle multidimensionnel, on pose un paramètre de difficulté pour chaque item dans chacune des dimensions, or dans un modèle longitudinal, la difficulté d'un item n'évolue pas au cours du temps, ce qui suppose donc qu'il n'y a qu'un seul paramètre de difficulté par item. Il est vrai, que dans les modèles multidimensionnels, seule une somme pondérée des paramètres de difficulté est estimable pour chaque item, ce qui fait qu'en pratique les modèles restent globalement similaires.

3.6 Synthèse

Ce chapitre fait état des modèles de l'IRT multidimensionnelle retrouvée dans la littérature et utilisés pour l'analyse de la qualité de vie. Ces modèles restent cependant complexes en ce qui concerne l'estimation de leurs paramètres mais surtout de l'interprétation de ces paramètres. Le chapitre suivant est donc consacré à un nouveau modèle multidimensionnel présenté en [68] qui permet une interprétation aisée de l'ensemble de ses paramètres, et permet l'utilisation de scores pour la mesure de chacun de ses traits latents.

Chapitre 4

Une proposition : le modèle (polytomique) multidimensionnel de Rasch marginalement exhaustif (M(P)MSRM)

4.1 Un nouveau modèle

Alors qu'il existe déjà un certain nombre de modèles, on peut se demander l'intérêt de définir un nouveau modèle. Le modèle que nous proposons est en effet un modèle spécifique que nous avons défini dans un cadre exploratoire (voir section 7.3). Ce modèle "corrige" certains défauts dans un tel cadre du modèle multidimensionnel de Rasch, du MPLT et du GMRM. Ces défauts sont principalement :

- ces modèles peuvent être difficile d'usage dans un cadre exploratoire, c'est-à-dire dans une analyse ayant pour but de définir les liens entre les items et les traits latents, en particulier pour le modèle multidimensionnel de Rasch qui suppose une connaissance de ces relations lors de la construction des items,
- pour le MPLT, le GMRM, et le MRCML le vecteur des sous-scores est exhaustif du trait latent multidimensionnel, mais ces deux modèles ne vérifient pas une propriété plus spécifique que nous appellerons l'exhaustivité marginale : chaque sous-score doit être exhaustif du trait latent correspondant, ce qui implique que marginalement aux autres items, tous les items reliés à une même dimension vérifient un modèle de Rasch ;
- pour le MPLT et le MRCML, dans un cadre unidimensionnel, le modèle n'est pas forcément un modèle de Rasch (cadre dichotomique), ni une extrapolation polytomique du modèle de Rasch (cadre polytomique), mais plutôt un OPLM.

Le modèle présenté ci-après a fait l'objet d'une large présentation en [68].

4.2 L'exhaustivité marginale

La propriété d'exhaustivité marginale est définie de la façon suivante :

Propriété 1 *L'exhaustivité marginale est définie par le fait qu'il existe un sous-score S_q^* exhaustif du trait latent Θ_q et ce, pour chacun des traits latents Θ_q , ($q = 1, \dots, Q$).*

4.3 Modèles marginaux

Propriété 2 *Soient $\mathbf{X} = (X_j)_{j=1, \dots, J_X}$ et $\mathbf{Y} = (Y_j)_{j=1, \dots, J_Y}$ deux vecteurs de réponses à deux ensembles distincts d'items dont les réponses dépendent d'un trait latent multidimensionnel $\boldsymbol{\theta} = (\theta_q)_{q=1, \dots, Q}$. Les items dont les réponses appartiennent à \mathbf{X} sont ceux qui dépendent du trait latent $\theta_r \in \boldsymbol{\theta}$. Soit Ω l'ensemble de tous les vecteurs \mathbf{Y} possibles. Soit $\boldsymbol{\theta}_r^* = (\theta_q)_{q=1, \dots, r-1, r+1, \dots, Q}$. Marginalement aux items de \mathbf{Y} et à $\boldsymbol{\theta}_r^*$, les items de \mathbf{X} suivent un modèle de Rasch relativement au trait latent θ_r . Analytiquement, en posant $\mathbf{x} = (x_j)_{j=1, \dots, J_X}$ cela correspond à :*

$$\int_{\mathbb{R}^{Q-1}} \sum_{\mathbf{y} \in \Omega} P(\mathbf{X} = \mathbf{x}, \mathbf{Y} = \mathbf{y} / \boldsymbol{\theta}; \boldsymbol{\delta}) G(\boldsymbol{\theta}_r^*) d\boldsymbol{\theta}_r^* = P(\mathbf{X} = \mathbf{x} / \theta_r, \boldsymbol{\delta}) = \prod_{j=1}^{J_X} \frac{\exp(x_j(\theta_r - \delta_j))}{1 + \exp(\theta_r - \delta_j)}$$

4.4 Définition

Le modèle multidimensionnel marginalement exhaustif de Rasch (Multidimensional Marginally Sufficient Rasch Model - MMSRM) est défini comme un modèle de l'IRT vérifiant la propriété d'exhaustivité marginale et la propriété 2.

4.5 Proposition

Soient Q ensembles distincts d'items $\mathbf{X}^{(q)}_{q=1, \dots, Q}$ vérifiant chacun un modèle de Rasch suivant le trait latent Θ_q , $q = 1, \dots, Q$. Soit J_q le nombre d'items de l'ensemble $\mathbf{X}^{(q)}$, $q = 1, \dots, Q$. Soit $\delta_j^{(q)}$ le paramètre de difficulté associé au j^e item de l'ensemble $\mathbf{X}^{(q)}$. Soit $\mathbf{X} = \bigcup_{q=1, \dots, Q} \mathbf{X}^{(q)}$. La fonction réponse du j^e item de l'ensemble $\mathbf{X}^{(q)}$ est donc :

$$P(X_{nj}^{(q)} = x_{nj}^{(q)} / \theta_{nq}; \delta_j^{(q)}) = \frac{\exp \left[x_{nj}^{(q)} (\theta_{nq} - \delta_j^{(q)}) \right]}{1 + \exp (\theta_{nq} - \delta_j^{(q)})} \quad (4.1)$$

Par l'hypothèse d'indépendance locale des items, définissant les modèles de l'IRT multidimensionnelle, on obtient :

$$P(\mathbf{X}_n = \mathbf{x}_n / \boldsymbol{\theta}_n; \boldsymbol{\delta}) = \prod_{q=1}^Q \prod_{j=1}^{J_q} \frac{\exp \left[x_{nj}^{(q)} (\theta_{nq} - \delta_j^{(q)}) \right]}{1 + \exp (\theta_{nq} - \delta_j^{(q)})} \quad (4.2)$$

Lorsque le trait latent multidimensionnel est considéré comme une variable aléatoire multivariée, il reste à définir la fonction de distribution du trait latent multidimensionnel que nous noterons $G(\boldsymbol{\theta})$. Il est alors possible d'écrire la loi jointe du vecteur \mathbf{X}_n et du trait latent $\boldsymbol{\theta}_n$:

$$f(\mathbf{X}_n = \mathbf{x}_n, \boldsymbol{\theta}_n / \boldsymbol{\delta}) = P(\mathbf{X}_n = \mathbf{x}_n / \boldsymbol{\theta}_n; \boldsymbol{\delta}) G(\boldsymbol{\theta}_n) \quad (4.3)$$

Par construction, comme chaque ensemble d'items $\mathbf{X}^{(q)}$ vérifie un modèle de Rasch, on sait que le sous-score $S_{nq} = \sum_{j=1}^{J_q} X_{nj}^{(q)}$ est exhaustif du trait latent Θ_q donc ce modèle vérifie la propriété d'exhaustivité marginale.

Toujours par construction, en utilisant la propriété d'indépendance local et les fonctions réponse aux items liés au trait latent θ_q définis à l'équation (4.1), on montre aisément que ces items, marginalement aux autres items et aux autres traits latents, suivent un modèle de Rasch :

$$P(\mathbf{X}_n^{(q)} = \mathbf{x}_n^{(q)} / \theta_{nq}; \boldsymbol{\delta}) = \prod_{j=1}^{J_q} \frac{\exp [x_{nj}^{(q)} (\theta_{nq} - \delta_j^{(q)})]}{1 + \exp (\theta_{nq} - \delta_j^{(q)})} \quad (4.4)$$

Donc le modèle proposé vérifie la propriété 2, et est donc un modèle multidimensionnel marginalement exhaustif de Rasch.

Le MMSRM modélise une structure simple au sens de Stout [140], puisque chaque item est lié à un seul trait latent.

4.6 Représentation graphique

En considérant le trait latent multidimensionnel comme un ensemble de paramètres fixes, on peut représenter un tel modèle comme à la figure 4.1.

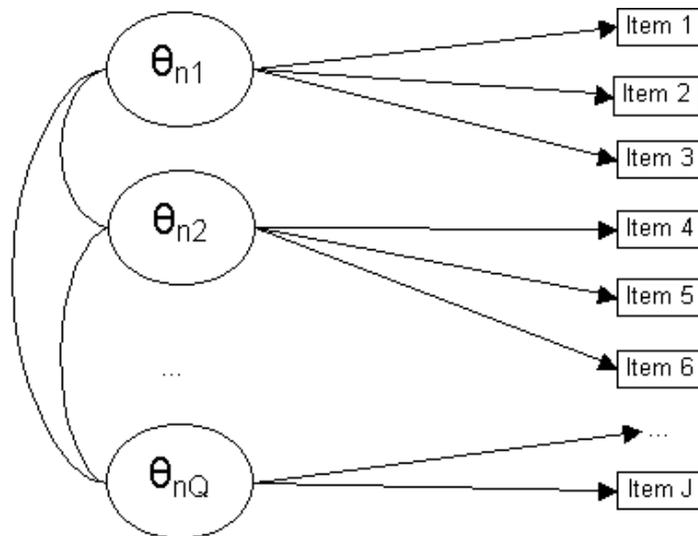


FIG. 4.1 – Représentation graphique du modèle multidimensionnel marginalement exhaustif de Rasch

La figure 4.2 représente la propriété d'exhaustivité marginale dans le cadre du MMSRM (en considérant les paramètres θ_{nq} , $q = 1, \dots, Q$ comme des paramètres fixes). Cette figure est à mettre en lien avec la figure 3.2 (page 49) présentant la propriété d'exhaustivité du vecteur

des sous-scores (\mathbf{s}_n) sur le trait latent multidimensionnel ($\boldsymbol{\theta}_n$).

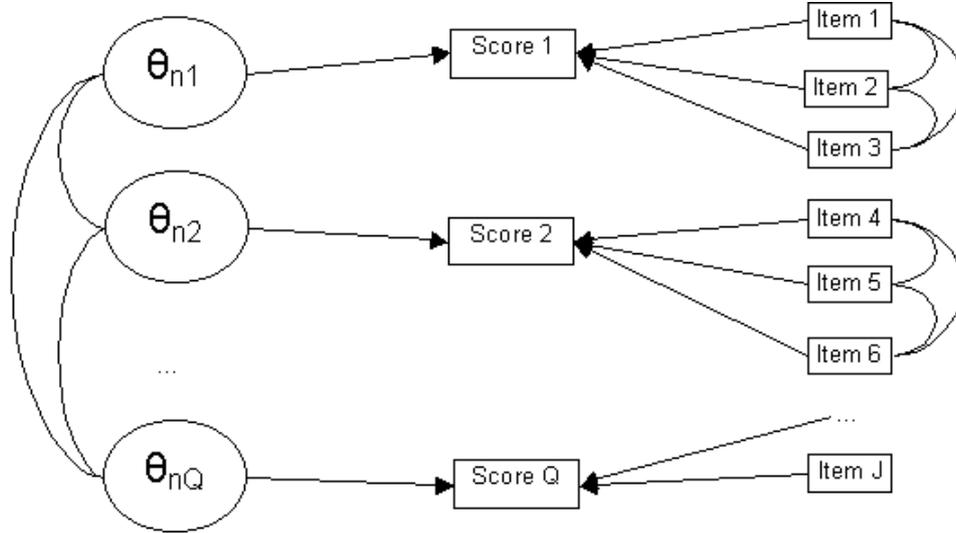


FIG. 4.2 – Représentation graphique de la propriété d'exhaustivité marginale dans le cadre du MMSRM

4.7 Estimation

Il existe deux manières de considérer les traits latents :

- soit on considère les paramètres θ_{nq} , $n = 1, \dots, N$, $q = 1, \dots, Q$ comme des paramètres fixes,
- soit on considère le vecteur $\boldsymbol{\theta}_n$ comme une réalisation d'une variable aléatoire multidimensionnelle $\boldsymbol{\Theta}$.

Dans le premier cas, il est possible de considérer ce modèle comme un cas particulier du MPLT dans le cadre dichotomique, sous les contraintes

$$\forall j = 1, \dots, J, \exists q_j / B_{jq_j} = 1 \text{ et } B_{jq} = 0 \text{ si } q \neq q_j,$$

où q_j est alors l'indice du seul trait latent auquel est relié l'item j . L'estimation peut alors être réalisée par maximum de vraisemblance conditionnelle telle que explicité à la section 3.4.3.4.

Dans le second cas, il faut définir la distribution du trait latent multidimensionnel : par extrapolation du modèle de Rasch, il est d'usage d'utiliser une loi multinormale centrée de matrice de variance $\boldsymbol{\Sigma}$, même si n'importe quelle fonction de distribution peut être utilisée à la place. L'estimation se fait alors par maximum de vraisemblance marginale, comme explicité à la section 3.4.3.4.

Dans ce dernier cas, le modèle est plus proche par définition du MRCML, dont il est un cas particulier en posant $\boldsymbol{\delta} = (\delta_j)_{j=1,\dots,J}$, $\mathbf{A} = \mathbf{I}_J$, $\mathbf{x}_n = (x_{nj})_{j=1,\dots,J}$ et

$$\mathbf{B} = \begin{pmatrix} B_{11} & \cdots & B_{1q} & \cdots & B_{1Q} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ B_{1q} & \cdots & B_{jq} & \cdots & B_{jQ} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ B_{J1} & \cdots & B_{Jq} & \cdots & B_{JQ} \end{pmatrix} \quad (4.5)$$

En posant par exemple les J_1 premiers items comme reliés à la première dimension ($B_{j1} = 1$), les J_2 items suivants reliés à la seconde dimension, etc..., la matrice \mathbf{B} se présente alors sous la forme :

$$\mathbf{B} = \begin{pmatrix} \mathbf{1}_{J_1} & & & \mathbf{0} & & \\ \mathbf{0} & \mathbf{1}_{J_2} & & \mathbf{0} & & \\ & \ddots & & \ddots & & \\ & & \mathbf{0} & & \mathbf{1}_{J_q} & \mathbf{0} \\ & & & \ddots & & \ddots \\ & & & \mathbf{0} & & \mathbf{1}_{J_Q} \end{pmatrix} \quad (4.6)$$

Ce modèle fait partie de ce que Adams, Wilson et Wang [2] nomment un modèle multidimensionnel entre items (multidimensional between items model), puisque qu'il revient à considérer plusieurs modèles unidimensionnels reliés ensemble avec cependant possibilité d'avoir des dépendances entre les diverses composantes du trait latent multidimensionnel. Dans un tel modèle, il est possible d'estimer l'ensemble des paramètres en deux temps : dans un premier temps, les paramètres caractérisant les items peuvent être estimés sous-échelle par sous-échelle par maximum de vraisemblance marginale ; dans un second temps, les paramètres définissant la distribution du trait latent multidimensionnel peuvent être estimés conditionnellement aux estimations des paramètres caractérisant les items. Cette façon de procéder permet un gain important en terme de temps de calcul, sans affecter les estimations des paramètres [2]. Cependant, dans le cadre du MMSRM, chaque modèle unidimensionnel est un modèle de Rasch (par construction) ce qui n'est pas le cas dans le modèle de Adams, Wilson et Wang.

4.8 Interprétation des paramètres de difficulté des items

Nous avons vu (section 3.4.3.8) que dans le MPLT ou dans le MRCML, l'ensemble des paramètres de difficulté des items δ_j n'est pas identifiable, et que seule la somme pondérée $\delta_j = \sum_{q=1}^Q \delta_{jq}$ est estimable. Avec les contraintes posées pour le MMSRM, il est aisé de voir que $\delta_j = \delta_{jq_j} = \delta_j^{(q)}$, c'est-à-dire que la somme identifiable δ_j correspond exactement à la difficulté de l'item j sur l'unique trait latent auquel cet item est relié (ce trait latent est indexé q_j). Ce paramètre a alors un sens, ce qui lui donne un intérêt supplémentaire par rapport aux deux autres modèles plus souples.

4.9 Nécessité d'une structure simple

Nous avons vu que le MMSRM défini à la section 4.7 est un cas particulier du MPLT avec une structure simple. Nous allons démontrer que cette structure simple est une nécessité pour

que le modèle vérifie la propriété d'exhaustivité marginale.

Remarquons tout d'abord que le modèle de Rasch est un cas particulier du MPLT unidimensionnel et dichotomique lorsque tous les coefficients B_j sont égaux à 1. Nous estimerons qu'un MPLT dichotomique vérifiant la propriété d'exhaustivité marginale doit vérifier le fait que les paramètres B_{jq} sont soit nuls, soit égaux à 1.

On se place dans un cadre bidimensionnel : dans ce cadre, il existe trois ensembles d'items. Le premier groupe dénommé A_1 est l'ensemble des items reliés au premier trait latent Θ_1 mais pas au second Θ_2 donc dans ce groupe, chaque item j vérifie $B_{j1} = 1$ et $B_{j2} = 0$. Le second groupe dénommé A_2 concerne les items reliés au second trait latent, mais pas au premier donc pour ces items, on a $B_{j1} = 0$ et $B_{j2} = 1$. Enfin, le troisième groupe dénommé A_{12} concerne les items j reliés au deux traits latents donc tels que $B_{j1} = B_{j2} = 1$. Puisqu'il n'est pas possible d'avoir des items reliés à aucun trait latent, les ensembles A_1 , A_2 et A_{12} sont disjoints entre eux et forment une partition des items.

Théorème 4.1 *Si l'ensemble A_{12} n'est pas vide, alors le modèle sous jacent n'est pas un MMSRM.*

Démonstration.

Prenons un item $k \in A_{12}$, alors $P(X_k = 1/\boldsymbol{\theta}, \delta_k) = \text{logit}(\theta_1 + \theta_2 - \delta_j)$. Or (voir [82, page 120] et [18]) :

$$\text{logit}(\theta_1 + \theta_2 - \delta_j) \approx \Phi(c \cdot (\theta_1 + \theta_2 - \delta_j)), \quad (4.7)$$

avec $c = 1.7^{-1}$ et $\Phi(\cdot)$ la fonction de répartition de la loi normale centrée réduite. Or si l'on a une variable aléatoire T distribuée selon une loi normale centrée de variance σ^2 dont la fonction de répartition est notée $G_T(t/\sigma)$, et a et b deux réels quelconques, alors (d'après [50] et [154]), nous avons :

$$\int_{\mathbb{R}} \Phi(a + bt)G_T(t/\sigma)dt = \Phi\left(|\sigma^2 b^2 + 1|^{-1/2}a\right) \quad (4.8)$$

Alors, en supposant le trait latent Θ_2 de fonction de distribution une loi normale centrée de variance σ_2^2 notée $G_2(\theta_2/\sigma_2)$, on a

$$\begin{aligned} E_{\Theta_2}(P(X_k/\boldsymbol{\theta}, \delta_k)) &= \int_{\mathbb{R}} \Phi(c \cdot (\theta_1 + \theta_2 - \delta_j)) G_2(\theta_2/\sigma_2) d\theta_2 \\ &= \Phi\left(|\sigma_2^2 c^2 + 1|^{-1/2} c(\theta_1 - \delta_j)\right) \end{aligned} \quad (4.9)$$

Donc on a

$$E_{\Theta_2}(P(X_k/\boldsymbol{\theta}, \delta_k)) \approx \text{logit}\left(|\sigma_2^2 c^2 + 1|^{-1/2}(\theta_1 - \delta_j)\right) \quad (4.10)$$

Or pour un item l de A_1 , on montre directement que

$$E_{\Theta_2}(P(X_l/\boldsymbol{\theta}, \delta_l)) \approx \text{logit}(\theta_1 - \delta_l) \quad (4.11)$$

Ainsi pour que marginalement à Θ_2 , les items reliés à Θ_1 suivent un modèle de Rasch, il faudrait que leurs pouvoirs discriminant soient tous égaux, c'est-à-dire que $|\sigma_2^2 c^2 + 1|^{-1/2} = 1$,

or $\sigma_2^2 > 0$, donc $|\sigma_2^2 c^2 + 1|^{-1/2} < 1$. Ainsi, si l'ensemble A_{12} n'est pas vide, le modèle ne peut pas être un MMSRM.

Une structure simple est donc nécessaire pour que le modèle puisse être un MMSRM. Cette démonstration peut être généralisée pour Q traits latents, en montrant que les items reliés à plusieurs traits latents ne peuvent pas, marginalement à tous les traits latents sauf un, suivre un modèle de Rasch avec les items reliés seulement au trait latent considéré.

□

4.10 Extension au cadre polytomique

Il est relativement aisé d'étendre le modèle proposé au cadre polytomique [67]. Au lieu de sélectionner Q ensembles d'items comme dans la section 4.5 qui vérifient un modèle de Rasch, on sélectionne Q ensembles d'items qui vérifient un modèle de crédit partiel (PCM). La propriété 2 devient :

Propriété 3 Soient $\mathbf{X} = (X_j)_{j=1,\dots,J_X}$ et $\mathbf{Y} = (Y_j)_{j=1,\dots,J_Y}$ deux vecteurs de réponses à deux ensembles distincts d'items dont les réponses dépendent d'un trait latent multidimensionnel $\boldsymbol{\theta} = (\theta_q)_{q=1,\dots,Q}$. Les items dont les réponses appartiennent à \mathbf{X} sont ceux qui dépendent du trait latent $\theta_r \in \boldsymbol{\theta}$. Soit Ω l'ensemble de tous les vecteurs \mathbf{Y} possibles. Soit $\boldsymbol{\theta}_r^* = (\theta_q)_{q=1,\dots,r-1,r+1,\dots,Q}$. Marginalement aux items de \mathbf{Y} et à $\boldsymbol{\theta}_r^*$, les items de \mathbf{X} suivent un modèle de crédit partiel relativement au trait latent θ_r .

Le modèle ainsi obtenu est appelé modèle polytomique multidimensionnel marginalement exhaustif de Rasch (Multidimensional Polytomous Marginally Sufficient Rasch Model - MPM-SRM) et vérifie par construction la propriété d'exhaustivité marginale.

Le modèle s'écrit alors :

$$P(X_{nj}^{(q)} = x / \theta_{nq}; \boldsymbol{\delta}_j^{(q)}) = \frac{\exp \left[x \left(\theta_{nq} - \sum_{y=1}^x \delta_{jx}^{(q)} \right) \right]}{1 + \sum_{z=1}^{m_j} \exp \left[x \left(\theta_{nq} - \sum_{y=1}^z \delta_{jy}^{(q)} \right) \right]} \quad (4.12)$$

avec $\boldsymbol{\delta}_j^{(q)} = (\delta_{jx}^{(q)})_{x=1,\dots,m_j}$.

En considérant les paramètres θ_{nq} , $n = 1, \dots, N$, $q = 1, \dots, Q$ comme des paramètres fixes, ce modèle est un cas particulier du MPLT en posant les contraintes : $B_{jq_j x} = x, \forall j = 1, \dots, J$, $x = 1, \dots, m_j$ et $B_{jq_x} = 0$, si $q \neq q_j$. Dans le cas où le trait latent multidimensionnel est traité comme une variable multivariée aléatoire, le modèle obtenu est un cas particulier du MRCML en posant les matrices \mathbf{A}_j , $j = 1, \dots, J$ de taille (m_j, m_j) :

$$\mathbf{A}_j = \begin{pmatrix} 1 & 0 & \cdots & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ \vdots & \cdots & \ddots & \ddots & \vdots \\ 1 & \cdots & \cdots & 1 & 0 \\ 1 & \cdots & \cdots & \cdots & 1 \end{pmatrix} \quad (4.13)$$

Puis en posant

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_1 & & & \mathbf{0} & & \\ \mathbf{0} & \mathbf{A}_2 & & & & \mathbf{0} \\ & \ddots & & & & \ddots \\ & & \mathbf{0} & & \mathbf{A}_j & & \mathbf{0} \\ & & & \ddots & & & \ddots \\ & & & & \mathbf{0} & & & \mathbf{A}_J \end{pmatrix} \quad (4.14)$$

4.11 Application sur trois échelles du Sickness Impact Profil (SIP)

Nous avons estimé les paramètres d'un MMSRM sur trois sous-échelles de la version française du SIP [22] dont nous connaissons les réponses de 483 patients dépressifs [60]. Les trois sous-échelles sont celles traitées par Hamon [54] : il s'agit des sous-échelles "communication" (9 items préfixés par "c"), "mobilité" (10 items préfixés par "m") et "comportement émotionnel" (9 items préfixés par "eb"). Le tableau 4.1 donne les énoncés des 28 questions. Toutes les questions sont à réponse dichotomique (oui/non).

Le MMSRM étudié considère que les items de chaque sous-échelles sont liées à un trait latent distinct : le trait latent est donc tridimensionnel. La distribution de ce trait latent multidimensionnel est supposée être une loi normale centrée multivariée de matrice de variance inconnue.

Les estimations des paramètres sont obtenues par maximum de vraisemblance marginale. On peut les comparer aux estimations obtenues dans chacun des trois modèles de Rasch unidimensionnel pour ce qui concerne les variances des traits latents : les deux estimations sont données dans le tableau 4.2. Pour faciliter les comparaisons entre les deux approches, seuls les 450 patients ayant répondu à l'ensemble des 28 items ont été retenus.

On note que les deux manières d'estimer les variances donnent des estimations proches. Les items peuvent être classés sans problème en fonction de leur difficulté dans chacune des sous-échelles.

Comparé à l'analyse par trois modèles de Rasch unidimensionnels, le MMSRM permet d'estimer les corrélations entre les trois traits latents, et d'obtenir une meilleure adéquation globale du modèle aux données. Dans cet exemple, les trois traits latents sont corrélés positivement et relativement fortement (entre 0.34 et 0.54). La log-vraisemblance du MMSRM est de -6542.79 contre une somme des trois log-vraisemblance obtenues dans les trois modèles de Rasch de -6605.23 : l'adéquation du MMSRM aux données est donc meilleure que l'adéquation de trois modèles unidimensionnels indépendants (statistique du rapport de vraisemblance de 124.08 pour 3 degrés de liberté soit $p < 0.001$).

Item	Énoncé
c1	J'ai du mal à écrire ou à taper à la machine
c2	Je communique surtout par gestes ; par exemple, je fais signe avec la tête, je montre du doigt, je m'exprime par signes
c3	Seules quelques personnes qui me connaissent bien, me comprennent lorsque je parle
c4	Quand je parle, je perds souvent le contrôle de ma voix ; par exemple, je me mets à parler fort ou à voix basse, ma voix tremble ou, tout à coup, elle n'est plus la même
c5	Je n'écris que pour signer mon nom
c6	Je ne discute avec quelqu'un qu'en étant très près de lui ou en le regardant
c7	J'ai des difficultés à parler ; par exemple, les mots ne me viennent pas, je me mets à bégayer, à balbutier, j'articule mal les mots
c8	Les gens comprennent mal ce que je dis
c9	Je ne m'exprime pas clairement quand je suis stressé(e)
m1	Je ne me débrouille qu'à l'intérieur de mon immeuble
m2	Je reste dans la même pièce
m3	Je reste d'avantage au lit
m4	Je reste au lit presque tout le temps
m5	Maintenant, je ne prends plus les transports en commun
m6	Je reste à la maison presque tout le temps
m7	Je ne vais que dans les lieux où je suis sur(e) de trouver des toilettes à proximité
m8	Je ne vais pas en ville
m9	Je ne suis hors de chez moi que durant de courtes périodes
m10	Je ne peux pas me débrouiller dans le noir ou dans les endroits non éclairés sans l'aide de quelqu'un
eb1	Je dis que je ne vaut rien ou que je ne sers à rien, par exemple, je dis que je suis un poids pour les autres
eb2	Tout d'un coup, je me mets à rire ou à pleurer
eb3	Il m'arrive souvent de gémir et de me plaindre parce que j'ai mal ou que je ne me sens pas bien
eb4	J'ai essayé de me suicider
eb5	J'agis avec nervosité, j'ai la bougeotte
eb6	Je ne cesse de me tenir ou de frotter les parties de mon corps qui me font mal ou qui me gênent
eb7	Je manque de patience ou d'indulgence avec moi-même ; par exemple, je dis souvent du mal de moi, je me traite de tous les noms, je me sens responsable de tout ce qui arrive
eb8	Je parle sans espoir de l'avenir
eb9	J'ai des frayeurs soudaines

TAB. 4.1 – Énoncés des items des trois sous-échelles du SIP choisies

	Estimation		Estimations		Estimations	
Paramètres de difficulté	c1	-0.106	m1	1.749	eb1	-0.712
	c2	2.512	m2	2.468	eb2	-0.273
	c3	2.401	m3	-0.003	eb3	0.099
	c4	1.097	m4	1.899	eb4	2.035
	c5	0.347	m5	1.223	eb5	-1.268
	c6	1.179	m6	-0.720	eb6	0.845
	c7	0.883	m7	3.791	eb7	-1.414
	c8	1.954	m8	0.790	eb8	-1.373
	c9	-0.154	m9	-0.467	eb9	-0.607
			m10	2.687		
Variances (MMSRM)	1.982		1.597		0.643	
Variances (Rasch)	2.267		1.650		0.664	
Corrélations	Corr(c,m)=0.525					
	Corr(c,eb)=0.538					
	Corr(m,eb)=0.342					

TAB. 4.2 – Estimations des paramètres du MMSRM pour les trois sous-échelles du SIP

4.12 Synthèse

Cette première partie nous a permis de définir la théorie de réponse aux items, et de présenter les principaux modèles unidimensionnels et multidimensionnels de cette théorie utilisés dans le cadre de l'analyse de la qualité de vie. Un nouveau modèle multidimensionnel simple à interpréter a été présenté. Dans la partie suivante, nous nous attacherons à la construction d'échelles de qualité de vie à partir d'un ensemble d'items, de façon à ce que chacune des échelles ait de bonnes propriétés psychométriques. Nous verrons tout d'abord les méthodes issues de l'analyse factorielle, puis les méthodes existantes en IRT non-paramétriques. Nous proposerons alors trois méthodes basées sur les modèles de l'IRT multidimensionnels pour construire de telles échelles.

Deuxième partie

La sélection d'échelles unidimensionnelles d'items

Chapitre 5

Les méthodes classiques de sélection d'items

5.1 But de ce chapitre

Ce chapitre présente très brièvement deux méthodes utilisées en psychométrie : l'analyse pas à pas du alpha de Cronbach et l'analyse factorielle. Ces deux méthodes sont sans doute les plus utilisées actuellement. Nous les traiterons cependant très brièvement car elles n'entrent pas dans le champ de l'IRT qui nous intéresse plus particulièrement.

5.2 La procédure pas à pas du alpha de Cronbach

Cette procédure [25] est simple à mettre en oeuvre [105]. Elle consiste, pour un ensemble d'items, de rechercher un sous-ensemble d'items (sous-échelle) le plus fiable possible, en se basant sur le coefficient alpha de Cronbach. En effet, une des conséquences de la formule de Spearman-Brown est que la fiabilité d'un ensemble de J items est monotone croissante sur J . Comme le coefficient alpha de Cronbach mesure la fiabilité, on peut construire une courbe donnant la valeur du coefficient alpha de Cronbach en fonction du nombre d'items dans l'échelle. L'idée principale étant d'avoir pour chaque valeur de J la valeur maximale du coefficient alpha de Cronbach parmi toutes les combinaisons possible de J items. Comme il est difficile de calculer le coefficient alpha de Cronbach pour un tel nombre de combinaisons (surtout si l'ensemble initial est de grande taille), on utilise l'algorithme suivant :

- à la première étape on calcule la valeur du coefficient alpha de Cronbach sur l'ensemble initial des items
- à chaque étape, on calcule la valeur du coefficient alpha de Cronbach en retirant un à un chacun des items, et on supprime, pour les étapes ultérieures, l'item qui permet d'obtenir la valeur du coefficient alpha de Cronbach maximale
- on arrête l'algorithme quand il ne reste que deux items

On trace alors la courbe des valeurs maximales obtenues du alpha de Cronbach pour chaque nombre possible d'items (de 2 à J) et on retient les items non supprimés avant une éventuelle décroissance de la courbe (dans le cas où la courbe est croissante de 2 à J , on retient les J items initiaux).

5.3 L'analyse factorielle exploratoire

L'analyse factorielle est sans doute la méthode la plus ancienne utilisée en psychométrie pour définir les liens entre les items et les traits latents. Elle a été introduite par Spearman [137] [138] et popularisée par Thurstone [143]. Deux méthodes d'analyse factorielle sont principalement utilisées en psychométrie : l'analyse en facteurs communs et spécifiques (AFCS, que l'on retrouve sous le nom d'analyse en facteurs, ou "factor analysis"), et l'analyse en composantes principales (ACP).

5.3.1 Principes de l'analyse en facteurs communs et spécifiques

L'AFCS connaît une grande renommée en science de l'éducation, et reste sans doute moins connue en recherche clinique (où l'ACP semble être plus souvent utilisée). Cette analyse repose sur une modélisation de la matrice de réponses aux J items par un ensemble de Q facteurs communs (un facteur commun est une variable continue qui influe sur les réponses d'au moins deux items différents) et J facteurs spécifique (un facteur spécifique influe sur les réponses d'un seul item).

5.3.1.1 L'AFCS pour déterminer des échelles d'items

Le principe d'utilisation de l'AFCS pour déterminer des échelles d'items repose sur le fait que les facteurs communs n'expliquent pas tous avec la même intensité les réponses de chacun des items. On va donc chercher, pour chaque item, le ou les facteurs communs qui expliquent le plus ses réponses. Aussi, l'ensemble des items qui sera relié de cette manière à un même facteur sera considéré comme unidimensionnel et formera une échelle.

5.3.1.2 Estimation des paramètres

Soient \mathbf{X} la matrice centrée de taille $(N \times J)$ des réponses des N individus aux J items X_j , $j = 1, \dots, J$, \mathbf{F} la matrice $(N \times Q)$ des coordonnées des N individus sur les Q facteurs communs \mathbf{f}_q , $q = 1, \dots, Q$, $\mathbf{\Gamma}$ la matrice $(J \times Q)$ composée de paramètres γ_{jq} inconnus et \mathbf{E} la matrice $(N \times J)$ de coordonnées des N individus sur les J facteurs spécifiques \mathbf{e}_j , $j = 1, \dots, J$.

Le modèle de l'analyse en facteurs s'écrit :

$$\mathbf{X} = \mathbf{F}\mathbf{\Gamma}' + \mathbf{E}, \quad (5.1)$$

ou, en d'autre terme, en posant X_{nj} la réponse de l'individu n à l'item j , f_{nq} la coordonnées de l'individu n sur le facteur commun \mathbf{f}_q et e_{nj} la coordonnée de l'individu n sur le facteur spécifique \mathbf{e}_j , alors :

$$X_{nj} = \sum_{q=1}^Q \gamma_{jq} f_{nq} + e_{nj} \quad (5.2)$$

On pose la contrainte que les facteurs communs sont standardisés et indépendants deux à deux, donc que $\mathbf{F}'\mathbf{F} = \mathbf{I}_Q$. En outre, on suppose que les facteurs communs sont indépendants des facteurs spécifiques et que les facteurs spécifiques sont indépendants deux à deux. Enfin,

les facteurs communs et spécifiques sont centrés, donc $E(\mathbf{f}_q) = E(\mathbf{e}_j) = \mathbf{0}$, $q = 1, \dots, Q$, $j = 1, \dots, J$ où \mathbf{f}_q est le q^e facteur commun (q^e colonne de la matrice \mathbf{F}) et \mathbf{e}_j est le j^e facteur spécifique (j^e colonne de la matrice \mathbf{E}) et $\mathbf{E}'\mathbf{E} = \mathbf{\Delta}$ diagonale.

Soient \mathbf{S} et $\mathbf{\Sigma}$ les matrices de variance-covariance empirique et théorique des variables X_j , $j = 1, \dots, J$ (on remarquera que dans le cas où les variables X_j sont standardisées au préalable, ces matrices correspondent aux matrices de corrélations), et $\mathbf{\Delta}$ la matrice de variance-covariance des facteurs spécifiques. Alors

$$\mathbf{\Sigma} = \mathbf{\Gamma}\mathbf{\Gamma}' + \mathbf{\Delta}, \quad (5.3)$$

car comme les facteurs communs et les facteurs propres sont indépendants, $\mathbf{\Gamma}\mathbf{F}'\mathbf{E} = \mathbf{E}'\mathbf{F}\mathbf{\Gamma}' = \mathbf{0}$.

Ainsi, on peut trouver une estimation $(\hat{\mathbf{\Gamma}}; \hat{\mathbf{\Delta}})$ de $(\mathbf{\Gamma}; \mathbf{\Delta})$ en résolvant le système

$$\mathbf{S} = \frac{1}{N}\mathbf{X}'\mathbf{X} = \hat{\mathbf{\Sigma}} = \hat{\mathbf{\Gamma}}\hat{\mathbf{\Gamma}}' + \hat{\mathbf{\Delta}} \quad (5.4)$$

Il existe une infinité de solutions de l'ensemble $(\mathbf{\Gamma}, \mathbf{F}, \mathbf{E})$ à ce système. On peut par exemple utiliser les Q premières composantes principales de la matrice de variance-covariance \mathbf{S} [81].

Pour résoudre le système d'équations, on peut utiliser le maximum de vraisemblance si les distributions des X_j sont supposées normales. Dans le cas contraire, notamment lorsque l'on a affaire à des items dichotomiques ou polytomiques, on utilise la méthode des moindres carrés, ou celle des moindres carrés généralisés [21] [118] qui ne nécessitent pas d'hypothèse sur la distribution des variables X_j .

Dans la méthode des moindres carrés, on minimise la quantité

$$\sum_{j=1}^J \sum_{k=1}^j (s_{jk} - \hat{\sigma}_{jk})^2, \quad (5.5)$$

où s_{jk} et $\hat{\sigma}_{jk}$ sont les éléments des matrices $\frac{N-1}{N}\mathbf{S}$ et $\hat{\mathbf{\Sigma}}$ respectivement. Dans le cadre des moindres carrés généralisés, la quantité à minimiser est

$$\frac{1}{2} \text{trace} \left\{ \left[\frac{N-1}{N}\mathbf{S} - \hat{\mathbf{\Sigma}} \right] \left[\frac{N-1}{N}\mathbf{S} \right]^{-1} \right\}^2 \quad (5.6)$$

Cependant, comme il existe une infinité de solutions au modèle, il est courant de rechercher la structure approximativement simple la plus proche des données, en effectuant une rotation des facteurs, de manière à ce que chaque item soit le mieux possible représenté sur un seul facteur commun, et le moins bien possible sur les autres facteurs communs, en maximisant pour chaque item j , $j = 1, \dots, J$ un paramètre $|\gamma_{jq}|$, $q = 1, \dots, Q$ (les γ_{jq} étant les éléments de la matrice $\mathbf{\Gamma}$), et en minimisant les autres paramètres $|\gamma_{jq}|$. En effet, à toute solution $(\mathbf{\Gamma}, \mathbf{F}, \mathbf{E})$ du modèle, alors $(\mathbf{\Gamma}\mathbf{A}, \mathbf{A}^{-1}\mathbf{F}, \mathbf{E})$ est aussi solution de ce système avec la matrice \mathbf{A} régulière ($\mathbf{A}\mathbf{A}' = \mathbf{A}'\mathbf{A} = \mathbf{I}_Q$). Les rotations de ce type sont appelées orthogonales (ou rigides) car elles

conservent l'orthogonalité des facteurs communs. Les rotations orthogonales les plus connues sont la rotation Varimax qui maximise la quantité

$$\sum_{q=1}^Q \left\{ \frac{\sum_{j=1}^J (b_{jq}^2/c_j^2)^2}{J} - \left[\frac{\sum_{j=1}^J (b_{jq}^2/c_j^2)^2}{J} \right]^2 \right\}, \quad (5.7)$$

la rotation Quartimax qui maximise la quantité

$$\frac{\sum_{q=1}^Q \sum_{j=1}^J (b_{jq}^2)^2}{JQ} - \left[\frac{\sum_{q=1}^Q \sum_{j=1}^J (b_{jq}^2)^2}{JQ} \right]^2, \quad (5.8)$$

et la rotation Orthomax qui maximise la quantité

$$\sum_{q=1}^Q \left[\sum_{j=1}^J b_{jq}^4 - \frac{\gamma}{J} \left(\sum_{j=1}^J b_{jq}^2 \right)^2 \right], \quad (5.9)$$

avec b_{jq} , $j = 1, \dots, J$, $q = 1, \dots, Q$ les éléments de la matrice $(J \times Q)$ $\mathbf{B} = \mathbf{\Gamma}\mathbf{A}$ (qui correspondent donc aux coordonnées des items sur les facteurs communs après la rotation effectuée), et $c_j^2 = \sum_{q=1}^Q b_{jq}^2$ (appelée communalité). γ est un réel tel que $0 \leq \gamma \leq 1$ [81].

5.3.2 L'analyse en composantes principales (ACP)

Nous considérerons ici l'ACP comme un cas particulier de l'AFCS, où la solution choisie est celle où les facteurs correspondent aux composantes principales (axes portés par les vecteurs propres de la matrice de variance-covariance). Dans cette analyse, il n'est pas fait d'hypothèse *a priori* sur le nombre de facteurs communs Q . On notera cependant deux différences importantes entre les deux analyses [37] : d'une part, en ACP, on cherche les Q variables latentes qui reconstituent au mieux les variances des variables X_j , tandis qu'en AFCS, on cherche les Q variables latentes qui reconstituent au mieux les covariances entre les variables X_j ; d'autre part, les solutions de l'ACP sont emboîtées (pour passer de k composantes principales à $k+1$, il suffit de trouver la nouvelle composante principale de variance maximale orthogonale au k premières), tandis que les solutions de l'AFCS ne le sont pas (il faut alors réestimer l'ensemble des paramètres).

5.3.3 Le cas des variables qualitatives

L'AFCS (et l'ACP) sont basées sur la matrice de variance-covariance (ou de corrélation) des réponses aux variables. Cependant, on a très souvent affaire à des variables à réponses dichotomiques ou polytomiques. L'utilisation de la covariance ou du coefficient de corrélation de Pearson n'est alors pas toujours adapté, par exemple en introduisant des facteurs artificiels si les difficultés des items sont variables.

Aussi, il est d'usage d'utiliser des coefficients d'association entre variables qualitatives, en particulier les coefficients de corrélation tétrachorique.

Ces coefficients se définissent ainsi : soient deux variables Y_1 et Y_2 distribuées selon une loi binormale centrée réduite de corrélation ρ , dont les réalisations sont notées y_1 et y_2 . Soient deux seuils s_1 et s_2 et soient deux variables dichotomiques X_1 et X_2 de réalisations x_1 et x_2 définies telles que $x_1 = \mathbb{I}_{[s_1; +\infty]}(y_1)$ et $x_2 = \mathbb{I}_{[s_2; +\infty]}(y_2)$. Soit enfin le tableau de contingence entre les variables aléatoires X_1 et X_2 donné dans le tableau 5.1.

		X_k		Total
		1	0	
X_j	1	A	B	$A + B$
	0	C	D	$C + D$
Total		$A + C$	$B + D$	N

TAB. 5.1 – Tableau de contingence entre les items dichotomiques X_1 et X_2

On peut estimer ρ à partir de X_1 et X_2 en résolvant

$$\frac{A}{N} = \int_{\hat{h}_1}^{+\infty} \int_{\hat{h}_2}^{+\infty} \frac{1}{2\pi(1-\rho^2)^{1/2}} \exp\left[-\frac{x_1^2 - 2\rho x_1 x_2 + x_2^2}{2(1-\rho^2)}\right] dx_1 dx_2, \quad (5.10)$$

où $\hat{h}_1 = \Phi^{-1}\left(\frac{A+C}{N}\right)$ et $\hat{h}_2 = \Phi^{-1}\left(\frac{A+B}{N}\right)$ avec $\Phi^{-1}(\cdot)$ la réciproque de la fonction de répartition de la loi normale centrée réduite.

Il existe une solution unique à cette équation notée r_t et appelée coefficient de corrélation tétrachorique entre les variables X_1 et X_2 . Ce coefficient a été introduit par Pearson [121] qui en propose plusieurs estimations très rapides à obtenir :

$$Q_1 = \frac{AD - BC}{AD + BC}, \quad (5.11)$$

$$Q_3 = \sin\left(\frac{\pi}{2} \frac{\sqrt{AD} - \sqrt{BC}}{\sqrt{AD} + \sqrt{BC}}\right), \quad (5.12)$$

$$Q_4 = \sin\left\{\frac{\pi}{2} \left(1 + \frac{2BCN}{(AD - BC)(B + C)}\right)^{-1}\right\}. \quad (5.13)$$

Yule [153] propose deux mesures, la mesure Q_1 (appelée coefficient d'association de Yule) et la mesure

$$Q_2 = \frac{\sqrt{AD} - \sqrt{BC}}{\sqrt{AD} + \sqrt{BC}}, \quad (5.14)$$

appelée coefficient de colligation de Yule.

Hamdan [53] montre que r_t est l'estimateur du maximum de vraisemblance de ρ . Le coefficient de corrélation polychorique permet de la même manière de mesurer l'association entre deux variables nominales ordonnées.

On peut utiliser la matrice des coefficients de corrélation tétrachorique en analyse factorielle exploratoire. Cette approche, si elle permet de résoudre le problème technique de mesure de

l'association entre variables catégorielles, n'est pas dispensée de problèmes. En effet, ce coefficient est instable quand les proportions marginales des variables X_1 et X_2 sont extrêmes, ou quand N est petit [106]. Ensuite, la matrice de corrélation tétrachorique peut ne pas être définie positive [89]. Enfin, les variances peuvent être estimées approximativement à 0 (problème connu sous le nom de cas d'Heywood [72]) : dans ce cas, cela signifie qu'il y a trop ou pas assez de facteurs communs dans le modèle, ou pas assez d'observations pour obtenir des estimations stables des paramètres ou que le modèle en facteurs communs n'est pas approprié aux données [76].

5.3.4 L'analyse en facteurs d'information complète (full information item factor analysis)

Bock, Gibbons et Muraki [16] proposent une méthode d'estimation des paramètres qui nécessite moins de calculs (donc de temps de calculs), grâce à la mise en oeuvre d'un algorithme EM, remplaçant la méthode classique des moindres carrés, plus coûteuse en temps de calcul. Cette méthode est basée sur les fréquences de chaque vecteur des réponses possibles (plutôt que sur les premiers moments joints des variables réponses de l'analyse factorielle classique), ce qui explique son nom de "full information". Cette méthode a connu un grand succès à une époque où les capacités informatiques ne permettaient de traiter qu'une vingtaine d'items avec la méthode classique, contre une soixantaine avec cette méthode [fin des années 80 [16]]. Cette méthode reste implémentée dans des logiciels plus récents, comme TESTFAC, même si l'intérêt est moins évident avec les nouvelles générations d'ordinateurs.

5.3.5 La procédure Varclus

La procédure Varclus [6] a été popularisée par le logiciel SAS [78]. Cette procédure cherche des ensembles de variables quantitatives unidimensionnelles (dans le sens où seule la première valeur propre de la matrice de variance-covariance est supérieure à 1).

L'algorithme de cette procédure est le suivant :

- à une première étape, on considère un seul ensemble contenant toutes les variables,
- à chaque étape, on scinde l'ensemble qui présente la plus forte valeur pour la seconde valeur propre (sous réserve que celle-ci soit supérieure à 1) en deux ensembles : pour ce faire, on effectue une ACP suivie d'une rotation quartimax sur les variables de cet ensemble et on affecte chaque variable à deux groupes en fonction des corrélations (au carré) entre chaque variable et chacune des deux premières composantes après rotation,
- on stoppe l'algorithme dès que dans chaque ensemble, la seconde valeur propre est inférieure à 1.

Cette procédure a l'avantage de présenter une méthode objective (non soumise à des choix de l'utilisateur) pour définir le nombre d'ensembles de variables et la composition de chacun de ces ensembles. Elle peut être adaptée à des variables qualitatives en utilisant par exemple la matrice de corrélation tétrachorique. Elle présente cependant les mêmes inconvénients face à de telles variables que les autres analyses factorielles.

5.3.6 La procédure de classification autour de variables latentes - CLV

Cette procédure proposée par Vigneau, Qannari et al. [148] [149] est proche de la procédure Varclus dont elle a les mêmes objectifs : trouver K ensembles de variables disjoints et unidimensionnels.

La méthode s'appuie sur la minimisation du critère :

$$E = \frac{1}{N} \sum_{k=1}^K \sum_{j=1}^J \delta_{kj} \|\mathbf{x}_j - \alpha_{kj} \mathbf{c}_k\|^2 \quad (5.15)$$

avec \mathbf{x}_j , $j = 1, \dots, J$ les variables à classer après centrage, \mathbf{c}_k , $k = 1, \dots, K$ les composantes latentes de chacun des K groupes et α_{kj} des coefficients à estimer. δ_{kj} est une indicatrice qui prend pour valeur 1 si la variable \mathbf{x}_j appartient au groupe k , et 0 sinon.

Cela revient à maximiser la quantité

$$T = N \sum_{k=1}^K \mathbf{c}_k' \mathbf{X}_k' \mathbf{X}_k \mathbf{c}_k \quad (5.16)$$

avec $\mathbf{c}_k^* = \frac{\mathbf{c}_k}{\sqrt{\mathbf{c}_k' \mathbf{c}_k}}$ et \mathbf{X}_k , $k = 1, \dots, K$ l'ensemble des variables du groupe k .

La procédure CLV se déroule en deux temps. Une première phase est constituée d'une classification ascendante hiérarchique des variables en se basant sur le critère T . Une seconde phase consiste à consolider la partition obtenue grâce à un algorithme de nuées dynamiques.

On montre que le critère T est égale à la somme des premières valeurs propres de tous les groupes. C'est un critère facile à calculer. Il est aussi facile à montrer que ce critère diminue lorsque l'on regroupe des variables ou des groupes de variables, donc dans le cadre de la classification hiérarchique ascendante, la stratégie consiste à agréger à chaque étape les deux groupes de variables conduisant à la plus petite diminution de ce critère. En outre, on peut utiliser la valeur de T à chaque étape pour détecter les regroupements ayant conduit à une forte diminution du critère, et donc pour choisir le nombre adéquat de groupes de variables.

L'étape de consolidation de la procédure consiste à réaffecter éventuellement des variables à un autre groupe en se basant sur la covariance au carré entre chaque variable et la première composante principale de chaque groupe. On arrête cette consolidation quand la partition devient stable.

La principale différence entre Varclus et CLV réside dans le fait que Varclus est une classification descendante (à chaque étape on sépare le groupe de variables le moins unidimensionnel, jusqu'à ce que l'on obtienne seulement des groupes unidimensionnels), tandis que CLV est une classification ascendante (on regroupe des variables potentiellement unidimensionnelles). En outre, l'algorithme de consolidation de la partition est plus complexes et demande plus de calculs dans le cadre de Varclus.

Les auteurs de la procédure CLV proposent en outre d'étendre CLV à d'autres cas, en particulier en considérant le sens de la relation entre les variables, et/ou en trouvant des

composantes latentes comme des combinaisons linéaires de variables connexes, ce qui permet d'interpréter aisément les composantes latentes trouvées (voir par exemple [148] [133]).

5.3.7 L'analyse parallèle modifiée révisée (Revised Modified Parallel Analysis)

L'analyse parallèle modifiée révisée est une méthode issue de l'analyse factorielle proposée pour construire des échelles d'items unidimensionnelles. Cette méthode a été proposée en 1997 par Budescu, Cohen et Ben-Simon [17], et s'appuie sur l'analyse parallèle modifiée (modified parallel analysis) de Drasgow et Lissak [29] issue elle-même sur l'analyse parallèle (parallel analysis) de Humphreys et Montanelli [75].

5.3.7.1 L'analyse parallèle

L'analyse parallèle permet de définir la dimension (c'est-à-dire le nombre de facteurs communs pertinent) d'une matrice de réponses à J variables quantitatives. On calcule la matrice de corrélation de cette matrice, puis on remplace les éléments diagonaux de cette matrice par le carré du coefficient de corrélation linéaire entre chaque variable et les $(J - 1)$ autres. La même opération est réalisée sur un ensemble de J variables simulées par une loi normale centrée réduite sur le même nombre d'individus. On compare alors les valeurs propres des deux matrices et on détermine la dimension de la matrice initiale comme égale au nombre de valeurs propres de la matrice initiale plus grandes que celles de la matrice simulée.

5.3.7.2 L'analyse parallèle modifiée

L'analyse parallèle ne s'applique qu'à des variables numériques car on suppose une relation linéaire entre les variables et les facteurs communs. L'analyse parallèle modifiée résout ce problème pour les variables dichotomiques en utilisant les corrélations tétrachoriques et en remplaçant les éléments diagonaux de la matrice artificielle par le plus grand coefficient de corrélation tétrachorique entre chaque variable et toutes les autres prises une à une. La matrice simulée est obtenue à l'aide d'un modèle de l'IRT unidimensionnel ¹. Dans cette approche, seules les deux premières valeurs propres des deux matrices sont comparées, ce qui permet de déterminer si l'ensemble des items est unidimensionnel ou non.

5.3.7.3 L'analyse parallèle modifiée révisée

Budescu, Cohen et Ben-Simon [17] critiquent sur cinq points l'analyse parallèle modifiée :

1. la procédure est non robuste aux petits effectifs,
2. toutes les différences entre les valeurs propres sont attribuées à la dimension de la matrice initiale,
3. la procédure ne calcule pas les variabilités d'échantillonnage,
4. la procédure compare seulement les deux premières valeurs propres, alors que les différences peuvent apparaître seulement sur les suivantes.
5. la procédure a un seul intérêt : vérifier l'unidimensionnalité.

¹nous renverrons à l'article de Drasgow et Lissak [29] pour le choix du modèle de l'IRT choisi

Pour résoudre le premier point, ils proposent d'utiliser la matrice des corrélations tétrachoriques attendue sous les hypothèses de l'IRT (unidimensionnalité, indépendance locale et monotonie) à l'aide du modèle logistique à trois paramètres (3-PLM) : des données simulées sont obtenues sous ce modèle, et la matrice de corrélation tétrachorique sur ces données simulées est alors calculée.

Pour résoudre le second problème, ils proposent d'utiliser les mêmes éléments diagonaux dans les deux matrices (en prenant les éléments diagonaux de la matrice simulée).

Pour résoudre le troisième problème, ils utilisent une méthode "Jackknife" pour estimer la variance des estimations des valeurs propres des matrices, en retirant, une à la fois, chacune des variables.

Pour dépasser le quatrième point, ils proposent un test ("gap test") pour tester l'égalité des rapports

$$RL_l^{-j} = \frac{\lambda_{l(obs)}^{-j}}{\lambda_{l(th)}^{-j}}, \quad j = 1, \dots, J, \quad l = 1, 2, 3,$$

où $\lambda_{l(obs)}^{-j}$ et $\lambda_{l(th)}^{-j}$ sont respectivement les l^e valeurs propres obtenues dans la matrice initiale et dans la matrice estimée en supprimant l'item j . Sous l'hypothèse d'unidimensionnalité des données (un seul facteur commun), les RL_1^{-j} , RL_2^{-j} et RL_3^{-j} sont identiquement distribués, ainsi, si un item j est influencé par un autre facteur commun que les autres items, il peut être détecté. Ainsi, les items perturbant les distributions des RL_l^{-j} , $l = 1, 2, 3$ sont rejetés, et l'on considère que les items non rejetés forment une échelle unidimensionnelle. On peut refaire l'analyse avec les items rejetés la première fois, pour former une seconde échelle, et ainsi de suite, jusqu'à ce qu'il n'y ait plus d'items disponibles. Aussi, cette analyse permet de sélectionner des ensembles unidimensionnels d'items, et donc de définir la dimension de la matrice initiale, contrairement à l'analyse parallèle modifiée qui vérifie seulement l'unidimensionnalité des données (point 5).

5.4 Synthèse

Les méthodes issues de l'analyse factorielle présentées dans ce chapitre sont sans conteste les plus répandues pour construire des échelles d'items. Le principal inconvénient de ces méthodes est de ne s'intéresser qu'à la notion d'unidimensionnalité des données. Dans le chapitre suivant, nous présenterons trois méthodes basées sur l'IRT, qui permettent de construire des échelles vérifiant les propriétés fondamentales de l'IRT.

Chapitre 6

Les procédures existantes en Théorie de Réponse aux Items non paramétrique

6.1 La classification hiérarchique avec mesures de proximité conditionnelles (HCA/CCPROX)

6.1.1 Définition générale

La classification hiérarchique avec mesures de proximité conditionnelles (Hierarchical Clusters Analysis/Conditional Proximity Measures - HCA/CCPROX) a été proposée par Roussos, Stout et Marden [132]. Elle consiste à regrouper les items selon une classification ascendante hiérarchique. Les items sont donc considérés dans cette analyse comme des observations.

Utiliser une classification hiérarchique sur des items dichotomiques nécessite de fixer deux choses : d'une part la mesure de proximité entre deux items, et d'autre part la mesure de proximité entre deux ensembles ("clusters") d'items.

6.1.2 Les mesures de proximité entre les items

Parmi les nombreuses mesures d'association existantes entre des items dichotomiques, les auteurs en ont retenu trois, que nous nommerons les mesures A , $A+D$ et $Corr$. Soit le tableau de contingence 6.1 entre deux items dichotomiques X_j et X_k :

		X_k	
		1	0
X_j	1	A	B
	0	C	D

TAB. 6.1 – Tableau de contingence entre les items dichotomiques X_j et X_k

Alors les mesures de proximité A (connue sous le nom de mesure de dissimilarité de Jaccard

[80]), $A + D$ (connue sous le nom de mesure de dissimilarité de Sokal et Michener [136]) et $Corr$ (popularisée par Guilford [49]) entre ces deux items sont :

$$p_A(X_j, X_k) = \sqrt{1 - \left(\frac{A}{A + B + C} \right)}, \quad (6.1)$$

$$p_{A+D}(X_j, X_k) = \sqrt{1 - \left(\frac{A + D}{A + B + C + D} \right)}, \quad (6.2)$$

$$p_{Corr}(X_j, X_k) = \sqrt{2(Corr(X_j, X_k))} = \sqrt{2 \left\{ 1 - \left(\frac{AD - BC}{(A + B)(C + D)(A + C)(B + D)} \right) \right\}}. \quad (6.3)$$

Roussos, Stout et Marden proposent en outre 4 nouvelles mesures de proximité définies conditionnellement à un score S obtenu sur l'ensemble des items. Ces mesures seront nommées $CCorr$, $CCov$ et MH . Soit le tableau de contingence 6.2 entre deux items dichotomiques X_j et X_k , pour les individus ayant un score s aux $(J - 2)$ autres items de l'échelle :

		X_k	
		1	0
X_j	1	A_s	B_s
	0	C_s	D_s

TAB. 6.2 – Tableau de contingence entre les items dichotomiques X_j et X_k pour les individus ayant un score $S = s$

Alors les mesures de proximité $CCorr$, $CCov$ et MH entre ces deux items sont :

$$\begin{aligned} p_{CCor}(X_j, X_k) &= \sqrt{2 \left(1 - \frac{1}{N} \sum_{s=0}^{J-2} N_s Corr_s(X_j, X_k) \right)} \quad (6.4) \\ &= \sqrt{2 \left(1 - \frac{1}{N} \sum_{s=0}^{J-2} N_s \frac{A_s D_s - B_s C_s}{\sqrt{(A_s + B_s)(C_s + D_s)(A_s + C_s)(B_s + D_s)}} \right)}, \end{aligned}$$

$$p_{CCov}(X_j, X_k) = -\frac{1}{N} \sum_{s=0}^{J-2} N_s Cov_s(X_j, X_k) + c_{Cov} = -\frac{1}{N} \sum_{s=0}^{J-2} \frac{A_s D_s - B_s C_s}{N_s} + c_{CCov}, \quad (6.5)$$

$$p_{MH}(X_j, X_k) = -\log \left(\frac{\sum_{s=0}^{J-2} \frac{A_s D_s}{N_s}}{\sum_{s=0}^{J-2} \frac{B_s C_s}{N_s}} \right) + c_{MH}, \quad (6.6)$$

avec $N_s = A_s + B_s + C_s + D_s$ et c_{Cov} et c_{MH} deux constantes.

Pour garantir de bonnes propriétés de ces mesures de proximité (en particulier que la matrice de proximité de l'ensemble des items soit définie positive), on ajoute dans le cadre des mesures $Ccov$ et MH les constantes c_{CCov} et c_{MH} afin que $p_{CCov}(X_j, X_k)$ et $p_{MH}(X_j, X_k)$ soient positives $\forall j, k = 1, \dots, J, j \neq k$.

L'idée des deux premières mesures de proximité vient du fait que, si l'ensemble des items vérifie une structure simple (voir section 3.3), et que θ_α est un trait latent composite mesuré par le score $S = \sum_{j=1}^J X_j$, alors la corrélation conditionnelle au trait latent composite entre deux items mesurant la même dimension est positive tandis que celle entre deux items mesurant deux traits latents différents est négative. Aussi, grâce aux mesures de proximité $CCov$ et $CCorr$, deux items reliés à la même dimension sont plus proches que deux items reliés à deux dimensions différentes (les mesures de proximité sont plus petites dans le premier cas).

Concernant la troisième mesure de proximité, elle s'appuie sur le calcul de l'odds-ratio ajusté de Mantel-Hanszel [96].

On notera qu'il n'est pas toujours possible de calculer les mesures $CCorr$ et MH (lorsque qu'il existe au moins un score s tel que $A_s + B_s = 0$ ou $C_s + D_s = 0$ ou $A_s + C_s = 0$ ou $B_s + D_s = 0$ pour $CCorr$ ou tel que $B_s = 0$ ou $C_s = 0$ pour MH).

6.1.3 Les mesures de proximité entre les ensembles d'items

Les auteurs s'appuient en outre sur quatre méthodes existantes pour mesurer la proximité de deux ensembles d'items :

- le lien unique (single linkage) (SL [42]),
- le lien complet (complete linkage) (CL [102]),
- la méthode des moyennes par paire non pondérée (unweighted pair-group method of averages) (UPGMA [136]),
- la méthode des moyennes par paire pondérée (weighted pair-group method of averages) (WPGMA [103]).

Soient C_1 , C_2 et C_3 trois ensembles d'items distincts comportant un item ou plus. On suppose que C_1 et C_2 ont été assemblés pour former un nouvel ensemble d'items. Soit $p(C_1 \cup C_2, C_3)$ la mesure de proximité entre les ensembles $C_1 \cup C_2$ et C_3 .

Par le SL, cette mesure est égale à la mesure de proximité minimale entre toutes les paires d'items possibles provenant de $C_1 \cup C_2$ d'une part et de C_3 d'autre part.

Par le CL, cette mesure est égale à la mesure de proximité maximale entre toutes les paires d'items possibles provenant de $C_1 \cup C_2$ d'une part et de C_3 d'autre part.

Par le UPGMA, cette mesure est égale à la mesure de proximité moyenne entre toutes les paires d'items possibles provenant de $C_1 \cup C_2$ d'une part et de C_3 d'autre part.

Par le WPGMA, cette mesure est égale à la moyenne pondérée des mesures de proximité entre toutes les paires d'items possibles provenant de $C_1 \cup C_2$ d'une part et de C_3 d'autre part. Les poids sont accordés de façon inversement proportionnelle aux tailles respectives de C_1 et C_2 de manière à ce que ces deux sous-ensembles soient pondérés de façon égale.

6.1.4 Résultats

Roussos, Stout et Marden [132] ont réalisé des simulations pour étudier empiriquement cette méthode. Les paramètres de simulations utilisés sont :

- la structure des items (Simple ou approximativement simple),
- le nombre de dimensions ($Q = 2$ ou $Q = 3$),
- le nombre d'individus ($N = 250$, $N = 1000$ ou $N = 3000$),
- la corrélation entre les traits latents ($\rho = 0.7$, $\rho = 0.8$ ou $\rho = 0.9$),
- le nombre d'items dans chaque dimension ($J_q = 5$, $J_q = 10$ ou $J_q = 20$),
- la mesure de proximité utilisée (A , $A + D$, $Corr$, $CCorr$, $CCov$ ou MH),
- la méthode d'agrégation utilisée (SL , CL , $UPGMA$ ou $WPGMA$).

Les données sont simulées par une extension multidimensionnelle du modèle de Birnbaum (voir section 3.5.1) :

$$P(X_{nj} = 1/\theta_n) = \frac{\exp \left[1.7 \left(\alpha'_j \theta_n + \delta_j \right) \right]}{1 + \exp \left[1.7 \left(\alpha'_j \theta_n + \delta_j \right) \right]} \quad (6.7)$$

Le vecteur α_j des pouvoirs discriminants de l'item j est choisi de manière à ce que l'indice multidimensionnel de discrimination défini par Reckase et McKinley [131]

$$MDISC_j = \sqrt{\sum_{q=1}^Q \alpha_{jq}^2}, \quad (6.8)$$

soit fixé parmi 5 valeurs (0.4, 0.8, 1.2, 1.6 et 2.0). Le paramètre de difficulté δ_j de l'item j est choisi de manière à ce que l'indice

$$MDL_j = \frac{-\delta_j}{MDISC_j}, \quad (6.9)$$

soit fixé parmi 5 valeurs (-1.5, -1.0, 0.0, 1.0 et 1.5).

Dans le cas d'une simulation d'une structure simple, l'item j est relié exclusivement à la dimension q_j , alors $\alpha_{jq_j} = MDISC_j$ et $\delta_j = -MDL_j \cdot MDISC_j$.

Dans le cas d'une simulation d'une structure approximativement simple à Q dimensions, $\alpha_{jq_j} = \sqrt{MDISC_j^2 \left(1 - (Q-1) [\cos(15^\circ)]^2 \right)}$ et $\alpha_{jq} = MDISC_j \cos(15^\circ)$ si $q \neq q_j$.

Dans le cadre des trois mesures classiques, c'est la mesure $Corr$ qui donne les meilleurs résultats. Dans tous les cas, les mesures de proximité conditionnelles donnent de meilleurs résultats que les mesures classiques. La méthode d'agrégation $UPGMA$ semble être celle qui donne les meilleurs résultats, et SL la moins bonne. La mesure de proximité $CCov$ semble être la meilleure des trois mesures de proximité conditionnelles proposées.

Le taux de classement correct atteint, pour une combinaison $UPGMA/CCov$, 95% dans le cadre d'une structure simple et 90% dans le cadre d'une structure approximativement simple dans les cas suivants : $N = 500/\rho = 0.7$, $N = 1000/\rho = 0.8$, $N = 3000/\rho = 0.9$.

6.2 La Mokken Scale Procedure (MSP)

6.2.1 Le Modèle Monotonement Homogène de Mokken (Mokken Monotone Homogeneous Model) (MMHM)

Le modèle Monotonement Homogène de Mokken [107] [109] [108] (que l'on peut retrouver sous le nom de modèle de Mokken [104] ou d'échelle de Mokken [110]) est un modèle non paramétrique (dans lequel les IRF des items ne sont pas définies par des paramètres - voir section 2.5) dans lequel les items vérifient les trois hypothèses fondamentales de l'IRT :

- unidimensionnalité,
- indépendance locale,
- monotonie.

Ce modèle a été étendu dans le cadre du modèle doublement monotone (doubly monotone) (par exemple [108]), un modèle non paramétrique où les IRF des items sont non sécantes (on peut se reporter à cet égard à une comparaison de ce modèle avec le modèle de Rasch réalisée par Meijer, Sijtsma et Smid [104]). Le modèle de Mokken a également été développé dans un cadre polytomique [115].

Dans le cadre de ce modèle, la principale question qui a été débattu concerne la manière de vérifier qu'un ensemble d'items vérifie un MMHM (voir sur ce thème [115]). En fonction des indicateurs utilisés pour mesurer l'adéquation, une procédure pour permettre la construction d'échelles d'items vérifiant un tel modèle à partir d'un ensemble initial d'items a été proposée (voir à cet égard [109], [71], [108]).

6.2.2 Les erreurs de Guttman

6.2.2.1 Définition dans un cadre dichotomique

Les erreurs de Guttman [52] sont définies pour une paire d'items dichotomiques : elles représentent le nombre de personnes ayant répondu négativement à l'item le plus facile (des deux items), et positivement à l'item le plus difficile. La difficulté d'un item dans un cadre non paramétrique est représenté par la probabilité empirique de réponses négatives à l'item.

Reprenons le tableau de contingence entre les items X_j et X_k (voir tableau 6.3).

		X_k		
		1	0	
X_j	1	A	B	$A + B$
	0	C	D	$C + D$
		$A + C$	$B + D$	N

TAB. 6.3 – Tableau de contingence entre les items dichotomiques X_j et X_k

L'item X_j est dit plus facile que l'item X_k si $P(X_j = 1) > P(X_k = 1)$ c'est-à-dire si $A + B > A + C$. Dans ce cas, le nombre d'erreurs de Guttman e_{jk} entre les items X_j et X_k

est égal à

$$e_{jk} = N.P(X_j = 0, X_k = 1) = C \quad (6.10)$$

Ainsi, si l'item X_j est le plus facile ($B > C$), alors $e_{jk} = C$, sinon (si $C > B$) $e_{jk} = B$.
Alors

$$e_{jk} = \min(B, C) \quad (6.11)$$

Soit $e_{jk}^{(0)}$ le nombre d'erreurs de Guttman attendues sous l'hypothèse d'indépendance des réponses aux items, alors

$$e_{jk}^{(0)} = N \frac{e_{jk} + D}{N} \frac{e_{jk} + A}{N} \quad (6.12)$$

6.2.2.2 Extension au cadre polytomique

Il est possible d'étendre les erreurs de Guttman dans le cadre des items polytomiques [112] [117]. Dans ce cas, on définit par $e_{j(r)k(s)}$ les erreurs de Guttman entre la modalité r de l'item j et la modalité s de l'item k . La modalité r de l'item j est plus facile que la modalité s de l'item k si $P(X_j \geq r) > P(X_k \geq s)$. Alors

$$e_{j(r)k(s)} = N. \min(P(X_j \geq r, X_k < s); P(X_j < r, X_k \geq s)) \quad (6.13)$$

Alors le nombre d'erreurs de Guttman de la paire d'items polytomiques (X_j, X_k) est

$$e_{jk} = \sum_{r=1}^{m_j} \sum_{s=1}^{m_k} e_{j(r)k(s)} \quad (6.14)$$

De la même manière, on peut définir le nombre $e_{j(r)k(s)}^{(0)}$ d'erreurs attendues de Guttman entre la modalité r de l'item j et la modalité s de l'item k sous l'hypothèse d'indépendance des deux items par

$$e_{j(r)k(s)}^{(0)} = N.P(X_j < r)P(X_k \geq s), \quad (6.15)$$

avec $P(X_j \geq r) > P(X_k \geq s)$ (la modalité r de l'item j est plus facile que la modalité s de l'item k). Alors le nombre d'erreurs de Guttman attendues sous l'hypothèse d'indépendance des deux items est

$$e_{jk}^{(0)} = \sum_{r=1}^{m_j} \sum_{s=1}^{m_k} e_{j(r)k(s)}^{(0)} \quad (6.16)$$

6.2.3 Les coefficients d'échelles de Loevinger

Les coefficients d'échelles suivants ont été proposés par Loevinger [94]. Nous les définirons à partir des erreurs de Guttman, en concordance avec Molenaar [112] ou avec le manuel du logiciel MSP [117].

6.2.3.1 Le coefficient d'échelle de Loevinger entre deux items

Soit H_{jk} le coefficient d'échelle entre deux items j et k . Alors

$$H_{jk} = 1 - \frac{e_{jk}}{e_{jk}^{(0)}} \quad (6.17)$$

Mokken présente ce coefficient comme le rapport entre la covariance σ_{jk} des items j et k et la covariance maximale possible à marges fixées dans le tableau de contingence des items j et k notée $\sigma_{jk(max)}$.

$$H_{jk} = \frac{\sigma_{jk}}{\sigma_{jk(max)}} \quad (6.18)$$

Démonstration.

La covariance entre les items X_j et X_k est égale à

$$\sigma_{jk} = \frac{A}{N} - \frac{A+B}{N} \frac{A+C}{N} = \frac{NA - (A+B)(A+C)}{N} \quad (6.19)$$

On crée (tableau 6.4) le nouveau tableau de contingence théorique entre les items X_j et X_k avec les marges fixées permettant de maximiser la covariance σ_{jk} :

		X_k		
		1	0	
X_j	1	A^*	B^*	$A+B$
	0	C^*	D^*	$C+D$
		$A+C$	$B+D$	N

TAB. 6.4 – Tableau de contingence théorique entre les items dichotomiques X_j et X_k pour maximiser la covariance

Dans ce cas, la covariance vaut

$$\sigma_{jk(max)} = \frac{A^*}{N} - \frac{A+B}{N} \frac{A+C}{N} = \frac{NA^* - (A+B)(A+C)}{N} \quad (6.20)$$

Pour que cette covariance soit la covariance maximale à marge fixée, il faut donc maximiser A^* , ainsi $A^* = A + \min(B; C) = A + e_{jk}$ et $\min(B^*, C^*) = 0$. Ainsi, on obtient $B^* = B - e_{jk}$, $C^* = C - e_{jk}$ et $D^* = D + e_{jk}$. Comme les marges sont fixées, on a $A+B = A^* + B^*$ et $A+C = A^* + C^*$ donc

$$\begin{aligned} (A+B)(A+C) &= (A^* + B^*)(A^* + C^*) \\ &= A^{*2} + A^*B^* + A^*C^* + B^*C^* \\ &= A^*(N - D^*) \end{aligned} \quad (6.21)$$

Ainsi

$$\begin{aligned} \frac{\sigma_{jk}}{\sigma_{jk(max)}} &= \frac{NA - A^*(N - D^*)}{NA^* - A^*(N - D^*)} \\ &= \frac{NA - A^*N + A^*D^*}{A^*D^*} \\ &= 1 + \frac{N(A - A^*)}{(A + e_{jk})(D + e_{jk})} \\ &= 1 - \frac{e_{jk}}{e_{jk}^{(0)}} \\ &= H_{jk} \end{aligned} \quad (6.22)$$

$$(6.23)$$

□

Hemker, Sitjsma et Molenaar [71] présentent en outre ces coefficients de cette manière. Nous privilégierons la première approche pour deux raisons : d'une part, elle ne pose pas le problème de définition d'une covariance entre deux variables catégorielles, et en outre, elle nous semble plus facile à interpréter.

Ce coefficient est simple à interpréter. Il est toujours inférieur à 1 (la valeur 1 étant atteinte quand il n'y a aucune erreur de Guttman observée entre les deux items).

Quand il est proche de 1, cela signifie qu'il y a peu d'erreurs de Guttman observées (comparativement au nombre attendu sous l'hypothèse d'indépendance). Quand il est proche de 0, cela signifie que les réponses aux deux items sont indépendantes. Quand il est négatif, cela signifie qu'il y a plus d'erreurs de Guttman observées que d'attendues sous le modèle d'indépendance. Dans ce cas, il faut inverser les réponses (pour un des items, la réponse positive est favorable au trait latent, tandis que pour l'autre item, la réponse positive est défavorable au même trait latent). Ainsi, sous réserve que les items j et k soient correctement codés, le coefficient H_{jk} de Loevinger ne doit pas être significativement négatif.

Mokken [107] montre que si les items X_j et X_k sont monotonement homogènes, alors $\sigma_{jk} > 0$ et $H_{jk} > 0$. Donc pour qu'un ensemble d'items soit monotonement homogène, il faut que toutes les paires d'items (j, k) vérifie $H_{jk} > 0$. Il existe un test pour tester la non nullité du coefficient H_{jk} . Il s'agit du test z [117]. La statistique de ce test vaut

$$z_{jk} = \frac{AD - BC}{N\sqrt{(A+B)(A+C)(B+D)(C+D)}} \quad (6.24)$$

et cette statistique suit sous l'hypothèse de nullité du coefficient H_{jk} une loi normale centrée réduite.

6.2.3.2 Le coefficient de Loevinger d'intégration d'un item dans une échelle

Soient S un ensemble d'items, et H_j^S le coefficient d'intégration d'un item j appartenant à S ($j \in S$) dans l'échelle formée par les items de l'ensemble S .

Alors on pose e_j^S le nombre d'erreurs de Guttman associées à l'item j par rapport à l'ensemble des items de S et $e_j^{S(0)}$ ce nombre attendu sous l'hypothèse d'indépendance des réponses entre l'item j et chacun des items de l'ensemble S . On a alors $e_j^S = \sum_{k \in S, k \neq j} e_{jk}$ et $e_j^{S(0)} = \sum_{k \in S, k \neq j} e_{jk}^{(0)}$,

$$H_j^S = 1 - \frac{e_j^S}{e_j^{S(0)}} = \frac{\sum_{k \in S, k \neq j} e_{jk}^{(0)} H_{jk}}{\sum_{k \in S, k \neq j} e_{jk}^{(0)}} \quad (6.25)$$

Cet indice s'interprète sous réserve de considérer que tous les items de S en dehors de l'item j vérifient de bonnes propriétés d'échelles (vérifient un MMHM). Dans ce cas, un indice H_j^S proche de 1 signifie que l'item j s'intègre bien dans l'échelle S , tandis qu'un coefficient H_j^S proche de 0 signifie que l'item j s'intègre mal dans l'échelle formée de l'ensemble des autres items de S .

6.2.3.3 Le coefficient d'échelle de Loevinger

Soit e^S le nombre d'erreurs de Guttman entre les items d'un ensemble S et $e^{S(0)}$ ce nombre attendu sous l'hypothèse d'indépendance des réponses des items de S , alors $e^S = \sum_{j \in S} \sum_{k \in S, k > j} e_{jk}$ et $e^{S(0)} = \sum_{j \in S} \sum_{k \in S, k > j} e_{jk}^{(0)}$. Alors le coefficient H^S d'échelle de Loevinger vaut

$$H^S = 1 - \frac{e^S}{e^{S(0)}} = \frac{\sum_{j \in S} \sum_{k \in S, k > j} e_{jk}^{(0)} H_j^S}{\sum_{j \in S} \sum_{k \in S, k > j} e_{jk}^{(0)}} = \frac{\sum_{j \in S} \sum_{k \in S, k > j} e_{jk}^{(0)} H_{jk}}{\sum_{j \in S} \sum_{k \in S, k > j} e_{jk}^{(0)}} \quad (6.26)$$

On montre [108] que $H^S \geq \min_{j \in S} H_j^S$. Un coefficient H^S proche de 1 signifie que l'échelle formée par l'ensemble des items de S a de bonnes propriétés d'échelles (cet ensemble d'items vérifie un MMHM), tandis qu'un coefficient H^S proche de 0 montre cette échelle a de mauvaises propriétés.

6.2.4 Lien entre les coefficients d'échelles de Loevinger et le MMHM

Mokken [107] montre que si un ensemble d'items S vérifie un MMHM, alors H^S est significativement positif (dans un MMHM, $\forall j, k, H_{jk} > 0$ et on en déduit directement que $H^S > 0$). Cependant, la réciproque n'étant pas vraie, il suppose qu'il existe un seuil $0 < c < 1$ tel que si $H^S > c$, alors l'ensemble des items de S vérifie un MMHM. Il propose en outre la classification suivante, basée sur des observations empiriques :

- si $0.3 \leq H^S < 0.4$, alors l'échelle S est dite "faible" ("weak"),
- si $0.4 \leq H^S < 0.5$, alors l'échelle S est dite "medium" ("medium"),
- si $0.5 \leq H^S$, alors l'échelle S est dite "forte" ("strong").

Il suggère donc d'utiliser les coefficients de Loevinger pour construire un ensemble d'items vérifiant un MMHM. Cette idée a été formalisée par Hemker, Sitjsma et Molenaar [71] pour proposer la "Mokken Scale Procedure", une méthode de sélection automatique d'items permettant de construire des échelles vérifiant un MMHM.

6.2.5 Algorithme

A une étape initiale, on choisit la valeur d'un seuil c et une paire d'items (j, k) qui vérifie $H_{jk} > c$ tel que H_{jk} soit significativement supérieur à 0.

A l'étape n , on intègre dans l'échelle $S^{(n-1)}$ en cours de formation l'item j qui vérifie :

1. $j \notin S^{(n-1)}$
2. $S^{(n)} \equiv S^{(n-1)} \cup j$
3. $j = \arg \max_{k \notin S^{(n-1)}} H^{S^{*(n)}}$ avec $S^{*(n)} \equiv S^{(n-1)} \cup k$
4. $H_j^{S^{(n)}} \geq c$
5. H_{jk} significativement supérieur à 0, $\forall k \in S^{(n-1)}$

La procédure s'arrête dès qu'il n'existe plus d'items vérifiant l'ensemble de ces conditions. Pour éviter des problèmes de multiplicité des tests, à chaque étape, une correction de Bonferroni est effectuée sur le niveau de signification des tests.

On peut remplacer la paire d'items initiale par un noyau d'items choisi *a priori* par l'utilisateur. Une seconde échelle peut ensuite être formée à partir des items non sélectionnés dans la première échelle, et ainsi de suite, jusqu'à ce qu'il ne reste plus d'items disponibles, ou que les items restant ne puissent pas être groupés pour former une nouvelle échelle.

Le seuil c peut être choisi en fonction des caractéristiques voulues de l'échelle (plus il est grand, et plus l'échelle sera "forte" (au sens de Mokken)), ou de manière à vérifier certaines contraintes, comme le nombre d'items dans la première dimension. Dans tous les cas, les auteurs préconisent de choisir $c \geq 0.3$.

6.2.6 Résultats

Lors d'études de simulations précédentes à leur article, Hemker, Sijtsma et Molenaar [71] relèvent que les résultats de la procédure peuvent être modifiés en fonction :

- de la variance du trait latent (plus cette variance est grande, et plus la valeur de H^S est grande),
- du nombre de modalités d'items polytomiques (sans que l'influence dans un sens ou dans l'autre sur les résultats de la procédure soit clairement identifiée),
- de la corrélation entre les différents traits latents (plus cette corrélation est faible et meilleurs sont les résultats),
- du pouvoir discriminant des items (plus les items sont discriminants, et plus facilement ils sont sélectionnés).

En outre, ils réalisent une étude de simulations avec les paramètres suivants :

- le nombre de dimensions simulées ($Q = 2$ ou $Q = 3$)
- la corrélation entre les traits latents ($\rho = 0.0$, $\rho = 0.2$, $\rho = 0.4$, $\rho = 0.6$, $\rho = 0.8$ ou $\rho = 1.0$),
- le niveau de discrimination moyen pour les items d'une même dimension ($\alpha_M = 1.0$, $\alpha_M = 1.5$ ou $\alpha_M = 2.0$),
- la constance des pouvoirs discriminants des items d'une même dimension (constant ou non (± 0.5 autour de la valeur moyenne)).

Les simulations sont réalisées sur $N = 2000$ individus, avec $J_q = 9$, $\forall q = 1, \dots, Q$ et 4 répétitions de chaque cas. Les auteurs estiment que la taille de l'échantillon choisie ($N = 2000$) est suffisamment grande pour obtenir une certaine constance des résultats. Les données sont simulées à l'aide d'un modèle à réponses graduées [98] (extension du modèle de Birnbaum pour items polytomiques), avec une structure simple.

Cette étude de simulation permet d'étudier en profondeur les valeurs acceptables du seuil c suivant les conditions initiales. Ils montrent ainsi que pour des items faiblement discriminant ($\alpha_M = 1.0$), le seuil c doit être inférieur à 0.3, pour des items moyennement discriminant ($\alpha_M = 1.5$), il doit être inférieur à 0.4 et pour des items fortement discriminant ($\alpha_M = 2.0$), il doit être inférieur à 0.55. En outre, ils obtiennent de bons résultats avec des seuils très bas quand les traits latents sont très peu corrélés, mais ce seuil minimal doit être relevé quand les corrélations entre les traits latents augmentent. Les résultats sont similaires, que les pouvoirs discriminants des items d'une même dimension soient constants ou non, et avec deux ou trois dimensions simulées.

6.3 La procédure DETECT

6.3.1 L'indice DETECT et les indices Iss et R

Soit un ensemble de J items dichotomiques X_j , $j = 1, \dots, J$ formant une partition P en K classes. Soient Θ_q , $q = 1, \dots, Q$, Q traits latents avec $\Theta = (\Theta)_{q=1, \dots, Q}$ et soient α_q , $q = 1, \dots, Q$ Q poids associés avec $\alpha = (\alpha_q)_{q=1, \dots, Q}$. Soit un trait latent composite $\Theta\alpha = \alpha'\Theta$. On prendra comme définition de $\alpha_q = \sum_{j=1}^J \alpha_{jq}$ où α_{jq} représente le pouvoir discriminant de l'item j sur le trait latent q . On note Q la dimension de l'ensemble des items (définie par les auteurs comme le nombre minimal de traits latents pour que soient vérifiées l'indépendance locale des items et la monotonie - voir section 3.1)

On définit l'indice DETECT [155] (Dimensionality Evaluation To Enumerate Contributing Traits) comme :

$$D\alpha(P) = \frac{2}{J(J-1)} \sum_{j=1}^{J-1} \sum_{k=j+1}^J \delta_{jk}(P) E[Cov(X_j, X_k/\Theta\alpha)] \quad (6.27)$$

avec $\delta_{jk}(P) = 1$ si les items j et k sont placés dans le même groupe de la partition P et -1 sinon.

Soit P^* la partition en K^* classes des items qui maximise l'indice DETECT

$$D\alpha(P^*) = \max_P D\alpha(P) \quad (6.28)$$

Les auteurs montrent que

Théorème 6.1 (Théorème 1 dans [155]) *Quand $Q \leq 2$, alors l'indice DETECT est maximisé sur la vraie partition des items (P^* est la vraie partition des items).*

Théorème 6.2 (Théorème 2 dans [155]) *Quand l'ensemble des items a une structure approximativement simple, alors $K^* \leq Q$*

Soit les indices

$$Iss(P) = \frac{2}{J(J-1)} \sum_{j=1}^{J-1} \sum_{k=j+1}^J \delta_{jk}(P) \mathbb{I}_{\mathbb{R}^+} (E[Cov(X_j, X_k/\Theta\alpha)]) \quad (6.29)$$

et

$$R(P) = \frac{D\alpha(P)}{\frac{2}{J(J-1)} \sum_{j=1}^{J-1} \sum_{k=j+1}^J \delta_{jk}(P) |E[Cov(X_j, X_k/\Theta\alpha)]|} \quad (6.30)$$

On note que

$$-1 \leq Iss(P) \leq 1 \text{ et } -1 \leq R(P) \leq 1.$$

Théorème 6.3 (Théorèmes 3 et 4 dans [155]) *Un ensemble d'items a une structure approximativement simple si et seulement si*

$$\max_P Iss(P) = 1, \text{ et } \max_P R(P) = 1. \quad (6.31)$$

Définition 6.1 (Condition de régularité de la matrice de covariance des traits latents)

Soit σ_{jk} , $j = 1, \dots, J$, $k = 1, \dots, J$, $j \neq k$ les éléments de la matrice Σ de covariance des traits latents. Cette matrice vérifie la condition de régularité si

$$\sigma_{jk} < \frac{\left(\sum_{l=1}^Q \alpha_l \sigma_{jl}\right) \left(\sum_{m=1}^Q \alpha_m \sigma_{mk}\right)}{\sum_{l=1}^Q \sum_{m=1}^Q \alpha_l \alpha_m \sigma_{lm}} \quad (6.32)$$

Théorème 6.4 (Théorème 5 dans [155]) Si les items ont une structure simple, alors P^* est la vraie partition si la condition de régularité de la matrice de covariance des traits latents est respectée.

Les auteurs estiment que le théorème 6.4 continue de s'appliquer pour une structure approximativement simple si la condition de régularité de la matrice de covariance des traits latents est fortement respectée.

6.3.2 Estimation pratique des indices

En pratique, on ne peut pas calculer l'indice DETECT, puisque nous ne connaissons pas les valeurs des paramètres α_j , $j = 1, \dots, J$, et que la variable Θ_{α} est latente. Les auteurs proposent d'utiliser l'estimateur

$$\hat{D}(P) = \frac{2}{J(J-1)} \sum_{j=1}^{J-1} \sum_{k=j+1}^J \delta_{jk}(P) \hat{Cov}_{jk}, \quad (6.33)$$

où \hat{Cov}_{jk} est un "bon" estimateur de $E[Cov(X_j, X_k/\Theta_{\alpha})]$.

Ils proposent d'utiliser l'estimateur

$$\hat{Cov}_{jk}^* = \frac{1}{2} \left[\hat{Cov}_{jk}(S) + \hat{Cov}_{jk}(T) \right], \quad (6.34)$$

avec $S = \sum_{j=1}^J X_j$ et $T_{jk} = \sum_{l=1, l \neq j, k}^J X_l$ et

$$\hat{Cov}_{jk}(S) = \sum_{l=0}^J \frac{N_l}{N} \hat{Cov}(X_j, X_k/S = l), \quad (6.35)$$

$$\hat{Cov}_{jk}(T) = \sum_{l=0}^{J-2} \frac{N_{jkl}}{N} \hat{Cov}(X_j, X_k/T_{jk} = l), \quad (6.36)$$

avec N_l le nombre d'individus tels que $S = l$ et N_{jkl} le nombre d'individus tels que $T_{jk} = l$.

6.3.3 Algorithme de la procédure DETECT

Le théorème 6.4 permet de supposer qu'il est possible, en maximisant l'indice DETECT, et sous réserve que les items aient une structure approximativement simple, de retrouver cette structure. Les auteurs proposent un algorithme génétique pour maximiser l'indice DETECT, et nomment la procédure "procédure DETECT". Pour une présentation plus précise des algorithmes génétiques, nous renvoyons à l'article [155].

A une étape initiale, on produira trois partitions d'items à K classes (K fixé par l'utilisateur) basées sur la classification hiérarchique avec mesures conditionnelles de proximité présentée à la section 6.1 avec les mesures de proximité entre les ensembles d'items SL (simple linkage), CL (Complete linkage) et $UPGMA$ (Unweighted pair-group method of averages), et en utilisant les mesures de proximité $Ccov$ (ils proposent d'adapter ces mesures en utilisant les mesures \hat{Cov}_{jk}^* à la place de $\sum_{s=0}^{J-2} N_S Cov_S(X_j, X_k)$). Ces trois partitions seront appelées "partitions parents".

A chaque étape t , on fait "muter" M items (tel que $J/10 \leq M \leq J/5$) choisis aléatoirement, de telle manière que l'on considère toutes les possibilités d'appartenance d'un item mutant à chacune des classes existantes. On obtient ainsi $R = 3M(K - 1) + 3$ partitions mutantes, dont 3 d'entre elles seront des clones des partitions parents, et on calcule les R indices $\hat{D}_r(t)$, $r = 1, \dots, R$. On choisit alors les trois partitions qui maximisent $\hat{D}_r(t)$ et on passe à l'étape suivante ($t + 1$) en considérant celles-ci comme les nouvelles partitions parents. On arrête l'algorithme quand $\max_r D_r(t) = \max_r D_r(t - 1)$.

Pour déterminer le nombre de dimensions à retenir, on définit $\hat{d}(K)$ comme la valeur maximale obtenue avec K classes d'items. On pose par définition $\hat{d}(0) = \hat{d}(1) = 0$. L'algorithme génétique est réalisé pour $K = 2$, puis pour $K = 3$, et ainsi de suite. On arrête la procédure lorsque l'on obtient $\hat{d}(K - 2) \geq \max(\hat{d}(K - 1); \hat{d}(K))$. La partition en $(K - 2)$ classes d'items qui permet de trouver $\hat{d}(K - 2)$ est alors choisie.

Une importante critique que l'on peut faire à la procédure DETECT est le caractère non-déterministe de l'algorithme. Ainsi, il sera possible d'obtenir des résultats différents en relançant plusieurs fois la procédure sur un même ensemble d'items (ceci n'a cependant pas été étudié par les auteurs).

6.3.4 Résultats

Les auteurs ont réalisé une étude de simulations faisant varier :

- le nombre de dimensions (2, 3 ou 4),
- le nombre d'individus (400 ou 800),
- le nombre d'items total (20 ou 40).

100 répliquions de chaque cas ont été simulées. Les réponses sont simulées à l'aide d'une extension multidimensionnelle du modèle de Birnbaum (identique à celle utilisée par Roussos, Stout et Marden [132] pour l'étude de simulation avec la HCA/CCPROX, voir équation (6.7)), avec une structure approximativement simple. Les paramètres de discriminations des items sont tirés aléatoirement selon une loi normale bornée (avec une moyenne plus faible pour les dimensions secondaires que pour la dimension principale pour chaque item). Les paramètres de difficultés sont simulés par une loi normale bornée. Dans tous les cas, plus de 85% des simulations permettent d'obtenir un résultat parfait, et dans plus de 95% des cas, un résultat correct (bonne détermination du nombre de dimensions simulées mais un ou deux items classés incorrectement).

6.4 Comparaison de HCA/CCPROX, MSP et DETECT

6.4.1 Étude de simulations

Dans un article récent [144], van Abswoude, van der Ark et Sijtsma comparent les trois procédures HCA/CCPROX, MSP et DETECT. Les paramètres utilisés dans les simulations sont :

- le modèle utilisé pour simuler les données (extrapolations multidimensionnelles compensatoires du 2-PLM et du 5-PAM [134]),
- le nombre de traits latents (2 ou 4),
- les corrélations entre les traits latents (6 niveaux : 0.0 à 1.0 par seuil de 0.2),
- le nombre d'items reliés à chaque trait latent (4 combinaisons),
- les pouvoirs discriminants des items (2 niveaux).

Certains paramètres restent en outre fixés sur l'ensemble des simulations : la taille de l'échantillon N à 2000 et la structure des items, qui sera une structure simple dans tous les cas. Les paramètres de difficultés des items sont en outre équidistants dans chaque dimension sur l'intervalle $[-2; 2]$.

Le 5-PAM (5 parameters acceleration model) est modélisé par :

$$P(X_{nj} = 1/\theta_n, \alpha_j, \delta_j, \xi_j, \gamma_j^{up}, \gamma_j^{low}) = \gamma_j^{low} + \left(\gamma_j^{up} - \gamma_j^{low} \right) \left\{ \frac{\exp [1.7\alpha_j (\theta_n - \delta_j)]}{1 + \exp [1.7\alpha_j (\theta_n - \delta_j)]} \right\}^{\xi_j}, \quad (6.37)$$

où γ_j^{low} et γ_j^{up} permettent asymptotiquement d'avoir une probabilité de réponse positive supérieure à 0 et inférieure à 1, et ξ_j est appelé paramètre d'accélération de l'item j et rend la fonction de réponse à l'item dissymétrique si $\xi_j \neq 1$. Dans les simulations, γ_j^{low} est tiré aléatoirement dans une loi uniforme sur l'intervalle $[0.0; 0.2]$, γ_j^{up} est tiré aléatoirement sur une loi uniforme sur l'intervalle $[0.8; 1.0]$ et ξ_j est tiré aléatoirement par une loi uniforme sur l'intervalle $[0.25; 7]$.

Concernant les pouvoirs discriminants pour chaque item relatif à la dimension à laquelle est relié cet item, ils sont choisis avec des valeurs basses, c'est-à-dire en les simulant par une loi normale de moyenne 0.75 et de variance 0.1, en tronquant à 0.5 et 1.25, ou avec une valeur haute, et sont dans ce cas simulés par une loi normale de moyenne 1.75 et de variance 0.1, tronquée à 1.5 et 2.0. Les plans suivants sont utilisés : $[bas; bas]$, $[bas; haut]$ et $[haut; haut]$ pour $Q = 2$ et $[bas; bas; bas; bas]$, $[bas; bas; haut; haut]$ et $[haut; haut; haut; haut]$ pour $Q = 4$.

Enfin, concernant le nombre d'items dans chaque dimension, les simulations sont réalisées avec les plans suivants : $[7; 7]$, $[7; 21]$, $[21; 7]$ et $[21; 21]$ pour $Q = 2$ et $[7; 7; 7; 7]$, $[7; 7; 21; 21]$, $[21; 21; 7; 7]$ et $[21; 21; 21; 21]$ pour $Q = 4$.

Chaque simulation est répliquée une seule fois.

6.4.2 Résultats

Les résultats retrouvés par cette étude sont les suivants :

- MSP est moins efficace que les deux autres méthodes pour retrouver la partition réelle des items quand la corrélation entre les traits latents est forte,
- Pour toutes les méthodes, mais en particulier pour MSP, de forts pouvoirs discriminants des items permettent de retrouver plus efficacement la partition des items,
- Les résultats sont peu sensibles au nombre d'items, cependant, MSP peut parfois, sur des échelles avec beaucoup d'items, créer deux sous-échelles avec d'une part les items avec un faible pouvoir discriminant et d'autre part les items avec un fort pouvoir discriminant,
- HCA/CCPROX est parfois impossible à réaliser avec beaucoup d'items (présence de 0 dans les tables de contingence) : des adaptations sont possibles,
- le nombre de dimensions affecte peu les résultats,
- l'utilisation de l'extrapolation multidimensionnelle du 5-PAM affecte de façon importante les résultats des trois procédures, en particulier pour DETECT et HCA/CCPROX.

6.5 Synthèse

Les trois procédures non paramétriques de l'IRT présentées dans ce chapitre offrent d'intéressantes alternatives aux méthodes factorielles pour construire des échelles d'items non seulement unidimensionnelles mais vérifiant en outre les autres propriétés fondamentales de l'IRT (monotonie et indépendance locale). Ces procédures ne sont cependant pas basées sur un modèle paramétrique, aussi, par exemple, la construction de score pour la mesure des traits latents dans chaque ensemble formé n'est basé que sur des conjectures. Dans le chapitre suivant, nous proposons donc différentes manières de construire des ensembles d'items de manière à ce que ceux-ci aient une bonne adéquation à un modèle de l'IRT (nous utiliserons essentiellement le modèle de Rasch).

Chapitre 7

Propositions : La classification d'items en sous-échelles basée sur l'adéquation à un modèle

7.1 Introduction

Les méthodes présentées dans les deux derniers chapitres sont intéressantes dans la plupart des cas pour détecter des échelles unidimensionnelles. Cependant, il est souvent souhaitable, outre l'unidimensionnalité, d'obtenir des échelles vérifiant d'autres propriétés. Les trois méthodes proposées au chapitre 6 permettent d'obtenir des échelles qui vérifient les trois propriétés fondamentales de l'IRT, c'est-à-dire, outre l'unidimensionnalité, l'indépendance locale et la monotonie. Ces propriétés peuvent se révéler insuffisantes.

Le modèle de Rasch a connu un véritable succès grâce à la propriété d'exhaustivité du score pour le trait latent. Dans ce modèle, il est aisé, et théoriquement justifiable, d'utiliser le score plutôt que le trait latent pour classer les individus, et les placer dans un continuum : une fois les estimations des valeurs individuelles du trait latent connues pour chaque valeur possible du score, il est possible d'estimer la valeur individuelle sur le trait latent pour chaque individu.

Aussi, il nous a semblé intéressant de proposer des méthodes pour construire des échelles vérifiant, outre les propriétés fondamentales de l'IRT, celles du modèle de Rasch. Deux méthodes sont proposées dans cette partie. Elles sont basées toutes les deux sur l'adéquation des données à un modèle de Rasch. La première méthode explicitée à la section 7.2, appelée "BackRasch", est basée sur une procédure pas à pas descendante, où à chaque étape, on retire l'item ayant la plus mauvaise adéquation à un modèle de Rasch comprenant tous les items non retirés de l'échelle. La seconde méthode, appelée "Raschfit" est basée sur une procédure pas à pas ascendante, dont l'algorithme, plus complexe, est explicité à la section 7.3.

7.2 La procédure "BackRasch", procédure pas à pas descendante basée sur le modèle de Rasch

7.2.1 Principes

Le principe de la procédure "BackRasch" [60] [61] est celui des procédures pas à pas descendantes (Backward). A chaque étape, l'item qui a la moins bonne adéquation au modèle de Rasch sous-jacent à l'ensemble des items restants, est éliminé, et ce jusqu'à ce qu'il n'y ait plus d'item ayant une mauvaise adéquation au modèle.

7.2.2 Algorithme

L'algorithme se passe en un maximum de $K = (J - 2)$ étapes :

- A l'étape initiale, on sélectionne l'ensemble des items dans le modèle, et on fixe le niveau de signification α (par exemple $\alpha = 0.05$),
- A l'étape k , on estime les paramètres du modèle de Rasch avec les items non supprimés, et on calcule des statistiques d'adéquations locales de chaque item au modèle (statistiques R_{1Cj} , R_{1Mj} ou Q_{1j} , que l'on dénommera de façon générique par F_j). L'item j_0 est celui qui a la plus mauvaise adéquation au modèle : $F_{j_0} = \max_j F_j$. L'item j_0 est supprimé de la sous-échelle si F_{j_0} est significatif au niveau α . Si F_{j_0} est non significatif, l'algorithme est arrêté.

La première sous-échelle est composée des items qui n'ont pas été supprimés. Dans un second temps, on peut construire une seconde sous-échelle en partant de l'ensemble des items supprimés de la première sous-échelle, et en relançant le même algorithme sur ce nouvel ensemble d'items, et ainsi de suite pour construire les sous-échelles suivantes.

Cette méthode est relativement aisée à extrapoler à d'autres modèles que le modèle de Rasch. Il suffit simplement d'utiliser un test d'adéquation locale adapté au modèle.

7.3 Une proposition : la procédure "Raschfit", procédure pas à pas ascendante basée sur le modèle de Rasch

7.3.1 Principes

Nous proposons [68] une méthode pas à pas ascendante permettant de sélectionner des items ayant une bonne adéquation à un modèle de Rasch. Le problème des méthodes pas à pas ascendantes est cependant de devoir estimer les paramètres de beaucoup de modèles différents. Classiquement, l'algorithme de ces méthodes est de la forme :

- on part avec un modèle nul (aucune variable explicative),
- à chaque étape on ajoute la variable qui apporte le plus d'information, conditionnellement au fait que les variables déjà sélectionnées en apportent (en général, on utilise comme mesure de l'information la différence entre les log-vraisemblances maximales des deux modèles multipliée par 2),
- on arrête l'algorithme quand aucune variable non sélectionnée ne peut apporter de l'information de manière significative (en utilisant les log-vraisemblances, l'écart doit être plus grand qu'un chi-deux à k degrés de liberté où k correspond au nombre de degrés

de liberté apporté par la variable).

Dans ce type de procédure, on doit estimer les paramètres d'autant de modèles à chaque étape qu'il ne reste d'items non sélectionnés, soit J modèles à la première étape, $J - 1$ à la seconde étape, etc... jusqu'à ce que l'algorithme s'arrête. Si tous les items sont sélectionnés, alors globalement $\frac{(J+1)J}{2}$ modèles sont estimés sur l'ensemble de la procédure.

Dans le cadre d'un modèle de réponse aux items, nous sommes confrontés à deux problèmes :

1. les modèles sont longs à estimer, et une procédure qui voit le nombre de modèles augmenter quadratiquement avec le nombre d'items disponibles peut se révéler inadaptée pour un nombre important d'items,
2. la variable réponse n'est pas unique pour chaque individu : il y a autant de variables réponses pour chaque individu que d'items dans le modèle, aussi il est difficile de comparer des modèles n'ayant pas le même nombre d'items car les vraisemblances obtenues ne porteront pas sur le même nombre d'observations.

Pour résoudre ces deux problèmes, nous proposons la procédure "Raschfit" qui permet d'estimer un nombre de modèles qui augmente seulement linéairement avec le nombre d'items, et qui compare à chaque étape deux modèles considérant le même nombre d'items.

Cette procédure est basée sur l'idée que, si un nouvel item mesure un autre trait latent que les items déjà sélectionnés, alors l'adéquation des données par un modèle de Rasch devrait être moins bonne que l'adéquation des mêmes données dans le cadre d'un modèle où les items déjà sélectionnés mesureront un trait latent et le nouvel item mesurera un second trait latent. Le modèle multidimensionnel choisi sera un MMSRM.

L'adéquation doit tenir compte du fait que le MMSRM bi-dimensionnel comprend plus de paramètres que le modèle de Rasch pour le même nombre d'items. Le critère d'information d'Akaike (AIC) [4] qui est un critère de parcimonie sera utilisé pour mesurer l'adéquation des données au modèle relativement au nombre de paramètres utilisés. AIC est calculé de la manière suivante :

$$AIC = -2l + 2K, \quad (7.1)$$

où l est la log-vraisemblance du modèle et K est le nombre de paramètres du modèle.

Pour n'avoir à estimer qu'un nombre de modèles croissant linéairement avec le nombre d'items initiaux, il faut avoir à tester le même nombre de modèles à chaque étape. Pour ce faire, les items seront ordonnés, et l'ajout d'un seul item sera testé à chaque étape. Les paramètres de deux modèles seront donc à estimer à chaque étape de l'algorithme.

Notons enfin que pour pouvoir estimer les paramètres d'un modèle de Rasch, il faut au minimum 2 items, aussi initialement, un noyau d'au moins deux items sera sélectionné.

Ainsi, en admettant que le noyau initial comporte 2 items, il y aura $2(J - 2)$ modèles à estimer, dont $(J - 2)$ modèles de Rasch et $(J - 2)$ MMSRM.

7.3.2 Représentation graphique des deux modèles testés à chaque étape

A chaque étape nous comparons deux modèles. La figure 7.1 représente le modèle unidimensionnel et la figure 7.2 représente le modèle bi-dimensionnel comparés à chaque étape de la procédure.

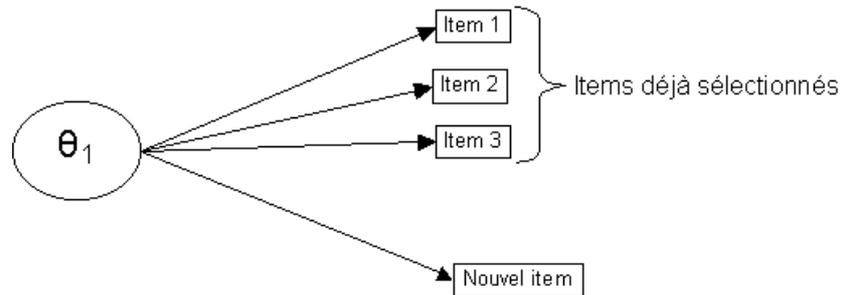


FIG. 7.1 – Modèle unidimensionnel avec trois items déjà sélectionnés et un nouvel item

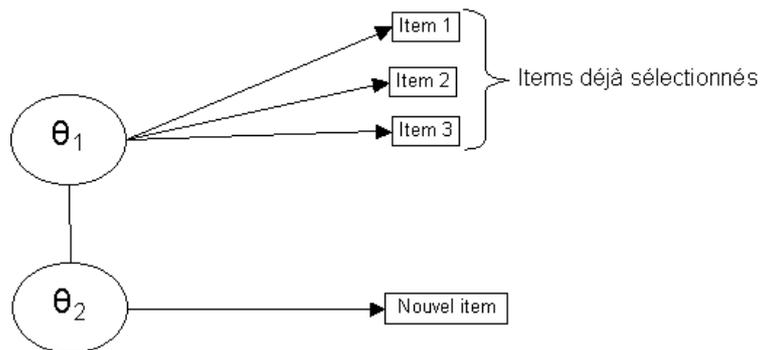


FIG. 7.2 – Modèle bi-dimensionnel avec trois items déjà sélectionnés et un nouvel item

7.3.3 Algorithme

L'algorithme de la procédure est le suivant :

- A l'étape initiale, un noyau d'items d'au moins deux items est choisi, et les autres items sont ordonnés,
- A chaque étape, le nouvel item est sélectionné dans la sous-échelle si le critère d'Akaike mesuré avec le modèle de Rasch avec cet item et les items sélectionnés est plus petit que celui mesuré dans un MMSRM où les items déjà sélectionnés sont reliés à un trait latent et le nouvel item à un second trait latent. Sinon, le nouvel item est rejeté.

L'algorithme s'arrête quand il ne reste plus d'items disponibles. On peut alors construire une seconde échelle avec les items rejetés de la première sous-échelle, et ainsi de suite, jusqu'à ce qu'il n'y ait plus d'items disponibles.

7.3.3.1 Noyau initial

Le noyau initial peut être choisi par l'utilisateur, en se basant soit sur des aspects sémantiques (un ou plusieurs expert(s) définira(ont) alors les items qui selon-lui(eux) mesurent le même concept), soit sur des informations spécifiques (tel sous-ensemble d'items a une bonne adéquation à un modèle de Rasch), soit en fonction de propriétés psychométriques.

Dans ce dernier cas, et en l'absence de toute autre information sur les items, nous proposons de choisir comme noyau la paire d'items (j, k) qui a le plus grand coefficient d'échelle de Loevinger : $(j, k) = \arg \max_{(l, m)} H_{lm}$ (on cherche alors les deux items les plus "corrélés" entre eux au sens de Loevinger).

Le choix du noyau initial d'items a sans doute son importance, que nous n'avons cependant pas mesurée.

7.3.3.2 Ordre des items

Dans l'algorithme, l'ajout d'un seul item est testé à chaque étape. Ceci sous-entend que l'ordre dans lequel vont être classés les items a une importance. Nous retiendrons trois cas : l'ordre des items est aléatoire, l'ordre des items est basé sur une connaissance (avis d'expert ou autre), l'ordre des items est basé sur les propriétés psychométriques des items.

Dans le dernier cas, nous proposons de classer les items dans l'ordre d'inclusion de ces items dans le cadre de la Mokken Scale Procedure (MSP) en retenant un seuil c minimum (0) et en n'effectuant aucun test de signification : ainsi, tous les items seront sélectionnés par cette procédure dans une seule sous-échelle. En outre, le noyau tel que défini à la section 7.3.3.1 sera utilisé dans la procédure MSP.

Nous retiendrons en outre deux cas, lorsque les items ne sont pas classés aléatoirement : les items sont classés de l'item qui s'intégrera de façon la plus probable au modèle de Rasch avec les items déjà sélectionnés jusqu'à l'item qui s'intégrera de façon la moins probable au modèle de Rasch avec les items sélectionnés (nous appellerons cet ordre le classement "positif"), et le classement inverse (appelé classement "négatif").

Le classement "positif" est sans doute le classement le plus robuste, car ce classement favorise l'intégration rapide (au plus tôt dans les étapes de l'algorithme) des items, ainsi les sous-échelles formées contiendront rapidement beaucoup d'items. En revanche, ce classement est celui qui rendra l'exécution de la procédure la plus longue possible, puisque la sous-échelle en cours sera rapidement importante en taille, donc les modèles contiendront plus de paramètres à estimer (ce qui sera un processus plus long). Le classement "négatif" a les avantages et inconvénients inversés par rapport au classement "positif" : la procédure sera moins robuste, mais la procédure sera plus rapide à exécuter.

7.3.3.3 Estimation des paramètres des modèles

Les traits latents sont considérés comme des variables aléatoires. A ce titre, on utilise la méthode du maximum de vraisemblance marginale pour estimer les paramètres du modèle.

Comme il l'est expliqué à la section 4.7, nous utiliserons une estimation en deux temps des paramètres du MMSRM : dans un premier temps, les paramètres de difficulté des items sont estimés dans des modèles unidimensionnels ; dans un second temps, on estime les paramètres de la distribution du trait latent multidimensionnel conditionnellement aux estimations obtenues pour les paramètres de difficulté du modèle. Cette façon de faire permet un grand gain en terme de temps de calcul.

On remarquera cependant que dans chaque MMSRM, un seul item est relié à la seconde dimension. Il est alors impossible d'estimer les paramètres du modèle de Rasch correspondant à cette dimension.

On peut estimer le paramètre de difficulté de l'item en question (que l'on indexera par j_0 - le paramètre de difficulté en question est donc δ_{j_0}) en donnant à chaque individu la moyenne du trait latent correspondant comme valeur individuelle de celui-ci : cette moyenne est fixée à 0 (voir section 2.6.8.1). On résoud alors l'équation :

$$\hat{\delta}_{j_0} = \max_{\delta_{j_0}} \prod_{n=1}^N P(X_{n0} = x_{nj_0} / \delta_{j_0}) = \max_{\delta_{j_0}} \prod_{n=1}^N \frac{\exp(-x_{nj_0} \delta_{j_0})}{1 + \exp(-\delta_{j_0})} \quad (7.2)$$

On montre facilement que

$$\hat{\delta}_{j_0} = -\log \left(\frac{t_{j_0}}{N - t_{j_0}} \right) \quad (7.3)$$

avec $t_{j_0} = \sum_{n=1}^N x_{nj_0}$.

Démonstration.

$$\begin{aligned} \hat{\delta}_{j_0} &= \max_{\delta_{j_0}} \prod_{n=1}^N \frac{\exp(-x_{nj_0} \delta_{j_0})}{1 + \exp(-\delta_{j_0})} \\ &= \max_{\delta_{j_0}} \log \left[\prod_{n=1}^N \frac{\exp(-x_{nj_0} \delta_{j_0})}{1 + \exp(-\delta_{j_0})} \right] \\ &= \max_{\delta_{j_0}} \log \left[\frac{\exp\left(-\sum_{n=1}^N x_{nj_0} \delta_{j_0}\right)}{\prod_{n=1}^N [1 + \exp(-\delta_{j_0})]} \right] \\ &= \max_{\delta_{j_0}} \{(-t_{j_0} \delta_{j_0}) - N \log [1 + \exp(-\delta_{j_0})]\} \\ &= \max_{\delta_{j_0}} f(\delta_{j_0}/t_{j_0}, N) \end{aligned} \quad (7.4)$$

$$\frac{df(\delta_{j_0}/t_{j_0}, N)}{d\delta_{j_0}} = -t_{j_0} - N \frac{-\exp(-\delta_{j_0})}{1 + \exp(-\delta_{j_0})} \quad (7.5)$$

Donc

$$-t_{j_0} + N \frac{\exp(-\hat{\delta}_{j_0})}{1 + \exp(-\hat{\delta}_{j_0})} = 0 \quad (7.6)$$

$$\frac{N \exp(-\hat{\delta}_{j_0})}{1 + \exp(-\hat{\delta}_{j_0})} = t_{j_0} \quad (7.7)$$

$$N \exp(-\hat{\delta}_{j_0}) = t_{j_0}(1 + \exp(-\hat{\delta}_{j_0})) \quad (7.8)$$

$$(N - t_{j_0}) \exp(-\hat{\delta}_{j_0}) = t_{j_0} \quad (7.9)$$

$$\exp(-\hat{\delta}_{j_0}) = \frac{t_{j_0}}{N - t_{j_0}} \quad (7.10)$$

$$\hat{\delta}_{j_0} = -\log\left(\frac{t_{j_0}}{N - t_{j_0}}\right) \quad (7.11)$$

□

7.3.3.4 Calcul du critère d'Akaike

Le critère d'Akaike est facile à calculer une fois connue la valeur de la log-vraisemblance l et le nombre de paramètres K (voir équation (7.1)).

Dans un modèle unidimensionnel (Rasch classique), il y a J paramètres de difficultés δ_j , $j = 1, \dots, J$ et 1 paramètre pour définir la distribution du trait latent (nous utiliserons une loi normale centrée, donc ce paramètre correspond à la variance du trait latent), donc $K = J + 1$. Dans un MMSRM à deux dimensions, il y a toujours J paramètres de difficulté δ_j , $j = 1, \dots, J$ et le trait latent bi-dimensionnel est supposé être distribué selon une loi multinormale centrée, donc trois paramètres caractérisent sa distribution : les deux variances des traits latents, et la covariance entre les traits latents. Donc dans ce modèle, $K = J + 3$.

7.3.4 Améliorations sur le temps de calcul : la procédure Raschfit-Fast

7.3.4.1 Principes

La procédure décrite dans les précédentes sections a un lourd inconvénient : estimer les paramètres d'un modèle bidimensionnel est un processus long. Aussi, nous proposons l'adaptation suivante présentée en [67] et [69].

Cette adaptation est basée sur l'idée que, si un ensemble de J items X_j , $j = 1, \dots, J$ a une bonne adéquation à un modèle de Rasch, alors les réponses de chacun de ces items seront mieux expliquées par le score $S = \sum_{j=1}^J X_j$ que par une constante. Ainsi, si l'on ajoute un nouvel item à cet ensemble, et que ce nouvel ensemble d'items vérifient un modèle de Rasch, alors la réponse au nouvel item sera mieux expliquée par le score obtenu sur les $(J + 1)$ items que par une constante, tandis que les réponses des J premiers items ne seront pas moins bien expliquées par le score calculé sur les $(J + 1)$ items que par celui calculé sur les J premiers items.

7.3.4.2 Représentation graphique

Le modèle "bi-dimensionnel" peut être représenté comme dans la figure 7.3. Ce modèle est à comparer au modèle présenté à la figure 7.1.

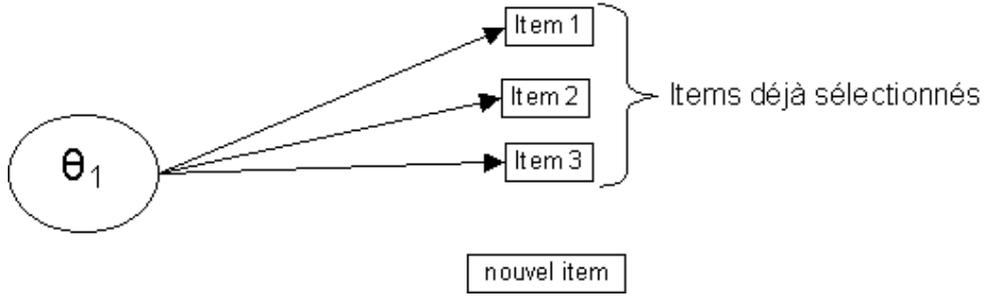


FIG. 7.3 – Modèle "bi-dimensionnel" avec trois items déjà sélectionnés et un nouvel item dans le cadre de la version rapide de la procédure "Raschfit"

7.3.4.3 Modèle

Dans cette adaptation de la procédure "Raschfit", nous utiliserons la propriété d'exhaustivité du score sur le trait latent : nous nous plaçons donc délibérément dans le cadre du modèle de Rasch à effets fixes (le trait latent n'est plus considéré comme une variable latente). Cette propriété implique que toute l'information disponible sur le trait latent est disponible dans la variable score S . Ainsi, si deux individus n_1 et n_2 ont la même valeur de score, alors les estimations des valeurs individuelles du trait latent pour ces deux individus seront égales, ou en d'autres termes :

$$s_{n_1} = s_{n_2} \Rightarrow \hat{\theta}_{n_1} = \hat{\theta}_{n_2} \quad (7.12)$$

Des estimations consistantes des valeurs individuelles du trait latent peuvent être obtenues conditionnellement aux paramètres de difficulté des items [73]. Dans le cas général, les valeurs des paramètres de difficulté sont inconnues mais peuvent être estimées de façon consistante par maximum de vraisemblance conditionnelle.

On notera alors θ_s la valeur individuelle du trait latent d'un individu ayant un score égal à s , $\hat{\delta}_j$ l'estimation par maximum de vraisemblance conditionnelle du paramètre de difficulté de l'item j et $\hat{\boldsymbol{\delta}} = (\hat{\delta}_j)_{j=1, \dots, J}$. On peut écrire la vraisemblance des réponses aux items d'un individu n :

$$\begin{aligned} L_{S_n}(\theta_n / \hat{\boldsymbol{\delta}}, \mathbf{x}_n) &= L_{S_n}(\theta_{s_n} / \hat{\boldsymbol{\delta}}, \mathbf{x}_n) \\ &= \prod_{j=1}^J \frac{\exp(x_{nj}(\theta_{s_n} - \hat{\delta}_j))}{1 + \exp(\theta_{s_n} - \hat{\delta}_j)} \end{aligned} \quad (7.13)$$

On ne peut cependant estimer θ_s que pour $s \neq 0$ et $s \neq J$, ainsi la vraisemblance d'un N-échantillon est :

$$L_S(\boldsymbol{\theta} / \hat{\boldsymbol{\delta}}, \mathbf{x}) = \prod_{\substack{n=1 \\ s_n \neq 0, s_n \neq J}}^N L_{S_n}(\theta_{s_n} / \hat{\boldsymbol{\delta}}, \mathbf{x}_n) \quad (7.14)$$

où $\boldsymbol{\theta} = (\theta_n)_{n=1, \dots, N}$.

Ce modèle contient J paramètres de difficulté δ_j , $j = 1, \dots, J$ et $(J - 1)$ paramètres θ_s , $s = 1, \dots, J - 1$, donc $K = 2J - 1$, donc

$$AIC = -\max_{\boldsymbol{\theta}} \left\{ \log \left[L_S(\boldsymbol{\theta}/\hat{\boldsymbol{\delta}}, \mathbf{x}) \right] \right\} + 2(2J - 1) \quad (7.15)$$

7.3.4.4 Adaptation de la procédure

On adapte la procédure "Raschfit" avec cette nouvelle vision. On appellera cette nouvelle version "Raschfit-Fast". L'étape initiale et la définition du noyau restent identique à la procédure de base.

A chaque étape ultérieure de l'algorithme, on compare un modèle de Rasch classique en ajoutant à un ensemble E composé de J items déjà sélectionnés un nouvel item j_0 , et un modèle où cet item j_0 ne sera pas intégré dans le modèle de Rasch.

Soient $S_n = \sum_{j \in E} X_{nj}$ et $S_n^* = S_n + X_{nj_0}$. On définit par F l'ensemble des individus sur lesquels nous estimerons les paramètres du modèle. Cet ensemble est défini ci dessous. $\hat{\delta}_j$ est l'estimation de δ_j , $j \in E$ dans un modèle de Rasch obtenue par maximum de vraisemblance conditionnelle avec l'ensemble des individus de F . $\hat{\delta}_j^*$ est l'estimation de δ_j , $j \in E \cup j_0$ dans un modèle de Rasch réalisée par maximum de vraisemblance conditionnelle avec l'ensemble des individus de F . $\hat{\boldsymbol{\delta}} = (\hat{\delta}_j)_{j \in E}$ et $\hat{\boldsymbol{\delta}}^* = (\hat{\delta}_j^*)_{j \in E \cup j_0}$

Le premier modèle est donc un modèle de Rasch classique. On maximisera donc sa vraisemblance conditionnelle :

$$L_{S^*}(\boldsymbol{\theta}/\hat{\boldsymbol{\delta}}^*, \mathbf{x}) = \prod_{n \in F} L_{S_n^*}(\theta_{s_n^*}/\hat{\boldsymbol{\delta}}^*, \mathbf{x}_n) \quad (7.16)$$

Ce modèle contient $K^* = 2J + 1$ paramètres. Le critère d'Akaike de ce modèle est donc

$$AIC^* = -\max_{\boldsymbol{\theta}} \log \left\{ L_{S^*}(\boldsymbol{\theta}/\hat{\boldsymbol{\delta}}^*, \mathbf{x}) \right\} + 2(2J + 1) \quad (7.17)$$

Le second modèle est un modèle où les J items de E vérifient un modèle de Rasch et où les réponses à l'item j_0 sont expliquées par un modèle logistique nul (contenant seulement une constante). Sa vraisemblance conditionnelle est donc :

$$\begin{aligned} L_S(\boldsymbol{\theta}/\hat{\boldsymbol{\delta}}, \mathbf{x}) &= \prod_{n \in F} \left\{ \frac{\exp(-x_{nj_0} \hat{\delta}_{j_0})}{1 + \exp(-\hat{\delta}_{j_0})} \prod_{j \in E} \frac{\exp(x_{nj}(\theta_{s_n} - \hat{\delta}_j))}{1 + \exp(\theta_{s_n} - \hat{\delta}_j)} \right\} \\ &= \prod_{n \in F} \left\{ \frac{\exp(-x_{nj_0} \hat{\delta}_{j_0})}{1 + \exp(-\hat{\delta}_{j_0})} L_{S_n}(\theta_{s_n}/\hat{\boldsymbol{\delta}}, \mathbf{x}_n) \right\} \end{aligned} \quad (7.18)$$

Ce modèle contient $K = 1 + 2J - 1 = 2J$ paramètres. Le critère d'Akaike de ce modèle est donc

$$AIC = -\max_{\boldsymbol{\theta}} \log \left\{ L_S(\boldsymbol{\theta}/\hat{\boldsymbol{\delta}}, \mathbf{x}) \right\} + 2(2J) \quad (7.19)$$

L'item j_0 est sélectionné dans la même dimension que les items de E si $AIC^* \leq AIC$.

F est l'ensemble des individus qui peuvent être utilisés pour l'estimation dans les deux modèles, donc $F = \{n/s_n \neq J, s_n^* \neq 0, s_n^* \neq J + 1\}$.

On note que les vraisemblances des deux modèles sont estimées en deux temps : dans un premier temps, les paramètres de difficultés sont estimés, puis la vraisemblance est estimée dans un second temps en considérant les estimations des paramètres de difficultés comme les vraies valeurs de ces paramètres. Ceci nous oblige donc à faire l'hypothèse que le critère d'Akaike issu de cette estimation de la vraisemblance a le même comportement que le critère d'Akaike "classique" résultant de l'estimation de la vraisemblance du modèle en une seule étape. Cette façon de procéder a cependant déjà été utilisée par Hoijtink, Rooks et Wilmink [74].

7.3.5 Extensions à d'autres modèles

L'extension de la procédure Raschfit est facilement extrapolable à d'autres modèles. La seule contrainte est que ces modèles utilisent une structure simple des items. Dans un cadre polytomique, on peut utiliser le MMPSRM [67]. On pourrait aussi utiliser des extrapolations multidimensionnelles de certains modèles comme le modèle de Birnbaum, sous la condition qu'un seul pouvoir discriminant soit non nul ($\alpha_{jq} = 0$ si $q \neq q_j$).

Notons seulement que si la procédure "Raschfit" est généralisable à n'importe quel modèle à structure simple, l'adaptation "Raschfit-fast" ne l'est qu'aux modèles de la famille de Rasch.

7.4 Étude de simulations

Nous avons réalisé une étude de simulations (voir [56], [66], [68] et [120]) permettant de comparer différentes méthodes de sélection d'items en échelles, et en particulier les procédures "BackRasch", "Raschfit" et "Raschfit-fast".

Nous avons testé plusieurs paramètres de simulation : le modèle utilisé dans les simulations (2 modèles), la corrélation entre les traits latents (6 niveaux), la structure des items (simple ou approximativement simple - 2 niveaux), le nombre d'items dans chaque dimension (3 possibilités) et le pouvoir discriminant des items (6 possibilités). Ceci entraîne un plan d'expérience avec $2 \times 6 \times 2 \times 3 \times 6 = 432$ plans différents. Cependant, seuls 360 d'entre eux étaient différents. En outre, étant donné le temps de calcul requis pour certaines procédures, en particulier pour "Raschfit", une seule réplique de chaque plan a été testée.

Les huit procédures testées sont "Raschfit" (classement négatif des items), "Raschfit-fast" (classement positif des items), "Backrasch", "MSP (c=0.3)", "MSP (c=0.2)", "AFCS-Varimax", "ACP-Varimax" et "HCA/CCPROX".

Les valeurs des paramètres, et l'ensemble du plan de simulations a été repris comme "proche" de l'étude de van Abswoude, van den Ark et Sijtsma [144], avec des différences notoires (nombre d'items par dimensions, ajout d'une valeur de pouvoir discriminant très basse...).

7.4.1 Les paramètres de simulations

7.4.1.1 Paramètres fixes

Les données ont été simulées avec $N = 2000$ individus. Ce nombre d'individus important permet de supposer une certaine robustesse des méthodes [71]. En pratique, il est courant de voir de tels nombres d'individus dans des études en sciences de l'éducation (voir des tailles d'échantillons encore plus grandes), mais dans le domaine de la qualité de vie, les effectifs sont généralement plus faibles (200 à 500 individus en général).

Nous avons toujours simulé $Q = 2$ traits latents par une loi multinormale centrée réduite. Le coefficient de corrélation entre les deux traits latents est un paramètre non fixe dans les simulations. Cependant, dans certains cas, il est égal à 1.0, ce qui signifie alors explicitement que les données sont unidimensionnelles. Nous n'avons pas effectué de simulations avec $Q > 2$ car dans les précédentes études de simulations citées ([71], [132] ou [144]), il n'apparaît pas de différence importantes sur la nature des résultats quand les données sont simulées avec plus de deux dimensions. Bien que cela ne soit en rien une preuve, nous supposons donc que les résultats sont du même ordre pour des dimensions supérieures.

Pour les paramètres de difficulté des items, les quantiles $\frac{j}{J_q+1}$, $j = 1, \dots, J_q$ de la loi normale centrée réduite ont été utilisés dans chacune des deux dimensions $q = 1, 2$, où J_q représente le nombre d'items relié à une dimension q .

7.4.1.2 Modèles

Deux modèles ont été utilisés dans cette étude de simulations. Il s'agit des extrapolations multidimensionnelles du 2-PLM et du 5-PAM (voir section 3.5.1). Nous appellerons par la suite ces extrapolations M2-PLM et M5-PAM. Le M2-PLM s'écrit :

$$P(X_{nj} = x_{nj}/\theta_n; \delta_j, \alpha_j) = \frac{\exp \left[1.7x_{nj} \left(\sum_{q=1}^Q \alpha_{jq}\theta_{nq} - \delta_j \right) \right]}{1 + \exp \left[1.7 \left(\sum_{q=1}^Q \alpha_{jq}\theta_{nq} - \delta_j \right) \right]} \quad (7.20)$$

et le M5-PAM :

$$P(X_{nj} = x_{nj}/\theta_n; \delta_j^*, \alpha_j^*, \xi_j, \gamma_j^{low}, \gamma_j^{up}) = \gamma_j^{low} + \left(\gamma_j^{up} - \gamma_j^{low} \right) \left\{ \frac{\exp \left[1.7x_{nj} \left(\sum_{q=1}^Q \alpha_{jq}^* \theta_{nq} - \delta_j^* \right) \right]}{1 + \exp \left[1.7 \left(\sum_{q=1}^Q \alpha_{jq}^* \theta_{nq} - \delta_j^* \right) \right]} \right\}^{\xi_j} \quad (7.21)$$

Afin de faciliter les comparaisons entre les deux modèles, nous avons choisi d'utiliser dans le M5-PAM des paramètres α_{jq}^* et δ_j^* tels que, pour chaque jeu de simulation, les réponses de fonction aux items dans le M2-PLM et le M5-PAM aient les mêmes pentes maximales, et les mêmes localisations de ces pentes maximales¹.

¹van Abswoude, van der Ark et Sijtsma [144] ont utilisé aussi ce principe, mais il semblerait que les calculs qu'ils ont effectué soient faux, ce qui laisse planer un doute sur leur analyse du comportement des procédures pour le M5-PAM

On obtient :

$$\alpha_{jq}^* = \frac{\alpha_{jq}}{4 \left(\gamma_j^{up} - \gamma_j^{low} \right) \left(\frac{\xi_j}{1+\xi_j} \right)^{\xi_j+1}} \quad (7.22)$$

Et

$$\delta_j^* = \frac{\delta_j}{4 \left(\gamma_j^{up} - \gamma_j^{low} \right) \left(\frac{\xi_j}{1+\xi_j} \right)^{\xi_j+1}} - \log(\xi_j) \quad (7.23)$$

La démonstration de ces résultats est donnée en annexe A.

7.4.1.3 Structure des items

Des structures simples et des structures approximativement simples ont été simulées. Pour simuler une structure simple, on pose q_j le trait latent relié à l'item j et $\alpha_{jq} = 0$ si $q \neq q_j$. Dans le cadre d'une structure approximativement simple, les paramètres α_{jq} avec $q \neq q_j$ sont fixés à une faible valeur (0.2).

Dans le cas des simulations obtenues par le modèle M2-PLM, et une structure simple, en effectuant le changement de variable $\tilde{\theta}_{nq_j} = 1.7\alpha_{jq_j}\theta_{nq_j}$ et $\tilde{\delta}_j = 1.7\delta_j$, on obtient le modèle

$$P(X_{nj} = x_{nj} / \tilde{\theta}_n; \tilde{\delta}_j) = \frac{\exp \left[x_{nj} \left(\tilde{\theta}_{nq_j} - \tilde{\delta}_j \right) \right]}{1 + \exp \left[\tilde{\theta}_{nq_j} - \tilde{\delta}_j \right]} \quad (7.24)$$

Ce modèle est donc un MMSRM. Nous étudierons donc en détail ce cas (voir section 7.4.2.3), en particulier pour la procédure "Raschfit" basée sur ce modèle.

7.4.1.4 Choix des modèles et des structures

Le choix de ces modèles a été fait pour étudier le comportement des diverses procédures, en particulier des trois procédures proposées basées sur le MMSRM (BackRasch, Raschfit et Raschfit-Fast), quand les données étudiées vérifient exactement ce modèle (M2-PLM - Structure simple), ou bien quand il existe des écarts plus ou moins important entre le véritable modèle sous-jacent aux données et le MMSRM, que ce soit au sujet de la structure des items qui peut être moins favorable qu'une structure simple (M2-PLM - structure approximativement simple), des fonctions réponses aux items qui peuvent être moins régulières que dans le MMSRM (M5-PAM - structure simple) ou des deux possibilités (M5-PAM - structure approximativement simple).

7.4.1.5 Corrélations entre les traits latents

Le coefficient de corrélation entre les deux traits latents peut prendre six valeurs différentes : 0.0, 0.2, 0.4, 0.6, 0.8 et 1.0. Dans le dernier cas, les deux traits latents sont identiques : les données sont dans ce cas unidimensionnelles.

7.4.1.6 Nombre d'items par dimension

On retiendra deux valeurs possibles pour le nombre d'items par dimension : 7 et 14. Trois plans seront testés : [7; 7], [7; 14] et [14; 7]. 7 items représentent une échelle de dimension modeste. A titre de comparaison, la plus petite échelle du Sickness Impact Profil (sur 8) comprend 7 items (Sommeil et Repos) mais une dizaine en moyenne. 14 items représente une échelle d'assez grande taille. A titre de comparaison, une seule échelle du SIP dépasse ce nombre avec 20 items (Vie sociale).

7.4.1.7 Pouvoirs discriminants des items

On retiendra trois valeurs possible pour le pouvoir discriminant de chaque item par rapport au trait latent auquel cet item est principalement relié : une valeur basse "l" ($\alpha_{jq_j} = 0.4$), une valeur moyenne "m" ($\alpha_{jq_j} = 0.7$) et une valeur haute "h" ($\alpha_{jq_j} = 1.7$). On note que l'on a simulé tous les items reliés principalement au même trait latent avec la même valeur de pouvoir discriminant. Six plans sont possibles : [l; l], [l; m], [l; h], [m; m], [m; h] et [h; h]. On note que les plans avec [7; 14] items et [14; 7] items sont identiques dans les cas [l; l], [m; m] et [h; h], aussi, ils n'ont été simulés qu'une seule fois.

7.4.2 Résultats

7.4.2.1 Classification des résultats

La classification des résultats de simulation n'est pas forcément une opération aussi simple que cela. Nous distinguerons cinq cas :

- 1 : la classification obtenue des items est celle qui a été simulée
- 2 : la classification obtenue des items n'est pas celle qui avait été simulée, mais il n'existe que des erreurs mineures de classement
- 3 : la classification obtenue des items n'est pas celle qui avait été simulée, mais il n'existe pas d'erreurs majeures de classement
- 4 : la classification obtenue des items n'est pas celle qui avait été simulée, et il existe des erreurs majeures de classement
- 5 : le résultat de la classification n'est pas spécifique (il ne permet pas de conclure).

Nous définirons les erreurs majeures de classement de la manière suivante :

Définition 7.1 (Erreur majeure de classement) *Une erreur majeure de classement entre deux items provient du fait que les deux items sont reliés exclusivement (cas de structures simples) ou principalement (cas de structures approximativement simples) à deux traits latents distincts, mais qu'une procédure de classification d'items classe ces deux items dans une même sous-échelle. Dans le cas où une seule dimension a été simulée, une erreur majeure de classement signifie que les items sont classés par la procédure de classification en fonction d'un paramètre caractérisant chaque sous-ensemble (les présentes simulations ne permettront d'étudier ainsi que l'influence du pouvoir discriminant des items, mais on peut supposer d'autres caractéristiques comme les difficultés des items) [Une erreur majeure de classement dans le cadre unidimensionnel sont les classements qui aboutiraient à un classement de type 1 ou 2 s'il y avait deux dimensions].*

Nous définirons les erreurs mineures de classement de la manière suivante :

Définition 7.2 (Erreur mineure de classement) Une erreur mineure de classement sur une dimension simulée signifie qu'un petit nombre d'items de cette dimensions (que nous fixerons subjectivement à 2 items pour les dimensions où $J_q = 7$ et à 3 items pour $J_q = 14$) ne sont pas classés avec les autres items et forment à eux seuls une nouvelle dimension, ou bien ne sont pas sélectionnés par la procédure.

Nous définirons un résultat non spécifique pour une classification de la manière suivante :

Définition 7.3 (Résultat non spécifique d'une classification) Un résultat non spécifique pour une classification concerne les classifications pour lesquelles il existe un grand nombre d'items dans au moins une dimension (que nous fixerons subjectivement à plus de 2 items pour les dimensions où $J_q = 7$ et à plus de 3 items pour $J_q = 14$) qui ne sont pas sélectionnés par la procédure et dans lesquelles il n'existe, pour les items sélectionnés, aucune erreur majeure de classement [On notera que, dans un cadre unidimensionnel, lorsque deux items provenant de chacun des deux sous-ensembles seront classés ensemble, les erreurs majeures de classement sont impossible et le résultat sera classé de type 3, même si peu d'items ont été sélectionnés (un tel résultat est présenté à la troisième ligne des classements de type 3 dans le tableau 7.2)].

Dans le tableau 7.1, sont présentés des résultats type pour des simulations avec deux ensembles de 7 items ($J_q = 7, \forall q = 1, 2$) dépendants de deux traits latents distincts.

Résultat	Premier ensemble d'items	Second ensemble d'items
1	1 1 1 1 1 1 1	2 2 2 2 2 2 2
	2 2 2 2 2 2 2	1 1 1 1 1 1 1
2	1 1 1 1 0 1 1	2 2 2 2 2 2 2
	2 2 2 2 3 3 2	1 1 1 1 1 1 1
	2 2 2 3 3 2 2	1 1 1 0 0 1 1
3	2 2 2 3 3 3 3	1 1 1 1 1 1 1
	1 1 1 2 2 2 1	3 3 3 4 4 4 4
4	1 1 1 2 2 2 2	1 1 1 2 2 2 2
	1 1 1 2 2 3 3	1 1 1 2 2 2 1
	1 1 1 1 1 1 1	1 1 1 1 1 1 1
	0 0 0 1 1 0 0	1 1 0 0 0 0 0
5	0 0 0 0 0 0 0	0 0 0 0 0 0 0
	0 0 0 0 0 0 0	1 1 1 1 1 1 1
	0 0 0 0 2 2 0	1 1 1 1 1 1 1
	0 0 0 0 1 1 1	0 0 2 2 0 0 0

TAB. 7.1 – Exemple de résultats types, avec deux ensembles de 7 items simulés reliés à deux traits latents distincts

Dans le tableau 7.2, sont présentés des résultats types pour des simulations avec deux ensembles de 7 items ($J_q = 7, \forall q = 1, 2$) dépendants d'un même trait latent (cas où $\rho = 1.0$).

Résultat	Premier ensemble d'items	Second ensemble d'items
1	1 1 1 1 1 1 1	1 1 1 1 1 1 1
2	1 1 1 1 0 1 1 1 1 1 1 1 1 1	1 0 0 1 1 1 1 1 1 1 0 0 1 1
3	2 2 2 1 1 1 1 1 1 1 2 2 2 1 0 0 0 0 1 1 0	2 2 1 2 1 1 1 1 1 2 2 3 3 3 1 1 1 1 1 1 1
4	1 1 1 1 1 1 1 1 1 1 2 2 2 1 1 1 1 1 1 1 1	2 2 2 2 2 2 2 3 3 3 4 4 4 4 2 2 2 3 3 3 3
5	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 2 2 2	0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1

TAB. 7.2 – Exemple de résultats type, avec deux ensembles de 7 items simulés reliés au même trait latent

7.4.2.2 Présentation des résultats

Les résultats détaillés par procédure sont détaillées dans les annexes B (voir page 169). Dans la présente section, nous présenterons les résultats agrégés en fonction du modèle, de la structure, de la corrélation entre les deux traits latents, et des pouvoirs discriminants dans chacune des dimensions utilisés dans les simulations. Le nombre d'items dans chaque dimension n'est pas retenu car il n'apparaît pas comme un facteur important.

Au vue des résultats, concernant le modèle (voir section 7.4.1.2) et la structure (voir section 7.4.1.3), nous retiendrons trois cas différents : le M2-PLM avec une structure simple (MMSRM), le M2-PLM avec une structure approximativement simple (perturbation du MMSRM) et le M5-PAM quelque soit la structure utilisée (perturbation importante du MMSRM).

Pour la corrélation entre les deux traits latents, nous retiendrons trois cas : une corrélation faible ($\rho \leq 0.4$), une corrélation forte ($0.6 \leq \rho \leq 0.8$) et le cas unidimensionnel ($\rho = 1.0$).

Enfin, pour les pouvoirs discriminants des items, nous retiendrons deux cas : au moins une des dimensions est simulées avec un pouvoir discriminant bas ($\alpha_{jq_j} = 0.4$), et l'autre cas.

7.4.2.3 M2-PLM - Structure simple : MMSRM

Dans ce premier cas (voir tableau 7.3) le plus favorable aux procédures basées sur le modèle de Rasch (Raschfit, Raschfit-fast, et Backrasch) puisque les données sont simulées par un tel modèle, on note en effet de très bon résultats pour celles-ci, en particulier pour Raschfit et Raschfit-Fast (pour les cas bi-dimensionnels). Backrasch produit cependant beaucoup d'erreurs (type 4) mais semble plus robuste lorsque les traits latents sont fortement corrélés. Les procédures HCA/CCPROX et AFCS-Varimax donnent elles-aussi de très bons résultats, en particulier pour des faibles corrélations des traits latents ($\rho \leq 0.4$). MSP ($c=0.3$) donne de

Type 1 ou 2/Type 4 (Type 5)	Corrélation entre les deux trait latents		
	$\rho \leq 0.4$	$0.6 \leq \rho \leq 0.8$	$\rho = 1.0$
Nombre de simulations	45	30	15
Raschfit	84%/11%	67%/30%	27%/47%
Raschfit-Fast	73%/7%	43%/40%	47%/33%
BackRasch	51%/44% (2%)	53%/47% (0%)	33%/53% (0%)
MSP (c=0.3)	44%/0% (53%)	10%/50% (37%)	53%/0% (47%)
MSP (c=0.2)	36%/13% (31%)	3%/70% (13%)	73%/0% (7%)
HCA/CCPROX	91%/7%	30%/60%	40%/0%
AFCS-Varimax	78%/0%	30%/40%	20%/0%
ACP-Varimax	31%/7%	23%/37%	13%/0%

TAB. 7.3 – Résultats agrégés des simulations avec le M2-PLM et une structure simple (Taux de résultats de type 1 ou 2/Taux de résultats de type 4 (Taux de résultats de type 5))

très bon résultats quand celui-ci est déterminé, mais il est souvent non déterminé (dès qu'une dimension a été simulée avec une valeur basse du pouvoir discriminant), mais elle perd ses qualités pour des hautes valeurs de corrélations entre les traits latents. MSP (c=0.2) semble plutôt moins bonne (moins de résultats de type 1 ou 2/plus de résultats de type 4), en particulier quand les traits latents sont fortement corrélés. L'ACP-Varimax donne des résultats très moyens.

Dans des cas unidimensionnels ($\rho = 1.0$), les procédures basées sur des modèles semblent les moins adaptées puisque ce sont elles qui ont les plus forts taux d'erreurs. MSP semble la procédure la plus à même de détecter une structure unidimensionnelle (avec cependant encore beaucoup de cas indéterminés, notamment quand une dimension est simulée avec un faible pouvoir discriminant). Les autres procédures détectent relativement peu souvent la réelle structure des items, mais produisent peu d'erreur (type 4).

7.4.2.4 M2-PLM - Structure approximativement simple

Lorsque le modèle utilisé dans les simulations vérifie une structure approximativement simple, les résultats de toutes les procédures se dégradent (voir tableau 7.4). Ceci est notamment vrai pour les procédures Raschfit-fast et Backrasch. HCA/CCPROX semble un peu plus robuste que les autres procédures à ces perturbations. MSP (c=0.3 ou c=0.2) reste encore très correctes quand le résultat est déterminé (peu souvent), et tant que la corrélation entre les deux traits latents reste faible. Raschfit continue d'avoir des résultats corrects, bien que l'on note beaucoup plus d'erreurs de classements (type 4) qu'avec des simulations par structure simple. Les méthodes basées sur l'analyse factorielle présentent peu de résultats corrects (type 1 ou 2), mais aussi peu de résultats incorrects (type 4).

Dans un cadre unidimensionnel ($\rho = 1.0$), les résultats sont globalement les mêmes qu'avec une structure simple, mais on note que pour les procédures basées sur des modèles (Raschfit, Raschfit-Fast et Backrasch), les résultats semblent meilleurs (moins de résultats de type 4).

Type 1 ou 2/Type 4 (Type 5)	Corrélation entre les deux trait latents		
	$\rho \leq 0.4$	$0.6 \leq \rho \leq 0.8$	$\rho = 1.0$
Nombre de simulations	45	30	15
Raschfit	60%/29%	40%/60%	27%/13%
Raschfit-Fast	18%/38%	10%/87%	27%/7%
BackRasch	2%/67% (18%)	0%/87% (10%)	7%/0% (27%)
MSP (c=0.3)	13%/0% (87%)	7%/7% (87%)	13%/0% (53%)
MSP (c=0.2)	42%/4% (49%)	7%/40% (50%)	53%/0% (13%)
HCA/CCPROX	80%/16%	30%/70%	47%/0%
AFCS-Varimax	47%/7%	10%/53%	0%/0%
ACP-Varimax	9%/20%	10%/53%	7%/7%

TAB. 7.4 – Résultats agrégés des simulations avec le M2-PLM et une structure approximativement simple (Taux de résultats de type 1 ou 2/Taux de résultats de type 4 (Taux de résultats de type 5))

7.4.2.5 M5-PAM

Type 1 ou 2/Type 4 (Type 5)	Corrélation entre les deux trait latents		
	$\rho \leq 0.4$	$0.6 \leq \rho \leq 0.8$	$\rho = 1.0$
Nombre de simulations	90	60	30
Raschfit	54%/39%	38%/58%	43%/20%
Raschfit-Fast	26%/56%	17%/75%	57%/7%
BackRasch	23%/69% (7%)	28%/62% (8%)	20%/40% (3%)
MSP (c=0.3)	18%/30% (49%)	0%/75% (23%)	57%/0% (13%)
MSP (c=0.2)	18%/61% (18%)	0%/95% (5%)	83%/0% (0%)
HCA/CCPROX	33%/64%	13%/87%	33%/0%
AFCS-Varimax	43%/44%	5%/85%	10%/0%
ACP-Varimax	24%/37%	7%/78%	17%/0%

TAB. 7.5 – Résultats agrégés des simulations avec le M5-PAM (Taux de résultats de type 1 ou 2/Taux de résultats de type 4 (Taux de résultats de type 5))

Avec des simulations réalisées par M5-PAM (voir tableau 7.5), les résultats sont sensiblement moins bons qu'avec les simulations réalisées par M2-PLM. Les erreurs (type 4) sont beaucoup plus importantes dans toutes les procédures, tandis que les bons résultats (type 1 ou 2) sont moins nombreux (à l'exception de ACP-Varimax). Cette dégradation est surtout sensible pour des corrélations faibles entre les traits latents (pour les corrélations fortes, les résultats étaient déjà médiocres avec le M2-PLM).

On peut noter en particulier les mauvais résultats dans ce cas des procédure non paramétriques (MSP, HCA/CCPROX) assez surprenantes, car les algorithmes sont moins liés à la forme des ICC que pour les procédures basées sur des modèles (Raschfit, Raschfit-Fast, Backrasch). Il est d'ailleurs étonnant de voir que ces dernières procédures produisent un taux de résultat corrects non négligeable, en particulier pour Raschfit).

7.4.2.6 Temps d'exécution

Le temps d'exécution est très variable d'une procédure à l'autre [67]. Nous avons retenu, pour chaque procédure, les temps d'exécution moyen, minimum et maximum et l'écart-type du temps d'exécution sur les procédures (pour Raschfit, Raschfit-Fast, HCA/CCPROX, il s'agit des moyennes sur l'ensemble des simulations, pour les autres procédures, il s'agit d'estimations basées sur 12 simulations tirées aléatoirement. Pour Raschfit et Raschfit-Fast, le temps a été compté en minutes puis rapporté en secondes). Nous avons distingué les cas avec 7 items dans chaque dimensions, et les cas avec 7 items dans une dimension et 14 dans l'autre. Les résultats sont présentés dans le tableau 7.6. Toutes les simulations ont été effectuées sur deux ordinateurs comparables en terme de vitesse de microprocesseurs (950 MHz ou 1 GHz) qui se différenciaient seulement sur la taille de la mémoire RAM (512 Mo et 256 Mo). Des essais ont montré que le type de processeur avait une grande importance (réduction de près de 40% du temps de calcul sur un ordinateur à 2GHz pour la procédure Raschfit) mais que la RAM influençait peu le temps de calcul (moins de 5% de différences de temps entre le 950MHz/512Mo et le 1GHz/256Mo).

Temps de calcul	[7;7]			[7;14] ou [14;7]		
	moyenne	écart-type	min./max.	moyenne	écart-type	min./max.
Raschfit	18190	3384	13740/39540	38829	9291	23460/64800
Raschfit-Fast	320	110	120/720	1216	530	240/2640
BackRasch	340	87	207/468	2037	213	1890/2410
MSP (c=0.3)	20	5	5/187	89	18	8/382
MSP (c=0.2)	24	6	6/237	107	22	10/467
HCA/CCPROX	89	5	79/123	353	81	287/905
AFCS-Varimax	1	1	1/3	2	1	1/4
ACP-Varimax	1	1	1/3	2	1	1/4

TAB. 7.6 – Temps de calculs moyens, minimum et maximum en secondes et écart-type pour chaque procédure sur les simulations en fonction du nombre d'items dans chaque dimension

La procédure "Raschfit" est de loin celle qui nécessite le plus de ressources : en moyenne, 5 heures pour 14 items et près de 11 heures pour 21 items, loin devant Raschfit-Fast et BackRasch (respectivement 5 minutes et 33 minutes en moyenne). Les procédures de l'IRT non paramétriques, bien que non instantannées, nécessitent beaucoup moins de ressources : 20 à 90 secondes en moyenne pour 14 items, 1 minute 30 seconde à 6 minutes en moyenne avec 21 items). Quand aux méthodes issues de l'analyse factorielles, elles sont quasi-instantannées.

7.4.2.7 Concordance entre les méthodes

Il est possible de calculer les kappa de concordance [23] entre les différentes méthodes prises deux à deux. On retient trois "jugements" possible : "correct" (type 1 ou 2), "moyen" (type 3) ou "mauvais" (type 4). Les résultats de type 5 sont supprimés, car ils ne concernent pas toutes les méthodes.

Ces concordances n'ont pas d'intérêts pour "estimer" la valeur des méthodes, mais permet de comparer les cohérences des résultats entre les différentes méthodes et observer par exemple

s'il s'agit toujours des mêmes simulations qui obtiennent le même type de classifications entre les différentes méthodes.

On obtient la matrice de concordance définie dans le tableau 7.7. Aucune pondération particulière n'ont été données à des "jugements" différents entre deux méthodes.

	Raschfit	Raschfit Fast	Backrasch	MSP (c=0.3)	MSP (c=0.2)	HCA CCPROX	AFCS Varimax
Raschfit-Fast	0.38						
BackRasch	0.30	0.34					
MSP (c=0.3)	0.27	0.38	0.18				
MSP (c=0.2)	0.18	0.27	0.08	0.64			
HCA/CCPROX	0.29	0.26	0.25	0.54	0.45		
AFCS-Varimax	0.29	0.28	0.17	0.46	0.35	0.48	
ACP-Varimax	0.24	0.32	0.18	0.29	0.20	0.28	0.49

TAB. 7.7 – Matrice des coefficients kappa de concordance entre les méthodes

Les concordances sont relativement faibles entre les méthodes : le kappa de concordance maximal vaut "seulement" 0.64, et en moyenne, les kappa de concordance valent 0.32. Les méthodes les plus concordantes sont MSP (c=0.3) avec MSP (c=0.2). On note aussi une forte concordance entre les deux méthodes issues de l'analyse factorielle. L'analyse des concordances rappelle les trois groupes de méthodes : celles basées sur l'adéquation à un modèle (Raschfit, Raschfit-Fast et BackRasch), celles basées sur les propriétés de l'IRT (MSP et HCA/CCPROX) et celles basées sur l'analyse factorielle (AFCS et ACP). On note que Raschfit-Fast est assez concordante avec MSP, et que HCA/CCPROX est concordante avec l'AFCS. En revanche, BackRasch semble la méthode la moins concordante avec les autres, en particulier avec MSP et les analyses factorielles.

7.4.2.8 Synthèse et adaptations

Nous pouvons donc conclure sur les différentes procédures de la façon suivante :

- **Raschfit** semble très bien adaptée dès lors que les données suivent un MMSRM (on peut penser que cela serait vrai avec d'autres modèles). Les résultats sont alors bons, même avec de fortes corrélations entre les traits latents. Dans un cadre unidimensionnel, cette procédure donne de bons résultats dès lors que les pouvoirs discriminants des items sont similaires (sinon, elle a tendance à classer en plusieurs échelles les groupes d'items ayant des pouvoirs discriminants différents, ce qui est logique avec les hypothèses du modèle de Rasch). Quand les données ne suivent plus un MMSRM, les résultats de cette procédure semblent très corrects comparés aux autres procédures. En revanche, cette procédure est très longue à tourner (5 heures avec deux fois 7 items et 10 heures avec 7 et 14 items en moyenne),
- **Raschfit-Fast** donne aussi de très bons résultats quand les données sont simulées avec un MMSRM. Elle semble moins robuste que Raschfit à de fortes corrélations entre les traits latents, ou à des écarts au modèle utilisé. Cette procédure permet cependant un

gain de temps considérable sur Raschfit (60 fois plus rapide avec deux fois 7 items et 30 fois plus rapide avec 7 et 14 items en moyenne),

- **Backrasch** donne des résultats beaucoup plus médiocres, même lorsque les données sont simulées par un MMSRM. En revanche, cette procédure semble plus à même de détecter un jeu de données unidimensionnel que Raschfit. La procédure prend environ autant de temps (voir plus) que Raschfit-Fast, sur laquelle elle n'a aucun avantage,
- **MSP** donne de bons résultats dans des cas très favorables (corrélations faibles entre les traits latents, forts pouvoirs discriminants des items), et ce quelque soit le modèle utilisé dans les simulations (même si l'on note une dégradation des résultats lorsque les modèles utilisés sont moins réguliers que le MMSRM). En revanche, cette procédure a tendance à fournir un grand nombre de résultats indéterminés, notamment dès que le pouvoir discriminant des items est faibles. En baissant un peu le seuil c (à 0.2), les résultats restent corrects, mais les résultats qui deviennent déterminés ont tendance à donner des mauvais résultats. C'est une procédure en revanche très rapide, ce qui lui donne un certain avantage de ce point de vue,
- **HCA/CCPROX** donne de très bons résultats lorsque les données sont simulées par MMSRM, mais aussi des résultats corrects quand le modèle utilisé est moins régulier. Les résultats sont très bons lorsque les pouvoirs discriminants des items sont forts. Cette procédure semble la plus robuste aux écarts aux MMSRM, mais aussi à de fortes corrélations des traits latents. C'est une procédure rapide (un peu plus lente que MSP mais environ 4 fois plus rapide que Raschfit-Fast ou BackRasch, ce qui en fait une procédure très intéressante,
- **AFCS-Varimax** donne des résultats corrects sur l'ensemble des simulations mais a tendance à dissocier en plusieurs échelles des items issus d'une même dimension en fonction de leur difficulté (les items faciles formant une dimensions, et les items difficiles une autre, soit un résultat de type 3). Les résultats de cette procédure sont quasi instantanés,
- **ACP-Varimax** donne des résultats décevant, mais peu souvent mauvais (beaucoup de type 3, comme pour AFCS-Varimax). Les résultats de cette procédure sont quasi instantannés.

Les conclusions que nous rapportons sur MSP, HCA/CCPROX, AFCS-Varimax et ACP-Varimax sont conformes à celles retrouvées dans la littérature ([144], [71], et [132] par exemple pour MSP et HCA-CCPROX ou [21] et [17] pour les analyses factorielles).

Deux adaptations nous semble en outre importante : la possibilité d'étendre les procédures facilement à des items polytomiques ordinaux, et la possibilité de choisir un noyau d'items pour construire les échelles. Le tableau 7.8 résume les caractéristiques de chaque procédure.

7.5 Applications

7.5.1 Application sur le Profil d'impact des maladies (Sickness Impact Profil) (SIP)

Nous avons utilisé les mêmes méthodes que celles des simulations, en ajoutant Raschfit-Fast avec un classement négatif des items, Varclus et CLV, soit un total de 11 procédures, sur les trois sous-échelles de la version française du SIP décrites à la partie 4.11. Cette répartition des méthodes équilibre les "groupes" de méthodes avec 4 méthodes basées sur l'adéquation

	Sélection basée sur	Propriétés de l'IRT	Adéquation au modèle de Rasch	Items polytomiques	Noyau initial
Raschfit	adéquation globale au modèle de Rasch	X	X	X ²	X
Raschfit-Fast	adéquation globale au modèle de Rasch	X	X	X ²	X
BackRasch	adéquation locale au modèle de Rasch	X	X	X ³	X ⁴
MSP	Monotonicité	X		X ⁵	X ⁵
HCA/CCPROX	Indépendance locale	X		X ⁶	X ⁷
AFCS-Varimax	Analyse factorielle			X ⁸	
ACP-Varimax	Analyse factorielle			X ⁸	

TAB. 7.8 – Caractéristiques de chacune des méthodes de sélection d'items

² voir section 7.3.5

³ en utilisant un Partial Credit Model

⁴ en interdisant le retrait de certains items lors de la procédure

⁵ prévu par les auteurs [71]

⁶ en adaptant les mesures de proximité aux items polytomiques

⁷ en faisant en sorte que ce noyau forme un cluster initial

⁸ en utilisant les corrélations polychoriques

des données à un modèle de Rasch, 3 basée sur les propriétés de l'IRT, et 4 basées sur une vision algébrique de l'unidimensionnalité (trouver des ensembles d'items tels que la première valeur propre dans chaque groupe soit la plus grande possible).

Toutes les procédures donnent des résultats sensiblement différents. Afin de trouver une classification synthétique, nous avons utilisé, pour chaque paire d'items, le nombre de classements dans un même groupe comme mesure de similarité.

Une classification hiérarchique ascendante par la méthode du "lien complet" (complete linkage) permet d'obtenir un dendrogramme représentant à chaque agrégation le nombre minimale de classements dans un même groupe parmi toutes les paires d'items de ce nouveau groupe. Un tel dendrogramme est représenté à la figure 7.4.

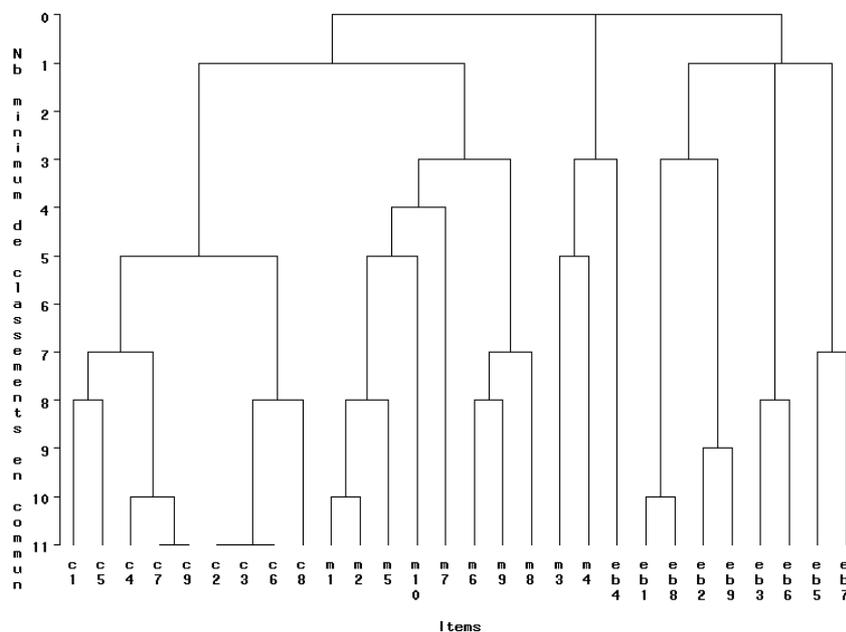


FIG. 7.4 – Dendrogramme issue de la classification des items en fonction du nombre de classements dans un même groupe par les 11 procédures

On scinde le dendrogramme en fonction du nombre minimal de classement dans un même groupe que l'on souhaite pour chaque paire d'items de chaque groupe. Par exemple, si l'on souhaite que dans chaque groupe, chaque paire d'items soit classée dans un même groupe par au moins 3 procédures, on obtient 6 groupes d'items décrit dans le tableau 7.9.

Dans cette classification "synthétique", tous les items de la sous-échelle "Communication" sont classés ensemble (et avec aucun autre item). On trouve cependant deux grands groupes d'items. D'une part les items c2, c3, c6 et c8, qui parlent plutôt des problèmes de compréhension des autres, et d'autres part des items c1, c4, c5, c7 et c9 qui parlent soit de la communication écrite (items c1 et c5) ou plutôt des difficultés à parler (items c4, c7 et c9).

Sous-échelles	Liste des items	Nombre de classements minimum en commun
1	c1-c9	5
2	m1, m2, m5-m10	3
3	m3, m4, eb4	3
4	eb1, eb2, eb8, eb9	3
5	eb3, eb6	8
6	eb5, eb7	7

TAB. 7.9 – Partition synthétique des 28 items des sous-échelles "communication", "mobilité" et "comportement émotionnelle"

Le deuxième groupe est constitué de tous les items de la sous-échelle "mobilité" sauf m3 et m4. On retient dans ce regroupement 4 grands groupes : les items m6, m8 et m9 qui parlent de la difficulté à quitter le domicile, les items m1, m2 et m5 qui envisage plutôt des difficultés lorsque la personne se trouve à l'extérieur de son domicile, l'item m7, plus isolé, qui parle de la nécessité d'avoir des toilettes à proximité, et l'item m10 qui parle des difficultés à se mouvoir dans l'obscurité.

Le troisième groupe est constitué des items m3, m4 et eb4. m3 et m4 parlent clairement des difficultés à quitter le lit (soit en se référant à une période antérieure non précisée pour m3, soit en considérant la part subjective de cette "activité" pour la personne. L'item eb4 parle plutôt de tendance suicidaire. Ce regroupement pourrait donc être une base à la mesure de troubles de type dépressif : perte de l'entrain aux activités habituelles, tendance suicidaire.

Le groupe suivant est constitué des items eb1, eb2, eb8 et eb9 avec deux regroupements très nets : d'une part eb1 et eb8 qui se situent eux aussi dans la mesure de troubles de type dépressif [perte de l'estime de soi (eb1), perte des attentes dans l'avenir (eb8)] avec de forts taux de réponses positives dans cette population de déprimés (respectivement 65% et 77% et 86% des personnes qui ont répondu positivement à au moins un de ces deux items), et d'autre part eb2 et eb9 qui envisagent plutôt des sautes d'humeurs.

Le cinquième groupe est composé des items eb3 et eb6 qui font tous les deux référence à des douleurs somatiques.

Le dernier groupe rassemble les items eb5 et eb7 qui parlent plutôt de troubles anxieux.

Notons que l'ensemble des procédures a des différences sensibles entre leurs résultats et la partition "type" proposée. Les procédures MSP ou HCA sont celles qui donnent des partitions des items les plus proches de celle-ci. Raschfit et Backrasch ont plutôt eu tendance à regrouper les échelles "communication" et "mobilité" (échelles fortement corrélées à 0,60) aux exceptions notables de m3, m4 et dans une moindre mesure m9. Les méthodes issues de l'analyse factorielle (ACP, AFCS et Varclus) ont plutôt eu tendance à scinder les échelles initiales en plusieurs sous-échelles.

CLV est une méthode à mettre à part, car elle trouve une partition sensiblement différente des autres, en particulier en ce qui concerne l'échelle "Comportement émotionnel", où cette méthode a regroupé ensemble 8 des 9 items composant cette échelle, alors que les 10 autres méthodes ont toutes scindé cette échelle en au moins 4 ensembles. CLV, en revanche trouve une structure des items très cohérente de la structure théorique du questionnaire en trois échelle, puisque, à l'exception de eb4 classé avec l'échelle mobilité, la structure est exactement retrouvée. Cela nous permet de supposer que les trois échelles du SIP sont bien unidimensionnelles et que les regroupements effectués par ce profil de qualité de vie permet de maximiser un critère proche de celui utilisé par CLV (à savoir la maximisation de la somme des premières valeurs propres issues des matrices de données de chaque groupe). Néanmoins, en tenant compte d'autres propriétés recherchées (monotonie, indépendance locale, exhaustivité du score sur le trait latent...), cette structure ne semble pas la plus adéquate.

Notons que tous les items ont une corrélation avec le score calculé avec les autres items du groupe auquel les items sont agrégés dans la partition synthétique (rest-scores) plus importante qu'avec tous les scores calculés dans les autres groupes. Néanmoins, on note de faibles corrélations ($<0,20$) entre les items et les rest-scores dans leur groupe pour les items m7 et eb4, ou de fortes corrélations ($>0,25$) entre certains items et d'autres échelles [items c1 (avec le groupe 2), c5 (avec le groupe 2), c6 (avec le groupe 2), m4 (avec le groupe 2), m8 (avec le groupe 1), eb1 (avec le groupe 1)].

En analysant les deux premières échelles du tableau 7.9 par un modèle de Rasch, on trouve, pour la première échelle, une mauvaise adéquation des items c1 et c5, et pour la seconde de m7. Ces trois items sont isolés des autres items de leur échelle par plusieurs procédures (voir dendrogramme 7.4).

D'un point de vue psychométrique, on peut s'interroger sur les différences entre la structure initiale du SIP, définie par des méthodes d'analyse factorielle et validée cliniquement, et la structure proposée dans cette analyse : il semblerait que la structure initiale, du moins pour la sous-échelle "comportement émotionnel", ne permette pas d'obtenir des scores avec de bonnes propriétés psychométriques, et que cette échelle soit plutôt un agglomérat de plusieurs concepts différents plus ou moins corrélés.

7.5.2 Application dans un autre champ : prédire le risque de récurrence de la tentative suicidaire par un score simple

7.5.2.1 L'étude

La méthode Raschfit a été appliquée dans un champ différent de celui de la qualité de vie, dans le but de trouver un score simple pour prédire le risque de récurrence de tentative de suicide, suite à un passage aux urgences pour cette raison. Ce travail s'est appuyé sur une étude réalisée par le service de psychiatrie du Centre Hospitalier de Dreux en 2000 auprès de 259 patients. Durant cette étude, les psychiatres ont relevé durant un an toutes les caractéristiques des patients reçus aux urgences pour tentative de suicide (caractéristiques sociales, caractéristiques de la tentative de suicide, antécédents psychiatriques personnels et familiaux, diagnostic psychiatrique). Un des buts de l'étude était de réussir à détecter les patients à

risque de récidives sans diagnostic psychiatrique (donc en permettant à des non psychiatres comme des psychologues de les détecter lors d'un entretien préalable). Deux méthodes ont été utilisées dans ce but : l'analyse discriminante, et la procédure Raschfit. L'usage de la procédure Raschfit apportait ici un avantage indéniable comparé à l'analyse discriminante : celui de permettre la création d'un score simple (non pondéré) pour mesurer le trait latent (dans ce cas la propension d'un patient à récidiver suite à une tentative de suicide). Cette étude a été présentée en [59] et [119] et publiée en [120].

Une revue de la littérature basée principalement sur les travaux de Chastang [19] a permis de déterminer les facteurs de risque de la récurrence suicidaire :

- l'existence d'antécédent de tentative de suicide,
- les antécédents traumatiques (deuils violents ou brutaux, violences dans l'enfance),
- les antécédents de suicides familiaux,
- la dépression, l'anxiété et les psychoses.

La présence de ces facteurs de risque chez chacun des patients a permis le classement des patients entre les personnes à risque de récurrence, et les personnes à moindre risque de récurrence. Notamment, toutes les personnes qui avaient un antécédent de tentative de suicide de moins d'un an ont été systématiquement classées à risque puisque ce facteur de risque est l'un des principaux facteurs de risque connus de la récurrence suicidaire [5]. Ainsi, 117 des 259 patients (45%) ont été classés à risque.

Nous avons donc cherché un score qui permette de retrouver au mieux ces patients. Deux méthodes ont été utilisées pour ce faire : l'analyse discriminante, et la méthode Raschfit.

7.5.2.2 Analyse discriminante

L'analyse discriminante permet de construire un score linéaire à partir des variables retenues. L'avantage de cette méthode est d'être particulièrement bien adaptée à l'objet de l'étude (expliquer une partition des individus en deux groupes par un ensemble de variables). Pour ce faire, nous avons dichotomisé toutes les variables, en regroupant, suite à une analyse descriptive, les modalités les plus favorables à un risque de récurrence d'une part, et les autres d'autre part. Nous avons cherché à minimiser la fonction de coût suivante :

$$cout(modele) = 4p + 2n_1 + n_0 \quad (7.25)$$

avec p le nombre de variables retenues, n_1 le nombre de patients à risque classés à moindre risque et n_0 le nombre de patients à moindre risque classés à risque.

Ce type d'analyse permet d'obtenir un score basées sur les variables suivantes (entre crochets, les coefficients linéaires associés aux modalités concernées) :

- situation professionnelle [non retraité : 0.4],
- antécédents de tentatives de suicide [oui, de moins d'un an : 2.6],
- antécédents traumatiques [deuils brutaux ou violences dans l'enfance : 2.3]
- antécédents traumatiques [deuils récents : 2.0]
- antécédents familiaux [suicides familiaux : 1.2]
- antécédents familiaux [psychiatriques : 0.6]

– constante : -2.1

Un score positif classera le sujet comme à risque de récurrence, un score négatif le classera comme à moindre risque.

7.5.2.3 Raschfit

La procédure Raschfit a été réalisée sur l'ensemble des items dichotomisés (les réponses "positives" seront choisies pour être les plus favorables à un classement à risque). Cette procédure permet de construire deux groupes de variables (voir tableau 7.10) qui mesurent chacune un trait latent par un score simple. Six items ne sont pas retenus car la procédure "Raschfit" les classe séparément.

	Variable	Modalités classées comme "réponses positives"
Score 1	1. Situation matrimoniale	Divorcé ou veuf
	2. Maladie non psychiatrique	Oui
	3. Addictions	Oui
	4. Antécédents de TS	Moins d'un an
	5. Deuil récent	Oui
	6. Violence subie dans l'enfance	Oui
	7. Suicide familial	Oui
	8. Antécédents psychiatriques familiaux	Oui
	9. Estime de soi	Altérée ou très altérée
Score 2	10. Sexe	Femmes
	11. Situation professionnelle	Actif, invalides ou scolarisés
	12. Deuil brutal	Oui
	13. Insertion	Stable
Variable non retenues	14. Classe d'âge	Moins de 55 ans
	15. Drogues	Oui
	16. Fugues	Oui
	17. Entourage	Famille non étayante ou réseau social
	18. Délinquance	Oui
	19. Antécédents psychiatriques	Oui

TAB. 7.10 – Variables regroupées avec la procédure Raschfit

La figure 4.1 représente le MMSRM obtenu où θ^* représente un trait latent multidimensionnel non précisé.

La figure 7.6 représente le MMSRM obtenu avec seulement les deux principaux traits latents, en introduisant les scores exhaustifs de chacun de ces deux traits latents et la variable "risque de récurrence", ainsi que les corrélations entre cette variable et les deux scores.

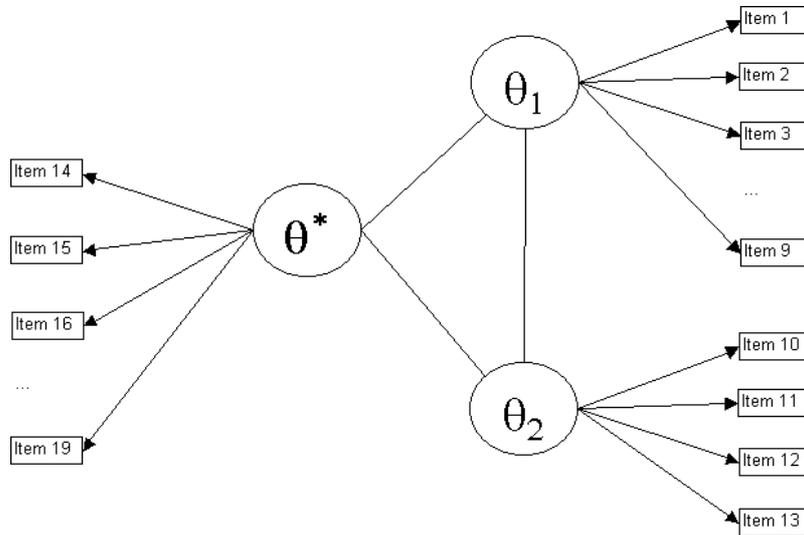


FIG. 7.5 – Représentation graphique du MMSRM obtenu avec les 19 items

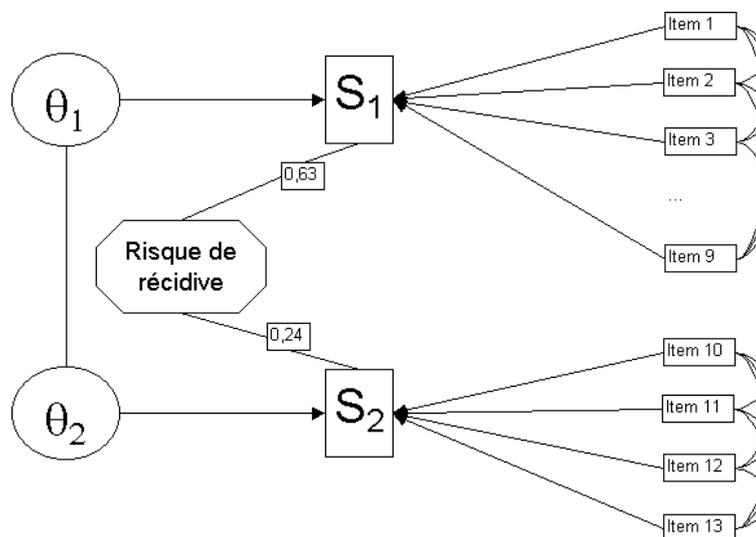


FIG. 7.6 – Représentation graphique du MMSRM avec les deux principaux traits latents et la variable "risque de récidive"

Avec ces deux scores (le premier en 9 points, le second en 4 points), nous avons tenté de retrouver nos deux groupes de patients (à risque de récurrence/à moindre risque de récurrence) en utilisant tous les seuils possibles, pour chacun des deux scores et pour la combinaison croisée des deux scores, en minimisant la même fonction de coût que pour l'analyse discriminante. Le premier score est celui qui permet de mieux retrouver le groupe des individus. C'est aussi celui qui est le plus corrélé à la variable de risque de récurrence (Coefficient de corrélation de 0.75). Nous avons alors "réduit" cette échelle, en supprimant une à une la variable dont la suppression permettait de diminuer la fonction de coût. L'algorithme a été stoppé dès que la fonction de coût ne pouvait plus être réduite.

Finalement, seuls quatre critères ont ainsi été gardés pour composer un score :

- antécédent de tentative de suicide de moins d'un an,
- décès récent,
- violence subie dans l'enfance,
- antécédents psychiatriques familiaux.

Le nouveau score sous-jacent a une corrélation de 0.75 avec la variable "risque de récurrence". La règle de décision permettant de retrouver au mieux la catégorie de patient est la suivante : "si le patient vérifie au moins un des quatre critères (si son score est supérieur ou égal à 1), alors il est classé à risque de récurrence."

7.5.2.4 Validation des scores

Les deux scores (le premier issu de l'analyse discriminante, le second de la procédure Raschfit) ont été testés sur trois échantillons :

- sur l'échantillon de travail (permettant d'évaluer la sensibilité et la spécificité),
- sur l'échantillon indépendant des tentatives de suicides ayant été suivies dans le reste de l'année d'une récurrence (ces observations n'avaient pas été utilisées dans les calculs) (permettant d'évaluer la sensibilité),
- sur le sous-échantillon de l'échantillon de travail des tentatives de suicides des trois premiers mois de l'année n'ayant pas été suivies dans le reste de l'année d'une récurrence (permettant d'évaluer la spécificité).

Les deux derniers échantillons sont appelés "échantillons témoins". Les deux scores évalués avec les outils utilisés classiquement en épidémiologie :

- Taux de classement correct (TCC) des patients
- Sensibilité (Se) : taux de classement correct des patients à risque
- Spécificité (Sp) : taux de classement correct des patients à moindre risque
- Valeur prédictive positive (VPP) : taux de classement correct chez les patients classés à risque
- Valeur prédictive négative (VPP) : taux de classement correct chez les patients classés à moindre risque

Le tableau 7.11 présente ces différents indices pour les deux scores proposés. Les chiffres en gras correspondent aux indicateurs obtenus sur l'échantillon de travail, tandis que les chiffres en italiques correspondent aux indicateurs obtenus sur les deux échantillons témoins. Les chiffres

suivis d'une astérisque sont des estimations basées sur la répartition initiale des patients dans les groupes "à risque" et "à moindre risque".

Echantillon de travail <i>Echantillons témoins</i>	Score issu de l'analyse discriminante	Score issu de la procédure Raschfit
Taux de classements corrects	93% 63%*	92% 68%*
Sensibilité	97% 65%	99% 69%
Spécificité	90% 61%	86% 66%
Valeur prédictive positive	89% 43%*	85% 47%*
Valeur prédictive négative	97% 80%*	99% 83%*

TAB. 7.11 – Indicateurs de validité des scores proposés

Nous voyons que les deux scores proposés ont de très bons indicateurs sur l'échantillon de travail. Ceux-ci se dégradent cependant de façon non négligeable sur les échantillons-témoins. Les indicateurs sont comparables entre les deux scores, que ce soit sur l'échantillon de travail ou sur les échantillons témoins (avec un léger avantage pour le score issu de la procédure Raschfit dans ce dernier cas). L'avantage du score issu de la procédure Raschfit est sa simplicité d'utilisation, notamment dans un cadre clinique, comparé au score issu de l'analyse discriminante qui nécessite des calculs (aussi légers soit-ils) ou bien l'usage d'un arbre de décision.

Ces scores sont des propositions, et ne peuvent être utilisés pour l'instant que comme hypothèse de recherche. Ils nécessitent une validation dans un cadre identique (population général de patients ayant effectués une tentative de suicide) avant toute utilisation clinique, comme le préconisent McGinn et al. [101] qui ont réalisé une classification en niveaux des règles de prédiction clinique, en fonction de la validation de ces règles et donc de leur preuve scientifique. Les scores proposés ici se placent ainsi au dernier niveau de cette classification en quatre niveaux.

7.6 Synthèse

Dans les deux premières parties, nous avons présentés un certain nombres d'outils (indices, modèles, procédures) provenant principalement de la théorie de réponse aux items. Dans la partie suivante, nous présentons des programmes SAS et Stata permettant le calcul de ces indices, l'estimation de ces paramètres ou la réalisation de ces procédures de façon automatisée.

Troisième partie

La Théorie de Réponse aux Items sous
SAS et Stata

Chapitre 8

Analyse de l'existant et nouveaux macros-programmes pour l'IRT

8.1 L'IRT et les logiciels généralistes

Les logiciels généralistes (SAS, Splus, Stata, R, SPSS...) ont peu développé les modèles de l'IRT. Diverses raisons peuvent expliquer ce manque :

- des raisons historiques : l'IRT a été développé principalement dans le champ des sciences de l'éducation, par des non-statisticiens (psychologues notamment), peu habitués à utiliser des logiciels statistiques généralistes. Ces utilisateurs ont donc développé leurs propres logiciels plutôt que de s'adapter aux logiciels généralistes existants,
- des raisons techniques : les modèles de l'IRT nécessitent de grandes ressources informatiques pour estimer les paramètres (que ce soit par CML ou MML), aussi, les logiciels spécifiques ont été développés dans des langages rapides (de bas niveau) comme le FORTRAN ou le C, plutôt que dans des langages de haut niveau plus lents comme les langages des logiciels généralistes.

Pourtant, depuis quelques années, la montée en puissance des ordinateurs permet désormais d'utiliser des méthodes utilisant de larges ressources avec les logiciels généralistes dans des temps raisonnables. Ainsi, depuis la toute fin des années 90, les logiciels généralistes proposent des procédures pour estimer les paramètres des modèles non linéaires à effet mixtes, ce qui permet donc l'estimation par MML des modèles de l'IRT. En outre, certains auteurs ont en parallèle montré qu'il était désormais possible d'estimer dans des temps raisonnables les paramètres de certains modèles par CML. Cela permet aujourd'hui le développement de l'IRT dans ces logiciels, qui avaient maintenu ce champ à l'écart.

L'usage des logiciels généralistes a pourtant de nombreux avantages puisqu'il permet à tous les possesseurs de ces logiciels d'utiliser ceux-ci sans contraintes budgétaires, et sans avoir à apprendre de nouveaux langages. De plus, contrairement à l'ensemble des logiciels spécifiques, il est possible de programmer l'ensemble des indices, tests ou modèles de l'IRT (avec certes parfois beaucoup de temps), ce qui évite l'usage de plusieurs logiciels spécialisés, avec toutes les contraintes de transferts de données de l'un à l'autre par exemple. Les possibilités de ces logiciels sont donc très importantes dans ce champ.

Nous nous sommes intéressés à deux logiciels généralistes particuliers : SAS et Stata. Dans la suite de ce chapitre, nous ferons le point sur ces logiciels en ce qui concerne l'existant dans les versions de base (accessibles directement à tous les utilisateurs de ces logiciels) et aux ajouts spécifiques trouvés dans la littérature. Dans un second temps, nous proposerons un certain nombre de programmes automatiques particuliers pour réaliser certaines procédures, calculer certains indices ou estimer certains modèles présentés précédemment. Nous expliciterons en détail les options de ces logiciels, et le cas échéant, expliciterons les programmes.

8.2 SAS

8.2.1 Version de base

La version de base de SAS ne prévoit aucune procédure spécifique concernant l'analyse d'items. On retiendra simplement la possibilité d'obtenir la valeur du alpha de Cronbach pour un ensemble d'items dichotomiques ou polytomiques dans le cadre de la PROC CORR (en ajoutant l'option ALPHA), ainsi que quelques indices associés (corrélation items-score, valeur moyenne des corrélations inter-items...).

8.2.2 Ajouts

Sous SAS, il existe peu d'ajouts concernant l'analyse d'items retrouvés sous formes de macros effectuant des analyses automatisés. Les seuls ajouts dans ce domaine retrouvés, en dehors des programmes présentés par la suite sont très récents (2003) et concernent un travail en ce sens de Christensen et Bjorner [20]. Six macro-programmes sont téléchargeables¹ qui sont détaillées dans le tableau 8.1.

Nom des programmes	Détail
%rasch	Estimation du modèle de Rasch par CML
%llrasch	Estimation du modèle log-linéaire de Rasch (covariables possibles) par CML
%pml	Réalisation du test d'unidimensionnalité de Martin-Löf
%latreg	Modèles de régressions avec une covariable latente
%latreg_2d et %latreg_st	Modèles de régressions avec deux covariables latentes

TAB. 8.1 – Macro-programmes de Christensen et Bjorner

8.3 Stata

8.3.1 Version de base

Tout comme pour SAS, Stata n'a pas prévu, dans sa version de base, de procédures spécifiques pour l'IRT. On notera l'existence de la procédure (module) **alpha** permettant de calculer la valeur du alpha de Cronbach sur des items dichotomiques ou polytomiques associés à quelques autres indices. Cette procédure est donc très proche de la procédure CORR de SAS de ce point de vue.

¹<http://www.biostat.ku.dk/rasch/macro.html> ou <http://freeirt.free.fr>

8.3.2 Ajouts

Dans sa philosophie, Stata est plus enclin à l'élaboration de procédures par ses utilisateurs que SAS. Aussi, les ajouts sont un peu plus anciens sous ce logiciel que sous SAS. Il est cependant intéressant de voir que plusieurs de ces ajouts concernent l'exportation de données Stata dans des formats lisibles par des logiciels spécifiques développés dans le cadre de l'IRT. C'est notamment le cas des procédures `to_msp`, `to_rumm` et `raschvt` qui permettent respectivement d'exporter des données vers les logiciels MSP, RUMM et WINSTEPS².

On ne peut en outre pas parler de modèles à variables latentes sous Stata sans nommer le programme `gllamm`, écrit par Rabe-Hesketh, Skrondal et Pickles (voir par exemple [123], [124] ou [125]). Ce programme permet d'estimer les paramètres des modèles linéaires généralisés à effets mixtes, dont le modèle de Rasch ou celui de Birnbaum [123]. Cependant, cette procédure entre dans un cadre très large, celui des modèles linéaires généralisés à effets mixtes, et n'offre qu'un processus d'estimation des paramètres. Un site Internet (<http://www.gllamm.org>) concerne ce programme.

8.4 Théorie classique et estimations des modèles de l'IRT sous SAS

8.4.1 Le macro-programme %AnaQol

8.4.1.1 Définition

Le macro-programme SAS `%AnaQol` (version 3.3) permet sous la même syntaxe de :

- calculer le alpha de Cronbach,
- réaliser une procédure pas à pas du alpha de Cronbach,
- dessiner les traces classiques ou cumulées d'items dichotomiques ou polytomiques en fonction du score ou du rest-score,
- dessiner les traces logistiques pour les items dichotomiques,
- estimer les paramètres du modèle de Rasch, de l'OPLM ou du modèle de Birnbaum pour les items dichotomiques,
- estimer les paramètres du Partial Credit Model ou du Rating Scale Model pour les items polytomiques,
- dessiner sur les mêmes graphiques les traces empiriques et théoriques de chaque item sous les modèles de Rasch et de Birnbaum,
- calculer les statistiques du test Q1 de van den Wollenberg sous le modèle de Rasch et celui de Wright-Panchapakesan sous le modèle de Birnbaum pour les items dichotomiques.

Ce programme a été détaillé en [70]. Il nécessite le programme `%gammasy` qui permet de calculer les valeurs des fonctions symétriques gamma.

²Le premier programme est proposé par Jeroen Weesie, les deux autres par Fred Wolfe. Ils sont téléchargeables en tapant "findit" depuis Stata. Aucune documentation supplémentaire au fichier hlp classique sous Stata n'a été retrouvée pour ces trois programmes

8.4.1.2 Syntaxe

La syntaxe du macro-programme **%AnaQol** est la suivante :

```
%AnaQol(DATASET=nom_de_la_table, ITEMS=liste_des_items,
ALPHA=yes/no, SBSALPHA=yes/no, TRACES=none/classical/cumulative/fit,
REF=score/restscore, DISPLAY=yes/no, LOGISTIC=yes/no, DETAILS=yes/no,
MODEL=none/rasch/birnbaum/oplm/pcm/rsm, B=liste_des_pouvoirs_discriminants);
```

Les options du programme **%AnaQol** sont répertoriées dans le tableau 8.2.

Options	Défaut	Description
DATASET=	<code>_last_</code>	Permet de choisir la table de travail
ITEMS=	<code>_numeric_</code>	Détermine les items à utiliser
OUT=	OUT	Définit le préfixe des tables de sorties
ALPHA=	NO	Calcule le alpha de Cronbach et les indices associés (<i>YES</i> ou <i>NO</i>)
SBSLPHA=	NO	Réalise une procédure pas-à-pas basée sur le alpha de Cronbach (<i>YES</i> ou <i>NO</i>)
TRACES=	NONE	Dessine les traces classiques (<i>CLASSICAL</i>) ou cumulées (<i>CUMULATIVE</i>) [différence visible seulement pour les items polytomiques], ou bien dessine sur le même graphique les traces théoriques et empiriques des items (<i>FIT</i>) (seulement avec un modèle pour items dichotomiques)
REF=	SCORE	Représente les traces en fonction du score (<i>SCORE</i>) ou du rest-score (<i>RESTSCORE</i>)
DISPLAY=	NO	Affiche sur les graphiques des traces les probabilités (<i>YES</i> ou <i>NO</i>)
LOGISTIC=	NO	Représente l'ensemble des traces logistiques (<i>YES</i> ou <i>NO</i>) [seulement pour des items dichotomiques]
DETAILS=	NO	Affiche le détail de l'algorithme d'estimation des paramètres d'un modèle (<i>YES</i> ou <i>NO</i>)
MODEL=	NONE	Modélise les réponses à des items dichotomiques par un modèle de Rasch (<i>RASCH</i>), un OPLM (<i>OPLM</i>) ou un modèle de Birnbaum (<i>BIRNBAUM</i>) ou celles d'items polytomiques par un Partial Credit Model (<i>PCM</i>) ou un Rating Scale Model (<i>RSM</i>)
B=		Liste les pouvoirs discriminant des items (dans l'ordre des items définis dans ITEMS=) [seulement si MODEL=OPLM]

TAB. 8.2 – Options du macro-programmes SAS **%AnaQol**

La syntaxe du macro-programme **%gammasym** est la suivante :

```
%gammasym(LIST=liste_de_valeurs);
```

Ce programme retourne les macro-variables locales "gamma#" pour chaque valeur # du score.

On notera que par défaut, le macro-programme **%AnaQol** permet d'obtenir les tables

`_rep` et `_dege` qui permettent, pour la première, d'obtenir les scores et rest-scores calculés pour chaque items, et pour la seconde d'obtenir une table mise en forme pour réaliser la procédure NLMIXED. Ce macro-programme peut sans problème être utilisé dans ces buts.

Un exemple de sortie du macro-programme `%AnaQol` est présenté en annexe C.2.

8.5 Théorie classique et estimations des modèles de l'IRT sous Stata

8.5.1 Le programme `traces`

8.5.1.1 Définition

Le programme Stata `traces` (version 3) permet d'obtenir diverses représentations graphiques des traces d'items dichotomiques ou polytomiques.

8.5.1.2 Syntaxe

La syntaxe du programme `traces` est la suivante :

```
. traces liste_des_items [, score restscore ci test cumulative logistic
repfile(répertoire) scorefiles(nom_de_fichier) restscorefiles(nom_de_fichier)
logisticfile(nom_de_fichier) nodraw nodrawcomb replace]
```

es options du programme `traces` sont répertoriées dans le tableau 8.3. Un exemple de sortie du programme `traces` est présenté en annexe C.3.

8.5.2 Le programme `raschtest`

8.5.2.1 Définition

Le programme Stata `raschtest` (version 7.2) permet d'estimer les paramètres d'un modèle de Rasch par maximum de vraisemblance marginale (en utilisant la procédure `xtlogit`), maximum de vraisemblance conditionnelle (en utilisant la procédure `clogit`) ou équations d'estimation généralisées (en utilisant la procédure `geekel2d` définie à la section 8.5.4). Ce module permet d'obtenir les tests du premier ordre (avec une estimation par CML, R1c, Q1 ou Wright-Panchapakesan, et avec les autres estimations R1m ou une version adaptée de Q1), le test de rapport de vraisemblance Z de Ansersen et le test U (avec estimation par CML), le test de cission, les indices OUTFIT et INFIT par items ou par individus. Le programme estime en outre les valeurs individuelles du trait latent associées à chaque niveau du score. Des regroupements de modalités peuvent être effectuées lorsque peu d'individus présentent certaines valeurs du score. Le programme est documenté en [63].

Le programme nécessite l'accès à plusieurs programmes qui ne font pas partie de la version de base de Stata :

- `gammasym` qui permet de calculer les valeurs des fonctions symétriques gamma,

Options	Description
score	Représente les traces des items en fonction du score (par défaut si les options restscore et logistic ne sont pas indiquées)
restscore	Représente les traces des items en fonction du rest-score
ci	Affiche sur les traces l'intervalle de confiance à 95%
test	Teste la nullité de la pente moyenne des traces
cumulative	Pour des items polytomiques, remplace les traces classiques par les traces cumulées
logistic	Représente les traces logistiques (pour des items dichotomiques)
reprofile	Permet de définir un répertoire pour sauver les graphiques
scorefiles	Permet de définir le nom générique des fichiers concernant les traces en fonction du score (et de les sauver)
restscorefiles	Permet de définir le nom générique des fichiers concernant les traces en fonction du rest-score (et de les sauver)
logisticfile	Permet de définir le nom du fichier des traces logistiques
nodraw	N'affiche pas les traces à l'écran
nodrawcomb	N'affiche pas le graphique des traces combinées (toutes les traces sur le même graphique)
replace	Permet d'écraser des fichiers déjà existants

TAB. 8.3 – Options du programmes Stata **traces**

- **gausshermite** qui permet d'approximer des intégrales avec des quadratures de Gauss-Hermite,
- **geekel2d** qui permet d'estimer les paramètres du modèle de Rasch par GEE,
- **gllamm** qui permet d'estimer les paramètres des modèles linéaires généralisés mixtes [123],
- **gllapred** qui permet d'estimer les estimations attendues *a posteriori* de Bayes pour le trait latent [123],
- **ghquadm** qui permet d'obtenir les noeuds et poids des quadratures de Gauss-Hermite,
- **elapse** qui permet de calculer le temps d'exécution d'un programme.

8.5.2.2 Syntaxe

La syntaxe du programme **raschtest** est la suivante :

```
. raschtest liste_des_items [if exp][in range] [, method(cml/mml/gee) test(R/Q/WP)
meandiff details group(liste_numérique) autogroup dirsave(répertoire) flessave pause
replace icc graph information splittest genlt(nouvelle_variable) genscore(nouvelle_variable)
genfit(liste_de_variables) comp(variable) trace explain
```

Les options de ce programme sont décrites dans le tableau 8.4.

Précisons simplement la syntaxe de l'option **group** : on indique dans cette option les valeurs seuils hauts des groupements de scores, ainsi **group**(1 2 3 4) signifie que l'on prendra les groupes de scores 0, 1, 2, 3, 4, 5 à J-1, et J. On note que les scores nul et parfait (0 et

Options	Description
method	Permet de choisir la méthode d'estimation des paramètre de difficulté des items (CML par défaut),
test	Permet de choisir le type du test du premier ordre (R par défaut),
meandiff	Permet, avec test(cml) de centrer les estimations des paramètres de difficulté des items,
details	Permet d'obtenir pour chaque groupe de scores et pour chaque item le nombre de réponses positives observées et attendues, ainsi que la contribution de chaque groupe de scores à la statistique de test du premier ordre,
group	Permet d'effectuer des regroupements de scores, si certains scores sont peu représentés par exemple,
autogroup	Permet d'effectuer automatiquement des regroupements de scores, avec au moins 30 individus par groupe,
dirsave	Permet de définir le répertoire dans lequel seront sauvés les graphiques (celui défini par <code>c(pwd)</code> par défaut)
filesave	Permet de sauvegarder les graphiques dans le répertoire précédent,
pause	Permet de ne pas dessiner les représentations graphique (intérêt seulement si les graphiques sont sauvés)
replace	Permet de remplacer éventuellement les fichiers graphiques,
icc	Représente, pour chaque item, les courbes caractéristiques observée et théorique,
graph	Représente sur un même graphique les distributions du scores, des paramètres de difficulté des items, et, éventuellement, la distribution du trait latent,
information	Représente la fonction d'information de l'échelle,
splittest	Réalise le split test pour chaque item,
gentl	Crée une nouvelle variable contenant l'estimation, pour chaque individu, du trait latent,
genscore	Crée une nouvelle variable contant le score de chaque individu,
genfit	Crée plusieurs nouvelles variables, dont l'indice OUTFIT, l'indice INFIT, et la contribution de chaque item à l'indice OUTFIT,
comp	Réalise un test de Student de comparaison de moyenne du trait latent pour deux groupes défini par la variable,
trace	Affiche, lors de l'exécution du module, les étapes réalisées,
explain	Affiche des informations sur les techniques d'estimation et les tests, et stoppe le module.

TAB. 8.4 – Options du programme Stata **raschtest**

J) restent toujours isolés. la syntaxe **group**(1 3 4) permettra de prendre les groupes 0, 1, 2 à 3, 4 à J-1, J. Avec une telle syntaxe, on notera, par exemple pour 6 items, que **group**(1 2 5) correspond exactement à **group**(1 2).

Le programme **gammasym** peut en outre être utilisé de façon autonome pour calculer la valeur de la fonction symétrique gamma. Sa syntaxe est la suivante :

```
. gammasym liste_des_paramètres
```

La liste des paramètres est composée des valeurs des paramètres δ_j , $j = 1, \dots, J$ à utiliser. Les valeurs des fonctions symétriques gamma sont retournées dans des scalaires $r(\text{gamma}\#)$ pour chaque niveau $\#$ possible du score (de 0 à J).

Le programme **gausshermite** peut en outre être utilisé de façon autonome pour estimer l'intégrale sur l'ensemble des réels d'une fonction $f(x)$ avec x distribué normalement. Sa syntaxe est la suivante :

```
. gausshermite fonction [, nodes(#) mu(#) sigma(#) display]
```

avec *fonction* une fonction en x (par exemple $x^2 + \log(x)$), l'option **nodes** permettant de définir le nombre de point de quadrature (12 par défaut), **mu** de définir la moyenne de x (0 par défaut), **sigma** l'écart type de x (1 par défaut) et **display** permettant d'afficher le résultat à l'écran.

Un exemple de sortie du programme **raschtest** est présenté en annexe C.4.

8.5.3 Le programme mmsrm

8.5.3.1 Définition

Le programme **mmsrm** permet d'estimer les paramètres d'un MMSRM par maximum de vraisemblance marginale. Ce programme utilise la procédure **gllamm** proposée par Rabe-Hesketh, Pickles et Skrondal [123] pour estimer les paramètres du modèle. La procédure se déroule en deux temps : dans un premier temps, les paramètres de difficulté des items sont estimés dans des modèles unidimensionnels puis les paramètres de la distribution du trait latent multidimensionnel sont estimés (voir section 4.7). Dans le cas où un seul trait latent est défini, la procédure renvoie au programme **raschtest**. Avec les versions actuelles de ce programme, il n'est possible de n'indiquer que deux traits latents au maximum.

8.5.3.2 Syntaxe

La syntaxe du programme **mmsrm** (version 2) est la suivante :

```
. mmsrm liste_des_items , partition(liste_de_nombre)[ trace nodetails]
```

L'option **partition** doit indiquer le nombre d'items reliés à chacun des traits latents : le premier chiffre indique le nombre d'items reliés à un premier trait latent, tandis que (l'éventuel) second chiffre indique le nombre d'items reliés à un second trait latent. Les items sont

Options	Description
partition	Définit la partition des items
trace	Permet d'obtenir le détail de la procédure d'estimation
nodetails	Supprime l'affichage des informations sur l'avancée de la procédure

TAB. 8.5 – Options du programme Stata **mmsrm**

pris dans l'ordre de *liste_des_items*.

Le programme utilise les modules **raschtest** (et les autres modules dont celui-ci a besoin) et **glamm** [123]. Un exemple de sortie du programme **mmsrm** est présenté en annexe C.5.

8.5.4 Le programme **geekel2d**

8.5.4.1 Définition

Le programme **geekel2d** permet d'estimer les paramètres du Multidimensional Polytomous Latent Trait Model (MPLT) restreint au cadre dichotomique par la méthode des GEE, avec une dimension (modèle de Rasch ou OPLM) ou deux dimensions (MMSRM, GMRM, MPLT). Le détail des calculs est donné à la section 3.4.3.6.

8.5.4.2 Syntaxe

La syntaxe du programme **geekel2d** (version 4.1) est la suivante :

```
. geekel2d liste_des_items [, coef(nom_de_matrice) nbit(#) critconv(#)
quad(#) ll novar]
```

Options	Description
coef	Définit le nom de la matrice B avec J lignes et Q colonnes (par défaut, cette matrice est considérée comme un vecteur de 1 et donc le modèle estimé est un modèle de Rasch)
nbit	Fixe le nombre maximal d'itérations (par défaut à 30)
critconv	Fixe la valeur du critère de convergence (par défaut à $1e - 15$)
quad	Définit le nombre de noeuds de quadrature à utiliser (par défaut à 12)
ll	Estime la log-vraisemblance marginale et le critère d'Akaike par quadratures de Gauss-Hermite
novar	N'estime pas les variances des estimations (plus rapide)

TAB. 8.6 – Options du programme Stata **geekel2d**

La matrice définie dans **coef** doit être sauvée au préalable. Cette matrice contient un item par ligne et une dimension par colonne. Si la matrice est un vecteur de 1, alors le modèle est un modèle de Rasch, si c'est un vecteur d'entiers, il s'agit d'un OPLM. Dans un cadre

bidimensionnel, lorsque la somme des coefficients de chaque ligne est égal à une constante, il s'agit du GMRM défini à la section 3.4.3.7, lorsque cette constante est égale à 1, il s'agit d'un MMSRM défini à la section 4.5. Dans les autres cas, il s'agit d'un MPLT. Bien que l'on puisse indiquer des réels dans la matrice, nous recommandons de n'utiliser que les premiers entiers, comme le font les auteurs du modèle [87].

Soit $\beta^{(k)}$ l'estimation du vecteur des paramètres du modèles à l'étape k , alors le critère de convergence à l'étape k est défini comme :

$$\epsilon^{(k)} = (\beta^{(k-1)} - \beta^{(k)})'(\beta^{(k-1)} - \beta^{(k)}) \quad (8.1)$$

avec $\beta^{(0)} = \mathbf{0}$.

Le programme utilise **ghquadm** pour calculer les noeuds et les poids de quadrature. Ce programme est disponible gratuitement sur internet ³. Un exemple de sortie du programme **geekel2d** est présenté en annexe C.6.

8.6 Les procédures de sélection d'échelles unidimensionnelles sous SAS

8.6.1 Le macro-programme %BackRasch

8.6.1.1 Définition

Le macro-programme **%BackRasch** permet de réaliser la procédure BackRasch sur un ensemble d'items. Il est possible de définir le niveau de signification et le nombre d'échelles à construire. Ce programme est détaillé en [61].

8.6.1.2 Syntaxe

La syntaxe du macro-programme **%BackRasch** (version 2) est la suivante. Les options sont définies dans le tableau 8.7.

```
%BackRasch(DATASET=nom_de_la_table, ITEMS=liste_des_items,
MAXDIM=1/#/max, P=0.05/#);
```

Le macro-programme **%BackRasch** ne crée pas de sorties spécifiques en dehors des informations obtenues dans les fenêtres log et output. Un exemple de sortie est présentée en annexe C.7. Ce macro-programme utilise **%AnaQol** (et les autres macro-programmes dont celui-ci a besoin).

8.6.2 Les macro-programmes %LoevH et %MSP

8.6.2.1 Définition

Le macro-programme SAS **%MSP** (version 3) permet de réaliser la Mokken Scale Procedure sur un ensemble d'items. Ce macro-programme utilise le macro-programme **%LoevH**

³on peut le trouver en tapant "findit ghquadm" depuis Stata

Options	Description
DATASET=	Nom de la table
ITEMS=	Noms des items à utiliser
MAXDIM=	Définit le nombre d'échelles maximales à construire. Par défaut ce nombre est 1, et il est possible d'indiquer <i>max</i> pour que le programme construise des échelles jusqu'à ce qu'il n'y ait plus d'items disponibles
P=	Indique le niveau de signification (par défaut 0.05)

TAB. 8.7 – Options du macro-programme SAS %BackRasch

(version 2) pour calculer les coefficients de Loevinger, cependant le macro-programme %LoevH n'est pas utilisable en tant que tel. Ces deux programmes sont détaillés en [62]. Le macro-programme %MSP tourne sous SAS/IML qui doit être lancé au préalable.

8.6.2.2 Syntaxe

La syntaxe du macro-programme %MSP est la suivante. Les options sont définies dans le tableau 8.8.

```
%MSP(DATASET=nom_de_la_table, ITEMS=liste_des_items, C=0.3/#/max);
```

Options	Description
DATASET=	Nom de la table
ITEMS=	Noms des items à utiliser
C=	Seuil <i>c</i> choisi

TAB. 8.8 – Options du macro-programme SAS %MSP

Un exemple de sortie du macro-programme %MSP est présenté en annexe C.8.

8.6.3 Le macro-programme %DETECT

8.6.3.1 Définition

Le macro-programme SAS %DETECT permet de calculer les indices DETECT, Iss et R ainsi que les corrélations entre les items et les (rest-)scores de la partition indiquée.

8.6.3.2 Syntaxe

La syntaxe du macro-programme SAS %DETECT (version 2) est la suivante. Les options sont définies dans le tableau 8.9.

```
%DETECT(DATASET=nom_de_la_table, OUT=nom_de_la_table_de_sortie,  
ITEMS=liste_des_items, PART=definition_de_la_partition);
```

Options	Description
DATASET=	Nom de la table
OUT=	Nom de la table de sortie
ITEMS=	Noms des items à utiliser
PART=	Définition de la partition (liste de nombre indiquant le nombre d'items reliés à chaque dimension, dans le même ordre que items=)

TAB. 8.9 – Options du macro-programme SAS %DETECT

Un exemple de sortie du macro-programme %DETECT est présenté en annexe C.9.

8.7 Les procédures de sélection d'échelles unidimensionnelles sous Stata

8.7.1 Le programme raschfit

8.7.1.1 Définition

La programme **raschfit** permet de réaliser la procédure Raschfit ou Raschfit-fast sur un ensemble d'items dichotomiques. Il est possible d'utiliser plusieurs manières pour classer les items : soit l'utilisateur classe lui-même les items (les premiers items formant le noyau de la première sous-échelles, les items étant ensuite pris un par un dans l'ordre défini, soit en utilisant la procédure MSP en prenant un seuil c égal à 0, et en n'effectuant aucun test. L'ordre des items, en dehors du noyau constitués des premiers items définis par la procédure (ou par l'utilisateur) suivra alors un classement positif (de l'item étant classé avec le noyau en premier à l'item regroupé en dernier avec l'échelle) ou négatif (ordre inverse en dehors du noyau).

8.7.1.2 Syntaxe

La syntaxe du programme Stata **raschfit** (version 2) est la suivante :

```
. raschfit liste_des_items , [kernel(#) nbscales(#) itemsorder(mot-clé)
nofast]
```

Les options du programme **raschfit** (version 2) sont définies dans le tableau 8.10.

raschfit utilise les modules **mssp**, **raschtest** et **mmsrm** (ainsi que les modules nécessaires à ces trois programmes).

Une seule sortie est prévue pour le programme **raschfit** : il s'agit d'un vecteur $r(\text{affect})$ qui comporte, pour chaque item, le numéro de la dimension à laquelle l'item a été affecté (0 si l'item n'a été affecté à aucune dimension). Un exemple de sortie du programme **raschfit** est présenté en Annexe 2 page 194.

Options	Description
kernel	Définit un noyau pour la première sous-échelle composée des # premiers items de <i>varlist</i> (par défaut 0)
nbscales	Définit le nombre maximal de sous-échelles à construire (par défaut 0)
itemsorder	Définit l'ordre des items dans la procédure, soit par le classement de l'utilisateur (<i>order</i>) soit dans un order positif obtenu avec MSP (<i>mSP</i>) soit dans un order négatif obtenu avec MSP (<i>mSPinv</i>). Lorsque <i>order</i> est défini, le noyau est composé des 2 premiers items de <i>varlist</i> (sauf indication contraire dans kernel (par défaut <i>mSPinv</i>))
nofast	Par défaut, Raschfit-fast est utilisée. Cette option permet d'effectuer la procédure Raschfit.

TAB. 8.10 – Options du programme Stata **raschfit**

8.7.2 Le programme **backrasch**

8.7.2.1 Définition

Le programme Stata **backrasch** permet de réaliser la procédure Backrasch sur un ensemble d'items dichotomiques. Il est possible de définir la méthode d'estimation (CML, MML ou GEE), le type de test à utiliser (R ou Q), le niveau de signification et le nombre maximal d'échelles à construire. Le programme est détaillé en [61].

8.7.2.2 Syntaxe

La syntaxe du programme Stata **backrasch** (version 2) est la suivante :

```
. backrasch liste_des_items [, method(cml/mml/gee) test(R/Q) p(#) nbscales(#)
nodetail noautogroup]
```

Les options du programme **backrasch** sont définies dans le tableau 8.11.

Options	Description
method	Permet de choisir la méthode d'estimation (CML par défaut)
test	Permet de choisir le test du premier ordre à utiliser (R par défaut)
p	Définit le niveau de signification (par défaut 0.05)
nbscales	Définit le nombre maximal d'échelles à construire (par défaut une seule)
nodetail	N'affiche pas toutes les informations pendant l'exécution de la commande
noautogroup	Evite le regroupement des scores (par défaut, le programme regroupe les scores en groupes de 30 individus minimum)

TAB. 8.11 – Options du programme Stata **backrasch**

Une seule sortie est prévue pour le programme **backrasch** : il s'agit d'un vecteur *r(selection)* qui comporte, pour chaque item, le numéro de la dimension à laquelle l'item a été affecté (0

si l'item n'a été affecté à aucune dimension).

Le programme **backrasch** utilise **raschtest** (et tous les programmes utilisés par celui-ci). Un exemple de sortie du programme **backrasch** est présenté en annexe C.11.

8.7.3 Les programmes **loevh** et **msh**

8.7.3.1 Définition

Le program Stata **loevh** (version 6) permet d'obtenir les erreurs de Guttman (par paire d'items, pour un item dans une échelle et pour l'ensemble de l'échelle) observées et théoriques (sous l'hypothèse d'indépendance deux à deux des items) au sein d'une échelle d'items dichotomiques or polytomiques ordonnés, ainsi que les différents coefficients de Loevinger (par paire d'items, d'intégration de chaque item dans l'échelle et pour l'échelle). Ce programme est détaillé en [62].

8.7.3.2 Syntaxe

Sa syntaxe est la suivante :

```
. loevh liste_des_items [, pairwise pair ppp pmm]
```

Les options de ce programme sont décrites dans le tableau 8.12.

Options	Description
pairwise	Supprime les individus ayant des valeurs manquantes dans chaque croisement d'items (au lieu de la suppression globale de tous les individus ayant au moins une donnée manquante)
pair	Affiche les erreurs de Guttman et les coefficients de Loevinger par paire d'item
ppp	Affiche la matrice P++
pmm	Affiche la matrice P- -

TAB. 8.12 – Options du programme Stata **loevh**

Par défaut, tous les individus ayant au moins une donnée manquante sont supprimés. Avec l'option **pairwise**, on calcule les erreurs de Guttman observées et théoriques par paire d'items avec tous les individus ayant répondu à au moins ces deux items. On utilise ainsi toute l'information disponible. La procédure par défaut correspond cependant à celle utilisée par le logiciel MSP.

Le programme Stata **msh** permet de réaliser sur un ensemble d'items la Mokken Scale Procedure. Ce programme est détaillé en [62]. La syntaxe de ce programme (version 6) est la suivante :

```
. msh liste des items [, c(#) kernel(#) p(#) notest pairwise nodetails]
```

Ce programme utilise le programme Stata **loevh** pour estimer les coefficients. Les options de ce programme sont décrites dans le tableau 8.13.

Options	Description
c	Définit la valeur du seuil c (par défaut 0.3)
kernel	Définit les $\#$ premiers items comme le noyau de la première sous-échelle
p	Définit la valeur du niveau de signification des tests (par défaut 0.05)
pairwise	Utilise l'option "pairwise" pour le calcul des coefficients H de Loevinger
notest	Supprime l'exécution des tests de nullité des coefficients H de Loevinger
nodetails	Supprime les informations sur l'algorithme
nobon	Empêche l'utilisation de corrections de Bonferroni pour les tests de nullité des coefficients de Loevinger
minvalue	Détermine la valeur minimale des coefficients de Loevinger pour chaque paire d'items lors de la construction d'une sous-échelle (par défaut 0)

TAB. 8.13 – Options du programme Stata **m**sp

Un exemple de sortie des programmes **loevh** et **m**sp est présenté en annexe C.12.

8.7.4 Le programme **hcaccprox**

8.7.4.1 Définition

Le programme **hcaccprox** permet de réaliser une classification hiérarchique avec mesures de proximités sur des items dichotomiques.

8.7.4.2 Syntaxe

La syntaxe de ce programme (version 2) est la suivante :

```
. hcaccprox liste_des_items [, prox(mot_clé) method(mot_clé)
partition(liste_de_nombres) measures details detect(#)]
```

Les options de ce programme sont décrites dans le tableau 8.14.

Ce programme utilise le programme **detect** quand l'option **detect** est utilisée.

Un exemple de sortie du programme **hcaccprox** est présenté en annexe C.13.

8.7.5 Le programme **detect**

8.7.5.1 Définition

Le programme **detect** permet de calculer, pour une partition d'items, les indices DETECT, Iss et R, ainsi que les corrélations entre chaque item et les scores et sous-scores.

Options	Description
prox	Définit la mesure de proximité à utiliser : <i>a</i> (Jaccard - voir équation (6.1)), <i>ad</i> (Sokal et Michener - voir équation (6.2)), <i>cor</i> (Guilford - voir équation (6.3)), <i>ccov</i> (voir équation (6.5)), <i>ccor</i> (voir équation (6.4)) ou <i>mh</i> (voir équation 6.6) (<i>ccov</i> par défaut)
method	Définit la méthode d'aggrégation des clusters : <i>single</i> (Lien unique), <i>complete</i> (lien complet), <i>UPGMA</i> (Méthode de la moyenne par paire non pondérée) (voir section 6.1.3) [<i>UPGMA</i> par défaut]
partition	Définit la liste des partitions à afficher par exemple 1 – 5 pour les partitions ayant entre 1 et 5 clusters, ou 2 3 4
measures	Affiche la matrice des mesures de proximités entre les items
details	Affiche le détail de l'algorithme d'aggrégation
detect	Calcule les coefficients DETECT, Iss et R pour les partitions en $K \leq \#$ clusters

TAB. 8.14 – Options du programme Stata **hcaccprox**

8.7.5.2 Syntaxe

La syntaxe de ce programme (version 3.1) est la suivante :

```
. detect liste_des_items, partition(liste_de_nombres) [noscores norestscores]
```

Les options de ce programme sont décrites dans le tableau 8.15.

Options	Description
partition	Définit la partition des items comme une suite de nombre. Le premier nombre indique le nombre d'items (les premiers de <i>varlist</i>) reliés au premier trait latent, le second nombre indique que les items suivants sont reliés au second trait latent, et ainsi de suite.
noscores	N'affiche pas la matrice de corrélation entre les items et les scores
norestscores	N'affiche pas la matrice de corrélation entre les items et les rest-scores

TAB. 8.15 – Options du programme Stata **detect**

Un exemple de sortie du programme **detect** est présenté en annexe C.14.

8.7.6 Le programme **simirt**

8.7.6.1 Définition

Le programme **simirt** permet de simuler des données selon un modèle de l'IRT (Rasch, Birnbaum, OPLM, 3-PLM, 4-PLM, 5-PAM). Il est possible de simuler deux ensembles d'items avec deux traits latents différents qui peuvent être ou non corrélés.

8.7.6.2 Syntaxe

La syntaxe de ce programme (version 1) est la suivante :

```
. simirt [, nbobs(#) dim(# [#]) mu(# [#]) cov(# [# #]) diff(liste_de_valeurs_ -
ou_expression) disc(liste_de_valeurs) pmin(liste_de_valeurs) pmax(liste_de_valeurs) acc(liste_
de_valeurs) clear store(filename) replace prefix(string [string]) draw ]
```

Les options de ce programme sont décrites dans le tableau 8.16.

Options	Description
nbobs	Définit le nombre d'individus à simuler,
dim	Définit le nombre d'items dans chaque dimension,
mu	Définit les moyennes des traits latents,
cov	Définit la matrice de covariance des traits latents,
diff	Définit les paramètres de difficulté des items,
disc	Définit les pouvoirs discriminants des items,
pmin	Définit les probabilités de réponses positives minimum des items,
pmax	Définit les probabilités de réponses positives maximum des items,
acc	Définit les paramètres d'accélération des items,
clear	Stipule que les données actuelles doivent être remplacées,
store	Définit le fichier dans lequel les données doivent être sauveées,
replace	Permet de remplacer éventuellement le fichier de sauvegarde,
prefix	Permet de définir le préfixe des items,
draw	Affiche les courbes caractéristiques théoriques des items.

TAB. 8.16 – Options du programme Stata **simirt**

8.8 Le site web AnaQol et le projet FreeIRT

8.8.1 AnaQol

Le site AnaQol (<http://anaqol.free.fr>) présente l'ensemble des programmes décrits au chapitre 8. Sur ce site, les mises à jour sont présentées, la syntaxe de chaque programme est expliquée, et il est possible de télécharger les différents programmes. Des références globales permettent de trouver la littérature indispensables à la bonne compréhension des programmes. Le site est développé en français et en anglais.

8.8.2 Le projet FreeIRT

Le projet FreeIRT est une collaboration avec Karl Bang Christensen, afin de mettre en commun le travail réalisé sous SAS et Stata. Le projet FreeIRT est une association informelle entre chercheurs proposant des programmes gratuitement et librement dans le domaine de l'IRT sur les logiciels SAS, Stata, Splus, R et SPSS. Le projet a été lancé en février 2004 et s'est ouvert au public dès mars 2004. L'adhésion est libre. Chacun peut proposer ses propres

programmes.

Les programmes peuvent être répertoriés à condition de vérifier un certain nombre de conditions de forme. Au minimum, les programmes doivent indiquer l'auteur, les différentes versions du programme, une adresse pour contacter l'auteur, la date et le numéro de la version, et une description sommaire de l'utilisation du programme. En outre, un certain nombre de règles complémentaires sont indiquées et souhaitées (manuel, références, options "noprint" ou " nodisplay", utilisation d'objets temporaires, sorties, gestion des données manquantes, possibilité de réaliser le programme sur un sous-ensemble d'observations, messages d'erreurs...).

Le site FreeIRT (<http://freeirt.free.fr>) est une interface graphique écrite en scripts PHP d'une base de données MySQL contenant toutes les informations sur les programmes, les versions et les auteurs. Une arborescence de thèmes couverts par les programmes a été créée. Il est ainsi possible d'effectuer des recherches par programme, par auteur, par logiciel ou par thème (mais aussi avec les combinaisons thèmes/logiciels). Le site est interactif dans le sens où chaque programmeur peut directement modifier les informations de la base de données en ce qui concerne ses propres informations ou les informations sur les programmes ou versions dont il est l'auteur principal, et mettre ainsi en temps réels à jour le site. En outre, il est possible d'enregistrer des références bibliographiques et de les associer à chaque programme.

Notons enfin qu'il est possible à tout visiteur d'émettre en temps réel des commentaires (problèmes rencontrés, conseils, souhaits) sur un programme et de notifier des erreurs. Dans le cas où des erreurs sont signalées, le programme reste téléchargeable, mais un message indique ces erreurs à chaque page où il est possible de télécharger le logiciel. Une nouvelle version ne peut être proposée que si elle corrige les erreurs notifiées.

Les sites AnaQol et FreeIRT ont donné lieu, entre janvier 2004 et mai 2005, à près de 5000 visites différentes et plus de 40000 pages vues. Parmi les programmes présentés dans ce documents, plus de 4500 fichiers de programmes ont été téléchargés (tous programmes et versions confondus).

Conclusion et perspectives

Dans ce travail, nous nous sommes intéressés à deux domaines : d'une part la création de sous-échelles d'items ayant de bonnes propriétés psychométriques, d'autres part, l'implémentation d'indices, modèles et procédures issues de l'IRT dans les logiciels SAS et Stata.

En ce qui concerne la création d'échelles d'items, nous avons proposé trois méthodes différentes. La première, appelée BackRasch correspond à une procédure pas à pas du descendant (Backward) sur un modèle de Rasch contenant initialement l'ensemble des items, en utilisant un test d'adéquation. Une seconde méthode, appelée "Raschtest", s'appuie sur un modèle multidimensionnel que nous avons défini, le modèle multidimensionnel marginalement exhaustif de Rasch (MMSRM) : le principe de cette méthode, basée sur une procédure pas à pas ascendante (Forward), est de comparer à chaque étape le modèle permettant l'ajout d'un item dans la même dimension que les items déjà sélectionné (par un modèle de Rasch) avec un modèle liant le nouvel item à une seconde dimensionnel (par un MMSRM). La comparaison des deux modèles est réalisée par le critère d'Akaike. Une dernière procédure, appelée "Raschtest- Fast", est une adaptation de "Raschtest" plus rapide, car basée principalement sur l'utilisation de modèles unidimensionnels.

Des simulations montrent dans les cas favorables une nette prédominance de "Raschfit" sur les deux autres méthodes ainsi que sur les méthodes retrouvées dans la littérature (issues de l'analyse factorielle, ou de la Théorie de Réponse aux Items non paramétriques). Dans des cas moins favorables (c'est-à-dire quand le modèle simulé n'est plus un MMSRM), la procédure "Raschfit" donne des résultats similaires aux autres procédures. En revanche, la procédure "Raschfit" requiert des temps d'exécution importants, car elle nécessite l'estimation des paramètres de plusieurs modèles multidimensionnels, ce qui donne en ce sens des avantages considérables aux deux autres procédures ou aux autres procédures de la littérature.

Ces méthodes ont été utilisées pour étudier la cohérence de la structure factorielle de trois sous-échelles du Sickness Impact Profile (SIP) : dans leur ensemble, les méthodes montrent de légères différences entre la structure du SIP pour deux des trois sous-échelles et une dernière sous-échelle incohérente d'un point de vue psychométrique. La procédure Raschfit a en outre été utilisée dans un autre contexte, pour créer un score simple (non pondéré) dans la prédiction de la récurrence suicidaire.

Ce travail pourrait être ultérieurement développé vers d'autres modèles, en particulier vers les modèles polytomiques. Des méthodes pour accélérer les processus d'estimations des paramètres des modèles, EM ou SEM par exemple, pourraient être des solutions à court terme pour développer les procédures basées sur des modèles longs à estimer. A terme, cet inconvénient

devrait cependant être compensé par l'augmentation des puissances des ordinateurs.

Concernant le deuxième thème de ce travail, à savoir l'implémentation sous SAS et Stata de programmes concernant l'IRT, il a aboutit à la conception de 7 macro-programmes SAS et 13 modules Stata permettant de calculer automatiquement un certain nombre d'indices, d'estimer les paramètres des modèles de l'IRT parmi les plus courants, et de réaliser un certain nombre de procédures, dont les procédures de sélection d'items proposées. Ce projet a été à l'initiative d'un projet plus vaste de répertorisation des programmes issus de l'IRT sous les logiciels statistiques généralistes (Projet FreeIRT). Un site Internet a été créé à l'occasion.

Ce travail doit être prolongé pour diversifier au maximum l'offre de programmes sous ces logiciels très utilisés et parfois seuls logiciels techniques disponibles pour les utilisateurs. Ce type de développement devrait permettre un gain de temps important pour les utilisateurs des programmes (pas de formation à d'autres logiciels qu'à celui utilisé habituellement, pas de transferts fastidieux des données d'un logiciel à l'autre) et un gain économique (pas de nouveau(x) logiciel(s) à acheter). À terme, ceci devrait permettre de développer la notoriété de l'IRT auprès des utilisateurs de logiciels généralistes. Ceci nécessite néanmoins encore un long travail de programmation et de validation pour obtenir des programmes d'aussi bonne qualité et aussi complets que les logiciels spécifiques actuels.

Bibliographie

- [1] T. A. ACKERMAN. Using multidimensional Item Response Theory to understand what items and tests are measuring. *Applied Measurement in Education*, 7(4) : 255–278, 1994.
- [2] R. J. ADAMS, M. R. WILSON, and W. WANG. The multidimensional random coefficient multinomial logit model. *Applied Psychological Measurement*, 21 : 1–23, 1997.
- [3] R. J. ADAMS, M. R. WILSON, and M. L. WU. Multilevel item response models : an approach to errors in variables regression. *Journal of educational and behavioral Statistics*, 22 : 46–73, 1997.
- [4] H. AKAIKE. A new look at the statistical model identification. *IEEE transactions on automatic control*, AC-19 : 716–723, 1974.
- [5] ANAES. Conférence de consensus. La crise suicidaire : reconnaître et prendre en charge, 2000.
- [6] M. R. ANDERBERG. *Cluster analysis for applications*. Academic Press, Inc, New-York, 1973.
- [7] E. B. ANDERSEN. Asymptotic properties of conditional maximum likelihood estimators. *Journal of the Royal Statistical Society Series B*, 32 : 283–301, 1970.
- [8] E. B. ANDERSEN. A goodness of fit test for the Rasch model. *Psychometrika*, 38(1) : 123–140, 1973.
- [9] E. B. ANDERSEN. Sufficient statistics and latent trait models. *Psychometrika*, 42(1) : 69–81, 1977.
- [10] E. B. ANDERSEN. Latent trait models and ability parameter estimation. *Applied Psychological Measurement*, 6(4) : 445–461, 1982.
- [11] E. B. ANDERSEN. *Polytomous Rasch models and their estimations*, chapter 15, pages 271–292. Springer-Verlag, New-York, 1995. In G. H. FISHER and I. W. MOLENAAR : Rasch Models, Foundations, Recent Developments and Applications.
- [12] E. B. ANDERSEN. *The Rating Scale Model*, chapter 4, pages 67–84. Springer Verlag, New York, 1997. In W. J. VAN DER LINDEN and R. K. HAMBLETON : Handbook of Modern Item Response Theory.
- [13] D. ANDRICH. A rating formulation for ordered response categories. *Psychometrika*, 43 : 561–573, 1978.
- [14] A. BINET and T. H. SIMON. *The development of intelligence in young children*. The training school, Vineland, NJ, 1916.
- [15] A. BIRNBAUM. *Some latent trait models and their use in inferring an examinee's ability*. Reading, MA :Addison-Wesley, 1968. In F. M. LORD and M. R. NOVICK (Eds.) : Statistical theories of mental test.

- [16] R. D. BOCK, R. GIBBONS, and E. MURAKI. Full-information item factor analysis. *Applied Psychological Measurement*, 12(3) : 261–280, 1988.
- [17] D. V. BUDESCU and Y. COHEN and A. BEN-SIMON. A revised modified parallel analysis for the construction of unidimensional item pools. *Applied Psychological Measurement*, 21(3) : 233–252, 1997.
- [18] G. CAMILLI. Origin of the scaling constant $d=1.7$, in item response theory. *Journal of Educational and Behavioral Statistics*, 19(3) : 293–295, 1994.
- [19] F. CHASTANG. *Description des facteurs de risque de la récurrence suicidaire. Description quantitative et détermination du poids relatif de ces différents facteurs*. PhD thesis, Université de Nancy 1 - Henri Poincaré, 1998.
- [20] K. B. CHRISTENSEN and J. B. BJORNER. SAS macros for Rasch based latent variable modelling, 2003. Unpublished document.
- [21] A. CHRISTOFFERSSON. Factor analysis of dichotomized variables. *Psychometrika*, 40(1) : 5–32, 1975.
- [22] A. J. CHWALOW, A. LURIE, K. BEAN, I. CHATELET, A. PARENT DU VENOT, D. DUSSEY, Y. DOUÇOT, and G. STRAUCH. A french version of the Sickness Impact Profile (SIP) : stages in the cross cultural validation of a generic quality of life scale. *Fundam. Clin. Pharmacol.*, 6 : 319–326, 1992.
- [23] J. COHEN. A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.*, 20 : 27–46, 1960.
- [24] L. CRONBACH. Coefficient alpha and the internal structure of tests. *Psychometrika*, 16 : 297–334, 1951.
- [25] F. CURT, M. MESBAH, J. LELLOUCH, and G. DELLATOLAS. Handedness scale : how many and wich items ? *Laterality*, 2(2) : 137–154, 1997.
- [26] J. DE LEEUW and N. D. VERHELST. Maximum likelihood estimation in generalized Rasch models. *Journal of Educational Statistics*, 11 : 183–196, 1986.
- [27] A. P. DEMPSTER, N. M. LAIRD, and D. B. RUBIN. Maximum likelihood from incomplete data via EM algorithm. *Journal of the royal statistical society - series B*, 39 : 1–38, 1977.
- [28] A. DIOT, A. MALLET, and R. BRUNO. Adaptation des quadratures de Gauss-Hermite pour améliorer l'estimation dans les modèles non linéaires à effets mixtes. *Journal de la Société Française de Statistique*, 43(1–2) : 195–204, 2002.
- [29] F. DRASGOW and R. I. LISSAK. Modified Parallel Analysis : a procedure for examining the latent dimensionality of dichotomous scored item responses. *Journal of Applied Psychology*, 68 : 363–373, 1983.
- [30] E. J. DUDEWICZ and S. N. MISHRA. *Modern Mathematical Statistics*. John Wiley & Sons, 1988.
- [31] S. E. EMBRETSON. A multidimensional latent trait model for measuring learning and change. *Psychometrika*, 56(3) : 495–515, 1991.
- [32] S. E. EMBRETSON and S. P. REISE. *Item Response Theory for psychologists*. Lawrence Erlbaum, Mahwah, NJ, 2000.
- [33] L. FAHRMEIR and G. TUTZ. *Multivariate statistical modelling based on generalized linear models*. Springer Verlag, New-York, 2001.

- [34] M.-L. FEDDAG. *Estimation par l'approche GEE pour le modèle logistique mixte multi-dimensionnel. Applications aux modèles de Rasch*. PhD thesis, Université de Bretagne Sud, 2003.
- [35] M. L. FEDDAG, I. GRAMA, and M. MESBAH. Generalized Estimating Equations (GEE) for mixed logistic models. *Communications in Statistics - Theory and methods*, 32(4) : 851–874, 2003.
- [36] G. H. FISHER. *Derivations of the Rasch model*, chapter 2, pages 15–38. Springer-Verlag, New-York, 1995. In G. H. FISHER and I. W. MOLENAAR : Rasch Models, Foundations, Recent developments and Applications.
- [37] J. FINE. Problèmes d'indétermination en analyse en facteurs et analyse en composantes principales optimale. *Revue de Statistique Appliquée*, 16(4) : 45–72, 1993.
- [38] G. H. FISCHER. On the existence and uniqueness of maximum likelihood estimates of the Rasch model. *Psychometrika*, 46 : 59–77, 1981.
- [39] G. H. FISHER. *The derivation of polytomous Rasch models*, chapter 16, pages 293–305. Springer-Verlag, New-York, 1995. In G. H. FISHER and I. W. MOLENAAR : Rasch Models, Foundations, Recent Developments and Applications.
- [40] G. H. FISHER. *The One Parameter Logistic Model*, chapter 12, pages 215–238. Springer-Verlag, New-York, 1995. In G. H. FISHER and I. W. MOLENAAR : Rasch Models, Foundations, Recent Developments and Applications.
- [41] G. H. FISHER and I. W. MOLENAAR. *Rasch Models, Recent developments and Applications*. Springer-Verlag, New-York, 1995.
- [42] K. FLOREK, J. LUKASZEWICZ, J. PERKAL, H. STEINHAUS, and S. ZUBRZYCKI. Sur la liaison et la division des points d'un ensemble fini. *Colloquium Mathematicum*, 2 : 282–285, 1951.
- [43] M. GHOSH. Inconsistent maximum likelihood estimators for the Rasch model. *Statistics & Probability Letters*, 23 : 165–170, 1995.
- [44] R. D. GIBBONS and D. R. HEDEKER. Full-information item bi-factor analysis. *Psychometrika*, 57 : 423–436, 1992.
- [45] C. A. W. GLAS. The derivation of some tests for the Rasch model from the multinomial distribution. *Psychometrika*, 53(4) : 525–546, 1988.
- [46] C. A. W. GLAS and J. L. ELIS. *RSP : Rasch Scaling Program*. iec ProGAMMA, Pays-Bas, 1993.
- [47] C. A. W. GLAS and N. D. VERHELST. *The Linear Logistic Test Model*, chapter 8, pages 131–155. Springer-Verlag, New-York, 1995. In G. H. FISHER and I. W. MOLENAAR : Rasch Models, Foundations, Recent Developments and Applications.
- [48] C. A. W. GLAS and N. D. VERHELST. *Testing the Rasch model*, chapter 5, pages 69–96. Springer-Verlag, New-York, 1995. In G. H. FISHER and I. W. MOLENAAR : Rasch Models, Foundations, Recent Developments and Applications.
- [49] J. P. GUILFORD. *Fundamental statistics in psychology and education*. McGraw-Hill, New-York, 1942.
- [50] A. R. GUILMOUR, R. D. ANDERSON, and A. AL. RAE. The analysis of binomial data by a generalized linear mixed model. *Biometrika*, 72 : 593–599, 1985.

- [51] C. GUILPIN. *Manuel de calcul numérique appliqué*. EDP Sciences, Les Ulis, 1999.
- [52] L. GUTTMAN. A basis for scaling qualitative data. *American Sociological Review*, 9 : 139–150, 1944.
- [53] M. A. HAMDAN. The equivalence of tetrachoric and maximum likelihood estimates in 2×2 tables. *Biometrika*, 57 : 212–215, 1970.
- [54] A. HAMON. *Modèle de Rasch et validation de questionnaires de Qualité de Vie*. PhD thesis, Laboratoire SABRES - Université de Bretagne Sud, 2000.
- [55] A. HAMON and S. IOVLEFF. Three stochastic versions of the EM algorithm for estimating longitudinal Rasch model. *Communications in Statistics - Simulation and computation*, 32(2) : 275–295, 2003.
- [56] J.-B. HARDOUIN. A process of items selection in subscales using a parametric multidimensional IRT model. In SMABS, editor, *Actes de la 23rd Biennial Conference of the Society for Multivariate Analysis in the Behavioural Sciences (SMABS) - Tilburg*, pages 68–69, 2 juillet 2002.
- [57] J.-B. HARDOUIN. A process of items selection in subscales using a parametric multidimensional IRT model. In *Actes du 2nd Euro-Japanese Workshop on Stochastic Modelling for Finance, Insurance, Production and Reliability - Chamonix*, pages 128–137, 18 septembre 2002.
- [58] J.-B. HARDOUIN. Éléments de la Théorie de Réponse aux Items à l’aide des logiciels SAS et Stata - Les macro-programmes IRT et MSP. In *Formation SFdS : Qualité de vie - Validation des indices et utilisation dans les essais cliniques*, pages 1–49, 17-18 octobre 2002.
- [59] J.-B. HARDOUIN. Création de scores non pondérés pour la mesure de traits latents. In *séminaire du laboratoire de Probabilité et Statistique de l’Université de Montpellier 2 et de l’ENSAM*, 13 janvier 2003.
- [60] J.-B. HARDOUIN. Classification of items into quality of life sub-scales. In *International Conference on : Statistics in Health Sciences - Nantes*, 24 juin 2004.
- [61] J.-B. HARDOUIN. Manual for the SAS macro-programs %BackRasch and for the Stata module BackRasch, juillet 2004. 8 pages.
- [62] J.-B. HARDOUIN. Manual for the SAS macro-programs %LoevH and %MSP and for the Stata modules LoevH and MSP, juillet 2004. 14 pages.
- [63] J.-B. HARDOUIN. The Rasch analysis : estimations and tests with the Raschtest module, 2005. soumis à The Stata Journal.
- [64] J.-B. HARDOUIN, M.-L. Feddag, and M. MESBAH. Estimation par GEE des paramètres du modèle polytomique multidimensionnel de Kelderman, octobre 2003. 28 pages - document technique non publié.
- [65] J.-B. HARDOUIN and M. MESBAH. Processus de sélection d’échelles d’items unidimensionnelles en phase exploratoire dans le cadre du modèle “objectif” de Rasch multidimensionnel. Technical Report UBS-SABRES/RR-2001-10-FR, Laboratoire SABRES - Université de Bretagne Sud, 2001.
- [66] J.-B. HARDOUIN and M. MESBAH. Construction of QoL subscales using a Multidimensional Item Response Model. In *Actes du Workshop Environment and Health Related Quality of Life : Statistical Perspectives - Vannes*, page 21, 8 novembre 2002.

- [67] J.-B. HARDOUIN and M. MESBAH. Classification of ordinal variables measuring the same latent trait using IRT models and GEE. In *Workshop en l'honneur de Marvin Zelen*. Université de Bordeaux 2, 22 septembre 2003.
- [68] J.-B. HARDOUIN and M. MESBAH. Clustering binary variables in subscales using an extended Rasch model and Akaike Information Criterion. *Communications in Statistics - Theory and Methods*, 33(6) : 1277–1294, 2004.
- [69] J.-B. HARDOUIN and M. MESBAH. Fast Clustering of Binary Variables in Subscales, 2004. en cours de rédaction.
- [70] J.-B. HARDOUIN and M. MESBAH. The SAS macro-program %AnaQol to estimate the parameters of Item Response Models, 2004. soumis à *Communications in Statistics - Simulation and Computation*.
- [71] B. T. HEMKER, K. SIJTSMA, and I. W. MOLENAAR. Selection of unidimensional scales from a multidimensional item bank in the polytomous Mokken IRT model. *Applied Psychological Measurement*, 19(4) : 337–352, 1995.
- [72] H. B. HEYWOOD. On finite sequences of real numbers. *Proceedings of the Royal Society, Series A*, 134 : 486–501, 1931.
- [73] H. HOIJTINK and A. BOOMSMA. *On person parameter estimation in the dichotomous Rasch model*, chapter 4, pages 39–52. Springer-Verlag, New-York, 1995. In G. H. FISHER and I. W. MOLENAAR : *Rasch Models, Foundations, Recent Developments and Applications*.
- [74] H. HOIJTINK, G. ROOKS, and F. W. WILMINK. Confirmatory factor analysis of items with a dichotomous response format using the multidimensional Rasch model. *Psychological Methods*, 4(3) : 300–314, 1999.
- [75] L. G. HUMPHREYS and R. G. MONTANELLI. An investigation of the parallel analysis criterion for determining the number of common factors. *Multivariate Behavioral Research*, 10 : 193–205, 1975.
- [76] SAS Institute Inc. *The FACTOR procedure - SAS/STAT online documentation version 8*. SAS Institute Inc., Carry, N. C., 1999.
- [77] SAS Institute Inc. *The NLMIXED procedure - SAS/STAT online documentation version 8*. SAS Institute Inc., Carry, N. C., 1999.
- [78] SAS Institute Inc. *The VARCLUS procedure - SAS/STAT online documentation version 8*. SAS Institute Inc., Carry, N. C., 1999.
- [79] INSERM. *Estimation par Maximum de Vraisemblance Marginale des paramètres du modèle de Rasch sous le logiciel SAS*, 15 juin 2001.
- [80] P. JACCARD. Nouvelles recherches sur la distribution florale. *Bulletin de la société Vaudoise des sciences naturelles*, 44 : 223–270, 1908.
- [81] J. D. JOBSON. *Principal Components Analysis and Factor Analysis*, chapter 9, pages 345–482. Springer-Verlag, New York, 1991. In *Applied Multivariate Data Analysis*.
- [82] N. L. JOHNSON, S. KOTZ, and N. BALAKRISHNAN. *Continuous Univariate Distributions*. Wiley Interscience, USA, 1994. volume 2.
- [83] B. W. JUNKER. Conditional association, essential independence and monotone unidimensional Item Response Model. *The annals of statistics*, 21(3) : 1359–1378, 1993.

- [84] H. KELDERMAN. Computing maximum likelihood estimates of loglinear models from marginal sums with special attention to loglinear item response theory. *Psychometrika*, 57(3) : 437–450, 1984.
- [85] H. KELDERMAN. Loglinear Rasch model tests. *Psychometrika*, 49(2) : 223–245, 1984.
- [86] H. KELDERMAN. *Loglinear multidimensional Item Response Models for polytomously scored items*, chapter 17, pages 287–304. Springer Verlag, New York, 1997. In W. J. VAN DER LINDEN and R. K. HAMBLETON : *Handbook of Modern Item Response Theory*.
- [87] H. KELDERMAN and C. P. M. RIJKES. Loglinear multidimensional IRT models for polytomously scored items. *Psychometrika*, 59(2) : 149–176, 1994.
- [88] H. KELDERMAN and R. STEEN. *LOGIMO*. IecProgramma, Groningen : Netherlands, 1993.
- [89] D. L. KNOL and M. P. F. BERGER. Empirical comparison between factor analysis and multidimensional Item Response Models. *Multivariate Behavioral Research*, 26(3) : 457–477, 1996.
- [90] S. KREINER. Graphical Rasch models. In INSERM, editor, *Méthodes statistiques pour la mesure d'indicateurs subjectifs en santé - Ateliers de formation 127*, 11-13 juin 2001.
- [91] A. LEPLÈGE. Note technique : invariance et modèles de mesure dans les sciences sociales, introduction à l'oeuvre de Georg Rasch. *Informations sur les Sciences Sociales*, 42(4) : 535–555, 2003.
- [92] K. Y. LIANG and S. L. ZEGER. Longitudinal data analysis using generalized linear models. *Biometrika*, 73 : 13–22, 1986.
- [93] J. M. LINACRE and B. D. WRIGHT. (dichotomous mean-square) chi square fit statistics. *Rasch measurement transactions*, 8(2) : 360, 1994.
- [94] J. LOEVINGER. The technic of homogeneous tests compared with some aspects of scale analysis and factor analysis. *Psychological Bulletin*, 45 : 507–529, 1948.
- [95] F. M. LORD and M. R. NOVICK. *Statistical theories of mental test*. Reading, MA :Addison-Wesley, 1968.
- [96] N. MANTEL and W. HANSZEL. Statistical aspects of the retrospective study of disease. *Journal of the National Cancer Institute*, 22 : 719–748, 1959.
- [97] P. MARTIN-LÖF. *Statistiska modeller [Statistical models]*. PhD thesis, Stockholm : Institutet för Försäkringsmatematik och Matematisk Statistisk vid Stockholms Universitet, 1973.
- [98] G. N. MASTERS. A Rasch model for partial credit scoring. *Psychometrika*, 47 : 149–174, 1982.
- [99] G. N. MASTERS and B. D. WRIGHT. *The Partial Credit Model*, chapter 6, pages 101–122. Springer Verlag, New York, 1997. In W. J. VAN DER LINDEN and R. K. HAMBLETON : *Handbook of Modern Item Response Theory*.
- [100] P. McCULLAGH and J. A. NELDER. *Generalized Linear Models*. Chapman and Hall, 1989. Second edition.
- [101] T. G. MCGINN, G. H. GUYATT, P. C. WYER, C. D. NAYLOR, and W. S. RICHARDSON. Users guides to the medical literature. xxii : How to use articles about clinical decision rules. *JAMA*, 284(1) :79–84, 2000.

- [102] L. L. MCQUITTY. Hierarchical linkage analysis for the isolation of types. *Educational and Psychological Measurement*, 20 : 55–67, 1960.
- [103] L. L. MCQUITTY. Similarity analysis by reciprocal pairs for discrete and continuous data. *Educational and Psychological Measurement*, 26 : 825–831, 1966.
- [104] R. R. MEIJER, K. SIJTSMA, and N. G. SMID. Theoretical and empirical comparison of the Mokken and the Rasch approach to IRT. *Applied Psychological Measurement*, 14(3) : 283–298, 1990.
- [105] M. MESBAH. Modèles et méthodes psychométriques classiques. In INSERM, editor, *Méthodes statistiques pour la mesure d'indicateurs subjectifs en santé - Ateliers de formation 127*, 11-13 juin 2001.
- [106] R. J. MISLEVY. Recent developments in the factor analysis of categorical variables. *Journal of Educational Statistics*, 11 : 3–31, 1986.
- [107] R. J. MOKKEN. *A theory and procedure of scale analysis*. De Gruyter, 1971.
- [108] R. J. MOKKEN. *Nonparametric models for dichotomous responses*, chapter 20, pages 351–368. Springer Verlag, New York, 1997. In W. J. VAN DER LINDEN and R. K. HAMBLETON : *Handbook of Modern Item Response Theory*.
- [109] R. J. MOKKEN and C. LEWIS. A nonparametric approach to the analysis of dichotomous item responses. *Applied Psychological Measurement*, 6(4) : 417–430, 1982.
- [110] R. J. MOKKEN, C. LEWIS, and K. SITJSMA. Rejoinder to 'the Mokken scale : a critical discussion'. *Applied Psychological Measurement*, 10(3) : 279–285, 1986.
- [111] I. W. MOLENAAR. Some improved diagnostics for failure of the Rasch model. *Psychometrika*, 48(1) : 49–72, 1983.
- [112] I. W. MOLENAAR. A weighted Loevinger H-coefficient extending Mokken scaling to multicategory items. *Kwantitatieve Methoden*, 12(37) : 97–117, 1991.
- [113] I. W. MOLENAAR. *Estimation of Item Parameters*, chapter 3, pages 39–52. Springer-Verlag, New-York, 1995. In G. H. FISHER and I. W. MOLENAAR : *Rasch Models, Foundations, Recent Developments and Applications*.
- [114] I. W. MOLENAAR. *Some Background for Item Response Theory and the Rasch Model*, chapter 1, pages 3–14. Springer-Verlag, New-York, 1995. In G. H. FISHER and I. W. MOLENAAR : *Rasch Models, Foundations, Recent Developments and Applications*.
- [115] I. W. MOLENAAR. *Nonparametric models for polytomous responses*, chapter 21, pages 369–380. Springer Verlag, New York, 1997. In W. J. VAN DER LINDEN and R. K. HAMBLETON : *Handbook of Modern Item Response Theory*.
- [116] I. W. MOLENAAR. Parametric and nonparametric item response theory models in health related quality of life measurement. In INSERM, editor, *Méthodes statistiques pour la mesure d'indicateurs subjectifs en santé - Ateliers de formation 127*, 11-13 juin 2001.
- [117] I. W. MOLENAAR, K. SIJTSMA, and P. BOER. *MSP5 for Windows : A program for Mokken scale analysis for polytomous items - version 5.0*. iec ProGAMMA, Groningen, The Netherlands, 2000.
- [118] B. MUTHÉN. Contributions to factor analysis of dichotomous variables. *Psychometrika*, 43(4) : 551–560, 1978.

- [119] C. OMNES and J.-B. HARDOUIN. Belote et rebelote, est-il possible de prédire avant le dix de der? In *Impact du suicide sur l'entourage, les soignants et les institution - Nancy*. XXXVèmes Journées du Groupement d'Étude sur la Prévention du Suicide (GEPS), 20 novembre 2003.
- [120] C. OMNES and J.-B. HARDOUIN. Belote et rebelote, est-il possible de prédire avant le dix de der? *Revue Française de psychiatrie et de psychologie médicale*, 76(7) : 27–32, 2004.
- [121] K. PEARSON. Mathematical contributions to the theory of evolution, vii : on the correlation of characters not quantitatively measurable. *Philos. Trans. R. Soc. Lond. A.*, 195 : 1–47, 1901.
- [122] J. PFANZAGL. On the consistency of conditionnal maximum likelihood estimators. *Annals of the Institute of Statistical Mathematics*, 45(4) : 703–719, 1993.
- [123] S. RABE-HESKETH, A. PICKLES, and A. SKRONDAL. The GLLAMM Manual. Technical report, Department of Biostatistics and Computing - Institute of Psychiatry - King's College - University of London, 2001. number 2001/01.
- [124] S. RABE-HESKETH, A. SKRONDAL, and A. PICKLES. Reliable estimation of generalized linear mixed models using adaptive quadrature. *The Stata Journal*, 2(1) : 1–21, 2002.
- [125] S. RABE-HESKETH, A. SKRONDAL, and A. PICKLES. Maximum likelihood estimation of generalized linear models with covariate measurement error. *The Stata Journal*, 3(4) :385–410, 2003.
- [126] G. RASCH. *Probabilistic models for some intelligence and attainment tests*. The Danish Institute of Educational Research, Copenhagen, 1960.
- [127] G. RASCH. *On general laws and the meaning of measurement in Psychology*, chapter IV, pages 321–333. University of California Press, Berkeley, CA, 1961. In Proceedings of the fourth Berkeley symposium on mathematical statistics and probability.
- [128] G. RASCH. *On specific objectivity. An attempt at formalizing the request for generality and validity of scientific statements*, pages 58–94. Munksgaard, Copenhagen, 1977. In M. BLEGVAD : The Danish yearbook of philosophy.
- [129] M. D. RECKASE. The difficulty of test items that measure more than one ability. *Applied Psychological Measurement*, 9 : 401–412, 1985.
- [130] M. D. RECKASE. *A linear logistic multidimensional model for dichotomous item response data*, chapter 16, pages 271–286. Springer Verlag, New York, 1997. In W. J. VAN DER LINDEN and R. K. HAMBLETON : Handbook of Modern Item Response Theory.
- [131] M. D. RECKASE and Robert L. MCKINLEY. The discriminating power of items that measure more than one dimension. *Applied Psychological Measurement*, 15(4) : 361–373, 1991.
- [132] L. A. ROUSSOS, W. F. STOUT, and J. I. MARDEN. Using new proximity measures with hierarchical cluster analysis to detect unidimensionality. *Journal of Educational Measurement*, 35(1) : 1–30, 1998.
- [133] K. SAHMER, E; VIGNEAU, and E. M. QANNARI. A cluster approach to analyze preference data : choice of the number of clusters, 2005. soumis à Food Quality and Preference.
- [134] F. SAMEJIMA. Acceleration model in the heterogeneous case of the general graded response model. *Psychometrika*, 60 : 549–572, 1957.

- [135] G. SAPORTA. *Probabilités, Analyse des données et Statistique*. TECHNIP, Paris, 1990.
- [136] R. R. SOKAL and C. D. MICHENER. A statistical method for evaluating systematic relationships. *University of Kansas Science Bulletin*, 38 : 1409–1438, 1958.
- [137] C. SPEARMAN. General intelligence, objectively determined and measures. *American Journal of Psychology*, 15 : 201–293, 1904.
- [138] C. E. SPEARMAN. *The abilities of man, their nature and measurement*. Macmillan, New York, 1927.
- [139] W. STEGELMANN. Expanding the Rasch model to a general model having more than one dimension. *Psychometrika*, 48(2) : 259–267, 1983.
- [140] W. F. STOUT. A new Item Response Theory modeling approach with applications to unidimensionality assessment and ability estimation. *Psychometrika*, 55(2) : 293–325, 1990.
- [141] A. H. STROUD and D. SECREST. *Gaussian Quadrature Formulas*. Englewood Cliffs NJ, Prentice-Hall, 1966.
- [142] B. F. SUTRADHAR and R. P. RAO. On marginal quasi-likelihood inference in generalized linear mixed models. *Journal of Multivariate Analysis*, 76 : 1–34, 2001.
- [143] L. L. THURSTONE. *Multiple Factor Analysis*. University Chicago Press, Chicago, 1947.
- [144] A. A. H. VAN ABSWOUDE, L. A. VAN DER ARK, and K. SIJTSMA. A comparative study of test data dimensionality assessment procedures under nonparametric IRT models. *Applied Psychological Measurement*, 28(1) : 3–24, 2004.
- [145] A. L. VAN DEN WOLLENBERG. On the Wright-Panchapakesan goodness of fit test for the Rasch model. Technical Report 80-MA-02, K.U. Nijmegen, 1980.
- [146] A. L. VAN DEN WOLLENBERG. Two new test statistics for the Rasch model. *Psychometrika*, 47(2) : 123–140, 1982.
- [147] W. J. VAN DER LINDEN and R. K. HAMBLETON. *Handbook of Modern Item Response Theory*. Springer Verlag, New York, 1997.
- [148] E. VIGNEAU and E. M. QANNARI. Clustering of variables around latent components. *Communications in Statistics - Simulation and Computation*, 32(4) : 1131–1150, 2003.
- [149] E. VIGNEAU, E. M. QANNARI, K. SAHMER, and D. LADIRAY. Classification de variables autour de composantes latentes, 2004. soumis à la Revue de Statistique Appliquée.
- [150] S. E. WHITELY. Multicomponent latent trait models for ability tests. *Psychometrika*, 45(4) : 479–494, 1980.
- [151] B. WRIGHT and N. PANCHAPAKESAN. A procedure for sample-free item analysis. *Educational and Psychological Measurement*, 29 : 23–48, 1969.
- [152] M. L. WU, R. J. ADAMS, and M. R. WILSON. *ConQuest : Generalised item response modelling software*. ACER, Melbourne, Australia, 1997.
- [153] G. U. YULE. On the methods of measuring association between two attributes. *Journal of the Royal Statistical Society*, 75 : 579–652, 1912.
- [154] S. L. ZEGER, K.-Y. LIANG, and P. S. ALBERT. Models for longitudinal data : A generalized estimating equation approach. *Biometrics*, 44 : 1049–1060, 1988.
- [155] J. ZHANG and W. F. STOUT. The theoretical DETECT index of dimensionality and its application to approximate simple structure. *Psychometrika*, 64(2) : 213–249, 1999.

- [156] A. H. ZWINDERMAN. A generalized Rasch model for manifest predictors. *Psychometrika*, 56 : 589–600, 1991.

Quatrième partie

Annexes

Annexe A

Comparaison du M2-PLM et du 5-PAM

Soient la fonction de réponse à l'item j sous le M2-PLM défini par :

$$P(X_{nj} = x_{nj}/\boldsymbol{\theta}_n; \delta_j, \boldsymbol{\alpha}_j) = \frac{\exp \left[1.7x_{nj} \left(\sum_{q=1}^Q \alpha_{jq}\theta_{nq} - \delta_j \right) \right]}{1 + \exp \left[1.7 \left(\sum_{q=1}^Q \alpha_{jq}\theta_{nq} - \delta_j \right) \right]} \quad (\text{A.1})$$

et celle sous le M5-PAM défini par :

$$P(X_{nj} = x_{nj}/\boldsymbol{\theta}_n; \delta_j^*, \boldsymbol{\alpha}_j^*, \xi_j, \gamma_j^{low}, \gamma_j^{up}) = \gamma_j^{low} + \left(\gamma_j^{up} - \gamma_j^{low} \right) \left\{ \frac{\exp \left[1.7x_{nj} \left(\sum_{q=1}^Q \alpha_{jq}^* \theta_{nq} - \delta_j^* \right) \right]}{1 + \exp \left[1.7 \left(\sum_{q=1}^Q \alpha_{jq}^* \theta_{nq} - \delta_j^* \right) \right]} \right\}^{\xi_j} \quad (\text{A.2})$$

Dans cette annexe, nous cherchons les valeurs α_{jq}^* , $\forall j = 1, \dots, J$, $\forall q = 1, \dots, Q$ de manière à ce que la fonction de réponse à l'item j ait sous ces deux modèles la même pente maximale, et que la pente soit maximale pour la même valeur de $\boldsymbol{\theta}_n$.

Soient $\tau_{jq(M2-PLM)}(\boldsymbol{\theta})$ la pente de l'ICC de l'item j sur le trait latent q en $\boldsymbol{\theta}$ sous le M2-PLM, et $\tau_{jq(M5-PAM)}(\boldsymbol{\theta})$ celle sous le M5-PAM, et soient $\tau_{jq(M2-PLM)}^*$ et $\tau_{jq(M5-PAM)}^*$ les valeurs maximales de ces pentes. Nous allons calculer ces pentes maximales.

Sous le M2-PLM, on a

$$\begin{aligned} \tau_{jq(M2-PLM)}(\boldsymbol{\theta}) &= \frac{dP(X_j = 1/\boldsymbol{\theta}; \delta_j, \boldsymbol{\alpha}_j)}{d\theta_q} \\ &= \frac{d}{d\theta_q} \frac{\exp \left[1.7 \left(\sum_{q=1}^Q \alpha_{jq}\theta_q - \delta_j \right) \right]}{1 + \exp \left[1.7 \left(\sum_{q=1}^Q \alpha_{jq}\theta_q - \delta_j \right) \right]} \\ &= \frac{1.7\alpha_{jq} \exp \left[1.7 \left(\sum_{q=1}^Q \alpha_{jq}\theta_q - \delta_j \right) \right]}{\left\{ 1 + \exp \left[1.7 \left(\sum_{q=1}^Q \alpha_{jq}\theta_q - \delta_j \right) \right] \right\}^2} \end{aligned} \quad (\text{A.3})$$

Sous le M2-PLM, on sait que la pente est maximée au point d'inflexion des fonctions de réponses aux items, c'est à dire au point où $\sum_{q=1}^Q \alpha_{jq} \theta_q - \delta_j = 0$. Alors

$$\tau_{jq(M2-PLM)}^* = \frac{1.7\alpha_{jq} \exp[1.7 \times 0]}{\{1 + \exp[1.7 \times 0]\}^2} = \frac{1.7\alpha_{jq}}{[1 + 1]^2} = \frac{1.7\alpha_{jq}}{4} \quad (\text{A.4})$$

Sous le M5-PAM, on a

$$\begin{aligned} \tau_{jq(M5-PAM)}(\boldsymbol{\theta}) &= \frac{dP(X_j = 1/\boldsymbol{\theta}; \delta_j^*, \boldsymbol{\alpha}_j^*, \xi_j, \gamma_j^{low}, \gamma_j^{up})}{d\theta_q} \\ &= \frac{d}{d\theta_q} \left[\gamma_j^{low} + (\gamma_j^{up} - \gamma_j^{low}) \left\{ \frac{\exp \left[1.7 \left(\sum_{q=1}^Q \alpha_{jq}^* \theta_q - \delta_j^* \right) \right]}{1 + \exp \left[1.7 \left(\sum_{q=1}^Q \alpha_{jq}^* \theta_q - \delta_j^* \right) \right]} \right\}^{\xi_j} \right] \\ &= 1.7\alpha_{jq}\xi_j (\gamma_j^{up} - \gamma_j^{low}) \left\{ \frac{\exp \left[1.7 \left(\sum_{q=1}^Q \alpha_{jq}^* \theta_q - \delta_j^* \right) \right]}{1 + \exp \left[1.7 \left(\sum_{q=1}^Q \alpha_{jq}^* \theta_q - \delta_j^* \right) \right]} \right\}^{\xi_j - 1} \end{aligned} \quad (\text{A.5})$$

Nous allons maximiser cette pente par rapport à θ_q en considérant le vecteur $\boldsymbol{\theta}_{\bar{q}} = (\theta_r)_{r=1, \dots, Q, r \neq q}$ fixé. Maximiser $\tau_{jq(M5-PAM)}(\boldsymbol{\theta})$ revient à maximiser $\log(\tau_{jq(M5-PAM)}(\boldsymbol{\theta}))$, donc on résoud l'équation :

$$\frac{d \log(\tau_{jq(M5-PAM)}(\boldsymbol{\theta}))}{d\theta_q} = 0 \quad (\text{A.6})$$

Posons $u_j = \sum_{q=1}^Q \alpha_{jq}^* \theta_q - \delta_j^*$, $v_j = \frac{\exp(u_j)}{1 + \exp(u_j)}$ et $u'_{jq} = \frac{du_j}{d\theta_q} = \alpha_{jq}^*$, alors

$$\begin{aligned} \log(\tau_{jq(M5-PAM)}(\boldsymbol{\theta})) &= \log(1.7\alpha_{jq}\xi_j(\gamma_j^{up} - \gamma_j^{low})) \\ &\quad - \log(1 + \exp(u_j)) + \xi_j [u_j - \log(1 + \exp(u_j))] \end{aligned} \quad (\text{A.7})$$

Alors

$$\begin{aligned} \frac{d \log(\tau_{jq(M5-PAM)}(\boldsymbol{\theta}))}{d\theta_q} &= -\frac{u'_{jq} \exp(u_j)}{1 + \exp(u_j)} + \xi_j u'_{jq} - \xi_j \frac{u'_{jq} \exp(u_j)}{1 + \exp(u_j)} \\ &= \alpha_{jq}^* (-v_j + \xi_j - \xi_j v_j) \end{aligned} \quad (\text{A.8})$$

Donc

$$\frac{d \log(\tau_{jq(M5-PAM)}(\boldsymbol{\theta}))}{d\theta_q} = 0 \Leftrightarrow -v_j + \xi_j - \xi_j v_j = 0 \quad (\text{A.9})$$

$$\Leftrightarrow v_j = \frac{\xi_j}{1 + \xi_j} \quad (\text{A.10})$$

$$\Leftrightarrow \exp(u_j) = \xi_j \quad (\text{A.11})$$

$$\Leftrightarrow u_j = \log(\xi_j) \quad (\text{A.12})$$

Alors

$$\begin{aligned} \tau_{jq(M5-PAM)}^* &= 1.7\alpha_{jq}^* \xi_j (\gamma_j^{up} - \gamma_j^{low}) \frac{1}{1 + \xi_j} \left(\frac{\xi_j}{1 + \xi_j} \right)^{\xi_j} \\ &= 1.7\alpha_{jq}^* (\gamma_j^{up} - \gamma_j^{low}) \left(\frac{\xi_j}{1 + \xi_j} \right)^{\xi_j + 1} \end{aligned} \quad (\text{A.13})$$

Donc pour que $\tau_{jq}^*(M2-PLM) = \tau_{jq}^*(M5-PAM)$, on doit avoir

$$\alpha_{jq}^* = \frac{\alpha_{jq}}{4 \left(\gamma_j^{up} - \gamma_j^{low} \right) \left(\frac{\xi_j}{1+\xi_j} \right)^{\xi_j+1}} \quad (\text{A.14})$$

Pour le M2-PLM, la pente est maximisée quand $\delta_j = \sum_{q=1}^Q \alpha_{jq} \theta_q$, tandis que pour le M5-PAM, elle est maximisée quand $\sum_{q=1}^Q \alpha_{jq}^* \theta_q - \delta_j^* = \log(\xi_j)$.

En posant $w_j = 4 \left(\gamma_j^{up} - \gamma_j^{low} \right) \left(\frac{\xi_j}{1+\xi_j} \right)^{\xi_j+1}$, on montre que

$$\sum_{q=1}^Q \alpha_{jq}^* \theta_q = \frac{1}{q_j} \sum_{q=1}^Q \alpha_{jq} \theta_q = \delta_j, \quad (\text{A.15})$$

donc on doit avoir

$$\frac{\delta_j}{w_j} - \delta_j^* = \log(\xi_j), \quad (\text{A.16})$$

soit

$$\delta_j^* = \frac{\delta_j}{w_j} - \log(\xi_j) \quad (\text{A.17})$$

Annexe B

Résultats des simulations

B.1 La procédure Raschfit

		(Pouvoir discriminant des items (l : bas ; m : moyen ; h : haut)) [Taille des deux dimensions simulées]														
Modèle	Corr.	(l;l)		(l;m)			(l;h)			(m;m)		(m;h)			(h;h)	
		[7;7]	[14;7]	[7;7]	[14;7]	[7;14]	[7;7]	[14;7]	[7;14]	[7;7]	[14;7]	[7;7]	[14;7]	[7;14]	[7;7]	[14;7]
2-PLM	0.0	2	3	1	1	1	2	1	2	1	1	1	1	1	1	2
SS	0.2	4	4	1	4	1	1	2	4	1	2	1	2	2	1	1
	0.4	2	4	1	1	3	1	1	1	1	2	1	1	2	1	2
	0.6	4	4	1	2	3	1	2	2	1	2	1	1	1	1	1
	0.8	4	4	4	4	4	1	1	4	4	1	1	1	1	1	1
	1.0	3	1	3	4	3	4	4	4	3	2	4	4	4	2	2
2-PLM	0.0	1	4	4	2	4	1	2	2	1	2	2	2	2	2	1
ASS	0.2	4	4	3	4	2	2	2	3	2	1	2	2	4	4	1
	0.4	4	4	3	4	4	1	3	3	1	1	2	2	4	2	2
	0.6	4	4	2	2	4	2	2	4	4	2	4	2	4	2	1
	0.8	4	4	4	4	1	1	2	4	4	4	4	2	4	4	4
	1.0	2	3	3	3	3	3	4	3	1	2	3	4	3	2	3
5-PAM	0.0	4	4	2	4	4	1	2	4	4	3	3	1	2	1	1
SS	0.2	4	4	4	4	4	1	1	2	4	2	1	2	3	2	2
	0.4	4	4	1	4	1	1	1	2	4	2	1	4	4	1	1
	0.6	4	4	1	4	4	1	2	4	4	4	2	3	1	1	4
	0.8	4	4	4	4	4	1	1	4	4	4	1	1	2	4	4
	1.0	2	2	3	3	3	4	4	3	1	2	4	3	3	1	2
5-PAM	0.0	4	4	4	4	3	2	4	2	2	4	2	2	1	1	1
ASS	0.2	4	4	4	4	3	2	2	2	4	4	2	2	2	1	2
	0.4	4	4	4	4	1	1	3	2	2	4	1	2	2	1	2
	0.6	4	4	4	4	4	2	3	2	4	4	2	2	4	1	2
	0.8	4	4	4	4	4	2	2	2	4	4	2	2	1	4	4
	1.0	2	3	3	3	3	4	4	4	2	2	2	3	2	1	2

TAB. B.1 – Résultat des 360 simulations avec la procédure Raschfit

B.2 La procédure Raschfit-Fast

		(Pouvoir discriminant des items (l : bas ; m : moyen ; h : haut))														
		[Taille des deux dimensions simulées]														
Modèle	Corr.	(l;l)		(l;m)			(l;h)			(m;m)		(m;h)			(h;h)	
		[7;7]	[14;7]	[7;7]	[14;7]	[7;14]	[7;7]	[14;7]	[7;14]	[7;7]	[14;7]	[7;7]	[14;7]	[7;14]	[7;7]	[14;7]
2-PLM	0.0	3	3	3	1	1	3	1	3	1	1	1	2	1	1	1
SS	0.2	1	3	4	1	1	1	1	3	2	2	1	1	1	1	1
	0.4	4	4	1	1	2	3	1	2	3	2	1	1	1	1	1
	0.6	4	4	3	3	3	2	1	1	2	4	2	2	4	1	1
	0.8	4	4	3	4	4	3	1	2	4	4	2	1	1	4	4
	1.0	2	1	3	2	2	4	4	4	3	2	4	4	3	1	1
2-PLM	0.0	4	1	4	2	4	3	4	3	3	3	3	2	4	3	2
ASS	0.2	4	4	4	4	3	4	3	3	3	3	3	3	4	4	3
	0.4	4	4	4	4	3	2	2	4	2	2	3	3	3	3	3
	0.6	4	4	4	4	4	2	2	4	4	4	4	4	4	4	3
	0.8	4	4	4	4	4	4	4	4	4	4	4	2	4	4	4
	1.0	3	3	3	3	3	3	4	3	3	1	3	3	2	2	2
5-PAM	0.0	4	4	4	4	4	3	2	1	4	1	1	1	2	1	1
SS	0.2	4	4	4	4	4	3	3	3	4	4	1	4	4	1	2
	0.4	4	4	4	4	4	2	4	4	4	4	2	2	1	1	1
	0.6	4	4	4	4	4	1	2	3	4	4	1	2	1	4	4
	0.8	4	4	4	4	4	1	1	1	4	4	4	4	4	4	4
	1.0	1	2	2	2	2	4	3	4	1	1	2	3	2	1	2
5-PAM	0.0	4	4	4	4	4	3	4	3	3	4	3	2	4	3	3
ASS	0.2	4	4	4	4	4	2	2	3	4	4	2	4	4	2	3
	0.4	4	4	4	4	4	2	4	3	4	4	3	3	4	3	3
	0.6	4	4	4	4	4	3	3	1	4	4	2	4	4	3	4
	0.8	4	4	4	4	4	4	3	4	4	4	4	4	4	4	4
	1.0	3	2	3	3	2	3	3	3	2	3	3	2	3	1	1

TAB. B.2 – Résultat des 360 simulations avec la procédure Raschfit-Fast

B.3 La procédure BackRasch

		(Pouvoir discriminant des items (l : bas ; m : moyen ; h : haut))														
		[Taille des deux dimensions simulées]														
Modèle	Corr.	(l;l)		(l;m)			(l;h)			(m;m)		(m;h)			(h;h)	
		[7;7]	[14;7]	[7;7]	[14;7]	[7;14]	[7;7]	[14;7]	[7;14]	[7;7]	[14;7]	[7;7]	[14;7]	[7;14]	[7;7]	[14;7]
2-PLM	0.0	4	4	1	4	2	3	4	2	4	2	1	2	2	5	1
SS	0.2	4	2	1	4	2	1	4	4	4	2	2	4	2	2	1
	0.4	4	4	2	4	4	4	4	4	4	2	2	4	2	1	2
	0.6	4	4	4	4	4	2	4	2	4	2	1	4	1	2	1
	0.8	4	4	1	4	2	1	2	2	4	4	2	4	2	2	2
	1.0	2	3	4	3	4	4	4	4	1	2	4	4	4	1	2
2-PLM	0.0	4	4	4	4	4	3	5	3	4	3	5	4	4	5	3
ASS	0.2	4	4	4	4	4	4	5	4	4	4	4	4	4	5	3
	0.4	4	4	2	4	4	4	4	4	4	3	4	4	5	5	5
	0.6	4	4	4	4	4	4	4	4	4	4	5	4	4	4	5
	0.8	4	4	4	4	3	4	4	4	4	4	4	4	4	4	5
	1.0	3	2	3	3	3	5	5	3	3	3	5	3	5	3	3
5-PAM	0.0	4	4	1	4	1	4	4	4	4	2	2	4	1	2	2
SS	0.2	4	4	1	4	2	1	4	1	4	4	2	4	1	1	1
	0.4	4	4	4	4	1	1	2	1	4	4	4	4	4	1	1
	0.6	4	4	1	4	1	1	2	1	4	4	1	4	1	1	1
	0.8	4	4	1	2	4	2	1	2	4	4	2	2	3	4	4
	1.0	1	1	4	3	4	4	4	4	2	2	4	4	4	3	2
5-PAM	0.0	4	4	4	4	4	4	4	4	4	4	4	5	4	5	3
ASS	0.2	4	4	4	4	4	4	4	4	4	4	4	4	4	5	5
	0.4	4	4	4	4	4	4	4	4	4	4	4	4	4	5	5
	0.6	4	4	4	4	4	5	4	5	4	4	4	4	4	4	5
	0.8	4	4	4	4	4	2	5	4	4	4	5	4	4	4	4
	1.0	3	3	3	3	3	5	4	4	2	3	4	4	3	3	3

TAB. B.3 – Résultat des 360 simulations avec la procédure BackRasch

B.4 La procédure MSP (c=0.3)

		(Pouvoir discriminant des items (l : bas ; m : moyen ; h : haut))														
		[Taille des deux dimensions simulées]														
Modèle	Corr.	(l;l)		(l;m)			(l;h)			(m;m)		(m;h)			(h;h)	
		[7;7]	[14;7]	[7;7]	[14;7]	[7;14]	[7;7]	[14;7]	[7;14]	[7;7]	[14;7]	[7;7]	[14;7]	[7;14]	[7;7]	[14;7]
2-PLM	0.0	5	5	5	5	5	5	5	5	2	2	1	1	2	1	1
SS	0.2	5	5	5	5	5	5	5	5	1	2	2	2	2	1	1
	0.4	5	5	5	5	5	5	5	5	2	2	1	3	2	1	1
	0.6	5	5	5	5	5	5	5	5	2	1	4	4	4	4	4
	0.8	4	4	5	4	5	5	4	4	2	3	4	4	4	4	4
	1.0	5	5	5	5	5	5	5	2	2	2	1	1	1	1	1
2-PLM	0.0	5	5	5	5	5	5	5	5	5	5	5	5	5	1	1
ASS	0.2	5	5	5	5	5	5	5	5	5	5	5	5	5	1	1
	0.4	5	5	5	5	5	5	5	5	5	5	5	5	5	1	1
	0.6	5	5	5	5	5	5	5	5	5	5	5	5	5	1	1
	0.8	5	5	5	5	5	5	5	5	5	5	5	5	5	4	4
	1.0	5	5	5	5	5	5	5	5	3	3	3	3	3	1	1
5-PAM	0.0	4	4	5	5	5	5	5	5	1	2	2	4	1	1	1
SS	0.2	5	4	5	5	5	4	4	4	1	1	4	4	4	1	1
	0.4	4	4	5	4	5	4	4	4	4	4	4	4	4	4	4
	0.6	4	4	5	4	4	4	4	4	4	4	4	4	4	4	4
	0.8	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
	1.0	3	3	2	3	1	2	1	1	1	1	1	1	1	1	1
5-PAM	0.0	5	5	5	5	5	5	5	5	3	5	5	5	5	1	1
ASS	0.2	5	5	5	4	4	5	5	5	5	5	5	5	5	1	1
	0.4	5	5	4	5	5	5	5	5	3	4	5	3	5	1	1
	0.6	5	5	4	5	5	5	5	5	5	4	5	4	3	4	4
	0.8	5	4	5	4	4	4	5	5	4	4	4	4	4	4	4
	1.0	5	3	3	5	5	5	3	3	3	3	1	2	1	1	1

TAB. B.4 – Résultat des 360 simulations avec la procédure MSP (c=0.3)

B.5 La procédure MSP (c=0.2)

		(Pouvoir discriminant des items (l : bas ; m : moyen ; h : haut))															
		[Taille des deux dimensions simulées]															
Modèle	Corr.	(l;l)		(l;m)			(l;h)			(m;m)		(m;h)			(h;h)		
		[7;7]	[14;7]	[7;7]	[14;7]	[7;14]	[7;7]	[14;7]	[7;14]	[7;7]	[14;7]	[7;7]	[14;7]	[7;14]	[7;7]	[14;7]	
2-PLM	0.0	5	5	5	5	3	3	3	5	1	1	1	1	1	1	1	1
SS	0.2	3	3	5	5	3	5	5	5	1	1	1	1	1	1	1	1
	0.4	5	3	3	5	5	5	4	3	1	1	4	4	4	4	4	4
	0.6	3	5	3	3	3	4	4	5	1	4	4	4	4	4	4	4
	0.8	4	4	5	4	5	4	4	4	4	4	4	4	4	4	4	4
	1.0	3	5	3	3	2	1	2	1	1	1	1	1	1	1	1	1
2-PLM	0.0	5	5	5	5	5	5	5	5	1	2	1	2	1	1	1	1
ASS	0.2	4	5	5	5	5	5	5	5	2	2	1	2	3	1	1	1
	0.4	5	5	5	4	5	5	5	5	2	3	1	2	1	1	1	1
	0.6	5	5	5	5	5	5	5	5	1	1	4	4	4	4	4	4
	0.8	5	5	5	4	5	5	5	5	4	3	4	4	4	4	4	4
	1.0	5	3	3	5	3	3	3	2	1	2	1	2	1	1	1	1
5-PAM	0.0	4	4	4	4	3	4	4	3	4	1	4	4	4	1	1	1
SS	0.2	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
	0.4	2	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
	0.6	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
	0.8	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
	1.0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
5-PAM	0.0	5	5	5	5	5	5	5	5	2	2	2	2	1	1	1	1
ASS	0.2	5	5	4	5	4	5	5	5	2	2	4	4	2	1	1	1
	0.4	5	4	4	4	5	4	4	4	4	4	4	4	4	4	4	4
	0.6	4	4	4	5	5	4	4	4	4	4	4	4	4	4	4	4
	0.8	5	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
	1.0	3	3	2	3	3	2	3	3	1	2	1	1	1	1	1	1

TAB. B.5 – Résultat des 360 simulations avec la procédure MSP (c=0.2)

B.6 La procédure HCA/CCPROX

		(Pouvoir discriminant des items (l : bas ; m : moyen ; h : haut))														
		[Taille des deux dimensions simulées]														
Modèle	Corr.	(l;l)		(l;m)			(l;h)			(m;m)		(m;h)			(h;h)	
		[7;7]	[14;7]	[7;7]	[14;7]	[7;14]	[7;7]	[14;7]	[7;14]	[7;7]	[14;7]	[7;7]	[14;7]	[7;14]	[7;7]	[14;7]
2-PLM	0.0	1	2	1	2	1	1	1	1	1	1	1	1	1	1	1
SS	0.2	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1
	0.4	2	2	1	2	2	4	4	4	1	1	1	3	2	1	1
	0.6	1	2	4	2	3	4	4	4	1	2	3	4	3	1	1
	0.8	4	4	4	4	4	4	4	4	4	4	4	4	4	1	2
	1.0	2	3	3	3	2	3	3	3	2	2	3	3	3	1	1
2-PLM	0.0	2	4	2	2	2	1	2	1	1	1	1	1	1	1	1
ASS	0.2	2	4	2	2	1	3	4	1	1	1	1	1	1	1	1
	0.4	2	4	3	4	2	4	4	1	1	1	1	2	1	1	1
	0.6	4	4	4	4	4	4	4	4	1	1	2	4	2	2	1
	0.8	4	4	4	4	2	4	4	4	2	4	4	4	4	4	2
	1.0	3	3	3	3	2	3	3	2	2	2	2	3	2	2	3
5-PAM	0.0	4	4	4	4	2	4	4	4	1	2	1	2	1	1	1
SS	0.2	4	4	4	4	4	4	4	4	1	2	2	4	2	1	1
	0.4	4	4	4	4	4	4	4	4	4	4	3	4	4	1	1
	0.6	4	4	4	4	4	4	4	4	4	4	4	4	4	1	1
	0.8	4	4	4	4	4	4	4	4	4	4	4	4	4	2	2
	1.0	2	2	3	3	2	3	3	3	2	2	3	3	3	1	2
5-PAM	0.0	4	4	3	4	4	4	4	4	1	2	2	2	4	1	1
ASS	0.2	4	4	4	4	4	4	4	4	2	4	2	4	4	1	1
	0.4	4	4	4	4	4	4	4	4	2	4	2	4	4	1	1
	0.6	4	4	4	4	4	4	4	4	4	4	4	4	4	1	2
	0.8	4	4	4	4	4	4	4	4	4	4	4	4	4	2	2
	1.0	3	2	3	3	3	3	3	3	3	2	3	3	2	3	3

TAB. B.6 – Résultat des 360 simulations avec la procédure HCA/CCPROX

B.7 La procédure AFCS-Varimax

		(Pouvoir discriminant des items (l : bas ; m : moyen ; h : haut)) [Taille des deux dimensions simulées]														
Modèle	Corr.	(l;l)		(l;m)			(l;h)			(m;m)		(m;h)			(h;h)	
		[7;7]	[14;7]	[7;7]	[14;7]	[7;14]	[7;7]	[14;7]	[7;14]	[7;7]	[14;7]	[7;7]	[14;7]	[7;14]	[7;7]	[14;7]
2-PLM	0.0	2	3	1	1	3	1	1	1	1	1	1	1	1	1	1
SS	0.2	2	3	1	3	1	1	3	2	1	1	1	1	1	1	1
	0.4	2	3	1	3	2	1	3	2	1	1	1	1	3	1	3
	0.6	3	3	1	3	4	2	4	2	1	3	1	1	3	1	3
	0.8	4	4	4	4	4	4	4	4	1	3	4	4	3	1	3
	1.0	3	3	2	3	3	2	3	3	3	3	3	1	3	3	3
2-PLM	0.0	3	3	3	3	3	3	3	3	1	1	1	2	3	1	1
ASS	0.2	4	4	3	3	1	2	3	3	1	1	1	2	3	1	1
	0.4	3	4	3	3	2	2	3	3	1	1	1	2	3	1	3
	0.6	3	4	4	4	3	2	4	3	1	3	3	2	3	3	3
	0.8	4	4	4	4	4	4	4	4	3	3	4	4	4	4	3
	1.0	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
5-PAM	0.0	4	4	1	4	3	1	4	1	1	1	1	2	1	1	1
SS	0.2	4	4	4	4	4	4	3	1	1	1	1	1	1	1	1
	0.4	4	4	4	4	4	4	4	4	1	3	1	1	3	1	3
	0.6	4	4	4	4	4	4	4	4	4	4	3	1	4	1	3
	0.8	4	4	4	4	4	4	4	4	4	4	4	4	4	3	3
	1.0	3	3	2	3	3	1	1	3	3	3	3	3	3	3	3
5-PAM	0.0	4	4	2	4	4	4	3	3	1	2	1	2	4	1	1
ASS	0.2	4	4	4	4	4	2	4	3	1	2	1	2	4	1	1
	0.4	4	4	4	4	4	4	4	4	1	3	2	3	4	1	3
	0.6	4	4	4	4	4	4	4	4	4	4	4	4	4	1	3
	0.8	4	4	4	4	4	4	4	4	4	4	4	4	4	4	3
	1.0	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3

TAB. B.7 – Résultat des 360 simulations avec la procédure AFCS-Varimax

B.8 La procédure ACP-Varimax

		(Pouvoir discriminant des items (l : bas ; m : moyen ; h : haut))														
		[Taille des deux dimensions simulées]														
Modèle	Corr.	(l;l)		(l;m)			(l;h)			(m;m)		(m;h)			(h;h)	
		[7;7]	[14;7]	[7;7]	[14;7]	[7;14]	[7;7]	[14;7]	[7;14]	[7;7]	[14;7]	[7;7]	[14;7]	[7;14]	[7;7]	[14;7]
2-PLM	0.0	3	4	1	3	3	3	3	3	1	3	1	3	3	1	3
SS	0.2	3	3	3	3	3	2	3	3	1	3	1	3	3	1	3
	0.4	4	4	1	3	3	1	3	3	1	3	1	1	3	1	3
	0.6	4	4	1	3	4	3	3	3	1	3	1	3	3	1	3
	0.8	4	4	3	4	4	4	3	4	1	4	1	4	3	1	3
	1.0	3	3	3	3	3	3	3	3	3	3	1	3	3	1	3
2-PLM	0.0	3	4	4	3	4	3	3	3	1	3	3	3	3	3	3
ASS	0.2	4	4	2	4	4	3	3	3	1	3	3	3	3	3	3
	0.4	3	4	4	3	3	3	3	3	1	3	3	3	3	3	3
	0.6	4	4	4	4	4	3	3	3	1	4	4	2	3	3	3
	0.8	4	4	4	4	4	4	3	4	1	4	4	3	3	4	3
	1.0	3	3	3	3	3	3	4	3	1	3	3	3	3	3	3
5-PAM	0.0	4	4	1	3	3	1	3	3	1	3	1	3	3	1	3
SS	0.2	3	4	4	3	4	1	3	3	1	3	1	3	3	1	3
	0.4	4	4	4	4	4	1	3	4	1	3	1	3	3	1	3
	0.6	4	4	4	3	3	4	4	4	4	4	1	3	4	1	3
	0.8	4	4	4	4	4	4	4	4	4	4	4	4	4	4	3
	1.0	3	3	1	3	3	1	3	3	1	3	1	3	3	1	3
5-PAM	0.0	4	4	4	4	4	2	3	3	4	4	1	3	4	1	3
ASS	0.2	4	4	3	4	4	2	3	3	4	4	1	3	4	1	3
	0.4	4	4	4	3	4	2	4	3	1	4	3	3	4	1	3
	0.6	4	4	4	4	4	4	4	4	4	4	3	3	4	1	3
	0.8	4	4	4	4	4	4	4	4	4	4	4	4	4	1	3
	1.0	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3

TAB. B.8 – Résultat des 360 simulations avec la procédure ACP-Varimax

Annexe C

Sorties des programmes

C.1 Les données

Dans les exemples de sorties suivantes, nous utiliserons les données de trois sous-échelles du SIP : "Communication" (9 items : c1 à c9), "Comportement émotionnel" (9 items : eb1 à eb9) et "Mobilité" (10 items : m1 à m10).

C.2 Le macro-programme SAS %AnaQol

```
%anaqol(dataset=essai,items=eb1-eb9,alpha=yes,sbsalpha=yes, model=rasch);
```

Indexes	Values
Average Inter-Items Correlation	0.104731
Cronbach alpha	0.512871

Cronbach alpha

Item	Item Test Correlation	Item Rest-score Correlation	Average	Cronbach
			Inter-items Correlations without the item	alpha without the item
EB1	0.53028	0.31413	0.09381	0.45301
EB2	0.46377	0.21248	0.10631	0.48761
EB3	0.48404	0.22612	0.10459	0.48306
EB4	0.23286	0.05020	0.12768	0.53936
EB5	0.43652	0.22539	0.10468	0.48331
EB6	0.39291	0.14328	0.11520	0.51019
EB7	0.51160	0.31074	0.09422	0.45419
EB8	0.50740	0.30433	0.09499	0.45642
EB9	0.49533	0.25416	0.10110	0.47363

Cronbach alpha step-by-step procedure

Deleted Items	Number of Items	Maximal Cronbach Alpha
	9	0.512871
EB4	8	0.539361
EB6	7	0.542532
EB3	6	0.531448
EB5	5	0.504783
EB7	4	0.479989
EB2	3	0.472456
EB9	2	0.486649
EB1	.	.
EB8	.	.

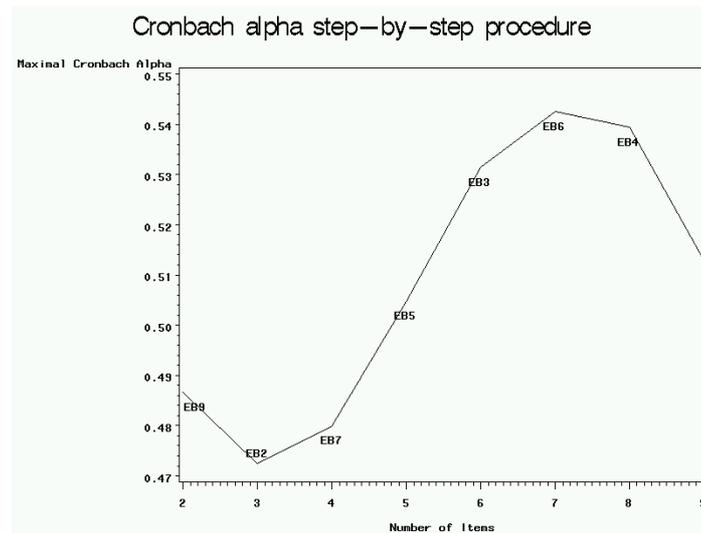


FIG. C.1 – Procédure pas à pas du alpha de Cronbach sur les items eb1 à eb9

Le coefficient alpha de Cronbach mesuré sur l'ensemble de l'échelle " comportement émotionnel" est très médiocre (0.51). La suppression de l'item eb4 permet de l'augmenter un peu (0.54). En dehors de cet item problématique (que nous avons d'ailleurs classé à part des autres items de cette échelle dans notre analyse du SIP), les autres items semblent former une échelle plutôt médiocre (le coefficient de Loevinger de l'échelle ainsi formée est par exemple de seulement 0.19).

Rasch model

Description	Value
-2 Log Likelihood	4682.8
AIC (smaller is better)	4702.8
AICC (smaller is better)	4702.9
BIC (smaller is better)	4743.9

Rasch model

Items	Difficulty parameters	Standard error of the difficulty parameters	Q1 Statistics	Degrees of freedom of the Q1 statistics	p-value of the Q1 statistics
EB1	-0.54494	0.10310	4.9675	7	0.66392
EB2	-0.09908	0.09943	5.2996	7	0.62346
EB3	0.23788	0.09906	10.0034	7	0.18838
EB4	2.20076	0.13988	12.6913	7	0.08000
EB5	-1.16349	0.11444	6.2758	7	0.50794
EB6	0.98271	0.10552	9.0701	7	0.24766
EB7	-1.23244	0.11619	8.5040	7	0.29025
EB8	-1.24656	0.11656	12.0016	7	0.10051
EB9	-0.43540	0.10186	5.5501	7	0.59314
GlobalQ1	.	.	66.1009	56	0.16738
Variance	0.66340	0.09787	.	.	.

Bien que l'échelle "comportement émotionnelle" semble médiocre, l'adéquation au modèle de Rasch est plutôt correcte : l'item eb4 est celui qui a la plus mauvaise adéquation au modèle de Rasch, mais le test Q1 qui lui est associé est peu significatif ($p=0.08$). Ceci peut expliquer que les procédures basées sur le modèle de Rasch (Raschfit, Raschfit-Fast, et BackRasch) étaient celles qui séparaient le moins les items de cette échelle (bien que les conclusions sur cette échelle ne soient pas identiques avec les trois échelles).

C.3 Le programme Stata traces

```
. traces c* , restscore ci test logistic repfiles(c:\graphs)
restscorefiles(restscores) logisticfile(logistic)
```

L'item c1 (figure C.2) a une trace plutôt plate, tandis que la trace l'item c4 (figure C.3) est plutôt accentuée. La figure C.4 montre bien de fortes différences entre les pentes des traces logistiques : le modèle de Rasch ne semble pas opportun dans une telle échelle.

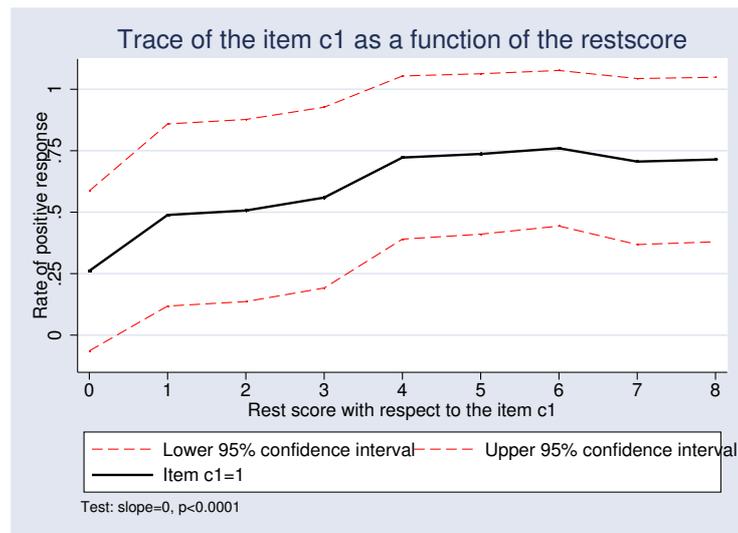


FIG. C.2 – Trace de l'item c1

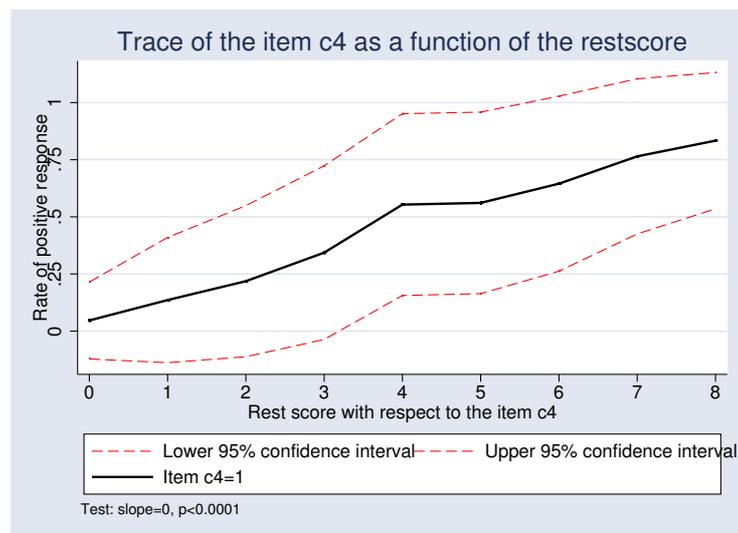


FIG. C.3 – Trace de l'item c4

C.4 Le programme Stata raschtest

Nous allons utiliser les 9 items de l'échelle "communication" du SIP. Les scores 7 et 8 sont regroupés car ils concernent peu d'individus (38 au total sur ces deux scores). L'estimation est réalisée dans un premier temps par maximum de vraisemblance conditionnelle, puis dans un second temps par maximum de vraisemblance marginale.

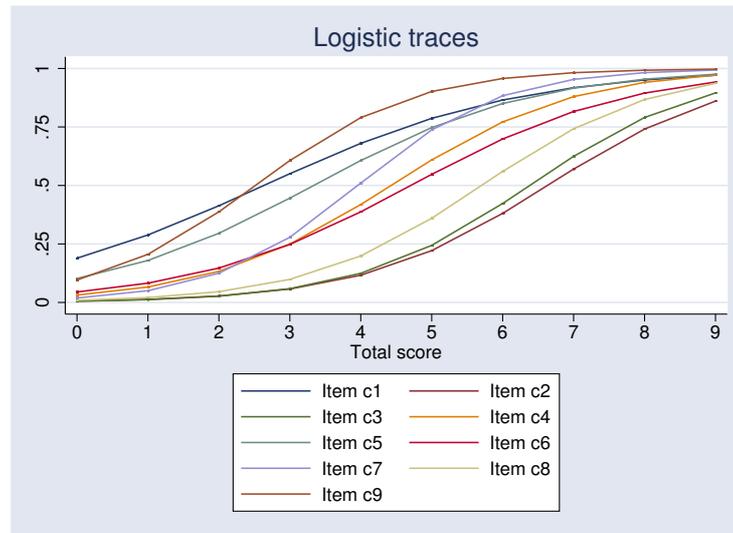


FIG. C.4 – Traces logistiques des items c1 à c9

```
. raschtest c1-c9, autogroup mean
```

Estimation method: Conditional maximum likelihood (CML)

Number of items: 9

Number of groups: 9 (7 of them are used to compute the statistics of test)

Number of individuals: 385 (11 individuals removed for missing values)

Conditional log-likelihood: -1202.5759

Log-likelihood: -2258.6276

Items	Difficulty parameters	std Err.	R1c	df	p-value	Outfit	Infit	U
c1	-1.23053	0.10870	25.411	6	0.0003	1.464	1.213	5.347
c2	1.38114	0.14085	4.979	6	0.5465	0.767	0.881	-1.091
c3	1.27237	0.13721	5.736	6	0.4534	0.812	0.820	-1.232
c4	-0.02176	0.11110	2.911	6	0.8200	0.890	0.932	-0.727
c5	-0.77800	0.10734	6.839	6	0.3360	1.191	1.071	3.032
c6	0.06070	0.11197	4.541	6	0.6038	1.079	1.028	1.488
c7	-0.23523	0.10929	16.854	6	0.0098	0.713	0.812	-3.111
c8	0.83095	0.12501	10.652	6	0.0998	0.757	0.851	-1.257
c9	-1.27964	0.10902	3.379	6	0.7600	0.887	0.947	-0.672
R1c test			R1c=	83.197	48	0.0012		
Andersen LR test			Z=	80.042	48	0.0025		

The mean of the difficulty parameters is fixed to 0

Group Score		Ability parameters		std Err.	Freq.	ll
0	0				82	
1	1	-2.39926		0.12764	74	-120.6067
2	2	-1.48907		0.09607	79	-222.1793
3	3	-0.84086		0.09684	63	-219.4348
4	4	-0.27809		0.10651	48	-178.6362
5	5	0.26419		0.10568	49	-202.6461
6	6	0.83209		0.13265	34	-121.2928
7	7	1.48943		0.17555	24	-97.7592
	8	2.41090		0.29488	14	
8	9				5	

Ces premières sorties concernent l'estimation par CML des paramètres du modèle. Le premier tableau présente, pour chaque item, les estimations des paramètres de difficulté de chacun des items, la contribution au test du premier ordre (par défaut, le test R1c), les indices OUTFIT et INFIT et la statistique du test U. En outre, on obtient la statistique du premier ordre, ainsi que le test Z de rapport de vraisemblance d'Andersen. Le second tableau donne les estimations, pour chaque valeur du score (ou regroupement de scores) à l'estimation du trait latent pour les individus ayant ce score. Ce second tableau indique en outre l'effectif de chaque score, et la vraisemblance associée à chaque groupe de scores (utilisée pour le test d'Andersen).

```
. raschtest c1-c9, autogroup method(mml) icc information
```

```
Estimation method: Marginal maximum likelihood (MML)
```

```
Number of items: 9
```

```
Number of groups: 9 (9 of them are used to compute the statistics of test)
```

```
Number of individuals: 472 (11 individuals removed for missing values)
```

```
Marginal log-likelihood: -2206.7816
```

Items	Difficulty		R1m	df	p-value	Outfit	Infit
	parameters	std Err.					
c1	-0.10599	0.13144	26.040	6	0.0002	1.367	1.275
c2	2.51187	0.17318	6.779	6	0.3418	1.282	1.432
c3	2.40141	0.16927	6.901	6	0.3301	1.374	1.348
c4	1.09654	0.13908	4.348	6	0.6297	1.190	1.177
c5	0.34266	0.13228	7.702	6	0.2608	1.247	1.209
c6	1.17910	0.14025	4.300	6	0.6362	1.441	1.283
c7	0.88323	0.13644	18.433	6	0.0052	0.960	1.044
c8	1.95427	0.15587	10.830	6	0.0938	1.193	1.272
c9	-0.15446	0.13150	12.557	6	0.0506	1.017	1.021
R1m test		R1m=	86.786	55	0.0040		
Sigma	1.51695	0.08266					

Group	Score	Ability		Freq.
		parameters	std Err.	
0	0	-1.83819	1.02033	82
1	1	-1.83819	0.78613	74
2	2	-1.01641	0.80462	79
3	3	-0.42959	0.79598	63
4	4	0.26284	0.54585	48
5	5	0.71403	0.60539	49
6	6	1.00437	0.83925	34
7	7	1.55340	0.70397	24
	8	2.19576	0.64902	14
8	9	3.22890	0.92297	5

Ces secondes sorties concernent les estimations par MML des paramètres du modèle. On note cette fois que la contrainte d'identifiabilité ne porte plus sur les paramètres de difficulté des items, mais sur la distribution du trait latent dont la moyenne est fixée à 0. Le test R1m est effectué, à la place du test R1c. Le test d'Andersen n'est pas possible avec ce type d'estimation, aussi il n'est pas effectué. On note un décalage des estimations des paramètres de

difficultés due à la contrainte d'identifiabilité qui est différente (et qui permet de fixer le niveau de référence), mais les écarts entre les paramètres de difficulté sont préservés.

Dans la seconde commande, nous avons demandé l'édition des traces empiriques et théoriques des items, ainsi que le graphique de l'information de l'échelle. Dans la figure C.5, nous avons représentés les traces de l'item 1 (ayant une mauvaise adéquation au modèle) et l'item 4 (ayant une bonne adéquation au modèle). Dans le graphique C.6, nous avons représenté le graphique d'information : l'échelle apporte beaucoup d'information pour les individus ayant une valeur du trait latent comprise entre 0 et 2 (valeurs plutôt favorables), mais peu pour les petites valeurs du trait latent (inférieures à 0).

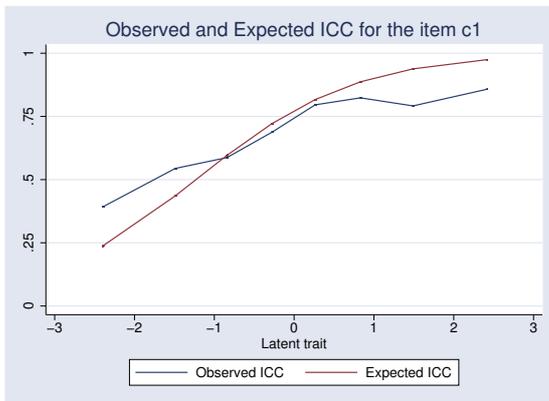


figure C.5a. traces de l'item c1

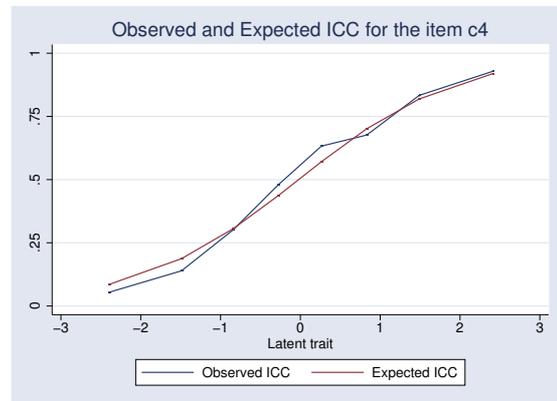


figure C.5b. traces de l'item c4

FIG. C.5 – Courbes caractéristiques des items c1 et c4 sous le modèle de Rasch

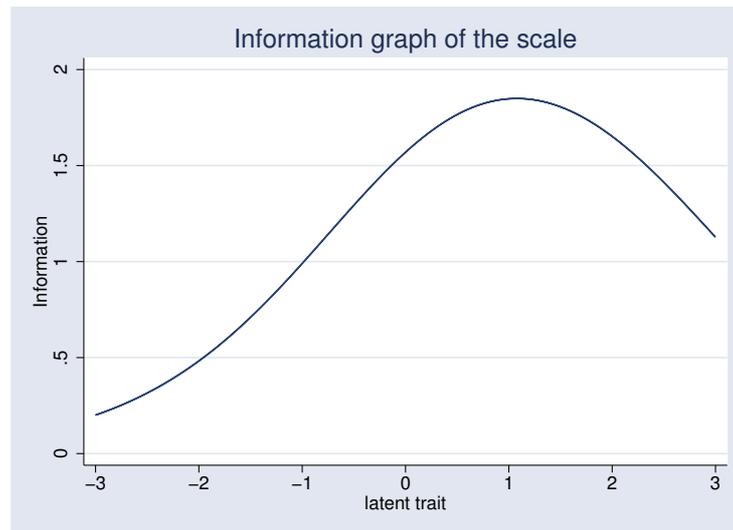


FIG. C.6 – Graphique de l'information des items c1 à c9 sous le modèle de Rasch

C.5 Le programme Stata mmsrm

```
. mmsrm c* eb* ,part(9 9) trace
Estimation of the difficulty parameters of the dimension 1.
Estimation of the difficulty parameters of the dimension 2.

Estimation of the parameters of the distribution of the multidimensional latent trait.
This process could be long to run. Be patient !
```

[...]

Variances and covariances of random effects

 ***level 2 (__000005)

```
var(1): 2.0732417 (.17981448)
cov(2,1): .68730267 (.08594218) cor(2,1): .59049852
var(2): .65344417 (.08806499)
```

 Log-likelihood:-4713.7784

Items	Parameters	std Err.
c1	-0.10599	0.13144
c2	2.51187	0.17318
c3	2.40141	0.16927
c4	1.09654	0.13908
c5	0.34266	0.13228
c6	1.17910	0.14025
c7	0.88323	0.13644
c8	1.95427	0.15587
c9	-0.15446	0.13150

eb1	-0.71235	0.11048
eb2	-0.27305	0.10655
eb3	0.09936	0.10594
eb4	2.03499	0.14568
eb5	-1.26833	0.12069
eb6	0.84476	0.11224
eb7	-1.41382	0.12441
eb8	-1.37321	0.12333
eb9	-0.60700	0.10921

Var1	2.07324	
Var2	0.65344	
cov12	0.68730	

Dans ces sorties, le tableau final permet d'obtenir les estimations des paramètres de difficultés de chacun des items, ainsi que les estimations des éléments de la matrice de variance-covariance des traits latents.

C.6 Le programme Stata `geekel2d`

```
. geekel2d eb1-eb9,l1
```

Initial step

First iteration

```
iteration: 1      Convergence index: .2577922957010083
iteration: 2      Convergence index: .0065367199339846
iteration: 3      Convergence index: .0001482492222776
iteration: 4      Convergence index: 5.31562191982e-07
iteration: 5      Convergence index: 1.03170257558e-08
iteration: 6      Convergence index: 1.07362851303e-10
iteration: 7      Convergence index: 1.58710594818e-12
iteration: 8      Convergence index: 1.98494631355e-14
```

-----RESUME-----

The algorithm converge at the 8th iteration

---ESTIMATIONS---

	Parameters	Standard errors
eb1:	-0.702047	0.109114
eb2:	-0.246575	0.105010
eb3:	0.128134	0.103842
eb4:	2.028494	0.153906
eb5:	-1.221470	0.120592
eb6:	0.863806	0.113395
eb7:	-1.411664	0.124082
eb8:	-1.371421	0.122656
eb9:	-0.570982	0.107606
var1:	0.675080	0.102185
l1:	-2445.3206	
AIC:	4910.6411	

La première partie des sorties concerne l'algorithme d'estimation, le dernier tableau indique les estimations des paramètres, l'estimation de la variance. L'option **ll** permet d'afficher la valeur de la log-vraisemblance maximale et du critère d'Akaike. Ces estimations peuvent être comparées à celles obtenues par maximum de vraisemblance marginale obtenues ci dessous avec le programme **raschtest**.

```
. raschtest eb1-eb9, method(mml) test(none)
```

```
Estimation method: Marginal maximum likelihood (MML)
```

```
Number of items: 9
```

```
Number of groups: 10
```

```
Number of individuals: 470 (13 individuals removed for missing values)
```

```
Marginal log-likelihood: -2445.1507
```

	Difficulty	
Items	parameters	std Err.

eb1	-0.71235	0.11048
eb2	-0.27305	0.10655
eb3	0.09936	0.10594
eb4	2.03499	0.14568
eb5	-1.26833	0.12069
eb6	0.84476	0.11224
eb7	-1.41382	0.12441
eb8	-1.37321	0.12333
eb9	-0.60700	0.10921

Sigma	0.81939	0.05885

C.7 Le macro-programme SAS %BackRasch

```
%backrasch(dataset=essai,items=m1-m10);
```

```
NOTE: The AnaQol macro run without error.
```

```
The item M3 is bad fitted (p= 0), so it is excluded
```

```
NOTE: The AnaQol macro run without error.
```

```
The item M7 is bad fitted (p=0.0005051561), so it is excluded
```

```
NOTE: The AnaQol macro run without error.
```

```
There is no more bad fitted items
```

Dans la fenêtre "log", on retrouve pour chaque échelle construite l'historique de l'algorithme de construction des échelles. Dans la fenêtre "output", les sorties du programmes %AnaQol obtenue à chaque étape de l'algorithme sont disponibles. Dans cet exemple, la procédure supprime les items M3 et M7. M3 est un item qui avait été isolé lors de l'analyse du SIP (avec M4), M7 est un item qui est peu souvent relié à M9 par exemple. L'item M4 n'est pas supprimé et a même plutôt une bonne adéquation au modèle.

Rasch model

Description	Value
-2 Log Likelihood	4300.8
AIC (smaller is better)	4322.8
AICC (smaller is better)	4322.8
BIC (smaller is better)	4368.0

Rasch model

Items	Difficulty parameters	Standard error of the difficulty parameters	Q1 Statistics	Degrees of freedom of the Q1 statistics	p-value of the Q1 statistics
M1	1.11244	0.13103	3.996	8	0.85746
M2	1.75837	0.15277	7.199	8	0.51532
M3	-0.67369	0.11037	104.200	8	0.00000
M4	1.21610	0.13391	5.062	8	0.75093
M5	0.55101	0.11890	5.273	8	0.72809
M6	-1.38748	0.11603	15.446	8	0.05103
M7	3.17766	0.24430	32.829	8	0.00007
M8	0.13045	0.11334	15.757	8	0.04599
M9	-1.15133	0.11329	18.376	8	0.01858
M10	1.95448	0.16135	3.060	8	0.93052
GlobalQ1	.	.	190.078	72	0.00000
Variance	1.64962	0.20048	.	.	.

Rasch model

Description	Value
-2 Log Likelihood	3630.6
AIC (smaller is better)	3650.6
AICC (smaller is better)	3650.7
BIC (smaller is better)	3691.7

Rasch model

Items	Difficulty parameters	Standard error of the difficulty parameters	Q1 Statistics	Degrees of freedom of the Q1 statistics	p-value of the Q1 statistics
M1	1.13304	0.13731	7.0440	7	0.42431
M2	1.81641	0.15920	6.9873	7	0.43021
M4	1.24314	0.14023	6.8456	7	0.44514
M5	0.53242	0.12499	18.7929	7	0.00886
M6	-1.57633	0.12451	5.8472	7	0.55770
M7	3.29264	0.25007	25.9928	7	0.00051
M8	0.07902	0.11937	7.0812	7	0.42047
M9	-1.31763	0.12100	9.7769	7	0.20157
M10	2.02274	0.16780	3.3887	7	0.84687
GlobalQ1	.	.	81.5615	56	0.01450
Variance	2.38496	0.29550	.	.	.

Rasch model

Description	Value
-2 Log Likelihood	3488.5
AIC (smaller is better)	3506.5
AICC (smaller is better)	3506.5
BIC (smaller is better)	3543.4

Rasch model

Items	Difficulty parameters	Standard error of the difficulty parameters	Q1 Statistics	Degrees of freedom of the Q1 statistics	p-value of the Q1 statistics
M1	1.29675	0.13736	7.3279	6	0.29158
M2	1.98487	0.15958	5.6703	6	0.46111
M4	1.40762	0.14033	4.6343	6	0.59149
M5	0.69251	0.12473	6.5466	6	0.36482
M6	-1.42955	0.12356	4.9062	6	0.55590
M8	0.23647	0.11888	10.3560	6	0.11044
M9	-1.16898	0.12005	9.6795	6	0.13881
M10	2.19287	0.16828	4.1326	6	0.65874
GlobalQ1	.	.	46.5967	42	0.28893
Variance	2.46328	0.30844	.	.	.

C.8 Les macro-programmes SAS %LoevH et %MSP

```
proc iml;
%msp(dataset=essai,items=c1 c2 c3 c4 c5 c6 c7 c8 c9 m1 m2 m3 m4 m5 m6 m7 m8 m9 m10);
c=0.3
```

The subscale 1 is composed of 9 items with H=0.4051

SCALE1	Hj
c9	0.4766
c8	0.4198
c7	0.4484
c3	0.4500
c2	0.4132
c4	0.3822
c6	0.3499
m7	0.3411
c5	0.3423

The subscale 2 is composed of 7 items with H=0.4266

SCALE2	Hj
m6	0.5451
m2	0.3909
m9	0.5328
m1	0.3859
m8	0.4164
m5	0.3696
m10	0.3053

The subscale 3 is composed of 2 items with H=0.5252

SCALE3	Hj
m4	0.5252
m3	0.5252

MSP retrouve des ensembles d'items relativement proches des échelles originelles : tous les items de l'échelle "communication" sont classés ensemble, avec l'item m7. 7 des 10 items de "mobilité" sont classés ensembles. Les items "atypiques" sont m3, m4 (tous deux classés à part dans notre analyse du SIP) et m7.

C.9 Le macro-programme SAS %DETECT

```
%detect(dataset=sip,out=out,items=c1-c9 m1-m10,part=9 10);
```

Main indexes	
Description	Indexes
=====	=====
DETECT Index	0.009742
Iss Index	0.555556
R Index	0.801691

Correlations between the items and the sub-scores and sub-rest-scores

Items	Score1	Score2	Rest-Score1	Rest-Score2
C1	0.486544	0.262118	0.297954	0.022884
C2	0.518973	0.157717	0.397490	-.009296
C3	0.561725	0.178994	0.442510	0.006835
C4	0.618276	0.128081	0.470540	-.095802
C5	0.572603	0.338221	0.401848	0.104976
C6	0.560908	0.266263	0.401852	0.046092
C7	0.693885	0.241348	0.561829	0.011781
C8	0.588040	0.177224	0.458543	-.014595
C9	0.627524	0.212893	0.469145	-.027827
M1	0.199234	0.518802	0.027657	0.355934
M2	0.241379	0.476642	0.100261	0.337770
M3	0.128690	0.363237	-.085963	0.130623
M4	0.111508	0.503060	-.056489	0.342241
M5	0.209532	0.510114	0.016567	0.322574
M6	0.204859	0.638040	-.001411	0.466905
M7	0.182721	0.235635	0.099935	0.143962
M8	0.239390	0.609342	0.035266	0.431349
M9	0.213057	0.677017	0.002619	0.513784
M10	0.247462	0.397173	0.115629	0.258921

L'indice Iss permet d'avoir une idée du pourcentage de covariance conditionnelles entre toutes les paires d'items qui vérifient une structure approximativement simple : 56% des covariances conditionnelles vérifient donc ce type de structure. L'indice R permet de mesurer la quantité de covariances conditionnelles qui vérifient cette structure : 80% de la covariance conditionnelle est "expliquée" par une structure approximativement simple. L'indice DETECT est difficile à interpréter seul, il est plus utile pour comparer différentes partitions. Par exemple, en comparant cette partition avec celle qui agrégerait l'item m10 aux items c1 à c9, on obtient DETECT=0.009355 (et aussi de plus petites valeurs de Iss et R). Ainsi, l'item m10 semble être plus proche de l'échelle composée des items m1 à m9 que de celle composée des items c1 à c9.

La table des corrélations entre les items et les scores peut s'utiliser en recherchant pour chaque item le score avec lequel cet item est le mieux corrélé. Afin de faciliter les comparaisons, il peut être judicieux de comparer les corrélations entre chaque item et le rest-score de l'échelle à laquelle l'item est associé (échelle 1 pour les items c1 à c9 et l'échelle 2 pour les items m1 à m10) avec les corrélations entre l'item et le score pour les échelles auxquels les items ne sont pas associés. Par exemple, on voit que la corrélation entre l'item c1 et le rest-score dans l'échelle 1 est de 0.297954 tandis que la corrélation entre ce même item et le score à l'échelle 2 est de 0.262118, donc cet item est mieux intégré dans l'échelle 1 que dans l'échelle 2. On observe ainsi un problème de classement de l'item m7, et plusieurs cas limites (c1, m3 et m10).

C.10 Le programme Stata raschfit

```
. raschfit m*, itemsorder(msp)
SCALE: 1
-----
The program is ordering the items
The kernel of the scale is m2 m6

iteration: 1
-----

The program is estimating the parameters of the models (N=289)
Unidimensional model: ll:      -201.8149   AIC:      413.6299
Bidimensional model: ll:      -207.1991   AIC:      422.3981
The item m9 is selected in the scale 1

iteration: 2
-----

The program is estimating the parameters of the models (N=307)
Unidimensional model: ll:      -536.0562   AIC:      1086.1125
Bidimensional model: ll:      -542.1203   AIC:      1096.2407
The item m1 is selected in the scale 1

iteration: 3
-----

The program is estimating the parameters of the models (N=324)
Unidimensional model: ll:      -604.2995   AIC:      1226.5989
Bidimensional model: ll:      -598.2478   AIC:      1212.4955
The item m7 is not selected in the scale 1

iteration: 4
-----

The program is estimating the parameters of the models (N=329)
Unidimensional model: ll:      -747.2427   AIC:      1512.4855
Bidimensional model: ll:      -757.9466   AIC:      1531.8933
The item m8 is selected in the scale 1

iteration: 5
-----

The program is estimating the parameters of the models (N=337)
Unidimensional model: ll:      -861.5987   AIC:      1745.1975
Bidimensional model: ll:      -857.3082   AIC:      1734.6163
The item m10 is not selected in the scale 1
```

```
iteration: 6
```

```
-----
```

```
The program is estimating the parameters of the models (N=343)
Unidimensional model: ll:      -934.5748   AIC:    1891.1497
Bidimensional model: ll:      -936.9886   AIC:    1893.9772
The item m5 is selected in the scale 1
```

```
iteration: 7
```

```
-----
```

```
The program is estimating the parameters of the models (N=353)
Unidimensional model: ll:     -1119.4353   AIC:    2264.8705
Bidimensional model: ll:     -1119.4904   AIC:    2262.9809
The item m4 is not selected in the scale 1
```

```
iteration: 8
```

```
-----
```

```
The program is estimating the parameters of the models (N=382)
Unidimensional model: ll:     -1300.5777   AIC:    2627.1554
Bidimensional model: ll:     -1185.1846   AIC:    2394.3691
The item m3 is not selected in the scale 1
```

La procédure Raschfit-Fast (classement positif) appliquée aux items de l'échelle "mobilité" rejette les items m3, m4, m7 et m10. On remarque que m3 et m4 n'ont pas été classés avec les autres items de cette échelle dans notre analyse du SIP, quant à m7 et m10, il ne sont classés que rarement avec certains autres items de cette échelle comme m9 : ce sont donc les 4 items qui posent le plus de problèmes.

C.11 Le programme Stata backrasch

```
. backrasch c*
```

```
subscale : 1
```

```
-----
```

```
The item c1 is removed of the scale 1 (p=0.0003)
The item c5 is removed of the scale 1 (p=0.0086)
The item c6 is removed of the scale 1 (p=0.0076)
No more item to remove of the scale 1
```

```
Number of selected items : 6
```

Estimation method: Conditional maximum likelihood (CML)

Number of items: 6

Number of groups: 6 (4 of them are used to compute the statistics of test)

Number of individuals: 301 (8 individuals removed for missing values)

Conditional log-likelihood: -536.6696

Log-likelihood: -1214.0285

	Items	Difficulty parameters	std Err.	R1c	df	p-value	Outfit	Infit	U
	c2	3.03011	0.21803	2.154	3	0.5410	0.961	0.902	0.853
	c3	2.88408	0.21301	0.435	3	0.9329	0.906	0.923	0.511
	c4	1.46000	0.18192	0.721	3	0.8682	0.992	0.978	0.881
	c7	1.25053	0.17934	4.030	3	0.2583	0.811	0.825	-1.816
	c8	2.41521	0.19952	5.806	3	0.1214	0.733	0.872	-1.031
*	c9	0.00000	.	1.127	3	0.7705	0.993	0.978	1.030
R1c test				R1c=	14.583	15	0.4819		
Andersen LR test				Z=	16.007	15	0.3816		

*: The difficulty parameter of this item had been fixed to 0

Group	Score	Ability parameters	std Err.	Freq.	ll
0	0			159	
1	1	-0.12728	0.12055	98	-112.8545
2	2	1.00081	0.11026	78	-163.3927
3	3	1.87885	0.11104	68	-148.6300
4	4	2.73076	0.16231	34	-103.7890
	5	3.77975	0.23956	23	
5	6			15	

On regarde avec ce programme si la première échelle obtenue lors de l'analyse du SIP vérifie un modèle de Rasch. On remarque que les items liés à l'écriture (c1 et c5) sont enlevés en premier. Enfin, l'item c6 est enlevé. Les estimations obtenues avec l'échelle finale sont données.

C.12 Les programmes Stata loevh et msp

Nous allons utiliser les échelles "communication" et "comportement émotionnel"

```
. loevH c1-c9
```

Item	Obs	Easyness P(Xj=1)	Observed Guttman errors	Expected Guttman errors	Loevinger H coeff	z-stat.	p-value
c1	472	0.5191	388	545.22	0.28837	11.6206	0.00000
c2	472	0.1377	197	337.94	0.41706	14.5056	0.00000
c3	472	0.1483	205	359.68	0.43005	15.4702	0.00000
c4	472	0.3157	359	576.19	0.37694	17.0728	0.00000
c5	472	0.4407	387	581.44	0.33441	14.4484	0.00000
c6	472	0.3030	379	566.11	0.33052	14.8534	0.00000
c7	472	0.3496	330	587.82	0.43860	19.8252	0.00000
c8	472	0.1970	254	438.04	0.42014	16.6049	0.00000
c9	472	0.5275	307	537.66	0.42900	17.0594	0.00000
Scale	472		1403	2265.05	0.38059	33.1333	0.00000

Le programme loevH permet d'afficher l'ensemble des coefficients de Loevinger reliés à l'échelle. Tous les coefficients de Loevinger sont supérieur à 0.3 sauf celui correspondant à c1.

```
. msp c1-c9 eb*
```

```
Scale: 1
```

```
-----  
Significance level: 0.000327
```

```
The two first items selected in the scale 1 are c8 and c9 (Hjk=0.7466)
```

```
Significance level: 0.000296
```

```
The item c7 is selected in the scale 1           Hj=0.6035       H=0.6381
```

```
Significance level: 0.000272
```

```
The item c2 is selected in the scale 1           Hj=0.4553       H=0.5638
```

```
Significance level: 0.000253
```

```
The item c3 is selected in the scale 1           Hj=0.4440       H=0.5178
```

```
Significance level: 0.000237
```

```
The item c4 is selected in the scale 1           Hj=0.4491       H=0.4908
```

```
Significance level: 0.000224
```

```
The item eb5 is selected in the scale 1          Hj=0.3950       H=0.4724
```

```
Significance level: 0.000214
```

```
The item eb7 is selected in the scale 1          Hj=0.3248       H=0.4426
```

```
Significance level: 0.000205
```

```
The item c6 is selected in the scale 1           Hj=0.3357       H=0.4130
```

```
Significance level: 0.000198
```

```
None new item can be selected in the scale 1 because all the Hj are lesser than .3
```

```
or none new item had all the related Hjk coefficients significantly greater than 0.
```

Item	Obs	Easyness P(X _j =1)	Observed Guttman errors	Expected Guttman errors	Loevinger H coeff	z-stat.	p-value
c6	462	0.3074	319	480.19	0.33567	13.4513	0.00000
eb7	462	0.7749	191	285.21	0.33032	8.5735	0.00000
eb5	462	0.7532	197	305.10	0.35432	9.5669	0.00000
c4	462	0.3139	286	483.50	0.40848	16.4010	0.00000
c3	462	0.1515	173	318.39	0.45665	15.1815	0.00000
c2	462	0.1385	171	295.76	0.42182	13.4804	0.00000
c7	462	0.3506	257	486.00	0.47119	18.5732	0.00000
c8	462	0.1991	212	380.73	0.44317	15.9788	0.00000
c9	462	0.5281	230	433.64	0.46961	15.9062	0.00000
Scale	462		1018	1734.26	0.41301	29.9767	0.00000

Scale: 2

Significance level: 0.001389

The two first items selected in the scale 2 are eb1 and eb8 (H_{jk}=0.4334)

Significance level: 0.001163

None new item can be selected in the scale 2 because all the H_j are lesser than .3
or none new item had all the related H_{jk} coefficients significantly greater than 0.

Item	Obs	Easyness P(X _j =1)	Observed Guttman errors	Expected Guttman errors	Loevinger H coeff	z-stat.	p-value
eb1	462	0.6494	39	68.83	0.43340	6.9173	0.00000
eb8	462	0.7706	39	68.83	0.43340	6.9173	0.00000
Scale	462		39	68.83	0.43340	6.9173	0.00000

Scale: 3

Significance level: 0.002381

The two first items selected in the scale 3 are eb3 and eb6 (H_{jk}=0.4066)

The following items are excluded at this step: eb2 eb4

Significance level: 0.002083

None new item can be selected in the scale 3 because all the H_j are lesser than .3
or none new item had all the related H_{jk} coefficients significantly greater than 0.

Item	Obs	Easyness P(Xj=1)	Observed Guttman errors	Expected Guttman errors	Loevinger H coeff	z-stat.	p-value
eb3	462	0.4762	46	77.52	0.40663	6.2934	0.00000
eb6	462	0.3203	46	77.52	0.40663	6.2934	0.00000
Scale	462		46	77.52	0.40663	6.2934	0.00000

Scale: 4

Significance level: 0.005000

The two first items selected in the scale 4 are c1 and c5 (Hjk=0.3485)

The following items are excluded at this step: eb4

Significance level: 0.004167

None new item can be selected in the scale 4 because all the Hj are lesser than .3
or none new item had all the related Hjk coefficients significantly greater than 0.

Item	Obs	Easyness P(Xj=1)	Observed Guttman errors	Expected Guttman errors	Loevinger H coeff	z-stat.	p-value
c1	462	0.5260	63	96.70	0.34851	6.3235	0.00000
c5	462	0.4416	63	96.70	0.34851	6.3235	0.00000
Scale	462		63	96.70	0.34851	6.3235	0.00000

Scale: 5

Significance level: 0.016667

None new scale can be constructed

because none pair of item, among the remaining items, verify Hjk>.3

La première partie des sorties concerne le détail des étapes de l'algorithme de MSP (cette partie est supprimée avec l'option **nodetails**). A chaque étape, l'item sélectionné est indiqué, ainsi que la valeur du coefficient d'intégration de cet item dans la nouvelle échelle (Hj) et le coefficient d'échelle (H).

La seconde partie indique pour chaque échelle construite le nombre d'items, le coefficient d'échelle, la liste des items composant cette échelle et le coefficient d'intégration de chaque item dans l'échelle.

C.13 Le programme Stata hcaccprox

```
. hcaccprox c* eb* m*, detect(10)
```

```
Indexes to test the 10 latest partitions of the items
```

	DETECT	Iss	R
Only one cluster:	-0.00082	-0.16402	-0.08200
2 clusters:	-0.00081	-0.17989	-0.08131
3 clusters:	0.00039	-0.09524	0.03924
4 clusters:	0.00384	0.17989	0.38574
5 clusters:	0.00472	0.24339	0.47399
6 clusters:	0.00477	0.23810	0.47962
7 clusters:	0.00467	0.22751	0.46948
8 clusters:	0.00437	0.21693	0.43934
9 clusters:	0.00613	0.38624	0.61583
10 clusters:	0.00582	0.35979	0.58477

Cette première commande permet de visualiser en une seule fois les caractéristiques des dix plus petites partitions des items. La partition en 9 clusters semble la plus pertinentes car elle maxime les indices DETECT, Iss et R. On peut obtenir la composition des 9 clusters en tapant :

```
. hcaccprox c* eb* m*, part(9)
```

```
Number of clusters : 9
```

```
Cluster 1: eb4
Cluster 2: m7
Cluster 3: m10
Cluster 4: eb3 eb6
Cluster 5: m3 m4
Cluster 6: eb5 eb7
Cluster 7: c1 c5 c8 c4 c7 c9 c2 c3 c6
Cluster 8: eb1 eb8 eb2 eb9
Cluster 9: m2 m8 m6 m9 m1 m5
```

C.14 Le programme Stata detect

```
. detect c* eb* m* , part(9 9 10) noscores
```

```
DETECT : 0.0066
Iss : 0.4656
R : 0.6675
```

Correlations Items-Rest-Scores			
Items	dim 1	dim 2	dim 3
c1	0.3129	-0.0853	0.0228
c2	0.3980	-0.0726	-0.0042
c3	0.4265	-0.0543	0.0296
c4	0.4763	-0.0055	-0.0846
c5	0.3962	-0.0750	0.1163
c6	0.4088	-0.0150	0.0392
c7	0.5633	0.0108	0.0044
c8	0.4612	-0.0061	-0.0023
c9	0.4760	0.0473	-0.0196
eb1	0.0720	0.3060	0.0186
eb2	-0.0938	0.2102	-0.2358
eb3	-0.0257	0.2463	-0.1314
eb4	-0.0698	0.0504	-0.0337
eb5	0.0191	0.2177	-0.1735
eb6	-0.0386	0.1606	-0.1117
eb7	-0.0046	0.3082	-0.1774
eb8	-0.0059	0.3138	0.0061
eb9	-0.1029	0.2592	-0.1310
m1	0.0276	-0.0947	0.3662
m2	0.1000	-0.0470	0.3414
m3	-0.0964	-0.1295	0.1383
m4	-0.0514	-0.0721	0.3482
m5	0.0151	-0.1219	0.3285
m6	-0.0026	-0.1947	0.4684
m7	0.0957	-0.0534	0.1750
m8	0.0487	-0.1857	0.4425
m9	0.0084	-0.1461	0.5143
m10	0.1154	0.0586	0.2499

Seulement 47% des covariances vérifient une structure simple ($I_{ss}=0.4656$), et seulement 67% de la covariance vérifie une telle structure ($R=0.6675$). Cela signifie que la structure initiale des trois sous-échelles du SIP ne vérifie pas une structure simple. En étudiant les corrélations entre les items et les sous-scores, on note en effet des corrélations positives (parfois assez importantes) entre des items et des sous-scores d'une autre dimension : par exemple, c5 et le sous-score "mobilité" (dimension 3) ont un coefficient de corrélation de 0.1163, m2 et le sous-score "communication" (dimension 1) ont un coefficient de corrélation de 0.1000, ou m10 et le sous-score "communication" ont un coefficient de corrélation de 0.1154. D'autres items ont aussi de faibles corrélations avec le sous-score de leur dimension : eb4 (0.0504), eb6 (0.1606), m3 (0.1383) ou m7 (0.1750).