

Modélisation conjointe de trajectoires socio-professionnelles individuelles et de la mortalité par cause

A Latouche¹, M. Karimi et G. Rey

19 septembre 2013

¹Cnam, Paris. aurelien.latouche@cnam.fr

Projet ILEM : Inégalités sociales, lieux de résidence et mortalité par cancers

- Projet Financé depuis 2012 par l'INCa
- Porteur du projet G. Rey (Inserm, CépiDc²)
- Données : Cohorte/Panel COSMOP-DADS

Objectif :

Analyser le lien entre les trajectoires sociales des individus et le niveau de mortalité pour différents cancers.

²Centre d'épidémiologie sur les causes médicales de décès

- Cohorte pour la Surveillance de la Mortalité par Profession.

Données de survie (+ cause de décès en 13 catégories ³)

- Déclaration Annuelle des Données Sociales

Données longitudinales : historique des professions

COSMOP+DADS : \Rightarrow Une modélisation conjointe est naturelle

³<http://www.iresp.net/communication/publication-dun-bulletin-trimestriel-de-4-pages/>

- Echantillon au 1/24^{ème} de la population soit 1,755,590 individus (55% hommes)
- Episodes salariés (en dehors du secteur public) de 1976-2002
 - catégorie 2 : artisans, commerçants et chefs d'entreprises
 - catégorie 3: cadres et professions intellectuelles supérieures
 - catégorie 4: professions intermédiaires
 - catégorie 5: employés
 - catégorie 6: ouvriers
- Commune de domicile à chaque épisode

Pas de niveau d'étude ni de revenu

Périodes sans déclaration de la profession

- Définitives
 - retraite
 - activité non salarié
- Temporaires
 - période de chômage
 - arrêt maladie
 - grossesse
 - période non salarié

On distingue 2 approches complémentaires

- Analyse d'historique d'événements (*Event history analysis*)
: succession temporelle d'événements

- Analyse de séquences : recherche de motifs (comparaison de mots)

Illustration sur COSMOP+DADS

Analyse Retrospective

- Analyse de molécules comme du texte : alphabet TGACT (Thymine-Guanine-Adenine- Cytosine-Thymine)
- La comparaison de mots permet la comparaison d'acides aminé, de protéine ou d'ADN
- L' objectif étant d'identifier des régions similaires : même rôle fonctionnel
 - ⇒ Transposer cette approche pour les trajectoires professionnelles

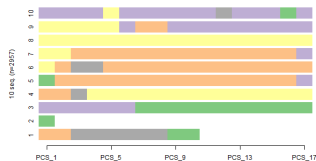
Rq : pas de lettre manquante dans un mot !

Une séquence est l'ensemble des états consécutifs d'un sujet.

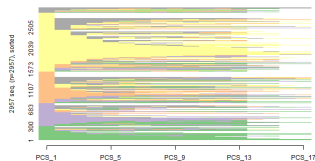
- L'analyse de séquences est une technique descriptive,
- une approche pour visualiser la structure de données
- pour identifier les régularités, les différences et les trajectoires-types.

Implémentation et prosélytisme : TraMineR

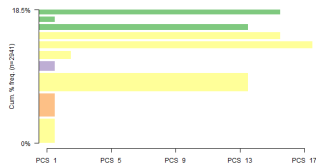
(a) 10 premières séquences



(b) toutes les séquences



(c) 10 séquences les plus fréquentes



Calcul de la distance entre les séquences

La distance est définie comme le coût minimum pour transformer une séquence en une autre.

- Insertion d'un élément dans la séquence
 - Suppression d'un élément de la séquence
- } *indel*

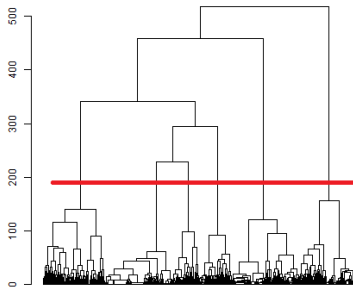
(Modifie l'ordre des événements dans les séquences)

- Substitution d'un élément par un autre
(préserve la structure du temps mais modifie la séquence des événements)

⇒ Faire une classification ascendantes hiérarchiques des séquences selon les distances obtenues

id	Catégories professionnelles occupées
1	5-5-*-*-*-*-*3-3
2	3
3	4-4-4-4-4-4-3-3-3-3-3-3-3-3-3-3-3
4	5-5*-6-6-6-6-6-6-6-6-6-6-6-6-6-6-6
5	3-5-5-5-5-5-5-5-5-5-5-5-5-5-5-4
6	6-5-*-*5-5-5-5-5-5-5-5-5-5-5-5-5

Illustration : Classification hiérarchique

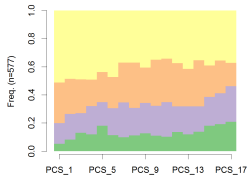


On sélectionne 6 classes

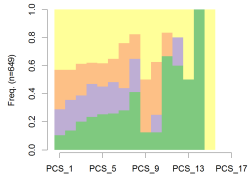
- Coupes instantanées de la distribution des individus de la classe entre les différents états à chaque instant

- Présente les proportions cumulées d'individus dans chacune des situations

Classe 1

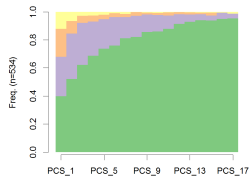


Classe 2

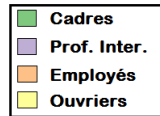
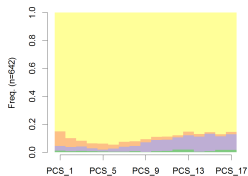


*Chronogrammes
de la typologie
en 6 classes*

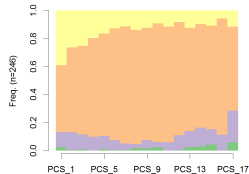
Classe 3



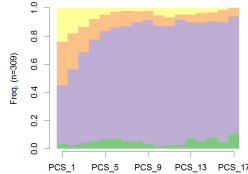
Classe 4



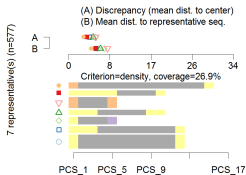
Classe 5



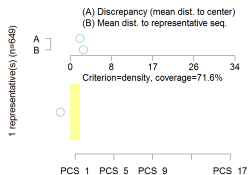
Classe 6



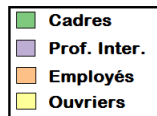
Classe 1



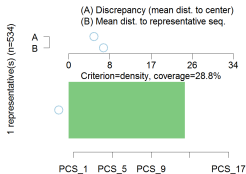
Classe 2



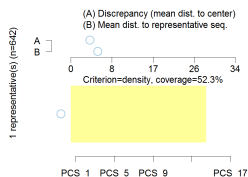
Séquences représentatives



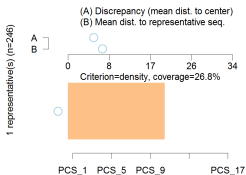
Classe 3



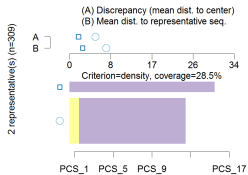
Classe 4



Classe 5



Classe 6



- Pas de méthode exacte pour choisir les coûts
- Choix du coût *indel* pour être cohérent avec le contexte social
- Pas de différence entre les trajectoires complètes et incomplètes
- Ne gère pas l'ordre et le temps

Analyse Prospective

Événements sont formalisés par les états d'un processus (markoviens ou non-markoviens)

La succession d'événements peut être modéliser par un processus X_t qui possède un nombre fini d'état \mathcal{S}

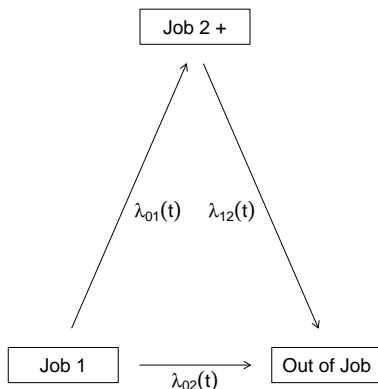
- Probabilités de transition :

$$P_{ij}(s, t) = P(X_t = j | X_s = i), i, j \in \mathcal{S}, s \leq t$$

- Intensité de transition :

$$\lambda_{ij}(t)dt = P(X_{t+dt} = j | X_t = i), i, j \in \mathcal{S}, i \neq j$$

- Modèles paramétrique, semi-paramétrique ou non-paramétrique pour modéliser les taux transitions et les taux d'occupations

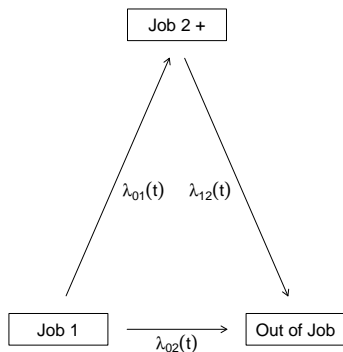


Certaines transitions ne sont pas possibles : matrice de transitions n'est pas symétrique

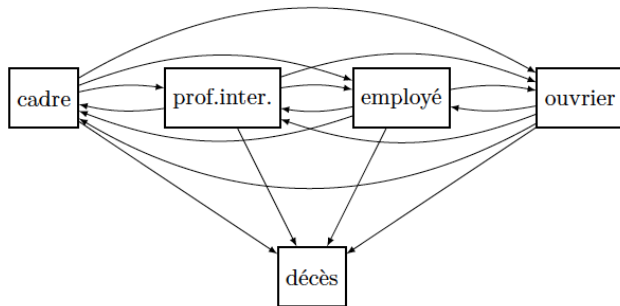
Matrice de transitions

	1	2	3
1	FALSE	TRUE	TRUE
2	FALSE	FALSE	TRUE
3	FALSE	FALSE	FALSE

Transitions possibles pour un modèle sain-malade-mort irréversible



Un exemple simple de modèle conjoint

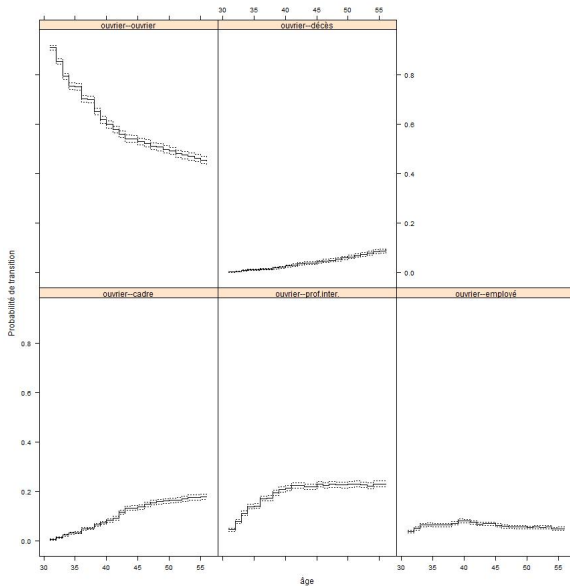


- Transitions vers les états transitoires modélisent les changements de la trajectoire professionnelle
- Transitions vers l'état absorbant modélisent la survenue du décès

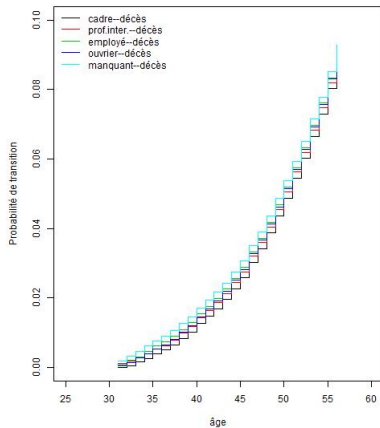
Implémentation : package `etm`

Approche Parsimonieuse ?

Probabilités de transitions : Ouvrier



Probabilité de transition vers le décès : Homme



Différences entre les 2 approches

Méthode	Analyse de séquence	EHA
Unité	séquence	événement
Métrique	matrice de distance	Matrice des taux transition
Inférence	retrospective	prospective
Mode d'inférence	Statique, non conditionnel	Dynamique, conditionnel
Type Inférence	Alignement	comparaison de taux
Niveau	population	individu

- Trajectoire professionnelle : Variables multinomiales répétées
- Trajectoire professionnelle : Covariable *interne*
- Utilisée telle que introduit un biais d'estimation de l'effet la profession sur la survie

Modélisation jointe où la composante longitudinale est associée au modèle de survie via des effets aléatoires (non observés) mais partagés.

- Effet Bi-directionnelle Trajectoire \leftrightarrow Survie
- Trajectoires *ascendantes* : Données répétées nominales ou ordinales ?

- Effet Bi-directionnelle Trajectoire \leftrightarrow Survie
- Trajectoires *ascendantes* : Données répétées nominales ou ordinales ?

Merci!

- Effet Bi-directionnelle Trajectoire \leftrightarrow Survie
- Trajectoires *ascendantes* : Données répétées nominales ou ordinales ?

Merci!

Questions ?

- Analyzing and Visualizing State Sequences in R with TraMineR. Gabadinho, A., G. Ritschard, N.S. Muller and M. Studer (2011). Journal of Statistical Software, 40(4), 1–37. <http://www.jstatsoft.org/v40/i04>
- Statistical analysis of life history calendar data. Eerola and Helske. Stat Med in Med Res. 2012
- Joint Modelling of Longitudinal and Time-to-Event Data: Challenges and Future Directions. D. Rizopoulos. 2012. Lifetime Data analysis.
- Joint Modelling of Longitudinal and Time-to-Event Data: An overview. Tsiatis. Davidian. Statistica Sinica. 2004